HAL
open science

Modèles bayésiens et application à l'estimation des caractéristiques de produits finis et au contrôle de la qualité

Qun Ying Zhu

To cite this version:

Qun Ying Zhu. Modèles bayésiens et application à l'estimation des caractéristiques de produits finis et au contrôle de la qualité. Sciences de la Terre. Ecole Nationale des Ponts et Chaussées, 1991. Français. NNT : . tel-00529487

HAL Id: tel-00529487

https://pastel.hal.science/tel-00529487v1

Submitted on 25 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NS 18779 (2)
p

Thèse présentée en vue de l'obtention du grade
de

DOCTORAT
de
l'Ecole Nationale des Ponts et Chaussées

MODELES BAYESIENS
ET APPLICATION A L'ESTIMATION DES CARACTERISTIQUES
DE PRODUITS FINIS ET AU CONTROLE DE LA QUALITE

par

Qun Ying ZHU

Soutenue devant le jury composé de

Président	Mme Laure ELIE
Rapporteur	Mme Laure ELIE
	M. Rudiger RACKWITZ
Examinateur	M. Vidal COHEN
	M. Bernard JACOB
	M. Jean Louis FAVRE

Paris, Juillet 1991

02

E.N.P.C.
INV01734

AVANT-PROPOS

Ce travail est réalisé dans le Laboratoire Central des Ponts et Chaussées (LCPC).

Je tiens tout d'abord à remercier Madame L. ELIE, professeur de l'Université Paris VII, qui m'a aidée et conseillée pendant toute la réalisation de cette étude et qui m'a consacrée beaucoup de temps et toute sa patience dans le soin apporté à la correction du mémoire. Je la remercie pour avoir accepté de présider mon jury de thèse. Qu'elle trouve ici l'expression de ma profonde gratitude.

Je remercie aussi Monsieur B. JACOB, chef de la section de la sécurité et du fonctionnement des structures (SFS), qui a proposé le sujet de thèse, m'a acceuillie dans son service et m'a ainsi permis de réaliser cette étude.

Je suis particulièrement reconnaissante à Monsieur V. COHEN, professeur de l'Ecole Nationale des Ponts et Chaussées et à Monsieur R. RACKWITZ, professeur de l'Université Technique de Munich, qui, malgré toutes leurs activités, ont accepté d'être membres du jury et ont prêté une attention aux modèles proposés dans cette thèse.

Je remercie ici tous ceux qui directement ou indirectement ont contribué à la réalisation de cette étude. Je pense en particulier aux membres de la section SFS, J. CARRACHILI, R. EYMARD, J.F. GORSE, F. GUERRIER et M. Y. LAU, aussi aux membres de la section de documentation du LCPC.

Ma reconnaissance va également à tous ceux qui m'ont aidé à accéder à toutes les informations, me permettant ainsi la mise en oeuvre des modèles adaptés au contrôle des boulons HR, des fils d'acier et des granulats.

SOMMAIRE

GÉNÉRALITÉS

I. PROBLÈME GÉNÉRAL 1

II. ASPECTS BAYÉSIENS 2

III. FORMALISME 4

IV. RÉSULTATS 6

V. APPLICATIONS 8

CHAPITRE I NOTIONS DE STATISTIQUE BAYÉSIENNE

1.0. PRELIMINAIRES 11

1.0.1. Modèle probabiliste et paramétrique
1.0.2. Echantillon
1.0.3. Statistique
1.0.4. Vraisemblance : fonction et principe

1.1. EXHAUSTIVITE 14

1.2. NOYAU D'UNE FONCTION 16

1.3. CONVEXITE 17

1.4. CONCEPTS BAYESIENS 18

1.4.1. Information a priori
1.4.2. Espace des paramètres
1.4.3. Formule de Bayes
1.4.4. Distribution a posteriori
1.4.5. Inférence statistique : l'approche décisionnelle

1.5. DE L'INFORMATION A PRIORI A LA LOI A PRIORI 25

1.5.1. Information subjective
et Vraisemblances relatives
1.5.2. Approximation et Estimation empirique
1.5.3. Distribution non-informative

— invariance de transformations
— loi non informative de Jeffrey

1.6. PRINCIPE DE CONJUGUAISON 33

1.6.1. Distributions conjuguées
1.6.2. Conjuguée naturelle et Opération associée

1.6.3. Opération linéaire de conjuguaison
1.6.4. Extention de la conjuguée naturelle

1.7. ESTIMATION BAYESIENNE 39

1.7.1. Inférence non-décisionnelle
1.7.2. Test et Région de confiance
1.7.3. Prédiction bayésienne

1.8. PROCESSUS DE CONTROLE 43

1.9. EXEMPLE 46

1.9.1. Vraisemblances relatives
1.9.2. Expériences antérieures
1.9.3. Comparaison et Représentation graphique

1.10. RESUME ET SCHEMA DU PROCESSUS 53

CHAPITRE II DÉTERMINATION DE LA LOI A PRIORI

2.0. PRESENTATION GENERALE 55

2.1. METHODE SEQUENTIELLE 57

2.1.1. Notion de séquentialité
2.1.2. Mesure uniforme du paramètre θ
2.1.3. Exemple du modèle binomiale-bêta (1)

2.2. METHODE DES SCORES D'EXPERT 65

2.2.1. Formalisme de scores d'expert
et Application logarithmique
2.2.2. Fonction de vraisemblance bayésienne
et Résolution générale
2.2.3. Statistique exhaustive correspondante
de la famille exponentielle

2.3. METHODE DES ECHANTILLONS EQUILIBRES 69

2.3.1. Exemple du modèle binomial-bêta (2)

2.4. METHODES DES FAMILLES PARAMETRIQUES 72

2.4.1. Méthode des moments
2.4.2. Exemple du modèle binomial-bêta (3)
2.4.3. Méthode des fractiles
2.4.4. Méthode du maximum de vraisemblance
et Exemple du modèle binomial-bêta (4)

2.5. METHODES DU QUASI-ECHANTILLON 79

2.5.1. Construction d'un quasi-échantillon du paramètre θ
2.5.2. Estimateur des paramètres de la loi a priori
2.5.3. Exemple du modèle binomial-bêta (5)
et Divers estimateurs du quasi-échantillon

2.6. EXEMPLE NUMERIQUE ET GRAPHIQUE 84

2.6.1. Résumé des estimateurs du modèle binomial-bêta
2.6.2. Caractéristiques a priori et a posteriori du taux
2.6.3. Résultats numériques et Comparaison des méthodes

CHAPITRE III FAMILLES PARTICULIÈRES DE DISTRIBUTIONS CONJUGUÉES ET MODÈLES BAYÉSIENS CORRESPONDANTS

3.0. OBJECTIF ET COMMENTAIRES 93

3.0.1. Deux familles générales et Leurs modèles bayésiens
3.0.2. Estimation des paramètres connus

3.1. DISTRIBUTION NORMALE (LAPLACE-GAUSS) 96

3.1.1. μ et σ inconnus
3.1.2. μ inconnu et σ connu
3.1.3. μ connu et σ inconnu

3.2. DISTRIBUTION LOG-NORMALE 106

3.3. DISTRIBUTION NORMALE INVERSE (WALD) 109

3.3.1. Couple des variables (μ,β^2)
3.3.2. Variable μ
3.3.3. Variable β^2

3.4. DISTRIBUTION GAMMA 114

3.4.1. α et β inconnus
3.4.2. α connu et β inconnu

3.5. DISTRIBUTION GAMMA GENERALISEE 120

3.6. DISTRIBUTION GAMMA INVERSE 121

3.7. DISTRIBUTIONS DE VALEURS EXTREMES : WEIBULL ET FRECHET 123

3.8. DISTRIBUTION DE PARETO 128

3.9. RESUME 131

CHAPITRE IV ESTIMATIONS DES PARAMÈTRES DE LA LOI A PRIORI

4.0. INTRODUCTION 135

4.1. ESTIMATION SEQUENTIELLE 136

4.2. ESTIMATION DES ECHANTILLONS EQUILIBRES 139

4.3. ESTIMATIONS DES FAMILLES PARAMETRIQUES 141

4.3.1. Distributions normale et log-normale
4.3.2. Distribution normale inverse
4.3.3. Distributions gamma et gamma inverse
4.3.4. Distributions de Weibull et de Fréchet
4.3.5. Distribution de Pareto

4.4. ESTIMATIONS DU QUASI-ECHANTILLON 148

4.4.1. Distributions normale et log-normale
4.4.2. Distribution normale inverse
4.4.3. Distributions gamma et gamma inverse
4.4.4. Distributions de Weibull et de Fréchet
4.4.5. Distribution de Pareto

4.5. MODELE NORMALE-GAMMA A QUATRE PARAMETRES 163

4.5.1. Forme mathématique et Estimateur sans biais
4.5.2. Non validation du modèle pour certaines méthodes

— méthode séquentielle
— méthode des échantillons équilibrés
— méthode des moments

4.5.3. Comparaison des estimateurs du quasi-échantillon

— méthode des double moments
— méthode des double maxima de vraisemblance
— signification des paramètres
et leurs contraintes

4.5.4. Modèle $NG(m,v,k)$ avec l'estimateur sans biais

4.6. ETUDE DE LA SENSIBILITE ET COMPARAISON DES METHODES 173

4.6.1. Etude de la sensibilité du modèle noraml-gamma
4.6.2. Comparaison des méthodes
4.6.3. Résumé des estimateurs des modèles normal et gamma

CHAPITRE V ETUDE DU CAS DES BOULONS HR

5.0. EXPOSE DE L'ENVIRONNEMENT 185

5.1. COEFFICIENT K 186

5.1.1. Définition et Principe du contrôle
5.1.2. Précision de l'analyse statistique
5.1.3. Prospection sur l'ensemble des données

5.2. APPROCHE BAYESIENNE 191

5.3. MISE EN OEUVRE DES MODELES 194

5.3.1. Quelques modèles pour la distribution normale de K
5.3.2. Prise en compte du passé

5.4. RESULTATS PREDICTIFS ET COMPARAISON DES MODELES 200

5.4.1. Ecart-type prédictif
5.4.2. Moyenne prédictive

5.5. COMPARAISON DES METHODES 205

5.5.1. Etude de la moyenne et la variance ensemble
5.5.2. Etude de la moyenne
5.5.3. Etude de la variance

5.6. MODELE SPECIFIQUE 220

5.6.1. Combinaison des modèles normal et gamma
5.6.2. Calcul des résultats prédictifs
et Représentation graphique
5.6.3. Enrichissment d'un échantillon
et Sensibilité du modèle

5.7. CONCLUSION 227

NOTATION 229

ANNEXES

A.1. DETERMINATIONS DE LA LOI A PRIORI Be(α,β) DU TAUX θ 235

Cas 1. Nombre de tirages constant k
Cas 2. Nombre d'unités défectueuse constant x

A.2. DETERMINATION DES ESTIMATEURS DES LOIS A PRIORI 240

Tableau 1. Estimation séquentielle
Tableau 2. Estimation des échantillons équilibrés
Tableau 3. Estimation des moments
Tableau 4. Estimation des double moments
Tableau 5. Estimation des moments-maxima
Tableau 6. Estimation des maxima-momants
Tableau 7. Estimation des double maxima de vraisemblance

A.3. HISTROGRAMMES DES BOULONS 1988 248

Figure 1. Valeurs K de la qualité HR1
Figure 2. Valeurs K de la qualité HR1Z
Figure 3. Valeurs K de la qualité HR2Z

A.4. RESULTATS DU COEFFICIENT K DES BOULONS 249

Tableau 1. Prédictions avant et après la réception du lot n°17 (par le modèle normal-gamma)
Tableau 2. Passage de l'a priori à a posteriori pour le lot n°17 (réalisé par le modèle gamma)
Tableau 3. Prédictions avant et après le réception du lot n°17 (par le modèle gamma)
Tableau 4. Prédictions avant et après la réception du lot n°18 (par le modèle normal-gamma)
Tableau 5. Passage de l'a priori à a posteriori pour le lot n°18 (réalisé par le modèle gamma)
Tableau 6. Prédictions avant et après le réception du lot n°18 (par le modèle gamma)

A.5. TABLE DE DISTRIBUTION DU COEFFICIENT DE CORRELATION 254

BIBLIOGRAPHIE 255

GENERALITES

I. Problème Général

Cette étude porte sur une propriété quantifiée par une variable, d'un matériau ou d'un produit dont les lots sont fabriqués dans des conditions spécifiques homogènes et répondent aux mêmes exigences ou à la même qualité. Citons par exemple des résistances de fils d'acier de mêmes spécifications, le coefficient K des boulons à haute résistance caractérisant sa qualité,... etc. Cette étude comporte les parties suivantes :

— développement de modèles descriptifs, suivant le point de vue bayésien;

— application à l'estimation et à la prévision de valeurs caractéristiques de X et de sa distribution en contexte aléatoire;

— application au contrôle de la qualité.

Dans un contexte économique, ces méthodes permettent d'optimiser les décisions concernant le contrôle de la qualité des produits ou leur emploi : acceptation d'un lot spécifique, détermination de valeurs caractéristiques, détermination de coefficients de sécurité et décision d'usages de matériaux dans un ouvrage, ... etc.

Dans les modèles classiques, l'analyse de la moyenne et de la variance (μ, σ^2) en fonction de la taille k d'un échantillon **x** conduit à trouver des statistiques significatives telle qu'un intervalle de confiance avec des risques (de 1er ou 2ème espèce) donnés, des fractiles dépendant des probabilités d'acceptation ou de rejet, plus généralement des courbes d'efficacité sur lesquelles est basée la procédure de contrôle soit à la fin d'une production, soit lors de la réception d'un lot livré à un client. Tout cela est fait sous l'hypothèse d'une distribution normale de la variable X, basée sur le théorème central limite, avec une moyenne et une variance estimées.

Le point de vue bayésien permet d'utiliser l'information antérieure (souvent des échantillons provenant de productions semblables) pour obtenir une connaissance a priori sur le paramètre $\theta = (\mu, \sigma^2)$, paramètre pouvant prendre différentes valeurs avec leurs probabilités d'occurence. L'analyse descriptive intègre ces valeurs et donne avec une bonne précision la distribution prévisionnelle p(x) au lieu d'une valeur $\hat{\theta}$, estimée la plus probable, dans l'approche classique. Les modèles bayésiens permettent de considérer des distributions autres que la distribution normale $N(\mu, \sigma^2)$. Or en réalité, bien que l'hypothèse de la distribution normale soit très fréquente en statistique, il arrive souvent qu'elle soit mal vérifiée en raison de non-symétrie, de troncature unilatérale ou d'aplatissement trop rapide ou trop lent des distributions, dont l'écart avec une distribution normale peut être non négligeable.

Dans le contrôle de qualité au sens usuel, c'est à dire sans critères économiques (fonction de coût, risque d'une décision, ...), on cherche les valeurs de critères, liées étroitement aux fractiles de la distribution prédictive développée ci-dessus, et la relation entre ces fractiles et des probabilités d'acceptation, appelée courbe d'acceptation dans le cas bayésien (et courbe d'efficacité dans le cas classique). Cette courbe indique comment varie la probabilité d'acceptation en fonction du niveau réel de la qualité. Une fois un plan d'échantillonnage fixé simple, multiple, ou séquentiel, le contrôle par attribut ou par quantile en découle systématiquement.

Après avoir traité les points précédents, la partie décisionnelle sera abordée avec les différentes exigences du contrôle, en élargissant le champ des décisions aux valeurs caractéristiques, jusqu'à l'introduction de la notion de fonction d'utilité et de règle d'optimisation. Il s'agit de trouver, en liaison avec des experts, les conséquences économiques des décisions, et plus précisément les fonctions de coût et les risques d'une décision, ou de mettre en oeuvre des fonctions simples pour voir comment fonctionnent les processus de décision.

II. Aspects Bayésiens Généraux

D'un point de vue pratique, l'analyse bayésienne n'est rien d'autre

qu'une méthode d'analyse statistique descriptive parmi les autres. En effet, c'est pour des raisons essentiellement méthodologiques que des extensions d'analyses classiques ou traditionnelles ont été élaborées dans le cadre de la théorie bayésienne, afin d'utiliser tout simplement toutes les ressources mathématiques, qui, alliées aux immenses possibilités ouvertes par le calcul automatique, en font une théorie intéressante. Bien que sophistiquée, cette méthodologie permet de couvrir une bonne partie des domaines d'applications des méthodes d'analyses habituelles : elle apporte à la fois des compléments appréciables dans la pratique expérimentale et des conclusions plus complètes ; elle permet d'améliorer les procédures existantes parfois mal adaptées à des situations particulières.

L'origine de cette idée remonte à Bayes qui l'a introduite dans son célèbre mémoire de l'année 1763 ([3]). D'une manière générale, cette analyse apporte, dans le domaine des modèles descriptifs, les idées suivantes :

— L'analyse statistique permet, dans la planification des expériences, de considérer toutes les informations quantitatives et qualitatives sur l'incertitude dans les modèles ; ici, il s'agit du paramètre θ.

— Le paramètre θ est considéré, dans l'approche classique, comme une grandeur inconnue (un vecteur dans le cas multidimensionnel), mais certaine. Dans l'approche bayésienne, il peut prendre plusieurs valeurs possibles, avec des probabilités associées, ce qui conduit à trouver des distributions sur l'espace des paramètres. Différents modèles sont adaptés à ces développements.

— Des informations sur ce paramètre θ avant échantillonnage permettent d'évaluer une loi de distribution, dite "a priori". L'analyse bayésienne consiste à déduire de cette loi de distribution, grâce au théorème de Bayes, une distribution dite "a posteriori", en ajustant la valeur du paramètre par un jugement probabiliste de l'incertitude compte tenu des données recueillies dans un échantillon $\mathbf{x}$.

Cette analyse bayésienne intègre, outre le plan d'échantillonnage traditionnel, les résultats d'expériences conçues antérieurement et les exploite de manière optimale lorsque les données expérimentales sont insuffisantes pour appliquer l'analyse fréquentiste.

De plus, cette analyse offre une possibilité suplémentaire : la loi a priori représente toutes les informations pertinentes provenant de données antérieures, mais il est possible d'y incorporer toute connaissance ou opinion, même très conjecturelle ou "subjective" que l'on peut avoir sur le problème étudié.

Après l'établissement d'une distribution a priori du paramètre inconnu θ, si des expériences complémentaires sont réalisées, on reprendra la distribution a posteriori de ce paramètre comme nouvelle distribution a priori. On aura ainsi une distribution réactualisée qui incorporera les apports des expériences successives, d'où la méthode dite séquentielle (cf. paragraphe 2.1). La recherche d'une stabilité de forme de ces distributions sera indispensable en pratique, notamment pour réaliser des calculs automatiques (cf. paragraphe 1.7), d'où la propriété de conjuguaison.

La distribution a priori vise à représenter un "état d'ignorance" sur le paramètre θ ; la distribution a posteriori correspondante pourra alors être interprétée comme résultant de l'apport propre des données.

Le cadre bayésien se présente donc comme une théorie formalisée de l'apprentissage par l'expérience.

III. FORMALISME

Un modèle caractérisé par un vecteur aléatoire X peut être décrit par un modèle paramétrique $Y = F(X,\theta)$, où X et Y sont deux variables aléatoires et θ est le paramètre (appelé état de la nature dans la théorie statistique bayésienne). Ici, nous considérons une variable aléatoire X à distribution continue admettant $F(x|\theta)$ comme fonction de répartition et $f(x|\theta)$ comme fonction de densité. Supposons connue la forme de cette distribution, tandis que le paramètre θ est inconnu et aléatoire. Nous cherchons dans un premier temps à connaître la distribution π de θ en exprimant une connaissance a priori sur l'espace Θ des valeurs possibles de θ, puis à évaluer ensuite X sous forme d'une distribution marginale, appelée distribution bayésienne ou prédictive.

Supposons que nous disposons comme information a priori d'une série d'observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ dans un processus séquentiel, et que toutes ses expériences recueillies soient indépendantes l'une de l'autre.

L'analyse classique ignore d'une certaine manière ces observations antérieures et n'utilise que l'observation $\mathbf{x}_{n+1}$ pour estimer une valeur θ_{n+1}, à partir de la fonction $F(\mathbf{x}_{n+1}|\theta)$ ou $f(\mathbf{x}_{n+1}|\theta)$, selon la méthode utilisée et le type d'information disponible. L'approche bayésienne au contraire utilise ces observations d'une manière optimale pour ajuster une distribution a priori notée par $\pi(\theta)$, puis pour calculer la distribution a posteriori notée par $\pi(\theta|\mathbf{x}_{n+1})$, grâce au théorème de Bayes :

$$\pi(\theta|\mathbf{x} = \mathbf{x}_{n+1}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_\Theta f(\mathbf{x}|\theta)\pi(\theta)d\theta} \qquad (*)$$

Alors, la moyenne bayésienne ou prédictive (a posteriori) $\mathbb{E}(\theta|\mathbf{x}_{n+1})$ déduite de $\pi(\theta|\mathbf{x}_{n+1})$ est un estimateur de θ au sens des moindres carrés.

Si le problème consiste tout simplement à estimer le paramètre θ ou sa distribution paramétrique ou non, on peut se reporter aux travaux sur les familles de distributions usuelles : normale (moyenne inconnue), uniforme, Poisson, négative binomiale, logarithmique, exponentielle et gamma (cf. [39] et [40]).

Nous nous plaçons ici dans le cadre paramétrique, et nous nous intéressons plutôt aux distributions $\pi(\theta)$ et $\pi(\theta|\mathbf{x})$, afin d'obtenir finalement les distributions prédictives a priori et a posteriori (dites aussi distributions bayésiennes) :

$$p(x) = \int_\Theta f(x|\theta)\pi(\theta)d\theta$$

$$p(x;\mathbf{x}) = \int_\Theta f(x|\theta)\pi(\theta|\mathbf{x})d\theta$$

sur lesquelles est basée l'analyse prévisionnelle. Ceci constitue un outil supplémentaire pour le processus de contrôle, permettant éventuellement de

réduire le volume d'essais ou la taille d'échantillons et de traiter le cas de petits échantillons sans pour autant perdre de précision ou de fiabilité.

IV. RÉSULTATS

Nous avons vu plus haut que l'évaluation de la valeur caractéristique de la variable X est, dans le plupart des cas, fondée sur l'hypothèse de la normalité. Les exploitations sont relativement faciles ; le modèle normal-gamma correspond au cas d'une distribution normale de X avec moyenne et variance inconnues : $\theta = (\mu, \sigma^2)$, le modèle normal dans le cas où la moyenne $\theta = \mu$ est inconnue et le modèle gamma se présente si la variance σ^2 ou la précision $\eta = 1/(2\sigma^2)$ est inconnue. Ce sont des résultats classiques.

Pour résoudre d'autres problèmes, des modèles issus de distributions non-normales sont investigués au chapitre III dans le cas de distributions à densités continues, par exemple log-normale, normale inverse, gamma, gamma inverse, de Weibull, de Frechet et de Pareto, etc. En effet, l'existence de distributions conjuguées, vérifiant une stabilité de forme pour faciliter le passage par le théorème de Bayes de la distribution a priori à l'a posteriori de θ après un nouvel échantillonnage, n'est valable que pour certaines familles de distributions (cf. paragraphes 1.1, 1.6 et 1.7).

Pour démarrer l'analyse bayésienne, il faut avant tout évaluer la distribution a priori de la variable θ, au lieu d'estimer une valeur certaine dans le cas classique. Nous avons vu que d'autres distributions peuvent être obtenues facilement, en utilisant le théorème de Bayes et la théorie des distributions conditionnelles. Ainsi, la distribution a priori joue un rôle primordial dans cette procédure d'analyse. Différentes techniques sont proposées pour établir une telle distribution et associées avec certaines conditions d'applicabilité en fonction de la nature des données. Citons par exemple, la méthode séquentielle, la méthode des scores d'expert, la méthode des échantillons équilibrés, les méthodes des familles paramétriques et les méthodes du "quasi-échantillon".

Parmi les méthodes développées, la méthode séquentielle peut déterminer à la fois la forme analytique de la distribution a priori et son paramètre

(cf. paragraphe 2.1). Pour utiliser d'autres méthodes, la connaissance de la distribution a priori est fondée sur une même notion fondamentale : la conjuguée de la fonction de vraisemblance. Il s'agit de donner à la distribution a priori $\pi(\theta)$ la forme de la fonction de vraisemblance d'un échantillon $\mathbf{x}$: $L(\theta;\mathbf{x}) = f(\mathbf{x}|\theta)$ (cf. paragraphe 1.7).

Nous allons montrer que les distributions ayant une telle forme vérifient la propriété de conjuguaison, à l'aide de l'utilisation de la théorie des statistiques exhaustives (cf. paragraphe 1.1). Il est clair que, d'après le théorème de Bayes (*), la propriété de conjuguaison dépend directement de la particularité de la distribution $f(x|\theta)$ de la variable X.

Une fois la forme de la distribution a priori connue, les méthodes des familles paramétriques et du quasi-échantillon permettent de déterminer son paramètre (cf. paragraphes 2.4 et 2.5).

La méthode des scores d'expert sera adaptée aux problèmes pour lesquels les expériences sont affectées de poids différents dans l'exploitation de l'information a priori, et donc dans la distribution a priori. Par exemple, l'information a priori mélange des expériences réalisées dans des lots de la production en cours et dans des lots de productions passées. Il convient d'introduire, avec des experts, des coefficients spécifiques $\{\alpha_i\}$. Dans ce cas, nous construirons une nouvelle fonction de vraisemblance au lieu de considérer la fonction de vraisemblance au sens usuel, et la méthode de la conjuguée naturelle permet de choisir une distribution a priori conjuguée à cette nouvelle fonction : $f'(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n|\theta) = \prod f(\mathbf{x}_i|\theta)^{\alpha_i}$. Nous l'appelons fonction de vraisemblance approchée (cf. paragraphe 2.2).

La combinaison de l'idée de scores d'expert avec l'une des autres méthodes nous donnera tous les moyens pour résoudre des problèmes où les poids accordés aux expériences sont différents.

Au contraire dans le cas où les techniques de production sont homogènes et où chacune des expériences doit représenter un poids identique dans la construction de la distribution a priori, d'où l'intérêt d'introduire la méthode des échantillons équlibrés permettant d'éliminer les effets dûs aux tailles différentes des échantillons (cf. paragraphe 2.3).

Remarquons que l'évaluation de la distribution a priori par la méthode séquentielle et la méthode des familles paramétriques, ainsi que de la distribution prédictive, est réalisée de façon directe ou indirecte à partir de l'ensemble des données mélangées (ou échantillon regroupé) $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n = \{x_j^i;\ j = 1, 2, \ldots, k_i \text{ et } i = 1, 2, \ldots, n\}$. Nous pouvons conclure que plus la taille d'un échantillon est grande, plus cet échantillon prend un poids important dans le résultat prédictif ; la correction de la distribution a priori par une nouvelle expérience sera en général moins importante et donc le modèle prédictif sera moins sensible.

Par contre, la méthode des échantillons équilibrés et la méthode du quasi-échantillon permettent d'évaluer de façon directe ou indirecte la loi de distribution a priori à partir des moyennes des statistiques telles que les moyennes $\{m_i\}$, les variances $\{v_i\}$, les précisions $\{h_i = 1/(2v_i)\}$, ...etc. Le modèle ainsi obtenu ne dépend plus des tailles des échantillons, les poids de l'information a priori (du passé) et de l'information provenant d'un nouvel échantillon (du présent) sont a priori équilibrés, puisqu'ils dépendent étroitement de paramètre de "taille" k, soit une sorte de moyenne des tailles $\{k_i\}$ pour l'information a priori, soit la taille de l'échantillon prélevé dans un lot à tester. Pour la méthode séquentielle, ce paramètre de taille est égal au nombre total des données mélangées, soit $k = \sum k_i$. Nous pouvons conclure que le modèle prédictif obtenu par la méthode des échantillons équilibrés ou par la méthode du quasi-échantillon dépend essentiellement des variations à l'intérieur des lots et de la dispersion des moyennes.

Dans l'approche paramétrique, la méthode du quasi-échantillon comprend deux étapes, dans lesquelles l'estimation des paramètres de la distribution considérée sera menée par l'une des méthodes classiques. Une telle estimation sera entâchée d'une erreur probabiliste selon la taille de l'échantillon. Par contre, la méthode des échantillons équilibrés prend en compte de façon intégrale la variation des paramètres à estimer, donc le résultat est relativement plus stable que celui obtenu par la méthode du quasi-échantillon.

Dans le chapitre IV, nous donnerons tous les estimateurs des modèles développés au chapitre III, nous étudierons à titre d'exemple la sensibilité du modèle normal-gamma et nous effectuerons une comparaison des estimateurs, et également des divers estimateurs des modèles normal et gamma.

V. APPLICATIONS

Bien entendu, toute étude dépend de la nature des données disponibles. Au vu des résultats obtenus et de la qualité demandée, on choisira un processus de traitement bayésien ou classique. Dans les cas de la vérification du contrôle en usine des fils d'aciers et des boulons HR à haute résistance, les contrôles sont effectués par mesure et les coefficients de variation des paramètres caractéristiques étudiés sont tous relativement petits.

Dans le cas des fils d'acier, la condition d'acceptation du contrôle de conformité est le respect d'une borne inférieure de résistance avec une probabilité donnée α, risque de 1er espèce s'il s'agit d'un fournisseur (contrôle final de production), ou de 2ème espèce s'il s'agit d'un client (contrôle de réception). Mais, le minimum des données est bien supérieure à la borne proposée dans la norme française. Nous concluons dans ce cas que la moyenne classique suffira pour assurer la qualité, et de plus nous proposons de diminuer le volume des contrôles, et donc leur coût d'essais, à la condition que les prélèvements du contrôle restent significatifs.

Dans le cas des boulons HR, le contrôle demande pour le moment uniquement l'estimation de la moyenne et de l'écart-type associé. Le nombre de prélèvements dans un lot est relativement petit par rapport à la taille du lot. Ainsi, nous introduisons la technique bayésienne dans l'estimation statistique des valeurs caractéristiques du coefficient K des boulons HR. Les contrôles précédents permettent de constater une bonne ou mauvaise stabilité de ces valeurs, et de donner une estimation a priori sur un lot à tester en supposant une continuité de productions homogènes ; le contrôle sur ce lot permet de justifier ou corriger ce préavis et de donner l'estimation a posteriori. Dans notre étude, nous proposons des modèles descriptifs adaptés au cas des boulons, à partir desquels la combinaison avec d'autres techniques permettra d'établir un processus complet du contrôle avec des règles bien définies, telles que la définition d'un plan d'échantillonnage convenable. Toutes les analyses ont pour but d'évaluer au mieux les vraies valeurs de X, et de diminuer autant que possible le volume d'essais sans perdre de précision dans l'analyse statistique ou d'augmenter cette précision sans réaliser plus d'essais.

On pourrait aussi, avec certaines précautions, déterminer les valeurs

caractéristiques d'un contrôle de conformité de spécifications, basées sur des données antérieures et des données prélevées dans un lot à étudier.

Les méthodes développées dans cette étude sont appliquées en supposant que les données représentent des tirages successifs de variables aléatoires indépendantes équidistribuées et à distribution continue.

Les applications numériques seront développées pour le contrôle des boulons HR. Les analyses seront faites en supposant la normalité de la distribution normale de la variable étudiée.

CHAPITRE I NOTIONS DE STATISTIQUE BAYÉSIENNE

Ce chapitre est consacré à exposer des notions élémentaires de la théorie bayésienne. Nous renvoyons le lecteur désireux de compléments dans ce domaine à l'ouvrage de J.O. Berger (1985).

1.0. PRELIMINAIRES

1.0.1. Modèle probabiliste et paramétrique

L'analyse statistique exige en général de disposer d'un ensemble d'observations. Pour les modéliser mathématiquement, nous considérons le modèle paramétrique $\left(\Omega,\ \mathcal{B},\ (\mathbb{P}_\theta)_{\theta\in\Theta}\right)$ défini de la façon suivante :

— l'ensemble Ω désigne tous les résultats possibles, considérés individuellement comme des réalisations d'une grandeur aléatoire observable, décrite par une variable aléatoire X;

— l'ensemble $\mathcal{B}$ est formé des événements possibles, représentés par des sous-ensembles de Ω , constituant une σ-algèbre de Ω;

— pour chaque θ de Θ, $\mathbb{P}_\theta$ est une distribution de probabilité sur l'espace probabilisable $(\Omega,\mathcal{B})$, attribuant une probabilité à chaque événement.

L'ensemble Θ des valeurs du paramètre θ est souvent multidimensionnel. Ce paramètre est appelé aussi l'état de la nature dans la théorie statistique de la décision.

Par exemple, supposons que la variable X suie une loi normale dont les paramètres de moyenne μ et de variance σ^2 sont à déterminer. L'inconnue θ du problème est le couple $\theta = (\mu,\sigma^2)$.

Désignons par μ_θ la loi de distribution de la variable aléatoire X dans l'état de la nature θ. Elle est la loi image de la probabilité $\mathbb{P}_\theta$ par X. Si la famille de lois $\{\mu_\theta \ ; \ \theta \in \Theta\}$ est dominée par une mesure σ-finie μ, nous noterons $f(x|\theta)$ la densité de probabilité de μ_θ relative à μ et nous avons :

$$f(x|\theta) = \frac{d\mu_\theta}{d\mu}(x) \tag{1.01}$$

Dans ce cas, le modèle correspondant est appelé modèle dominé ; et la loi de distribution de densité $f(x|\theta)$ est dite ici sous-jacente. Nous emploierons dans la suite la terme "X suit la loi f" ou "X $\approx\approx\approx$ f" au lieu de "X suit la loi de densité f" par souci de brièveté.

Les grandeurs aléatoires que l'on considère seront à valeurs dans un espace euclidien $\mathcal{H}$ qui sera le plus souvent $\mathbb{R}$ ou $\mathbb{R}^+$. Nous noterons X la variable aléatoire désignant une telle grandeur. Elle sera une application définie sur $(\Omega, \mathcal{B}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ à valeurs dans $\mathcal{H} = \mathbb{R}$ ou $\mathbb{R}^+$ muni de la tribu borélienne $\mathcal{B}(\mathbb{R})$ ou $\mathcal{B}(\mathbb{R}^+)$. L'espace Ω sera par la suite soit $\mathcal{H}$, soit $\mathcal{H}^n$, si nous considérons un échantillon de la variable X, de taille finie $n < +\infty$, et dont la loi correspondante est définie dans la section suivante.

1.0.2. Echantillon

Considérons l'expérience qui consiste à tirer n valeurs consécutives $x_1, x_2, \ldots, x_n$ d'une variable aléatoire X, ces réalisations étant indépendantes.

Ces réalisations qui constituent un échantillon de X peuvent être aussi considérées comme une réalisation d'un n-uplet de variables aléatoires $(X_1, X_2, \ldots, X_n)$ indépendantes et de même loi, celle de X. Cet n-uplet est appelé un n-échantillon aléatoire de la variable sous-jacente X, ou plus simplement un échantillon aléatoire de la variable X. S'il n'y a pas de confusion possible dans le contexte, la notation **X** représentera aussi cet n-échantillon aléatoire de X.

L'espace de probabilité $(\Omega, \mathcal{B}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ sera alors constitué de la façon suivante: Ω est l'ensemble des échantillons, noté par $\mathcal{H}^n$; $\mathcal{B}$ est la tribu des boréliens de $\mathcal{H}^n$; $\mathbb{P}_\theta$ est la distribution de probabilité produit des lois de

probabilité de chacun des X_i, $i = 1,2, \ldots,n$.

Dans ce cas , la loi de distribution de X sera le produit $\mu_\theta = \prod \mu_{\theta i}$ avec $\mu_{\theta i}$ la loi de la variable X_i, pour tout $i = 1,2,\ldots,n$. Dans le cas d'un modèle dominé, si nous notons $f(x_i|\theta)$ la densité de probabilité de $\mu_{\theta i}$ relative à μ pour $i = 1,2,\ldots,n$, la densité de cet échantillon obéit à la règle de produit :

$$f(\mathbf{x}|\theta) = \prod_i f(x_i|\theta) \tag{1.02}$$

où $\mathbf{x} = (x_1,x_2,\ldots,x_n)$.

1.0.3. Statistique

Il est d'usage dans la pratique de résumer les n valeurs d'un échantillon $\mathbf{x} = (x_1,x_2,\ldots,x_n)$ par quelques caractéristiques simples telles que leur moyenne arithmétique, leur moment centré d'ordre 2, leur plus grande valeur, leur plus petite valeur, leur moyenne géométrique, ...etc. Ces caractéristiques sont elles-mêmes des réalisations de variables aléatoires fonctions de $\mathbf{X} = (X_1,X_2,\ldots,X_n)$.

Une statistique T d'une variable aléatoire X est une fonction mesurable à valeurs dans un espace Λ, d'un n-échantillon aléatoire de X, constitué de n tirages indépendants de X. Nous supposerons que l'espace Λ est de dimension finie p : $\Lambda \subseteq \mathbb{R}^p$.

Pour simplifier la présentation, un échantillon $\mathbf{x}$ peut être une valeur de la variable aléatoire X lorsque n = 1.

1.0.4. Vraisemblance : fonction et principe

Soit $\mathbf{x} = (x_1,x_2,\ldots,x_n)$ un échantillon de la variable aléatoire X. Dans la plupart des problèmes, nous cherchons à déterminer le paramètre inconnu θ de manière à maximiser la probabilité d'observer l'échantillon $\mathbf{x}$ dans la famille de distributions possibles $\mathbb{P}_\theta$. Nous considérons alors la densité de cet échantillon $f(\mathbf{x}|\theta)$ comme une fonction de θ, appelée dans ce cas fonction de vraisemblance de θ en $\mathbf{x}$, et notée par $L(\theta;\mathbf{x}) = f(\mathbf{x}|\theta)$.

Le principe de vraisemblance repose sur les hypothèses suivantes :

(1) l'information consernant le paramètre θ, tirée de l'observation (ou de l'échantillon) $\mathbf{x}$, est contenue dans la vraisemblance $L(\theta;\mathbf{x})$;

(2) si pour deux observations $\mathbf{x}_1$ et $\mathbf{x}_2$, il existe une constante c telle que, pour tout θ, $L(\theta;\mathbf{x}_1) = L(\theta;\mathbf{x}_2)$, elles apportent la même information sur θ et doivent conduire à la même inférence.

(cf. [5] et [8] pour une étude plus détaillée).

Dans l'approche classique, on cherche à déterminer une (des) valeur(s) de θ qui maximise la fonction de vraisemblance $L(\theta;\mathbf{x})$, appelée estimateur du maximum de vraisemblance. Ceci est l'une des approches compatibles avec le principe de vraisemblance. Nous verrons que l'approche bayésienne intégre de manière automatique la notion de vraisemblance (cf. formule de Bayes (1.43)) avec $L(\theta;\mathbf{x}) = f(\mathbf{x}|\theta)$. Nous utiliserons souvent $f(\mathbf{x}|\theta)$ comme la fonction de vraisemblance au lieu de $L(\theta;\mathbf{x})$ pour alléger l'exposé.

1.1. EXHAUSTIVITE

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon nous apporte une certaine information sur ce paramètre. Lorsque l'on résume cet échantillon $\mathbf{x}$ par une statistique, il ne faut pas perdre cette information sur ce paramètre θ ; une statistique qui la conserve sera qualifiée d'exhaustive.

En termes mathématiques, une statistique T est dite exhaustive par rapport à θ, si la distribution conditionnelle $\mathbb{P}_t = \mathbb{P}_\theta[\cdot|T(X) = t]$ d'un n échantillon aléatoire $\mathbf{X}$, sachant $T(X) = t$, est indépendante du paramètre θ. Cette distribution conditionnelle est portée par l'ensemble $\Omega_t = \{\mathbf{x} \in \Omega;\ t = T(\mathbf{x})\}$.

Si le modèle est dominé et que la statistique T est exhaustive, la fonction de vraisemblance de l'échantillon $\mathbf{x}$ se factorise sous certaines conditions de régularité de la manière suivante :

$$f(\mathbf{x}|\theta) = h(\mathbf{x})g_\theta\big(T(\mathbf{x})\big) = h(\mathbf{x})g\big(T(\mathbf{x})|\theta\big) \tag{1.11}$$

où $h(\mathbf{x})$ est une fonction indépendante du paramètre θ. La notation $g(T(\mathbf{x})|\theta)$ indique que g dépend de θ, mais sa dépendance en $\mathbf{x}$ est seulement fonction de $\mathbf{t} = T(\mathbf{x})$. Le plus souvent $h(\mathbf{x})$ peut être choisie comme une densité de $\mathbb{P}_t$ sur Ω_t et g comme une densité de la loi de $T(\mathbf{x})$. Dans ce cas, nous disons que la distribution conditionnelle $f(x|\theta)$ admet une statistique exhaustive T.

Ainsi, toute l'information sur θ contenue dans $f(\mathbf{x}|\theta)$ se retrouve dans $g(T(\mathbf{x})|\theta)$. Nous pouvons donc dire qu'une fois $\mathbf{t} = T(\mathbf{x})$ connue, aucune valeur de l'échantillon ou d'autres statistiques ne nous apporteront des renseignements supplémentaires sur θ. Ce principe d'exhaustivité est très répandu dans l'analyse statistique.

Le principe de factorisation nous donne donc un moyen de reconnaître si une statistique est exhaustive, mais permet difficilement de la construire, ou même de savoir s'il en existe.

Pour trouver des statistiques exhaustives, il existe un résultat intéressant montré par Pitman (1936) et Koopman (1936) (cf. [2], [11] et [25]) : parmi les familles de distributions satisfaisant certaines conditions de régularité, une statistique exhaustive de dimension constante p ne peut exister que dans la famille exponentielle de la forme suivante :

$$f(x|\theta) = a(x)b(\theta)\exp\left[\sum_i c_i(x)d_i(\theta)\right] \tag{1.12}$$

où a et $(c_i)_{1\le i\le p}$ sont des fonctions de $\Omega = \mathcal{H}^n$ dans $\mathbb{R}$; b et $(d_i)_{1\le i\le p}$ sont des fonctions de Θ dans $\mathbb{R}$; et les fonctions a, b sont positives, vérifiant: $\int_\Omega f(x|\theta)dx = 1$.

Une statistique exhaustive particulière est définie par :

$$\tilde{T}(X) = \left\{ \sum_i c_1(X_i),\ \sum_i c_2(X_i),\ \ldots\ ,\ \sum_i c_p(X_i);\ n\right\} \tag{1.13}$$

ou une statistique exhaustive "normalisée" par :

$$T(X) = \left\{\frac{1}{n}\sum_i c_1(X_i), \frac{1}{n}\sum_i c_2(X_i),\ \ldots\ , \frac{1}{n}\sum_i c_p(X_i);\ n\right\} \tag{1.14}$$

Prenons un exemple : soit $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ un échantillon de taille n d'une distrubution gamma de paramètres α inconnu et β connu. Sa fonction de vraisemblance s'écrit :

$$f(\mathbf{x}|\alpha) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_i x_i\right)^{\alpha-1} e^{-\beta\Sigma x_i} = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n m_G^{n\alpha} e^{-\beta n m} m_G^{-n} \qquad (1.15)$$

où $m = \frac{1}{n}\sum x_i$ et $m_G = (\prod x_i)^{1/n}$ sont la moyenne arithmétique et la moyenne géométrique. Le théorème montre que $\{\prod x_i, n\}$ et $\{m_G, n\}$ sont des statistiques exhaustives par rapport à la variable α.

L'utilisation de la notion d'exhaustivité nous aidera à trouver une loi de distribution du paramètre θ dans l'approche bayésienne et à simplifier les procédures de recueil des données.

1.2. NOYAU D'UNE FONCTION

Un <u>noyau d'une fonction</u> $D(\theta)$ est une fonction $K(\theta)$ telle que :

$$D(\theta) = cte \cdot K(\theta) \qquad (1.21)$$

où cte est une constante par rapport à la variable θ. Désignons désormais $D(\theta) \propto K(\theta)$. Remarquons que le symbole propotionnel "$\propto$" est une relation d'équivalence, sauf si l'une des fonctions est singulière.

Lorsque le modèle considéré est dominé (cf. paragraphe 1.0.1), et s'il existe une statistique exhaustive T, la fonction de vraisemblance s'écrit :

$$f(\mathbf{x}|\theta) = h(\mathbf{x})g\big(T(\mathbf{x})|\theta\big) \propto g\big(T(\mathbf{x})|\theta\big) \qquad (1.22)$$

où l'échantillon $\mathbf{x}$ est donné a priori, et $g\big(T(\mathbf{x})|\theta\big)$ est donc un noyau de la fonction de vraisemblance (cf. formule (1.11)).

Reprenons l'exemple de la loi gamma de paramètres α inconnu et β connu. Nous avons vu que $T(\mathbf{x}) = (m_G, n)$ est une statistique exhaustive d'un échantillon $\mathbf{x}$. Il apparaît alors que le noyau de la fonction de vraisemblance est

$$g\left(T(\mathbf{x})\,|\,\alpha\right) = \left(\Gamma(\alpha)\right)^{-n}\left(\beta m_G\right)^{n\alpha} \tag{1.23}$$

Nous verrons dans la suite que le symbole proportionnel "$\propto$" joue un rôle important pour la simplification des calculs mathématiques et la commodité de la présentation, car il permet de ne pas normaliser les distributions considérées.

1.3. CONVEXITE

Rappelons quelques définitions et quelques résultats élémentaires.

Un ensemble $\Omega \subset \mathcal{H}^n$ est appelé <u>convexe</u> si toute "segment" reliant deux points quelconques x, y de Ω est inclus dans Ω, c'est à dire que, pour tout $\alpha \in [0,1]$, on a : $\forall x,\ y \in \Omega \quad \alpha x+(1-\alpha)y \in \Omega$.

Soit une fonction f(x) définie sur un ensemble convexe Ω à valeurs dans $\mathbb{R}$. La fonction f(x) est dite convexe sur cet ensemble si, quel que soit la constante $\alpha \in [0,1]$, on a : $\forall x,\ y \in \Omega \quad f[\alpha x+(1-\alpha)y] \le \alpha f(x) + (1-\alpha)f(y)$.

Si, pour tout couple de deux points non égaux $x \neq y$ et tout $\alpha \neq 0$ ou 1, l'inégalité est stricte, alors f est strictement convexe.

De manière analogue, la fonction f est dite <u>concave</u> si l'inégalité est vérifiée dans l'autre sens : $\forall x,\ y \in \Omega \quad f[\alpha x+(1-\alpha)y] \ge \alpha f(x) + (1-\alpha)f(y)$.

Si f(x) est convexe, -f(x) est concave et vice-versa. Une fonction linéaire est à la fois convexe et concave et réciproquement.

Par exemple, les fonctions x^2, $|x|$ et e^x définies sur $\mathbb{R}$ sont convexes, parmi elles, x^2 et e^x sont strictement convexes. Par contre, les fonctions $-x^2$ et logx (définie sur $(0,\infty)$) sont strictement concaves, et donc concaves.

Vérifier directement la convexité ou la concavité de la définition n'est pas toujours facile en pratique. Si la fonction f(x) est définie sur un ensemble convexe ouvert $\Omega \subset \mathcal{H}$, deux fois dérivable en tout point de Ω, on utilise le critère :

— pour que f soit convexe (ou concave), il faut et il suffit que, pour tout point $x \in \Omega$, la dérivée seconde satisfasse: $f''(x) \geq 0$ (ou $f''(x) \leq 0$).

Dans le cas où Ω est multidimensionnel, on vérifiera que la matrice carrée des dérivées partielles du second ordre est positive ou négative.

1.4. CONCEPTS BAYESIENS

Dans les problèmes pratiques, nous rencontrons des niveaux différents d'incertitudes, qui influencent les comportements des systèmes tels que les matériaux, ... etc. L'évaluation de ces incertitudes est un problème fréquemment étudié. Nous nous bornerons à constater les incertitudes au sens statistique.

L'analyse statistique suppose un ensemble des connaissances qualitatives et quantitatives, qui font apparaître en clair au moins une partie des incertitudes. Cet ensemble sera appelé l'information a priori. Dans l'étude présente, nous considérons l'incertitude portant sur la forme des modèles et pouvant être représentée par des quantités numériques inconnues.

1.4.1. Information a priori

En général, ce paramètre θ peut être évalué d'une façon indirecte dans le cadre d'un plan d'expérience ou sur des échantillons constitués. Il offre deux points de vue d'interprétation : dans l'approche classique, le paramètre θ est une valeur certaine, mais inconnue ; dans l'approche bayésienne θ apparait comme une variable aléatoire.

Le paramètre θ pourra certes, dans certains cas simples, être estimé par un échantillon de taille assez grande. Mais cette méthode est souvent impraticable pour une raison de coût. Par exemple, si l'on veux estimer la proportion d'unités défectueuses dans un contrôle de réception d'un lot de production et si le contrôle est destructif, alors il n'est pas de question de considérer un grand échantillon. De plus, dans certains cas, on ne pourra

pas avoir suffisamment d'informations pour que les résultats obtenus par les méthodes statistiques soient significatifs. Il est donc suggéré d'acquérir par ailleurs des connaissances sur ce paramètre θ, par exemple des observations antérieures, des expériences réalisables, des connaissances sur la structure et aussi des intuitions raisonnables ... etc, en supposant le tout obtenu sous les mêmes conditions fixées (ou similaires). Cet ensemble des connaissances constitue l'information a priori.

Le statisticien bayésien qui cherche à mieux connaître la valeur du paramètre θ commence par préciser l'espace Θ des valeurs possibles de θ avant d'entreprendre une quelconque observation. Ensuite, il exprime sa connaissance a priori sous forme d'une loi de probabilité Π sur Θ, puis combine la nouvelle observation $\mathbf{x}$ avec cet a priori dans l'analyse envisagée. L'utilisation de la distribution a priori sur le paramètre demeure en fait la meilleure manière d'incorporer des informations supplémentaires à un modèle statistique.

1.4.2. Espace des paramètres

Soit l'espace de probabilité $\left(\Omega,\ \mathcal{B},\ (\mathbb{P}_\theta)_{\theta\in\Theta}\right)$. La considération du fait que le paramètre θ est un élément aléatoire à valeurs dans l'espace Θ impose que Θ reçoive une structure adéquate d'espace probabilisable. Il s'agit en effet d'associer à Θ une σ-algèbre $\mathcal{A}$ formée d'événements. Dire que θ appartient à un certain événement $A \in \mathcal{A}$ (ou $A \subseteq \Theta$) est une façon d'exprimer une hypothèse.

A chaque événement A (donc à chaque hypothèse), nous pouvons attribuer une probabilité $\Pi(A)$. Nous obtenons alors un espace de probabilité des paramètres $(\Theta, \mathcal{A}, \Pi)$.

Pratiquement, l'espace Θ est de dimension finie, sous-ensemble d'un espace euclidien de dimension p. Prenons la loi gamma de paramètres α et β inconnus. On a alors $\theta = (\alpha,\beta)$ et l'espace des paramètres est inclus dans $\mathbb{R}^2$ soit

$$\Theta = \mathbb{R}^{+*}\times\mathbb{R}^{+*} = \left\{(\alpha,\beta) \in \mathbb{R}\times\mathbb{R}\ ;\ \alpha > 0,\ \beta > 0\right\} \qquad (1.40)$$

1.4.3. Formule de Bayes

Envisagons un espace de probabilité $\left(\Omega,\ \mathcal{B},\ (\mathbb{P}_\theta)_{\theta\in\Theta}\right)$. Soient A et B deux événements, donc sous-ensembles de Ω appartenant à la σ-algèbre $\mathcal{B}$: A, B $\in \mathcal{B}$. La probabilité conditionnelle $\mathbb{P}(A|B)$ est la probabilité pour que l'événement B s'étant produit, l'événement A advienne. Elle est définie par la relation suivante :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \tag{1.41}$$

pour tous A et B $\in \mathcal{B}$. Cette formule n'a de sens que si $\mathbb{P}(B) \neq 0$. Elle est équivalente à : $\mathbb{P}(A\cap B) = \mathbb{P}(B)\mathbb{P}(A|B)$, d'où la symétrie : $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$.

La formule (1.41), trouvée par Bayes dans le cas fini, s'étend au cas dénombrable. Soit $A_1, A_2, \ldots, A_n, \ldots$ une suite dénombrable d'événements disjoints tels que $\bigcup_n A_n = \Omega$ et soit B un événement observable tel que $\mathbb{P}(B) \neq 0$. On a :

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(B|A_n)\mathbb{P}(A_n)}{\sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)} \tag{1.42}$$

A l'exception de singularités, cette formule se transpose aux densités de probabilité dans le cas d'un modèle dominé.

Supposons que :

— une variable aléatoire X suive une loi de distribution admettant une densité de probabilité $f(x|\theta)$, où θ représente une valeur inconnue du paramètre;

— le paramètre inconnu θ soit lui-même considéré comme une variable aléatoire suivant une distribution a priori qui admet une densité de probabilité $\pi(\theta)$, sur l'espace de probabilité $(\Theta, \mathcal{A}, \Pi)$.

Alors la distribution de θ, après une observation **x** (soit une valeur, soit un échantillon) de la variable X, dite distribution conditionnelle de θ sachant **x**, a une densité définie par :

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_\Theta f(\mathbf{x}|\theta)\pi(\theta)d\theta} \tag{1.43}$$

ÉCOLE NATIONALE DES PONTS & CHAUSSÉES CENTRE PÉDAGOGIQUE DE DOCUMENTATION ET DE COMMUNICATION

où $p(\mathbf{x}) = \int_\Theta f(\mathbf{x}|\theta)\pi(\theta)d\theta$ est la distribution marginale de X en $\mathbf{x}$. Le terme de renormalisation $p(\mathbf{x})$ assure que : $\int_\Theta \pi(\theta|\mathbf{x})d\theta = 1$.

1.4.4. Distribution a posteriori

Nous savons que même si chaque état de la nature entraînait nécessairement un comportement défini, cela n'exclurait pas que plusieurs états puissent induire un même comportement. L'étude du problème inverse apparaît donc intéressante.

D'une manière formelle, la transition entre l'ensemble Θ des paramètres et l'ensemble Ω des observations est associée à une probabilité conditionnelle $\mathbb{P}(\mathbf{x}|\theta) = \mathbb{P}_\theta(\mathbf{x})$. Grâce au théorème de Bayes (1.43), le passage inverse de Ω vers Θ est réalisable et permet de calculer la probabilité conditionnelle $\mathbb{P}(\theta|\mathbf{x})$, qui peut être interprétée comme la probabilité d'être dans un état θ, après l'observation du résultat $\mathbf{x}$. Il faut avant tout disposer de la distribution a priori Π du paramètre θ, dont la densité est notée par π. La probabilité conditionnelle $\mathbb{P}(\theta|\mathbf{x})$ est appelée la <u>distribution a posteriori</u> et sa densité de probabilité est notée par $\pi(\theta|\mathbf{x})$.

D'un autre point de vue, la distribution a posteriori du paramètre θ représente l'actualisation de l'information a priori contenue dans la loi du paramètre $\pi(\theta)$, au vue de l'information contenue dans $\mathbf{x}$.

Cette distribution a posteriori joue alors un rôle fondamental dans l'approche bayésienne. D'après ce qui vient d'être introduit, la distribution a posteriori est une conséquence méthodologique si l'on connait π, d'où l'intérêt de déterminer cette distribution a priori.

1.4.5. Inférence statistique : l'approche décisionnelle

En pratique l'inférence statistique conduit à une décision finale prise par le "décideur", et il est important de pouvoir comparer les différentes

décisions au moyen d'un critère d'évaluation, qui va apparaître sous forme d'un coût.

Si nous notons $\mathcal{D}$ l'ensemble des décisions possibles, on appelera coût une fonction L de $\Theta\times\mathcal{D}$ dans $\mathbb{R}^+$. Une règle de décision δ est une application de Ω dans $\mathcal{D}$, et son coût moyen (ou risque) sera

$$R(\theta,\delta) = \mathbb{E}_\theta[L(\theta,\delta)] \tag{1.44}$$

Dans le cadre de l'estimation, l'ensemble des décisions est l'espace des paramètres Θ, et une règle de décision est un estimateur. Une fonction de coût couramment utilisée est la fonction de coût quadratique :

$$L(\theta,a) = \|\theta-a\|^2$$

avec $(\theta,a) \in \Theta\times\Theta$.

Dans l'approche classique, la règle de décision optimale δ doit minimiser $R(\theta,\delta)$ pour tout $\theta \in \Theta$. Mais en fait, il est difficile de trouver une telle décision.

Dans l'approche bayésienne, le paramètre θ suit une loi a priori, et suite à une observation x, sa loi a priori est réactualisée et transformée en loi a posteriori $\pi(\theta|x)$. On appelle alors fonction de coût a posteriori :

$$L(\pi,\delta|x) = \int_\Theta L\big(\theta,\delta(x)\big)\pi(\theta|x)d\theta \tag{1.45}$$

Pour chaque valeur de x, on cherche donc la décision $\delta(x)$ qui minimise ce coût et on construit ici une règle de décision.

Remarquons que le risque de Bayes, défini par

$$r(\pi,\delta) = \mathbb{E}^\pi[R(\theta,\delta)] = \int_\Theta R(\theta,\delta)\pi(\theta)d\theta \tag{1.46}$$

satisfait $$r(\pi,\delta) = \int_\Theta\int_\Omega L\big(\pi,\delta|x\big)p(x)dx$$

En effet comme $L \geq 0$, d'après le théorème de Fubini (cf. [18]), on obtient :

$$\begin{aligned} r(\pi,\delta) &= \int_\Theta R(\theta,\delta)\pi(\theta)d\theta \\ &= \int_\Theta \left(\int_\Omega L(\theta,\delta|x) f(x|\theta)dx \right) \pi(\theta)d\theta \\ &= \int_\Omega \left(\int_\Theta L(\theta,\delta|x) f(x|\theta)\pi(\theta)d\theta \right) dx \\ &= \int_\Omega \left(\int_\Theta L(\theta,\delta(x))\pi(\theta|x)d\theta \right) p(x)dx \\ &= \int_\Omega L(\theta,\delta|x)p(x)dx \end{aligned}$$

Toute solution δ^π minimisant $r(\pi,\delta)$, si elle existe, est appelée estimateur de Bayes associé au coût L. Son risque, dit risque de Bayes, est :

$$r(\pi) = r(\pi,\delta^\pi) = \underset{\delta\in\Delta}{\mathrm{Min}}\{r(\pi,\delta)\} \tag{1.47}$$

dépend du choix de π. Si l'on sait construire une règle de décision minimisant pour tout x, la fonction de coût a posteriori, alors on obtiendra un estimateur de Bayes.

Si nous considérons maintenant le coût quadratique $L(\theta,a) = \|\theta-a\|^2$, l'estimateur de Bayes δ^π du paramètre θ ($\in \mathbb{R}$), associé à une distribution a priori π, est la moyenne a posteriori :

$$\delta^\pi(x) = \mathbb{E}^\pi[\theta|x] = \int_\Theta \theta\pi(\theta|x)d\theta = \frac{\int_\Theta \theta f(x|\theta)\pi(\theta)d\theta}{\int_\Theta f(x|\theta)\pi(\theta)d\theta} \tag{1.48}$$

En effet comme

$$\mathbb{E}^\pi\left([\theta-\delta(x)]^2|x\right) = \mathbb{E}^\pi[\theta^2|x] - 2\delta(x)\mathbb{E}^\pi[\theta|x] + \delta^2(x) \quad ,$$

le minimum du coût a posteriori est effectivement atteint par

$$\delta^{\pi}(x) = \mathbb{E}^{\pi}[\theta|x] \quad .$$

Pour un vecteur des paramètres $\theta \in \Theta \subseteq \mathbb{R}^p$ avec $p > 1$, nous obtenons le même résultat associé au coût quadratique définie par $L(\theta,a) = (\theta-a)^t Q(\theta-a)$, où Q est une matrice symétrique definie positive et "^{t}" est l'opérateur transposé.

En géneral, l'estimateur bayésien dépend du choix $\pi(\theta)$ et varie avec le coût. On cherche donc un estimateur performant pour une classe des coûts car il est rare que l'on puisse déterminer exactement la fonction du coût (cf. [5], [6] et [19]). Par exemple, une classe de coûts quadratiques est définie par $L(\|\theta-\delta\|_Q)$ pour une norme $\|\cdot\|_Q$ donnée et toutes les fonctions croissantes L (cf. [22]). Il existe cependant une famille particulière de distributions telles a posteriori que l'estimateur de Bayes ne dépende pas du coût choisi; ce cont les distributions suivantes :

$$\log\pi(\theta|x) \propto \log f(x|\theta) + \log\pi(\theta) - \log p(x)$$

$$= A_1(x)e^{\alpha\theta} + A_2(x)e^{-\alpha\theta} + A_3(x)$$

(sous quelques hypothèses de régularité). Nous obtenons donc :

$$f(x|\theta) = \frac{B(x)}{\pi(\theta)}\exp\left[A_1(x)e^{\alpha\theta} + A_2(x)e^{-\alpha\theta}\right] \tag{1.49}$$

c'est la famille exponentielle (cf. formule (1.12)). La plupart des lois usuelles appartiennent à la famille exponentielle (cf. [11], [27] et [28] pour l'étude approfondie de la famille exponentielle).

Dans notre étude, la distribution $f(x|\theta)$ de la variable X est supposée l'une des distributions usuelles et nous restons donc dans le cadre de la famille exponentielle, qui possède de nombreuses propriétés intéressantes, par exemple l'exhaustivité (cf. paragraphe 1.1) : ceci permettra d'utiliser plus aisément le principe de conjuguaison, choisis pour les calculs automatiques que nous présenterons au paragraphe 1.6.

1.5. DE L'INFORMATION A PRIORI A LA LOI A PRIORI

L'importance de la distribution a priori π n'est pas tant dans le fait que le paramètre θ puisse ou ne puisse pas être conceptuellement envisagé comme une variable aléatoire, que dans le fait que π représente un moyen efficace de résumer de l'information a priori disponible sur le paramètre θ, ainsi que l'incertitude sur cette information, donc de permettre une appréciation quantitative de θ.

Avant de faire la détermination de la distribution a priori π, il faut distinguer deux sortes d'informations a priori : la première est constituée par la connaissance des résultats d'expériences aléatoires régies par une suite de variables aléatoires, suivant une distribution $f(x|\theta)$ dont θ est le paramètre, on dira que l'information est fondée sur des données antérieures. Dans le cas contraire, on dira qu'elle est subjective. Cependant, il peut arriver qu'aucune information a priori n'est disponible, dans ce cas on peut utiliser des distributions non-informatives, qui viseront à ne privilégier aucune valeur du paramètre θ (cf. paragraphe 1.5.3).

Le détermination de la distribution a priori π constitue une difficulté de mise en oeuvre de la méthodologie bayésienne : il convient de choisir une distribution a priori qui soit effectivement le reflet d'une information détenue sur le paramètre θ à estimer. Donc, on propose des recherches "automatiques" de la distribution a priori selon le type de l'information a priori telle que cette distribution ne soit pas trop éloignée de la "vraie" distribution.

1.5.1. Information subjective et Vraisemblances relatives

Il est clair que, si nous ne disposons pas de données antérieures, nous devons déterminer en partie la distribution a priori Π à l'aide d'information subjective résultant d'expériences professionnelles, d'intuitions raisonnables ...etc. Lorsque Θ est un ensemble discret fini, le problème est relativement simple en utilisant une méthode possible: un expert compare les vraisemblances intuitives des éléments de Θ de la manière suivante :

" θ_i a k fois plus de chances d'être réalisé que θ_j" .

Nous écrivons alors $\pi(\theta_i) = k \times \pi(\theta_j)$. Ensuite nous chercherons à obtenir le maximum de relations indépendantes et non-contradictoires.

Si le nombre d'éléments de Θ est n et le nombre de relations est n-1, la distribution a priori $\{\pi(\theta_i)\}$ sera déterminée en tenant compte de la relation : $\sum \pi(\theta_i) = 1$, qui assure que π est une probabilité. Cette méthode est appelée <u>méthode des vraisemblances relatives</u>.

Dans le cas où le nombre de relations est insuffisant, nous pourrons prendre une distribution a priori de θ, dite <u>distribution non-informative</u>, qui donne la même chance (ou probabilité) d'occurence à chaque élément de Θ, et que nous étudierons au paragraphe 1.5.3.

Lorsque Θ est un ensemble borné, une méthode fréquemment utilisée est de discrétiser l'espace Θ en un nombre fini de sous-ensembles auquels on applique la méthode des vraisemblances relatives, on obtient un histogramme que l'on peut au besoin interpoler. Mais, l'adéquation de l'histrogramme à des probabilités classiques, peut conduire à plusieures formes de densités, ayant éventuellement des inférences très différentes.

Si l'espace Θ n'est pas borné, le problème de la détermination des queues reste ouvert car il est difficile d'évaluer subjectivement les probabilités des régions extrêmes ; or la forme et les propriétés des estimateurs résultants en dépendent (cf. [6] et [16]).

1.5.2. Approximation et Estimation empirique

Lorsque l'information disponible ne porte pas sur le paramètre inconnu θ, mais sur des observations $x_1, x_2, \ldots, x_n$ de la variable aléatoire X d'une distribution dont la forme est supposée connue $P(x|\theta)$ (fonction de probabilité), on cherche à approcher les valeurs caractéristiques de $\Pi(\theta)$ à partir de la distribution marginale $P(x)$. En effet, une telle observation fournit une information sur θ par la relation :

$$P(x) = \sum_{i=1}^{n} \Pi(\theta_i) P(x|\theta_i) \tag{1.50}$$

A partir des connaissances de P(x), nous trouvons les valeurs $\{\Pi(\theta_i)\}$. Une telle procédure $\{(\theta_i, \pi(\theta_i)\}$ permettra de faire une approximation d'une distribution a priori quelconque.

En terme de densités, nous avons :

$$p(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta \tag{1.51}$$

Des approximations des moments et des fractiles ou quantiles de la loi de distribution p peuvent remonter à celles de la distribution a priori π. Ainsi, l'incertitude sur π peut se présenter par l'une des familles de distributions possibles suivantes :

— famille des moments : $\mathfrak{F}_M = \{\pi;\ a_i \leq \mathbb{E}^{\pi}(\theta^i) \leq b_i,\ i = 1,2,\ldots,q\}$

— famille des quantiles : $\mathfrak{F}_Q = \{\pi;\ a_i \leq \int_{I_i} \pi(\theta)d\theta \leq b_i,\ i = 1,2,\ldots,q\}$
(ou fractiles)

où $\{I_i\}$ est une partition de Θ.

L'utilisation de la distribution marginale p ou P peut se faire suivant différentes méthodes. Prenons par exemple une distance (ou une mesure) entre deux distributions, on peut chercher une distribution π en minimisant la distance entre les deux distributions :

$$p_{\pi}(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta \quad \text{et} \quad p(x)$$

celle-ci étant estimée de manière empirique.

Plus généralement, on cherche Π telle que la distance entre P et P_{Π} est minimisée, à partir de la relation :

$$P_{\Pi}(x) = \int_{\Theta} P(x|\theta)d\Pi(\theta) \tag{1.52}$$

Mais, nous n'entrons ici pas dans le détail de cette démarche pour les cas continu ou discret. De nombreux travaux abondent sur ce sujet, nous

envoyons à [5], [20], [31], [34], [38] et [43] pour les approches plus approfondies.

L'avantage de ces méthodes est que l'on peut faire une approximation d'une distribution a priori quelconque, même si l'on connait pas le forme de p ou P, puisque ces lois sont estimées de façon empirique. Ces estimations du paramètre θ et de sa distribution π sont appelées donc estimations empiriques.

Dans le cas cas continu, une autre approche souvent utilisée est de se restreindre à une famille paramétrique de densités et d'évaluer les paramètres par l'une des estimations naturelles (ou classiques), estimations des moments, estimations des fractiles (quantiles), y compris celle de médiane lorsque le paramètre ξ est unidimensionnel, estimation du maximum de vraisemblance (ou du mode, valeur qui rend maximum la quantité de vraisemblance) ... etc. La restriction de la forme entraine éventuellement une élimination de certaines distributions possibles adéquates à l'information a priori, mais cette approche a l'avantage de simplifier des calculs numériques et est appelée approche paramétrique.

Le paramètre ξ de la distribution a priori π est appelé hyperparamètre. L'estimation ponctuelle de ξ pose un problème d'imprécision pour certains bayésiens "purs". Une approche dite hierarchique propose de manière systématique un modèle aléatoire π_2 pour le hyperparamètre ξ. Le modèle devient :

$$\left\{\begin{array}{lcl} x|\theta & \approx\approx\approx & f(x|\theta) \\ \theta|\xi & \approx\approx\approx & \pi_1(\theta|\xi) \\ \xi & \approx\approx\approx & \pi_2(\xi) \end{array}\right\} \Longrightarrow \pi(\theta) = \int_{\Xi} \pi_1(\theta|\xi)\pi_2(\xi)d\xi \tag{1.53}$$

Donc, la distribution a priori π du paramètre θ est obtenue en intégrant sur l'espace Ξ des hyperparamètres puisque le paramètre ξ est inconnu. Dans la plupart des cas, cet espace Ξ est multidimensionnel, donc introduit une plus grande complexité que la distribution du modèle $f(x|\theta)$. En général, il n'y pas d'information a priori portant sur l'hyperparamètre ξ. On adopte souvent pour π_2 une distribution non-informative qui ne donne pas une préférence à des valeurs particulières (cf. paragraphe 1.5.3). L'analyse bayésienne fournie ainsi des outils pour faire face aux possibles imprécisions sur ces

lois a priori. Mais un problème important de l'approche est que parfois bien que les distributions a priori π et des observations f soient parfaitement connues, la distribution a posteriori ne peut être déterminée de manière explicite. L'interprétation de ces divers niveaux de confiance ne semble pas très évidente pour des problèmes pratiques.

Nous n'étudierons donc que l'approche paramétrique, ainsi que ses dérivées. La forme de la distribution a priori π sera déterminée selon les principes de conjuguaison naturelle que nous présenterons au paragraphe 1.6.

En un mot, dans le cas de l'information subjective, la distribution a priori reflète un degré de confiance personnelle.

1.5.3. Distribution non-informative

Lorsqu'aucune information n'est disponible, il est impossible de bâtir une distribution a priori sur des considérations subjectives. Nous supposons que la distribution de la variable aléatoire X appartienne à une famille de distributions $f(x|\theta)$. Nous chercherons une distribution du paramètre θ, qui ne "privilégie" aucune valeur du paramètre θ vis-à-vis du modèle $f(x|\theta)$ de la variable X.

On considère en premier lieu les transformations du problème, qui ne doivent pas influencer intuitivement la loi a priori non-informatif, c'est à dire que "si deux problèmes ont la même structure, alors la loi a priori non-informative doit être le même".

Le mathématicien pense naturellement à développer cette idée dans le domaine des théories d'invariance. Il s'agit de trouver une distribution invariante par certaine opération. Si la famille de distributions est stable par une classe de transformations, on pourra chercher une distribution a priori invariante par cette famille de transformations.

Prenons l'exemple d'une famille de distributions stable par les transformations affines. Si la variable X a pour densité f, la variable $Y = \sigma X+\mu$ a pour densité $\sigma^{-1}f\left(\frac{x-\mu}{\sigma}\right)$. Soit une famille avec f connue et $(\mu,\sigma) \in \mathbb{R}\times\mathbb{R}^{+*}$; la distribution normale est un exemple typique de ce genre. Remarquons que

si f est une densité de distribution réduite, les paramètres μ et σ ne sont que l'espérance mathématique et l'écart-type.

Pour la distribution a priori non-informative Π , nous pouvons chercher une mesure sur $\mathbb{R}\times\mathbb{R}^{+*}$ invariante par les transformations affines. Plus précisément, si la distribution de la variable Y est définie par les paramètres μ et σ, la distribution de bY + a sera caractérisée par les paramètres $b\mu + a$ et $b\sigma$ avec $b > 0$ et $a \in \mathbb{R}$. Il est alors naturel de chercher une distribution a priori qui soit invariante par la transformation $\Phi_{(a,b)}(\mu,\sigma) = (b\mu+a, b\sigma)$. Pour tout $A\times B \in \mathcal{B}(\mathbb{R}\times\mathbb{R}^{+*})$, nous avons :

$$\Pi(A,B) = \mathbb{P}^{\pi}\{(\mu,\sigma) \in A\times B\} = \mathbb{P}^{\pi}\{(b\mu+a, b\sigma) \in A\times B\}$$

$$= \mathbb{P}^{\pi}\left\{(\mu,\sigma) \in \frac{A-a}{b}\times\frac{B}{b}\right\} = \Pi\left(\frac{A-a}{b}\times\frac{B}{b}\right)$$

où $\frac{A-a}{b} = \left\{\frac{\mu-a}{b} \; ; \; \mu \in A\right\}$ et $\frac{B}{b} = \left\{\frac{\sigma}{b} \; ; \; \sigma \in B\right\}$.

En terme de densité, nous obtenons :

$$\int_{A\times B} \pi(\mu,\sigma)d\mu d\sigma = \int_{\frac{A-a}{b}\times\frac{B}{b}} \pi(\mu,\sigma)d\mu d\sigma = \int_{A\times B} \pi(b\mu+a, b\sigma)b^2 d\mu d\sigma$$

Comme l'égalité est valable pout tout $A\times B \in \mathcal{B}(\mathbb{R}\times\mathbb{R}^{+*})$, nous en déduisons :

$$\pi(\mu,\sigma) = b^2\pi(b\mu+a, b\sigma)$$

qui est valable sur $\mathbb{R}\times\mathbb{R}^{+*}$. Posons $\mu = 0$, $\sigma = \frac{1}{b}$ et $a = 0$, nous avons :

$$\pi(0,\frac{1}{b}) = b^2\pi(0,1) \quad .$$

d'où $\pi(0,b) = b^{-2}\pi(0,1)$. Posons $\pi(0,1) = 1$ par la convention, l'a priori non-informatif raisonnable est

$$\pi(\mu,\sigma) = \sigma^{-2} \tag{1.54}$$

Mais c'est une loi impropre, puisque $\int_0^{+\infty} \sigma^{-2}d\sigma = +\infty$.

Remarque

Toute transformation affine est caractérisée par $(a,b) \in \mathbb{R}\times\mathbb{R}^{+*}$ où a est le paramètre de translation et b le paramètre d'homothétie. L'ensemble des transformations affines étant un groupe, nous pouvons munir $\mathbb{R}\times\mathbb{R}^{+*}$ de cette structure de groupe. La mesure invariante Π ci-dessus n'est autre que la mesure de Haar à gauche sur ce groupe, ce qui est naturel car nous cherchons une distribution invariante à gauche par les transformations affines. Cette mesure de Haar à gauche est différente de la mesure de Haar à droite sur $\mathbb{R}\times\mathbb{R}^{+*}$ qui est le produit des mesures de Haar sur $\mathbb{R}$ et sur $\mathbb{R}^{+*}$ et elle a pour densité $\pi(\mu,\sigma) = \sigma^{-1}$, (cet a priori peut être aussi obtenu si les variables μ et σ sont considérées comme indépendantes).

Deux cas usuels sont présenté ci-dessous :

— si la famille de densité est invariante par translation, c'est à dire de la forme $f(x-\theta)$, alors $\pi(\theta) = 1$ (mesure de Haar sur $\mathbb{R}$) ;

— si la famille de densités est invariante par changement d'échelle, c'est à dire de la forme $\theta^{-1}f\left(\frac{x}{\theta}\right)$, alors $\pi(\theta) = \theta^{-1}$.

Cette solution fait intervenir un groupe d'invariance, mais parfois le choix n'est pas unique (cf. [27]). Jeffrey (1961) propose une approche plus globale qui évite de prendre en compte les structures d'invariance du modèle et la loi a priori non informative de Jeffrey est basée sur l'information de Fisher, donnée par

$$I(\theta) = \mathbb{E}_\theta\left(\frac{\partial \log f(x|\theta)}{\partial\theta}\right)^2 = -\,\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right) \tag{1.55}$$

dans le cas unidimensionnel. La distribution associée est

$$\pi(\theta) = I^{1/2}(\theta) \tag{1.56}$$

Elle vérifie effectivement l'invariance par reparamétrisation représentée par une transformation bijective $h : \theta \in \Theta \longrightarrow h(\theta) \in h(\Theta)$, puisque

$$I(\theta) = I(h(\theta))(h'(\theta))^2 \quad .$$

Dans le cas multidimensionnel, l'information de Fisher est généralisée par :

$$I(\theta) = - \mathbb{E}_\theta\left(\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j}\log f(x|\theta)\right]_{1\le i,\, j\le p}\right) \tag{1.57}$$

pour i et j = 1,2,...,p et la distribution de Jeffrey est

$$\pi(\theta) = \left[\det\left(I(\theta)\right)\right]^{1/2} \tag{1.58}$$

Prenons l'exemple de la distribution normale $N(\mu,\sigma^2)$ avec $\theta = (\mu,\sigma)$:

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

La distribution de Jeffey est $\pi(\mu,\sigma) \propto 1/\sigma^2$, puisque

$$I(\theta) = \mathbb{E}_\theta\begin{pmatrix} 1/\sigma^2 & 2(x-\mu)/\sigma^3 \\ 2(x-\mu)/\sigma^3 & -1/\sigma^2+3(x-\mu)^2/\sigma^4 \end{pmatrix} = \begin{pmatrix} 1/\sigma^2 & \\ & 2/\sigma^2 \end{pmatrix} \tag{1.59}$$

Par contre, si nous prenons $\theta = (\mu,\sigma^2)$ au lieu de $\theta = (\mu,\sigma)$, la distribution associée est $\pi(\mu,\sigma^2) \propto 1/\sigma^3$.

Une telle méthode arrive parfois à une distribution impropre mais cette généralisation de la distribution a priori n'est pas très gênante tant que la distribution a posteriori est définie (cf. paragraphe 2.1.3). De plus, on peut, techniquement, travailler avec des mesures non bornées (cf. [8], [10] et [47] pour ce point). Il convient alors de traiter les distributions impropres avec précaution.

Notons que cette approche non-informative ne repose pas sur une détermination de π en fonction de l'information a priori, mais l'information de Fisher dépend des facteurs de proportionalité dans la vraisemblance $f(x|\theta)$, et donc π dépend de $f(x|\theta)$.

Nous allons présenter ci-dessous une autre approche systématique qui dépend principalement du modèle considéré $f(x|\theta)$. Elle a pour objectif un calcul aisé des distributions a priori et a posteriori.

1.6. PRINCIPE DE CONJUGUAISON

En pratique, la recherche directe de la distribution a priori ne semble pas évidente. On cherche à obtenir des distributions a priori raisonnables d'une manière "objective" : l'approche étudiée maintenant n'utilise que le modèle considéré et prend en compte des considérations plus techniques que dans les approches précédentes.

1.6.1. Distributions conjuguées

En général, les densité $p(\mathbf{x})$ et $\pi(\theta|\mathbf{x})$ ne sont pas faciles à calculer. Quelquefois $\pi(\theta|\mathbf{x})$ ne peut être évaluée de manière explicite. La complexité augmente lorsque la dimension de l'espace Θ s'accroît. Il convient en effet de choisir une distribution a priori qui permette facilement d'exploiter la distribution a posteriori dès le recueil d'une nouvelle information $\mathbf{x}$ sur le paramètre θ. C'est la raison pour laquelle l'on s'intéresse aux couples de distributions "conjuguées", au sens défini ci-après.

Soit $\mathcal{F}$ une famille de distributions de densité $f(\mathbf{x}|\theta)$, indexée par θ. Une famille $\mathcal{K}$ de distributions a priori de densité $\pi(\theta)$ est dite <u>conjuguée</u> par rapport à $\mathcal{F}$, si la distribution a posteriori de densité $\pi(\theta|\mathbf{x})$ reste dans la même famille $\mathcal{K}$ pour tout $\pi \in \mathcal{K}$ et tout $f \in \mathcal{F}$. Autrement dit, la distribution a posteriori garde la même forme que la distribution a priori. Dans ce cas, il n'y a généralement pas besoin de calculer explicitement $p(x)$ parce que les noyaux de fonctions $f(\mathbf{x}|\theta)\pi(\theta)$, $\pi(\theta|\mathbf{x})$ sont les mêmes à un facteur constant près :

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta) \tag{1.60}$$

Lorsque la famille de distributions conjuguées $\mathcal{K}$ est paramétrée, le passage de la distribution a priori à la distribution a posteriori se réduit à un changement de leurs paramètres. Dans ce cas, la distribution a posteriori est toujours calculable. Une telle considération est fondée sur le principe suivant : l'information apportée par des observations $\mathbf{x}$ sur θ est limitée et dont la modification par $\mathbf{x}$ ne doit pas conduire à une remise en cause de la forme de $\pi(\theta)$, mais seulement de ses paramètres (cf. [36] pour l'origine de cette idée et [6]).

1.6.2. Conjuguée naturelle et Opération associée

En ce qui concerne la distribution a priori $\pi(\theta)$, il existe une méthode artificielle mais naturelle, pour mettre en évidence une famille de lois de distribution conjuguées par rapport à la famille $\mathcal{F} = \{f(\mathbf{x}|\theta),\ \theta \in \Theta\}$. Elle permet une certaine simplification des calculs et une cohérence des résultats lorsqu'il existe une statistique exhaustive T à valeurs dans $\mathbb{R}^p$.

Nous savons que

$$f(\mathbf{x}|\theta) \propto g\big(T(\mathbf{x})|\theta\big) \quad .$$

L'idée est de choisir une distribution a priori $\pi(\theta)$ de la forme $g(\mathbf{t}|\theta)$ pour un $\mathbf{t}$ fixé. Plus précisément

$$\pi(\theta) \propto g(\mathbf{t}|\theta) \tag{1.61}$$

La densité $\pi(\theta|\mathbf{x})$ de la distribution a posteriori satisfait :

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta) \quad .$$

Si nous demandons que cette distribution soit de la même forme que $\pi(\theta)$, à savoir dans la famille $\mathcal{K} = \{g(\mathbf{t}|\theta),\ \mathbf{t} \in \mathbb{R}^p\}$, il doit exister un $\mathbf{t}'' \in \mathbb{R}^p$ tel que

$$\pi(\theta|\mathbf{x}) \propto g(\mathbf{t}''|\theta)$$

par suite,

$$g(\mathbf{t}''|\theta) \propto g\big(T(\mathbf{x})|\theta\big)g(\mathbf{t}|\theta) \tag{1.62}$$

Ceci entraîne que la fonction g doit avoir une forme particulière puisque la famille de distributions doit être stable par une certaine multiplication et donc f doit avoir une forme particulière.

Ce principe de construction de la distribution a priori est appelée le <u>principe de la conjuguée naturelle</u>.

Au paragraphe 1.5.2, nous avons donné grâce à l'estimation bayésenne empirique une méthode pour obtenir une approximation d'une loi a priori

quelconque ; mais la distribution π obtenue satisfait rarement la propriété de conjuguaison. Dans la suite, nous nous plaçons toujours dans le cadre de la conjuguée naturelle.

Notons que cette méthode de construction de la distribution a priori sera justifiée par un autre point de vue que nous exposerons au chapitre II.

1.6.3. Opération linéaire de conjuguaison

Nous chercherons au chapitre III les distributions conjuguées à partir de la famille exponentielle de distributions continues $f(x|\theta)$, car une statistique exhaustive de dimension finie existe pour la famille exponentielle (cf. paragraphe 1.1).

Dans ce cas, si nous supposons $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ un échantillon quelconque, la distribution a priori est définie par

$$\pi(\theta) \propto b(\theta)^k \exp\left[\sum_{l=1}^{p} d_l(\theta)\left(\sum_{i=1}^{k} c_l(x_i)\right)\right] = g\left(\tilde{T}(\mathbf{x})|\theta\right) \qquad (1.63)$$

où $\tilde{T}(\mathbf{x})$ est une statistique exhaustive (cf. formule (1.13)), définie par :

$$\tilde{T}(\mathbf{x}) = \left\{ \sum_i c_1(x_i),\ \sum_i c_2(x_i),\ \ldots,\ \sum_i c_p(x_i),\ k \right\}$$

Après réalisation d'un nouvel échantillon $\mathbf{x}' = (x'_1, x'_2, \ldots, x'_{k'})$, la distribution a posteriori devient :

$$\pi(\theta|\mathbf{x}') \propto f(\mathbf{x}'|\theta)\pi(\theta) \propto g\left(\tilde{T}(\mathbf{x}')|\theta\right) \times g\left(\tilde{T}(\mathbf{x})|\theta\right)$$

Désignons par $\mathbf{x}'' = \mathbf{x} \cup \mathbf{x}' = (x_1, x_2, \ldots, x_k, x'_1, x'_2, \ldots, x'_{k'})$ l'échantillon composé de $\mathbf{x}$ et $\mathbf{x}'$, nous avons

$$g\left(\tilde{T}(\mathbf{x} \cup \mathbf{x}')|\theta\right) = b(\theta)^{k+k'} \exp\left[\sum_{l=1}^{p} d_l(\theta)\left(\sum_{i=1}^{k} c_l(x_i) + \sum_{i=1}^{k'} c_l(x')\right)\right]$$

il en résulte que

$$g\left(\tilde{T}(\mathbf{x}\cup\mathbf{x}')\,|\,\theta\right) = g\left(\tilde{T}(\mathbf{x})\,|\,\theta\right)\times g\left(\tilde{T}(\mathbf{x}')\,|\,\theta\right) = g\left(\tilde{T}(\mathbf{x})+\tilde{T}(\mathbf{x}')\,|\,\theta\right) \tag{1.64}$$

d'où l'opérateur de conjuguaison est <u>linéaire</u> :

$$\tilde{T}(\mathbf{x}\cup\mathbf{x}') = \tilde{T}(\mathbf{x}) + \tilde{T}(\mathbf{x}') \tag{1.65}$$

Pour simplifier la notation, nous utiliserons dans la suite la statistique exhaustive normalisée de $\tilde{T}(\mathbf{x})$ suivante :

$$T(\mathbf{x}) = \left\{\frac{1}{k}\sum_i c_1(x_i),\ \frac{1}{k}\sum_i c_2(x_i),\ \ldots,\ \frac{1}{k}\sum_i c_p(x_i);\ k\right\}$$

$$= \left\{\overline{c_1(\mathbf{x})},\ \overline{c_2(\mathbf{x})},\ \ldots,\ \overline{c_p(\mathbf{x})};\ k\right\} \tag{1.66}$$

L'opération de conjuguaison s'écrit alors :

$$T(\mathbf{x}\cup\mathbf{x}') = \left\{\left\{\frac{k\overline{c_l(\mathbf{x})} + k'\overline{c_l(\mathbf{x}')}}{k + k'}\right\}_{l=1,2,\ldots,p};\ k+k'\right\} \tag{1.67}$$

1.6.4. Extension de la conjuguée naturelle

La méthode de la conjuguée naturelle est très critiquée par des bayésiens car elle obéit à des contraintes techniques plutôt qu'à des impératifs d'adéquation à l'information a priori. Pour mesurer les conséquences de ce choix sur l'inférence a posteriori, de nombreux travaux ont abordés l'analyse de la sensibilité (ou de la robustesse) (cf. [5] et [7]).

Pour pallier à cette critique, certains bayésiens font appel au modèle hierarchique (cf. paragraphe 1.5.2) et considèrent une distribution a priori non-informative sur le paramètre ξ de la distribution conjuguée $\pi(\theta|\xi)$. La distribution a priori non-conditionnelle $\pi(\theta)$ est obtenue en intégrant l'hyperparamètre ξ, elle diffère en général d'une distribution conjuguée, mais elle est souvent plus robuste que la distribution conjuguée originale (cf. références citées ci-dessus).

Une autre méthode est la suivante : considérons l'exemple de la loi binomiale $B(k,\theta)$ avec $k \in \mathbb{N}^+$, dont la densité est : $f(x|\theta) = C_k^x\ \theta^x(1-\theta)^{k-x}$,

qui appartient à la famille de distributions exponentielles. Nous obtenons comme distribution conjuguée du paramètre θ une loi bêta : $\pi(\theta) \propto \theta^x(1-\theta)^{k-x}$ de paramètres $\alpha = x+1$ et $\beta = k-x+1$. Si, au travers des observations $\mathbf{x}$, on apercevoit que la distribution de θ semble bimodale, l'utilisation directe de la loi de distribution conjuguée bêta ne serait plus adaptée. Diaconia et Ylvisaker (1985) (cf. [6]) propose utiliser un mélange de deux distributions conjuguées :

$$\pi(\theta) = p\mathrm{Be}(\alpha_1,\beta_1) + (1-p)\mathrm{Be}(\alpha_2,\beta_2) \qquad \text{pour } 0 < p < 1.$$

On est donc conduit dans ce cas à utiliser des mélanges de distributions conjuguées. Ceci élargit le champ d'application sans entraîner de difficultés techniques. En effet, si $\mathcal{K}$ est une famille conjuguée, les mélanges linéaires de distributions conjuguées forment une autre famille conjuguée :

$$\tilde{\mathcal{K}} = \left\{ \sum\omega_i\pi(\theta|\xi_i) \ ; \ \sum\omega_i = 1 \text{ et } \pi \in \mathcal{K} \right\} \tag{1.68}$$

Dans le cas exponentiel, si l'on désigne par $T(\mathbf{x})$ une statistique exhaustive des observations $\mathbf{x}$, la distribution a posteriori est aussi un mélangé de distributions conjuguées :

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\sum\omega_i\pi(\theta|\xi_i)}{\int_\Theta f(\mathbf{x}|\theta)\sum\omega_i\pi(\theta|\xi_i)d\theta} = \sum\omega_i'(\mathbf{x})\pi(\theta|\xi_i+T(\mathbf{x})) \tag{1.69}$$

où le symble "+" est une opération linéaire.

Enfin, il est aussi possible d'approcher certaines distributions classiques f. qui ne possèdent pas des distributions conjuguées par des mélanges de distributions simples, principalement des familles exponentielles. On dit alors que f est un mélange caché (cf. [39] pour une étude approfondie).

Prenons l'exemple de la loi de Student non centrée $T(\theta,1,\upsilon)$. Désignons par $f(x|\theta)$ sa densité :

$$f(x|\theta) = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\upsilon\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)}\left(1 + \frac{(x-\theta)^2}{\upsilon}\right)^{-\frac{\upsilon+1}{2}}$$

où θ est paramètre de position à étudier. Alors, cette loi peut s'écrire comme un mélange d'une distribution normale par une distribution gamma :

$$X/\sigma \approx\approx\approx N(\theta,\sigma^2) \qquad \text{et} \qquad \sigma^{-2} \approx\approx\approx G(\upsilon/2,\upsilon/2)$$

avec les densités suivantes :

$$f_1(x|\theta,\sigma^2), = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$$

$$f_2(\sigma^{-2}) = \frac{\left(\frac{\upsilon}{2}\right)^{\upsilon/2}}{\Gamma\left(\frac{\upsilon}{2}\right)}\left(\sigma^{-2}\right)^{\frac{\upsilon}{2}-1}\exp\left(-\frac{\upsilon}{2}\sigma^{-2}\right)$$

Il n'est pas difficile de vérifier l'égalité (cf. paragraphe 3.1.1) :

$$f(x|\theta) = \int_{\mathbb{R}^{+*}} f_1(x|\theta,\sigma^2) f_2(\sigma^{-2}) d(\sigma^{-2}) \qquad .$$

On peut alors prendre comme distribution a priori sur θ la distribution normale $N(\mu,\tau^2)$ (cf. paragraphe 3.1.2 et [36]) et travailler conditionnellement à la variance σ^2 (inconnue).

En fait, il a été montré que les mélanges finis de distributions conjuguées de la famille exponentielle peuvent être utilisés pour approcher une loi de distribution a priori quelconque, à une ε-distance de Prohorov, du fait que les mélanges finis de masses de Dirac sont denses pour la distance de Prohorov et que l'on peut toujours approcher une masse de Dirac par des distributions conjuguées (cf. [6], [11] et [5] pour plus de détails). Une approximation par mélange continu est parfois intéressante lorsqu'il faut résoudre des problèmes de queues épaisses (cf. [37]).

Nous restons dans le cadre de la conjuguaison puisque les distributions usuelles sont adaptées à la plupart des problèmes. Pour les praticiens, la simplification des calculs est importante.

Ainsi, on dispose de deux familles conjuguées : naturelle et mélangée. Pour d'autre extensions de la famille de distributions a priori et l'analyse de la sensibilité, nous renvoyons à la référence [7].

1.7. ESTIMATION BAYESIENNE

1.7.1. Inférence non-décisionnelle

Lorsque l'on dispose de la distribution a priori π et des observations $\mathbf{x}$ de $f(\mathbf{x}|\theta)$, on peut utiliser la distribution actualisée de θ, $\pi(\theta|\mathbf{x})$, comme une distribution de probabilité au sens usuel pour décrire les propriétés de la variable aléatoire θ. Ainsi, l'estimation bayésienne du paramètre θ est fondée sur la distribution a posteriori $\pi(\theta|\mathbf{x})$ et ne dépend des observations $\mathbf{x}$ qu'au travers de la vraisemblance $f(\mathbf{x}|\theta)$. Le principe de vraisemblance est donc satisfait par l'approche bayésienne (cf. paragraphe 1.0.4).

Par exemple pour l'estimation pontuelle de θ, on définit l'estimateur du maximum de vraisemblance de θ comme le mode de $\pi(\theta|\mathbf{x})$, qui est donc le mode de $f(\mathbf{x}|\theta)\pi(\theta)$. Les propriétés de l'estimateur du maximum de vraisemblance classique (efficacité, convergence,...etc) sont conservées sous quelques hypothèses de régularité sur f et π (cf. [11]).

Egalement, on peut définir d'autres estimateurs classiques (estimateur des moments, estimateur des quantiles, ...), la conservation de propriétés asymptotiques de ces estimateurs est naturelle, car, lorsque la taille de l'échantillon tend vers l'infini, l'information apportée par cet échantillon devient prépondérante par rapport à l'information a priori. Par conséquent, les estimateurs bayésiens sont "asymptotiquement" équivalents aux estimateurs classiques. Nous verrons au chapitre III concrètement ce comportement asymptotique au travers des exemples donnés.

Si l'on cherche à estimer une fonction du paramètre $h(\theta)$, une fonction de coût classique est $L(\theta,a) = \|h(\theta) - a\|^2$ avec $a \in \mathbb{R}^k$ et $\theta \in \Theta$. La fonction de coût a posteriori (cf. formule 1.45) est

$$L(\pi,\delta|x) = \int_\Theta \big(h(\theta) - \delta(x)\big)^2 \pi(\theta|x)d\theta \tag{1.70}$$

Pour chaque valeur x, elle est minimisée pour

$$\delta^\pi(x) = \tilde{h}(\theta) = \int_\Theta h(\theta)\pi(\theta|x)d\theta = \mathbb{E}^\pi[h(\theta)|x] \tag{1.71}$$

Une estimation de Bayes est donc cette règle δ^{π} (cf. paragraphe 1.4.5). De plus, l'estimateur a priori $\delta_0^{\pi} = \mathbb{E}^{\pi}[h(\theta)]$ donnera un préavis avant échantillonnage. Lorsque la loi a priori tient compte des échantillons antérieurs, l'estimation a priori reflètera cette information a priori.

En sus d'un estimateur, le statisticien cherche à connaître une évaluation de la précision de l'estimation obtenue. Généralement dans l'approche classique, cette précision de performance est caractérisée par l'erreur quadratique, dite aussi coût quadratique (cf. paragraphe 1.4.5). Dans l'approche bayésienne, l'estimation est directe : $\delta^{\pi}(\mathbf{x}) = \mathbb{E}^{\pi}[h(\theta)|\mathbf{x}]$. De plus, l'erreur de cette estimation est égale à la variance a posteriori de cet estimateur :

$$\mathbb{V}^{\pi}[h(\theta)|\mathbf{x}] = \mathbb{E}^{\pi}[(\delta^{\pi}(\mathbf{x}) - h(\theta))^2|\mathbf{x}] \tag{1.72}$$

Il en va de même dans un cadre multidimensionnel, où la matrice de variance-covariance fournit des indications sur les performaneces d'un estimateur. L'approche bayésienne permet donc de construire de telles statistiques.

1.7.2. Test et Région de confiance

Dans l'approche inférentielle, on cherche parfois à vérifier simplement des hypothèses sur le paramètre θ, au lieu de déterminer la valeur exacte de θ (ou des fonctions de θ, $h(\theta)$). Dans l'approche classique, on distingue des problèmes d'estimation et de tests. Mais, ces deux problèmes inférentiels peuvent se ramener tous à un problème d'estimation dans le cadre décisionnel.

Par exemple, étant donné Θ_0 un sous ensemble d'intérêt de Θ, éventuellement réduit à un point, on s'intéresse à vérifier si θ, le vrai paramètre, appartient à Θ_0. L'espace de décisions est donc $\{\text{oui}, \text{non}\}$, équivalente de manière formelle à $\{0,1\}$. Le test de l'hypothèse nulle $H_0 = \{\theta \in \Theta_0\}$ (contre l'hypothèse alternative $H_1 = \{\theta \notin \Theta_0\}$) peut être envisagé comme celui de l'estimation de la fonction indicatrice :

$$I_{\{\Theta_0\}}(\theta) = \begin{cases} 1 & \text{si } \theta \in \Theta_0 \\ 0 & \text{sinon} \end{cases} \tag{1.73}$$

Plus généralement, l'hypothèse alternative contre laquelle on veut tester H_0 est définie par $H_1 = \{\theta \in \Theta_1\}$ avec $\Theta_0 \cap \Theta_1 = 0$. Le complémentaire de $\Theta_0 \cup \Theta_1$ n'a pas d'intérêt pour le problème. Dans le cas bayésien, cela revient à supposer a priori que $\pi(\theta \notin \Theta_0 \cup \Theta_1) = 0$. Pour proposer des estimateurs bayésiens associés, il suffit de disposer d'un critère de décision, c'est à dire une fonction de coût. Si l'on prend le coût "0-1" de Nyemann et Pearson sur lequel est fondée la théorie des tests classiques :

$$L(\theta,\gamma) = \begin{cases} 1 & \text{si } \gamma \neq I_{\{\Theta_0\}}(\theta) \\ 0 & \text{sinon} \end{cases} \tag{1.74}$$

la solution bayésienne est

$$\gamma^{\pi}(\mathbf{x}) = \begin{cases} 1 & \text{si } \mathbb{P}^{\pi}\{\theta\in\Theta_0|\mathbf{x}\} > \mathbb{P}^{\pi}\{\theta\in\Theta_1|\mathbf{x}\} \\ 0 & \text{sinon} \end{cases} \tag{1.75}$$

Si la solution des tests peut être envisagée comme estimation de $I_{\{\Theta_0\}}$ à valeurs dans l'intervalle [0,1], on peut considérer par exemple le coût quadratique, dans ce cas la solution bayésienne devient :

$$\gamma^{\pi}(\mathbf{x}) = \mathbb{P}^{\pi}\{\theta \in \Theta_0|\mathbf{x}\} \tag{1.76}$$

Ceci est la probabilité a posteriori du sous-ensemble Θ_0.

Pour un estimateur de θ, il est courant en statistique d'associer un intervalle de confiance (ou une région de confiance dans le cas multidimensionnel). Dans l'approche bayésienne, un intervalle de confiance est appelé un intervalle crédible, qui est par définition un sous-ensemble Θ_0 de Θ tel que .

$$1 - \alpha \leq \mathbb{P}\{\Theta_0|\mathbf{x}\} = \begin{cases} \int_{\Theta_0} \pi(\theta|\mathbf{x})d\theta \\ \sum_{\theta\in\Theta_0} \pi(\theta|\mathbf{x}) \end{cases} \tag{1.77}$$

Dans le cas de la distribution binomiale $B(k,\theta)$, si nous admettons comme loi a priroi du paramètre θ une distribution conjuguée $Be(\alpha,\beta)$, étant donné un échantillon $\mathbf{x} = (x,k)$, la distribution a posteriori de θ esi aussi

une loi betâ $Be(\alpha+x,\beta+k-x)$. Pour obtenir des intervalles de confiance, il suffit d'utiliser la table de la fonction de répartition de la loi betâ.

1.7.3. Prédiction bayésienne

La distribution a posteriori $\pi(\theta|\mathbf{x})$, ainsi que celle a priori $\pi(\theta)$, peut aussi s'appliquer à la prédiction. Si la variable observable X suit une loi de densité $f(x|\theta)$, et que la variable à étudier Y suit une loi de densité $g(y|\theta)$, alors à partir des observations de X, la distribution prédictive de Y est une moyenne sur les valeurs de θ suivant la loi de densité actualisée $\pi(\theta|\mathbf{x})$:

$$g(y|\mathbf{x}) = \int_\Theta g(y|\theta)\pi(\theta|\mathbf{x})d\theta = \frac{\int_\Theta g(y|\theta)f(\mathbf{x}|\theta)\pi(\theta)d\theta}{\int_\Theta f(\mathbf{x}|\theta)\pi(\theta)d\theta} \qquad (1.78)$$

On peut alors calculer des valeurs caractéristiques comme moyenne, variance, des fractiles associés à des probabibilités données a priori.

En particulier, X = Y et la distribution prédictive de la variable X est la distribution marginale p(x), qui décrit le comportement de X. Dans les paragraphes précédents, nous avons vu que la distribution p(x) n'était intervenue que dans la construction de la distribution a posteriori et dans la détermination de la distribution a priori (cf. paragraphe 1.5.2). Il est intéressant d'intégrer la distribution conditionnelle $f(x|\theta)$ à une distribution a priori ou a posteriori de la variable θ, de façon à obtenir des distributions composées de la variable X.

Ces distributions sont aussi appelées <u>distributions bayésiennes</u> a priori (et a posteriori) de la variable X et notées par :

$$\begin{aligned} p(x) &= \int_\Theta f(x|\theta)\pi(\theta)d\theta \\ p(x|\mathbf{x}) &= \int_\Theta f(x|\theta)\pi(\theta|\mathbf{x})d\theta \end{aligned} \qquad (1.79)$$

où $\mathbf{x}$ est un échantillon prélevé dans un lot à tester. Ce sont en somme des moyennes de toutes les distributions $f(x|\theta)$ pondérées par les probabilités des diverses valeurs de θ.

Il est important de remarquer qu'une nouvelle information $\mathbf{x}$ est incorporée dans la distribution bayésienne, c'est à dire qu'il n'est pas possible d'obtenir la nouvelle distribution $p(x|\mathbf{x})$ par une "méthode séquentielle directe". Il faut procéder indirectement : recalculer la loi de distribution a priori $\pi(\theta)$ à partir de toutes les informations avant la réception de $\mathbf{x}$, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, puis la distribution a posteriori $\pi(\theta|\mathbf{x})$ après la réception de $\mathbf{x} = \mathbf{x}_{n+1}$ et enfin la nouvelle distribution bayésienne.

La distribution prédictive a priori $p(x)$ nous fournit un préavis sur le niveau de la qualité avant d'effectuer l'échantillonnage, et la comparaison entre deux distributions prédictives peut permettre de connaître l'influence de l'échantillon $\mathbf{x}$ du lot à examiner.

Il est clair que les distributions marginales $p(x)$ et $p(x|\mathbf{x})$ effacent totalement le paramètre d'incertitude θ. Plus nous disposons de données, plus $p(x)$ se rapprochera de la vraie distribution de X, celle du paramètre θ se concentrant autour de la (les) vraie(s) valeur(s) de θ.

Dans l'approche classique, la distribution de la variable X n'est que la fonction de densité $f(x|\theta)$ avec θ prenant une valeur certaine, mais inconnue (elle est déterminée souvent par des valeurs observées de X). Cela revient à dire que la probabilité a priori du paramètre θ se réduit à cette valeur $\theta_\circ$, c'est à dire la distribution de Dirac : $\delta_{\theta_\circ}(\theta) = 1_{\{\theta=\theta_\circ\}}(\theta)$.

1.8. PROCESSUS DE CONTROLE

La présente étude a pour objectif principal d'évaluer la distribution prédictive de la variable X, sur laquelle est basée l'estimation des valeurs caractéristiques comme la moyenne M, l'écart-type S, un fractile V_p associé à une probabilité p donnée, parce que le critère d'acceptation du contrôle statistique est souvent lié à la queue de la distribution de X, et il dépend donc du couple (V_p, p). Nous présentons ici quelques notions concernant le

contrôle statistique.

Le contrôle statistique se divise en deux catégories selon que le caractère examiné est quantitatif ou qualitatif, le premier est appelé contrôle par mesure et le deuxième contrôle par attribut (pièce bonne ou mauvaise).

Le contrôle de matériaux dans le domaine génie civil est effectué essentiellement par mesure. La décision du contrôle est déterminée principalement par deux valeurs de critère : une valeur caractéristique V_P et un pourcentage p (ou une probabilité) de défectueux.

Nous nous plaçons dans le cas d'un seuil inférieur de spécification. D'autre cas se traitent de façon analogue. Par exemple, dans le cas où le contrôle exige un intervalle de tolérance symétrique autour de la moyenne m on cherche à déterminer une valeur λ telle que la probabilité de tomber au dehors de l'intervalle $(m - \lambda s,\ m + \lambda s)$ est inférieure à p où s est l'écart-type. On peut également chercher à déterminer la fractile V_P correspondant à la probabilité p/2, donc nous avons $V_P = m - \lambda s$, d'où $\lambda = (V_P - m)/s$ dans le cas symétrique.

Rappelons qu'un fractile d'une distribution est la valeur telle que sa fonction de répartition est égale à la probabilité donnée :

$$F(V_P) = \mathbb{P}(x < V_P) = p$$

ou $\mathbb{P}(x > V_P) = 1 - F(V_P) = 1 - p$ dans le cas symétrique.

Il existe deux façons de formuler la règle de critère : une règle la plus utilisée est appelée contrôle par pourcentage, et se formule ainsi: "un lot est caractérisé par le pourcentage p des individus prélevés et dont la caractéristique mesurée est inférieure à une valeur carcatéristique garantie V_G".

Avec la deuxième formulation, on se donne la proportion p_G et un lot est caractérisé par la plus petite valeur $V = V_{P_G}$ telle qu'une proportion p_G des individus prélevés dans ce lot présente des caractéristiques inférieures à V. Il s'agit d'un contrôle de quantile. Remarquons que dans la première

formulation, c'est V_G qu l'on se donne et c'est la proportion p qui caractérise le lot.

Dans le modèle probabiliste, ces deux formulations reviennent à la même chose. Le caratère X est considéré comme une variable aléatoire suivant une loi de distribution. Il s'agit en effet de trouver le couple (V_p, p) de cette loi de distribution, leur relation étant la fractile V_p d'ordre p ou la probabilité p au fractile V_p. La courbe (V_p, p) est appelée courbe d'efficacité dans le cas classique et courbe d'acceptation dans le cas bayésien.

Dans le cas où l'on procède au contrôle statistique sur un échantillon, le résultat du caractère V estimé par un échantillon est entâché de probabilités d'erreurs, dites risques statistiques, dû au fait d'une informamation incomplète. On introduit donc α la probabilité de déclarer qu'un lot est non conforme alors qu'en réalité il est conforme, c'est une protection pour le producteur ou le fournisseur. Quant au client, il voudrait savoir la probabilité de déclarer un lot conforme alors qu'en réalité il ne l'est pas. Les deux risques sont appelés respectivement risques de premier espèce α et de second espèce β.

L'analyse bayésienne cherche à évaluer par approximation la vraie distribution du caractère X. Nous avons vu au paragraphe précédent que cette distribution est une moyenne sur les valeurs possibles du paramètre θ selon la densité $\pi(\theta)$. Cette loi résume la connaissance sur θ pendant toute la période de tests, l'actualisation de cette connaissance est exprimée par la distribution a posteriori $\pi(\theta|\mathbf{x})$.

Surtout quand il s'agit d'un contrôle d'une longue série de lots présentés par le même producteur, les techniques bayésiennes pourront assurer des lots de bonne qualité en étudiant l'évolution du niveau de la qualité dans le temps, en traitant de façon chronologique les productions ; et la protection du client est ainsi assurée indirectement moyennant la courbe d'acceptation.

Si nous arrivons à définir des critères de changements et tolérances du niveau de la qualité, nous pouvons établir un processus optimal du contrôle dans le temps. Il permet de définir la fréquence d'échantillonnage, pour réduire autant que possible le volume d'essais sans perte de la fiabilité.

Dans le contexte "économique", nous pourrons éventuellement étendre la notion de coût de la qualité, celle-ci définie non par une valeur V avec un coût "0-1" (décision binaire), mais de manière à accorder différents poids α_j aux différentes valeurs V^j ou aux différents intervalles disjoints I^j autour de ces valeurs, pour en favoriser certains. Dans le modèle probabiliste, un caractère contrôlé a pour probabilité p_j de tomber dans chacun des intervalles. Ainsi, la qualité peut être quantifiée par :

$$\sum \alpha_j p^j = \sum \alpha_j \mathbb{P}\{x \in I^j\}$$

Le contrôle obtenu sera optimal au coût choisi. Cette idée est facilement généralisée en plusieurs dimensions lorsque la qualité dépend de plusieurs caractères à contrôler.

En un mot, tout processus du contrôle dépend de l'évaluation de la distribution prédictive de la variable X dans le lot.

1.9. EXEMPLE

Nous illustrerons les notions qui viennent d'être introduites par l'exemple d'un projet de digue anti-tempête.

Supposons que pour protéger un petit port, on envisage deux variantes de digue anti-tempête. La variante a_1 assure une certaine protection, mais limite l'ouverture. La seconde a_2 restreint moins le passage, mais ne pourra empêcher l'entrée de vagues dangereuses dans le cas d'une tempête sévère venant de l'est.

La performance du projet dépend en partie du nombre de telles tempêtes se produisant pendant la durée du service, soit 50 ans par exemple. Ce qui nous intéresse est d'essayer de prévoir le nombre de tempêtes à venir en 50 ans, nombre noté par X.

De telles tempêtes sont considérées comme des événements poissonniens où le nombre moyen μ est le paramètre inconnu, et le nombre X suit une distribution de Poisson :

$$\mathbb{P}\{X = x\} = f(x|\mu) = e^{-\mu}\frac{\mu^x}{x!} \qquad x = 0,1,2,\ldots \qquad (1.90)$$

Cette loi est exponentielle, puisque nous avons

$$f(x|\mu) = \frac{1}{x!}\exp\left(x\log\mu - \mu\right) \qquad (1.91)$$

Nous pouvons ici raisonnner de deux manières différentes selon le type d'information a priori.

1.9.1. Vraisemblances relatives

Nous supposons que l'ingénieur ne dispose malheureusement pas d'annales des tempêtes passées. Il doit donc fonder son estimation du nombre moyen sur toute information connexe et sur le jugement professionnel. Son consultant météorologue, à partir d'analyses météorologiques et de données régionales, suggère que le nombre moyen μ doit être entre 0.25 et 1, les valeurs basses étant un peu plus probables que les valeurs élevées. Il a réussi à établir les 3 relations suivantes pour la vraisemblance relative :

— 0.25 a trois fois plus de chance d'être la valeur vraie de μ que 1,

— 0.50 a deux fois plus de chance d'être la valeur vraie de μ que 0.75,

— 0.25 a 1.5 fois plus de chance d'être la valeur vraie de μ que 0.75.

Si l'on fixe $\Pi(1) = 1$, on obtient d'après les renseignements précédents $\Pi(0.25) = 3$, $\Pi(0.50) = 4$, $\Pi(0.75) = 2$. La fonction Π est donc une distribution de la moyenne μ sur $\Theta = \{0.25, 0.50, 0.75, 1.00\}$. L'ingénieur décide alors d'adopter pour $\Pi(\mu)$ la distribution ainsi définie.

μ	0.25	0.50	0.75	1.00
Proba	0.3	0.4	0.2	0.1

Tableau 1.1

Il est important de remarquer que plus on considère des valeurs entre

0.25 et 1, plus de renseignements seront exigés.

Si nous voulons une distribution continue sur un intervalle par exemple [0, 1.25], nous pourrons relier les valeurs successives par des segments, et puis éventuellement approximer cette courbe par la densité continue d'une distribution usuelle.

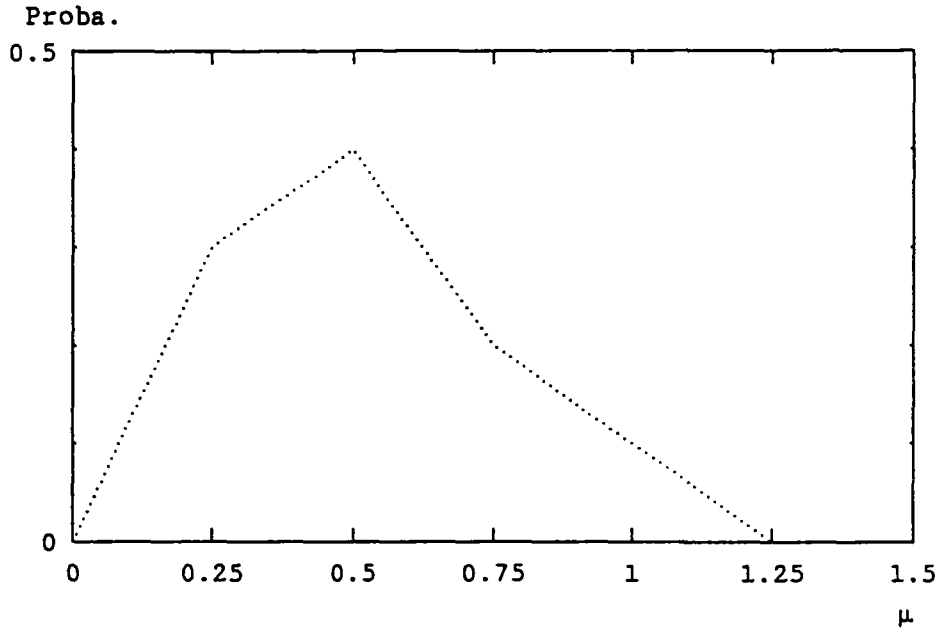


Figure 1.1 Probabilité a priori de la moyenne de tempêtes

La variable du nombre X de tempêtes dangereuses pour les 50 ans à venir a donc pour distribution discrète sur $\mathbb{R}^+$:

$$p(x) = \sum_{i=1}^{4} f(x|\mu_i)\pi(\mu_i)$$

$$= \frac{1}{10x!}\left[3\times 0.25^x e^{-0.25} + 4\times 0.5^x e^{-0.50} + 2\times 0.75^x e^{-0.75} + e^{-1}\right] \qquad (1.92)$$

d'où la moyenne 0.5250, puisque l'on a :

$$\mathbb{E}(X) = \mathbb{E}^P(X) = \sum_{x=0}^{\infty} xp(x)$$

$$= \sum_{x=0}^{\infty} x\left(\sum_{i=1}^{4} f(x|\mu_i)\pi(\mu_i)\right)$$

$$= \sum_{i=1}^{4} \left(\sum_{x=0}^{\infty} x f(x|\mu_i) \right) \pi(\mu_i)$$

$$= \sum_{i=1}^{4} \mu_i \pi(\mu_i)$$

$$= \mathbb{E}^{\pi}(\mu) = 0.5250 \tag{1.93}$$

Alors il y a 60.75% de chance de ne pas avoir de tempête dangereuse dans les 50 ans et 28.74% d'avoir une telle tempête, c'est à dire

$$\mathbb{P}\{X = 0\} = p(0) = 0.6075$$

et

$$\mathbb{P}\{X = 1\} = p(1) = 0.2874 \ .$$

Leur somme $\mathbb{P}\{X \leq 1\} = p(0) + p(1) = 0.8949$ est la probabilité d'avoir au plus une telle tempête.

1.9.2. Expériences antérieures

L'ingénieur se livre à une recherche de documentation d'archives. Il constate que l'on n'a pas relevé dans les 100 dernières années de tempête critique; auparavant on ne tenait pas d'archives météorologiques. Cependant, l'examen des journaux et documents historiques montre que dans les 225 dernières annnées, une seule tempête critique s'est produite, il y a exactement 150 ans. Si nous notons X le nombre de tempêtes dangereuses pour la période de 50 ans, on obtient l'échantillon : $x_1 = x_2 = 0$, $x_3 = 1$, $x_4 = x_5 = 0$.

Rappelons que les conséquences de sa décision dépendent de x_6 (le nombre de tempêtes critiques dans les 50 à venir).

Désignons par $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ l'échantillon composé (ou total). La méthode de la conjuquée naturelle permet d'établir une distribution a priori du nombre moyen μ, qui est conjuguée de la fonction de vraisemblance de $\mathbf{x}$:

$$\pi(\mu) \propto f(x|\mu) = \prod_{i=1}^{5} e^{-\mu} \frac{\mu^{x_i}}{x_i!} = e^{-5\mu} \frac{\mu^{\Sigma x_i}}{\Pi x_i!} = \mu e^{-5\mu} \tag{1.94}$$

La distribution a priori est alors une loi gamma G(2,5). Il en déduit que la probabilité de tomber dans l'un des intervalles de même longueur 0.25 autour des valeurs 0.25, 0.50, 0.75, 1.00, est figurée dans le tableau 1.2.

$\mu \in$	[0,0.125)	$I_{0.25}$	$I_{0.50}$	$I_{0.75}$	$I_{1.00}$	[1.125,∞)
proba	0.130	0.429	0.260	0.113	0.044	0.024

Tableau 1.2

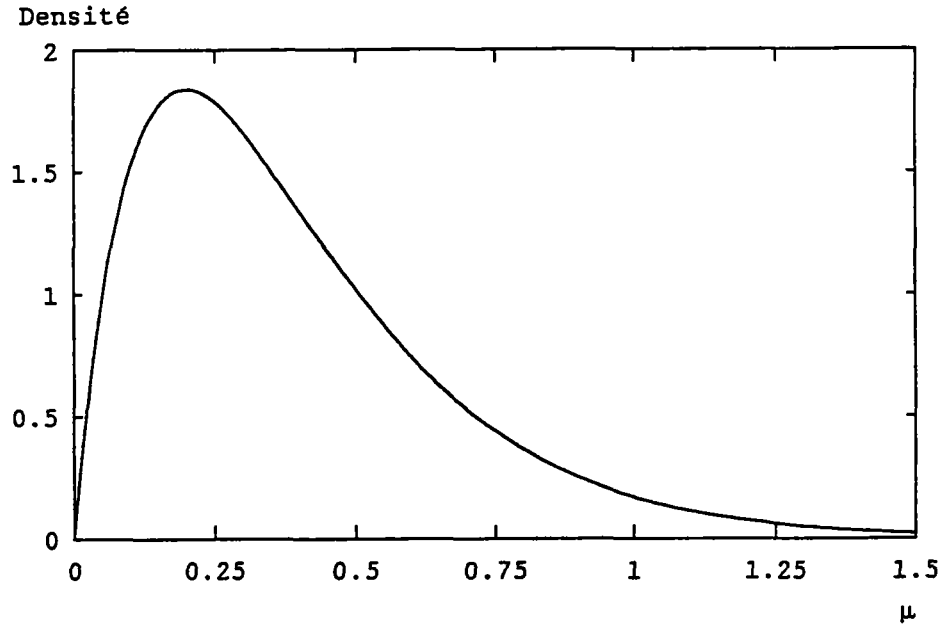


Figure 1.2 Densité de probabilité de la moyenne de tempêtes

La distribution prédictive de X est une moyenne de toutes les lois de distribution possibles $f(x|\mu)$, pondérée par π la distribution de μ; elle est aussi une distribution discrète, x = 1,2,3,... :

$$p(x) = \int_{\mathbb{R}^+} f(x|\mu)\pi(\mu)d\mu$$

$$= \frac{5^2}{x!}\times\frac{\Gamma(x+2)}{6^{x+2}} = \left(\frac{5}{6}\right)^2(x+1)6^{-x} \qquad (1.95)$$

La moyenne prédictive est $\mathbb{E}(X) = 0.40$. Nous en concluons alors qu'il y a 69.44% de chance de ne pas avoir de tempêtes critiques dans les 50 ans à venir, et 23.15% qu'il y ait exactement une tempête, c'est à dire

$$\mathbb{P}\{X = 0\} = p(0) = 0.6944$$

et

$$\mathbb{P}\{X = 1\} = p(1) = 0.2315 .$$

Leur somme $\mathbb{P}\{X \leq 1\} = p(0) + p(1) = 0.9259$ est la probabilité d'avoir au plus une telle tempête.

1.9.3. Comparaison et Représentation graphique

Malgré la différence des types de l'information a priori pris dans les deux cas, les deux distributions a priori de la moyenne μ sont très voisines lorsque nous comparons leurs fonctions de répartition :

$\mu_\circ$		0.25	0.50	0.75	1.00	1.25
$\mathbb{P}^\pi\{\mu < \mu_\circ\}$	cas 1	0.3	0.7	0.9	1.0	1.0
	cas 2	0.355	0.713	0.888	0.968	0.986

Tableau 1.3

Dans le deuxième cas, la fonction de répartition de la distribution a priori gamma est donnée par

$$\mathbb{P}^\pi\{\mu \leq \mu_\circ\} = \int_0^{\mu_\circ} \mu\pi(\mu)d\mu = 1 - e^{-5\mu_\circ}(5\mu_\circ + 1) \qquad (1.96)$$

Il est clair que dans le deuxième cas, la probabilité d'avoir au plus une tempête critique est légèrement plus grande que celle dans le premier cas. Nous illustrons respectivement la distribution et la densité de probabilité a priori de la moyenne des tempêtes dans les deux cas, également les probabilités d'avoir x tempêtes et d'avoir au plus x tempêtes.

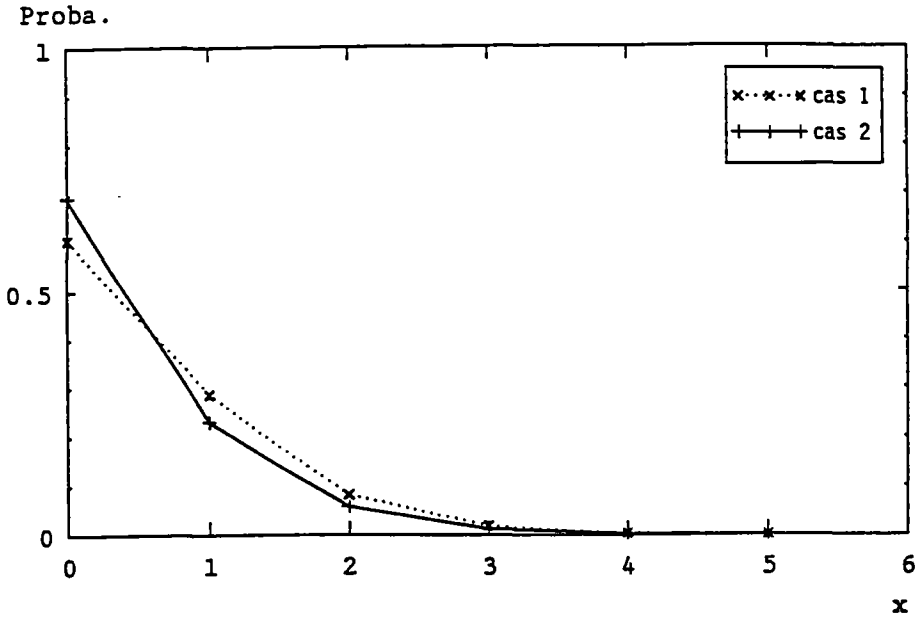


Figure 1.3 Probabilité d'avoir x tempêtes

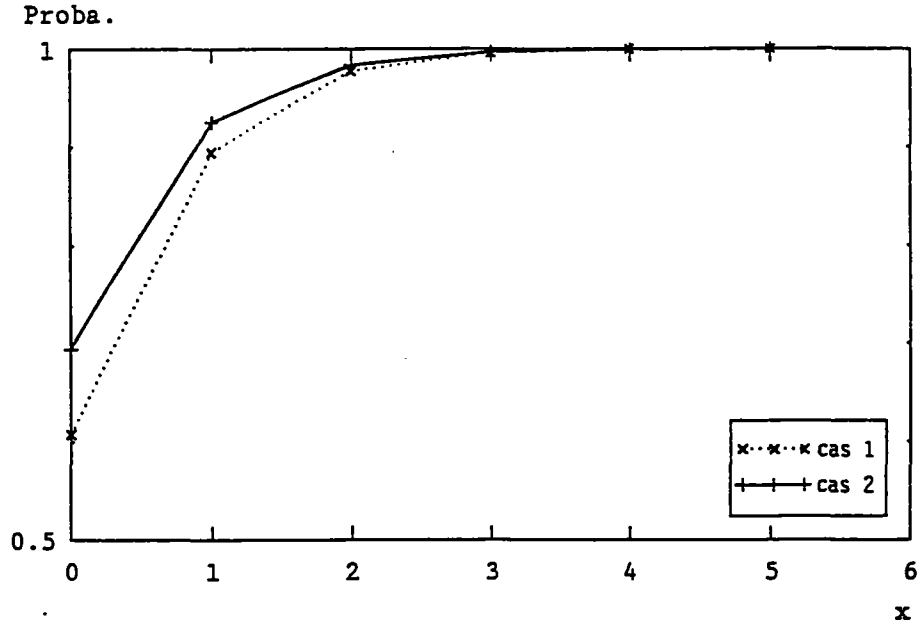


Figure 1.4 Probabilité d'avoir au plus x tempêtes

1.10. RESUME ET SCHEMA DU PROCESSUS

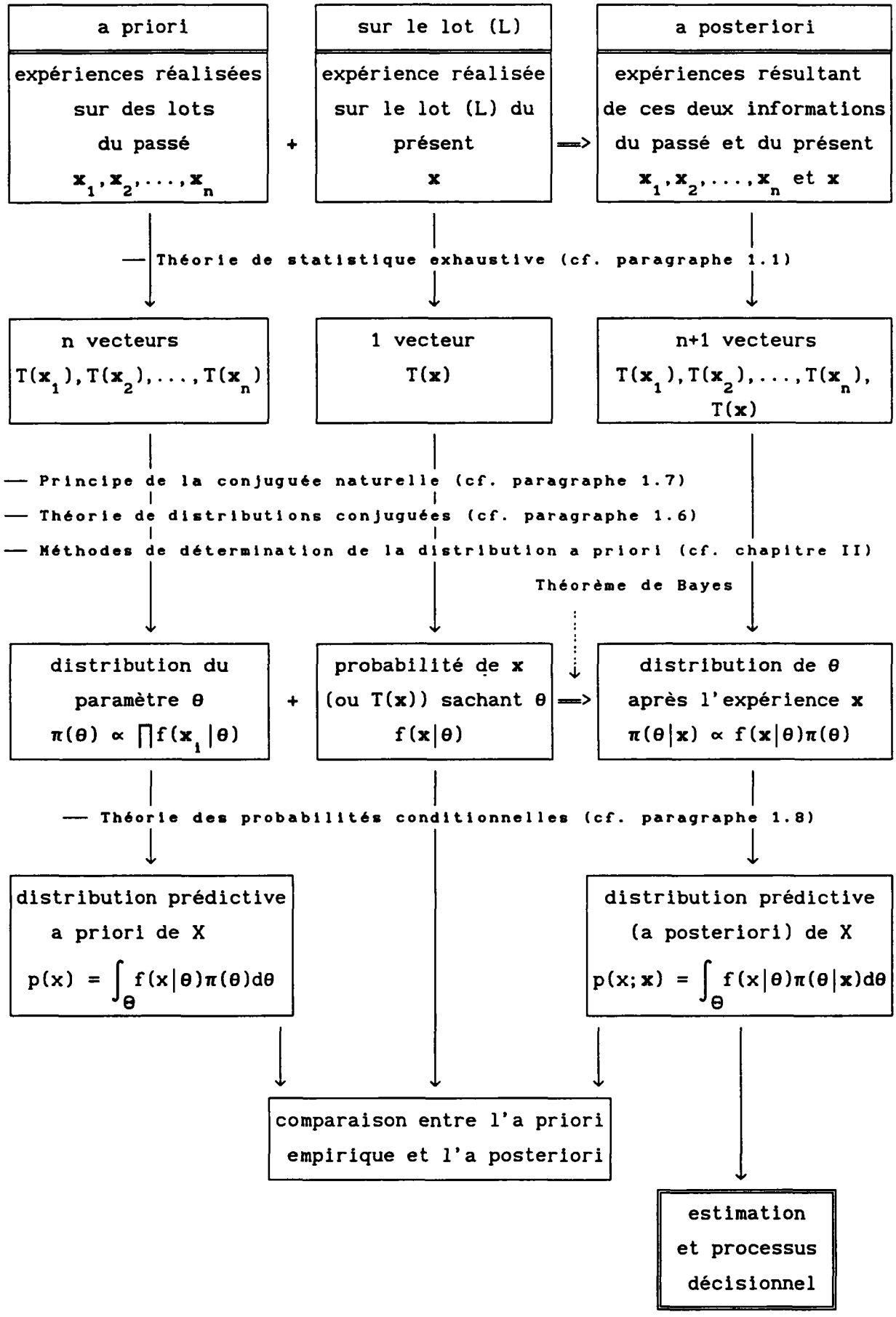
Pour éclaircir la terminologie employée dans ce rapport, on utilise les conventions suivantes dans les cas des modèles dominés :

— $f(x|\theta)$ représentera soit la densité de la distribution conditionnelle de la variable X, soit la distribution d'un échantillon de X ; nous supposerons souvent que la forme de la distribution est connue et appartient à une famille exponentielle de distributions normale, log-normale, gamma, ... ;

— $\pi_0(\theta) = \pi(\theta)$ et $\pi_1(\theta) = \pi(\theta|\mathbf{x})$ désigneront les densités des distributions a priori et a posteriori du paramètre θ, paramètre de la distribution de X;

— $p_0(x) = p(x)$ et $p_1(x) = p(x|\mathbf{x})$ désigneront les densités des distributions prédictives a priori et a posteriori de la variable X; ce sont ces distributions bayésiennes que nous chercherons à déterminer, afin d'établir des processus décisionnels tels qu'un contrôle de fiabilité ou de qualité ... etc.

Supposons ici que l'information a priori est fondée sur des expériences (ou des échantillons) antérieures $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, et que l'information sur un lot à tester est une expérience $\mathbf{x}$. Considérons que la distribution de la variable X à étudier $f(x|\theta)$ possède une statistique exhaustive T. La représentation schématique permet de visualiser l'ensemble de la méthodologie bayésienne appliquée à l'estimation d'un paramètre caractéristique X et aussi de résumer les différentes phases alliées aux outils mathématiques.


a priori
expériences réalisées sur des lots du passé
$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$
+
sur le lot (L)
expérience réalisée sur le lot (L) du présent
$\mathbf{x}$
==>
a posteriori
expériences résultant de ces deux informations du passé et du présent
$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ et $\mathbf{x}$
— Théorie de statistique exhaustive (cf. paragraphe 1.1)
n vecteurs
$T(\mathbf{x}_1), T(\mathbf{x}_2), \ldots, T(\mathbf{x}_n)$
1 vecteur
$T(\mathbf{x})$
n+1 vecteurs
$T(\mathbf{x}_1), T(\mathbf{x}_2), \ldots, T(\mathbf{x}_n), T(\mathbf{x})$
— Principe de la conjuguée naturelle (cf. paragraphe 1.7)
— Théorie de distributions conjuguées (cf. paragraphe 1.6)
— Méthodes de détermination de la distribution a priori (cf. chapitre II)
Théorème de Bayes
distribution du paramètre θ
$\pi(\theta) \propto \prod f(\mathbf{x}_i|\theta)$
+
probabilité de $\mathbf{x}$ (ou $T(\mathbf{x})$) sachant θ
$f(\mathbf{x}|\theta)$
==>
distribution de θ après l'expérience $\mathbf{x}$
$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta)$
— Théorie des probabilités conditionnelles (cf. paragraphe 1.8)
distribution prédictive a priori de X
$p(x) = \int_\Theta f(x|\theta)\pi(\theta)d\theta$
distribution prédictive (a posteriori) de X
$p(x;\mathbf{x}) = \int_\Theta f(x|\theta)\pi(\theta|\mathbf{x})d\theta$
comparaison entre l'a priori empirique et l'a posteriori
estimation et processus décisionnel

CHAPITRE II DÉTERMINATION DE LA LOI A PRIORI

2.0. PRESENTATION GENERALE

Considérons un modèle dominé dans lequel la loi de densité sous-jacente $f(x|\theta)$ est paramétrée par θ et un échantillon $\mathbf{x}$ de la variable aléatoire X ayant cette loi de distribution.

Nous avons vu que l'interprétation de l'information a priori sur le paramètre θ a une importance primordiale dans l'approche bayésienne. Il existe plusieurs méthodes pour déterminer la distribution a priori du paramètre θ selon le type d'information disponible. Nous allons étudier et développer quelques unes d'entre elles qui peuvent être utilisées avec profit lorsque nous nous limitons au cas où l'information résulte d'observations passées du processus aléatoire. L'évaluation de la distribution a priori est réalisée à l'aide des idées classiques et les analyses associées sont plutôt à la frontière entre statistiques bayésienne et classique; ainsi dans ce domaine, l'étude est peu développée.

Nous acceptons par la suite l'hypothèse que l'on connaît avec certitude la forme analytique de la densité $f(x|\theta)$, mais que nous ne connaissons pas avec précision son paramètre θ. Nous disposons d'une série de résultats d'expériences, identiquement distribués suivant une loi paramétrée par θ.

La première méthode présentée est la méthode séquentielle avec au départ soit une mesure uniforme sur l'ensemble Θ, soit une distribution non-informative de la variable aléatoire θ (cf. paragraphe 1.5.3). Nous allons montrer qu'elle permet de retrouver la même forme de distribution et le même paramètre que celle de la conjuguée naturelle avec la fonction de vraisemblance (cf. paragraphe 1.6.2).

Malgré des restrictions sur la famille de distributions conjuguées (cf. paragraphe 1.6), la propriété de conjuguaison permettra d'avoir une

stabilité de la forme des distributions; ainsi le passage de la distribution a priori à l'a posteriori se réduit à un changement de paramètres, afin de réaliser facilement les calculs.

Si nous connaissons la forme analytique de la distribution a priori, son paramètre sera déterminé par l'une des familles de méthodes suivantes :

— méthodes des familles paramétriques
- méthode des moments
- méthode des fractiles
- méthode du maximum de vraisemblance
- ...

— méthodes "du quasi-échantillon"
- méthode des doubles moments
- méthode des doubles maxima de vraisemblance
- ...

Dans cette étude, les méthodes des familles paramétriques sont toutes basées toutes sur la relation entre les distributions a priori et marginales $p(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$; nous les appliquerons au cas où les observations regroupées sont supposées reparties suivant la loi de distribution p.

Dans les méthodes du quasi-échantillon, nous commencerons par traduire chacun des échantillons en une valeur du paramètre inconnu θ ; nous obtenons ainsi un quasi-échantillon de θ :

observations du passé $\quad x_1, x_2, \ldots, x_n$

valeurs évaluées $\quad \theta_1, \theta_2, \ldots, \theta_n$

A partir de cet échantillon, la distribution a priori π sera déterminée par l'une des méthodes classiques.

Nous présenterons aussi la méthode des scores d'expert, qui permettra de traiter des problèmes où les échantillons ne sont pas affectés d'un poids identique dans l'exploitation de l'information a priori. Dans ce cas, l'information a priori sera constituée par l'ensemble des échantillons avec leurs scores (ou poids) accordés $\{(x_i, \alpha_i)\}$. Nous allons utiliser l'idée des scores d'expert pour définir la fonction de vraisemblance au sens bayésien. Nous choisirons la distribution a priori conjuguée de cette nouvelle fonction. A partir de l'ensemble $\{(x_i, \alpha_i)\}$, nous pourrons également généraliser d'autres méthodes citées ci-dessus avec cette idéee des scores d'expert.

Ainsi, leurs combinaisons donnent des outils pour résoudre les cas où les poids des échantillons sont soit identiques soit différents.

En nous inspirant de la méthode séquentielle et de la méthode des scores d'expert, nous proposerons une méthode, appelée méthode des échantillons équilibrés. En fait, les poids pris en compte dans l'estimation de la distribution a priori par ces échantillons sont différents, lorsque leurs tailles ne sont pas identiques, même si l'on n'a pas l'intention de distinguer ces échantillons comme dans la méthode séquentielle. La méthode des échantillons équilibrés, comme son nom l'indique, a pour but d'équilibrer réellement les poids dans l'exploitation de la distribution a priori.

Une fois la distribution a priori déterminée, nous examinerons successivement (au chapitre III) :

— la distribution a posteriori de θ,

— la distribution prédictive de la variable aléatoire X,

et puis nous passerons éventuellement à l'étape de l'analyse décisionnelle.

2.1. METHODE SEQUENTIELLE

2.1.1. Notion de séquentialité

Soit $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ une série d'échantillons successifs, de tailles k_i, relatifs à des d'expériences aléatoires identiques.

Soit $\pi(\theta)$ une distribution a priori. L'hypothèse séquentielle consiste à considérer la distribution a posteriori $\pi(\theta|\mathbf{x}_i)$ comme une nouvelle distribution a priori pour l'échantillon suivant $\mathbf{x}_{i+1}$. La distribution de θ est alors réajustée lors de la réalisation de $\mathbf{x}_{i+1}$ par la formule :

$$\pi(\theta|\mathbf{x}_{i+1}) \propto f(\mathbf{x}_{i+1}|\theta)\pi(\theta|\mathbf{x}_i) \tag{2.10}$$

Partant du premier résultat $\mathbf{x}_1$, nous calculons les distributions de θ l'une après l'autre et nous avons finalement :

$$\pi(\theta|x_1,x_2,\ldots,x_n) = \frac{f(x_1|\theta)f(x_2|\theta)\cdots f(x_i|\theta)\pi(\theta)}{\int_\Theta f(x_1|\theta)f(x_2|\theta)\cdots f(x_i|\theta)\pi(\theta)d\theta}$$

$$= \frac{\prod_j f(x_j|\theta)\pi(\theta)}{\int_\Theta \prod_j f(x_j|\theta)\pi(\theta)d\theta} \qquad (2.11)$$

Remarquons que, s'il n'existe pas de stabilité de la forme de la suite de distributions $\pi_i(\theta) = \pi(\theta|x_1,x_2,\ldots,x_i)$, $i = 1,2,\ldots,n$, la formule n'est guère exploitable de manière analytique sauf s'il existe une statistique exhaustive.

La loi de distribution $\pi_n(\theta)$ est maintenant la distribution a priori par rapport à l'échantillon suivant et nous notons $\pi(\theta) = \pi_n(\theta)$.

Soit x un nouvel échantillon , la distribution a posteriori $\pi(\theta|x)$ peut être considérée comme la (n+1)-ième distribution de la séquence :

$$\pi(\theta|x) = \pi_{n+1}(\theta) = \pi_n(\theta|x) = \pi(\theta|x_1,x_2,\ldots,x_n,x) \qquad (2.12)$$

Nous observons que cette méthode incorpore facilement les données consécutives et que quelle que soit la forme de la distribution a priori, si le nombre n des échantillons est grand, la modification de la distribution a posteriori est très faible. Nous somme donc amenés, pour tenir compte des connaissances récentes, à limiter le nombre n, c'est à dire à ne pas trop cumuler d'informations disponibles.

2.1.2. Mesure uniforme du paramètre θ

Pour utiliser la formule séquentielle, il nous manque une distribution a priori comme point de départ.

Supposons maintenant que nous ne possédions pour tout renseignement que les échantillons antérieurs. Avant le premier échantillon, l'information a priori se traduira par le fait que nous ne voudrons privilégier aucune valeur du paramètre inconnu θ. Dans ce cas, il est naturelle que nous prenions

l'a priori π comme une mesure uniforme sur l'espace mesurable $(\Theta, \mathcal{A})$.

Malheureusement, lorsque Θ est un ensemble de mesure non finie, la mesure uniforme est une distribution impropre, puisque la probabilité totale (ou la masse) est infinie : $m(\Theta) = \int_{\Theta} \pi(\theta)d\theta = +\infty$.

En cas unidimensionnel, et sans perdre de généralité, nous prendrons : $\Theta = (a,b)$ avec a et b $\in \mathbb{R} \cup \{-\infty , +\infty\}$; la formule séquentielle devient :

$$\pi(\theta|\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_i) = \lim_{\substack{x \longrightarrow a \\ y \longrightarrow b}} \frac{\frac{1}{y-x}\prod_j f(\mathbf{x}_j|\theta)}{\int_x^y \frac{1}{y-x}\prod_j f(\mathbf{x}_j|\theta)d\theta}$$

$$= \frac{\prod_j f(\mathbf{x}_j|\theta)}{\lim_{\substack{x \longrightarrow a \\ y \longrightarrow b}} \int_x^y \prod_j f(\mathbf{x}_j|\theta)d\theta}$$

$$\propto \prod_j f(\mathbf{x}_j|\theta) \qquad (2.13)$$

Bien entendu, nous devons supposer l'existence des limites mentionnées dans les expressions. Cette condition est vérifiée pour la famille des distributions exponentielles, ce qui est le cas en général. Cette formule peut se généraliser dans le cas multidimensionnel.

Ces distributions $\{\pi_i(\theta) = \pi(\theta|\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_i) \ ; \ 1 \le i \le n\}$ appartiennent à une famille $\mathcal{F}$ de distributions conjuguées par rapport à la famille de fonctions de vraisemblance $\mathcal{K} = \{f(\mathbf{x}|\theta) \ ; \ \mathbf{x} \in \mathbb{R}^k, \ k = 1, 2, \ldots \text{ et } \theta \in \Theta\}$. En effet, un échantillon composé des n échantillons $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n$. Compte tenu du fait que

$$f(\mathbf{x}_i|\theta) = \prod_j f(x_i^j|\theta) \qquad j = 1, 2, \ldots , k_i \qquad ,$$

nous avons

$$\pi_1(\theta) \propto f(\mathbf{x}_1|\theta) \propto \prod_j f(x_1^j|\theta)$$

$$\pi_n(\theta) \propto \prod_i f(\mathbf{x}_i|\theta) = \prod_i \prod_j f(\mathbf{x}_i^j|\theta) = f(\mathbf{x}|\theta) \qquad (2.14)$$

d'où $\mathcal{F} = \{f(\mathbf{x}|\theta);\ \mathbf{x} \in \mathbb{R}^k,\ k = 1,2,\ldots \text{ et } \theta \in \Theta\}$.

Si la distribution sous-jacente $f(x|\theta)$ admet une statistique exhaustive (sous entendu dans un modèle dominé): $f(\mathbf{x}|\theta) \propto g\big(T(\mathbf{x})|\theta\big)$ pour un échantillon quelconque $\mathbf{x}$, alors il existe une statistique exhaustive pour l'échantillon composé (ou total) des n échantillons antérieurs: $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n$, qui nous permet d'écrire à la fois les deux formules suivantes :

$$\pi_n(\theta) \propto \prod_i f(\mathbf{x}_i|\theta) \propto \prod_i g\big(T(\mathbf{x}_i)|\theta\big)$$

$$\pi_n(\theta) \propto f(\mathbf{x}|\theta) \propto g\big(T(\mathbf{x})|\theta\big) = g\big(T(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n)|\theta\big)$$

Pour la famille exponentielle, on obtient l'égalité suivante :

$$g\big(\tilde{T}(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n)|\theta\big) = \prod_i g\big(\tilde{T}(\mathbf{x}_i)|\theta\big) = g\Big(\sum_i \tilde{T}(\mathbf{x}_i)|\theta\Big)$$

d'où l'opération de conjuguaison linéaire :

$$\tilde{T}(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n) = \sum_i \tilde{T}(\mathbf{x}_i)$$

(cf. définition de $\tilde{T}$ au paragraphe 1.1 et paragraphe 2.3).

Nous obtenons donc la même forme de distribution a priori que par la méthode de la conjuguée naturelle (cf. paragraphe 1.6.3), ce qui en renforce l'intérêt.

2.1.3. Exemple du modèle binomial-bêta (1)

Prenons l'exemple d'une variable binomiale $X(k,\theta)$ avec $k \in \mathbb{N}^+$.

Soit un ensemble de N objets dont chacun est caractérisé par l'absence ou la présence d'une propriété A (par exemple, une malfaçon dans la fabrication). Désignons par θ la proportion d'objets présentant la propriété A. Le problème est d'estimer la vraie valeur de θ.

Considérons une variable aléatoire X, qui représente le nombre d'objets

ayant la propriété A parmi k objets tirés au sort indépendamment. Alors la variable X admet une distribution binomiale de paramètres k et θ, avec $k \geq 0$ et $0 \leq \theta \leq 1$.

Un résultat $x \in \{0, 1, \ldots, k\}$ a pour probabilité :

$$f(x|\theta) = C_k^x \, \theta^x (1-\theta)^{k-x} \tag{2.15}$$

où $C_k^x = \dfrac{k!}{x!\,(k-x)!} = \dfrac{\Gamma(k+1)}{\Gamma(x+1)\Gamma(k-x+1)}$.

L'inconnue est le taux θ à valeurs dans l'intervalle [0,1]. Au lieu d'estimer θ par la valeur x/k, nous considérons que θ est une variable aléatoire. (L'estimation du taux par $\theta = x/k$ est la méthode du maximum de vraisemblance, et l'estimateur s'appelle taux empirique.)

A partir des n expériences antérieures $\mathbf{x}_1 = \{(x_i, k_i)\}$, la question qui se pose est : quelle inférence peut-on faire sur le taux θ avant d'envisager une nouvelle expérience ; et après avoir obtenu un nouveau résultat $\mathbf{x} = (x, k)$, quelle inférence peut-on faire sur sur le taux? Il s'agit en effet de choisir dans un premier temps la distribution a priori du taux θ à partir des n expériences et de déterminer en second lieu la distribution du taux θ conditionnelle à $\mathbf{x} = (x, k)$.

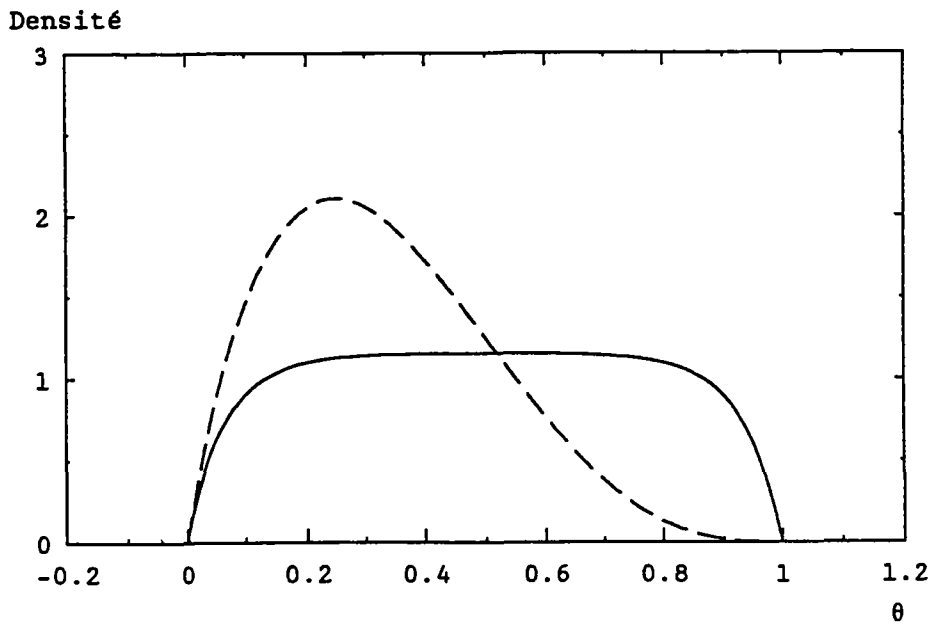


Figure 2.1 Distributions faible et forte

Nous pouvons, en première approximation, accepter que la distribution a priori de θ soit une distribution uniforme sur [0,1]: $\pi(\theta) = I_{[0,1]}(\theta)$. Mais il est raisonnable d'espérer que les valeurs petites du taux θ soient plus probables que les grandes, si θ désigne le taux de défectueux en contrôle de la qualité. Donc, c'est une distribution "forte" et asymétrique que nous voulons obtenir.

Nous savons que chaque expérience $\mathbf{x}_i = (x_i, k_i)$ a pour probabilité :

$$f(\mathbf{x}_i|\theta) = C_{k_i}^{x_i}\ \theta^{x_i}\ (1-\theta)^{k_i-x_i} \quad .$$

Si au départ nous acceptons la distribution uniforme π pour la variable θ, après la première expérience $\mathbf{x}_1$, cette distribution de θ est modifiée, et nous obtenons la distribution a posteriori $\pi_1(\theta)$ de forme suivante :

$$\pi_1(\theta) = \frac{C_n^{x_1}\ \theta^{x_1}\ (1-\theta)^{k_1-x_1}}{\int_0^1 C_n^{x_1}\ \theta^{x_1}\ (1-\theta)^{k_1-x_1} d\theta} \propto \theta^{x_1}(1-\theta)^{k_1-x_1} \quad .$$

ceci est un noyau d'une distribution bêta de paramètres $\begin{cases} \alpha = x_1 + 1 \\ \beta = k_1 - x_1 + 1 \end{cases}$.

D'après la formule séquentielle, nous obtiendrons finalement la distribution du taux θ de la manière suivante :

$$\pi_n(\theta) \propto \prod_i \theta^{x_i}\ (1-\theta)^{k_i-x_i} = \theta^{\Sigma x_i}\ (1-\theta)^{\Sigma k_i - \Sigma x_i}$$

Nous retrouvons $\pi(\theta) = \pi_n(\theta)$ distribution bêta de paramètres :

$$\begin{cases} \alpha = \sum_i x_i + 1 \\ \beta = \sum_i k_i - \sum x_i + 1 \end{cases} \quad .$$

Il est clair que la famille de distributions bêta est conjuguée par rapport à la famille de distributions binomiales. Après avoir acquis une nouvelle expérience $\mathbf{x}$, la distribution a posteriori de θ est la (n+1)-ième distribution de la séquence $\{\pi_i\}$: $\pi(\theta|\mathbf{x} = \mathbf{x}_{n+1}) = \pi_{n+1}(\theta)$.

Nous savons que le taux d'un échantillon $\mathbf{x}_i = (x_i, k_i)$ peut être estimé par la méthode du maximum de vraisemblance, soit : $\theta_i = x_i/k_i$. Nous pouvons aussi appliquer cette méthode pour l'estimation du taux a priori, soit :

$$\hat{\theta}_\pi = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{\sum x_i}{\sum k_i} = \frac{\sum k_i \theta_i}{\sum k_i} \tag{2.16}$$

Donc, la valeur $\hat{\theta}_\pi = \hat{\theta}$ peut être considérée comme l'estimateur empirique du taux θ de l'échantillon total $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n$, mais aussi comme la moyenne des taux empirique $\{\theta_i\}$ pondérés par leurs tailles $\{k_i\}$.

Si nous désignons la taille de l'échantillon total par $KT = \sum k_i$, alors, les paramètres de la distribution bêta peuvent s'écrire :

$$\begin{cases} \alpha = KT \cdot \hat{\theta} + 1 \\ \beta = KT \cdot (1 - \hat{\theta}) + 1 \end{cases} \tag{2.17}$$

La moyenne (a priori) du taux θ devient :

$$\bar{\theta}_\pi = \mathbb{E}^\pi(\theta) = \frac{\alpha}{\alpha + \beta} = \frac{\sum x_i + 1}{\sum k_i + 2} = \frac{\sum k_i \theta_i + 1}{\sum k_i + 2} \tag{2.18}$$

Si le rapport $\sum x_i / \sum (k_i - x_i) = \hat{\theta}/(1 - \hat{\theta})$ est inférieur à 1, nous avons rendu asymétrique la distribution a priori π.

Il est à noter que pour $\mathbf{x}_i = (x_i, k_i)$ un échantillon de la variable binomiale $X(k, \theta)$, une statistique exhaustive est $\tilde{T}(\mathbf{x}_i) = (x_i, k_i)$. Compte tenu du fait que

$$\tilde{T}(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n) = \left(\sum x_i, \sum k_i\right) = \sum \mathbf{x}_i = \sum \tilde{T}(\mathbf{x}_i)$$

nous retrouvons que l'opération est linéaire. Si l'on prend la statistique exhaustive "normalisée" $T(\mathbf{x}_i) = (\theta_i = x_i/k_i,\ k_i)$, alors on a :

$$T(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n) = \left(\frac{\sum x_i}{\sum k_i}, \sum k_i\right) = \left(\frac{\sum k_i \theta_i}{\sum k_i}, \sum k_i\right) \tag{2.19}$$

Pour la deuxième approximation, nous intégrons une loi de distribution a priori non-informative autre que la mesure uniforme, par exemple celle de Jeffrey (cf. paragraphe 1.5.3). D'après la forme de la loi binomiale $X(k,\theta)$, nous obtenons comme information de Fisher :

$$I(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right) = \mathbb{E}_\theta\left(\frac{x}{\theta^2} + \frac{k-x}{(1-\theta)^2}\right)$$

Comme $\mathbb{E}_\theta(X) = k\theta$, la distribution de Jeffey pour ce problème est

$$\pi(\theta) \propto \left(\frac{k}{\theta(1-\theta)}\right)^{1/2} \propto [\theta(1-\theta)]^{-1/2} \quad .$$

Nous considérons une autre distribution non-informative du taux obtenue par les tranformations plus compliquées : $\pi(\theta) = \theta^{-1}(1-\theta)^{-1}$ (cf. [47], [5]).

Avec la formule séquentielle (2.11), les distributions de paramètre θ de la distribution binomiale associées à ces trois a priori ci-dessus sont respectivement :

$$\pi^1 = \mathrm{Be}\left(\sum x_i+1,\ \sum k_i-\sum x_i+1\right)$$

$$\pi^2 = \mathrm{Be}\left(\sum x_i+\frac{1}{2},\ \sum k_i-\sum x_i+\frac{1}{2}\right)$$

$$\pi^3 = \mathrm{Be}\left(\sum x_i,\ \sum k_i-\sum x_i\right)$$

Nous obtenons facilement les estimateurs des moments et du maximum de vraisemblance de la distribution bêta $\mathrm{Be}(\alpha,\beta)$, ainsi les estimateurs correspondant aux trois distributions a priori :

Méthode	Estimation	Distribution a priori		
		$\pi = \pi^1$	$\pi = \pi^2$	$\pi = \pi^3$
Moments	$\bar{\theta}_\pi = \frac{\alpha}{\alpha+\beta}$	$\frac{nm+1}{nKM+2} = \frac{KT\cdot\hat{\theta}+1}{KT+2}$	$\frac{m+1/2}{nKM+1} = \frac{KT\cdot\hat{\theta}+1/2}{KT+1}$	$\frac{m}{KM} = \hat{\theta}$
Maximum	$\hat{\theta}_\pi = \frac{\alpha-1}{\alpha+\beta-2}$	$\frac{m}{KM} = \hat{\theta}$	$\frac{m-1/2}{nKM-1} = \frac{KT\cdot\hat{\theta}-1/2}{KT-1}$	$\frac{nm-1}{nKM-2} = \frac{KT\cdot\hat{\theta}-1}{KT-2}$

Tableau 2.1 Estimation bayésienne du taux a priori

en notant $m = \frac{1}{n}\sum x_i$ et $KM = \frac{1}{n}\sum k_i$, on a donc $KT = n \cdot KM$ et l'estimateur du maximum de vraisemblance est obtenu à partir de l'équation : $\frac{\partial \pi}{\partial \theta} = 0$.

Il est clair que si la somme $\sum x_i$ est assez grande, les trois distributions sont très proches l'une de l'autre et le choix de la distribution a une influence très faible sur le résultat. Dans le cas contraire, ces trois distributions sont différentes et le choix d'une distribution a priori est délicat. Lorsque n = 1, les deux estimateurs deviennent respectivement :

$$\delta(x) = \begin{cases} \frac{x+1}{k+2} \\ \frac{x+1/2}{k+1} \\ \frac{x}{k} \end{cases} \qquad \hat{\delta}(x) = \begin{cases} \frac{x}{k} \\ \mathrm{Max}\left(\frac{x-1/2}{k-1}, 0\right) \\ \mathrm{Max}\left(\frac{x-1}{k-1}, 0\right) \end{cases}$$

Pour obtenir une solution optimale, il faudra ajouter un critère de décision supplémentaire. Nous renvoyons à [8] pour des exemples où une analyse permet de choisir efficacement entre plusieurs estimateurs.

2.2. METHODE DES SCORES D'EXPERT

Dans un problème statistique, l'information est en général constituée par différentes expériences. A partir de ces expériences reproductibles, on peut, comme nous l'avons fait dans le paragraphe précédent, établir des modèles probabilistes sans distinguer les poids présentés par ces expériences. Cependant, il existe bien des cas où l'on ne peut pas accorder un poids "identique" à chaque expérience selon les avis d'experts, et donc on cherche une méthode qui rend compte des poids (ou scores) associés aux expériences, pour trouver une distribution de probabilité (ou un modèle) représentant au mieux l'information donnée.

Il existe beaucoup de traveaux théoriques et pratiques récents sur ce genre de problèmes. Nous n'entrons pas ici dans le cadre mathématique de cette théorie, et nous renvoyons à l'article de C. Geneste et J.V. Zidek (1986) ([19]) qui donne un résumé des idées originales et des développements de cette approche.

2.2.1. Formalisme de scores d'expert et Application logarithmique

Etant donnée une suite de densités de probabilité f_1, $f_2, \ldots,$ f_n et une suite de scores correspondants α_1, α_2, ..., α_n $(0 \leq \alpha_i \leq 1)$, le problème est de trouver une application (ou une opération) T: $(f_1, f_2, \ldots, f_n) \longrightarrow f$ telle que la fonction $f=T(f_1, f_2, \ldots, f_n)$ soit une densité de probabilité. Plusieurs axiomes (ou conditions) de régularisation sont proposés pour assurer que T possède des propriétés raisonnables telles que la symétrie (en colonnes), la conservation de la forme des distributions, la stabilité $T(f,f,\ldots,f) = f$, la cohérence bayésienne ... etc (cf. [29] pour la définition).

Une application T naturelle à considérer est l'application linéaire :

$$T(f_1, f_2, \ldots, f_n) = \sum_i \alpha_i f_i \tag{2.20}$$

Pour que cette fonction soit une densité de probabilité, il faut réaliser la condition: $\sum \alpha_i = 1$. Il est facile de vérifier que cette application linéaire T satisfait les conditions de symétrie, stabilité et continuité, mais elle ne conserve pas en général la forme des distributions.

Par contre, considérons une autre application simple T :

$$T(f_1, f_2, \ldots, f_n) = N \sum_i f_i^{\alpha_i} \tag{2.21}$$

où N est une constante de normalisation. Cette application est linéaire en logarithme :

$$\log\big(T(f_1, f_2, \ldots, f_n)\big) = \sum_i \alpha_i \log(f_i) \tag{2.22}$$

Par rapport à la première application, la seconde possède une propriété supplémentaire : elle conserve la forme des lois de distribution de type exponentiel (cf. paragraphe 1.1).

2.2.2. Fonction de vraisemblance bayésienne et Résolution générale

Dans notre étude, l'information a priori est une suite d'échantillons antérieurs $\mathbf{x}_1$, $\mathbf{x}_2, \ldots,$ $\mathbf{x}_n$ prélevés dans des lots successifs. Nous savons que

notre analyse bayésienne repose sur le principe de la conjuguée naturelle de la fonction de vraisemblance ; nous avons donc besoin d'établir une fonction de vraisemblance dans le cas de poids accordés aux échantillons et aux lots correspondants tant identiques que différents.

Nous allons construire une "fonction de vraisemblance approchée" qui va tenir compte des poids accordés aux échantillons. Les poids seront représentés par les valeurs numériques $\{\alpha_i > 0\}$ telles que $\sum\alpha_i = 1$. En pratique, ces valeurs sont souvent liées étroitement aux tailles des échantillons ou à la date d'échantillonnage. Par exemple, les échantillons prélevés dans des lots récents peuvent peser plus que ceux qui sont prélevés dans des lots plus anciens.

Compte tenu de l'importance de la stabilité de forme dans notre analyse bayésienne, nous allons rechercher une fonction de vraisemblance qui conservera la forme des distributions, puis choisir une distribution a priori comme conjuguée de cette nouvelle fonction.

Rappelons que la fonction de vraisemblance au sens usuel de ces n échantillons est le produit de chaque fonction de vraisemblance de l'échantillon $\mathbf{x}_i$, $f(\mathbf{x}_i|\theta)$, $i = 1, 2, \ldots, n$. Elle est aussi égale à la fonction de vraisemblance de l'échantillon total $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n$:

$$f(\mathbf{x}|\theta) = L(\theta;\mathbf{x}) = L(\theta;\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n) = \prod_i f(\mathbf{x}_i|\theta) \tag{2.23}$$

Nous allons considérer comme fonction de vraisemblance approchée la fonction suivante (cf formule (2.21)) :

$$f'(\mathbf{x}|\theta) = L'(\theta;\mathbf{x}) = L'(\theta;\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n) \propto \prod_i f(\mathbf{x}_i|\theta)^{\alpha_i} \tag{2.24}$$

Nous l'appelons <u>fonction de vraisemblance bayésienne</u>.

D'après le principe de la conjuguée naturelle (cf paragraphe 1.6), nous obtenons donc la loi de distribution a priori :

$$\pi(\theta) \propto f'(\mathbf{x}|\theta) \propto \prod_i f(\mathbf{x}_i|\theta)^{\alpha_i} \tag{2.25}$$

Il est clair que la fonction de vraisemblance au sens usuel (classique) des n échantillons est un cas "particulier" avec $\alpha_i = 1$, $i = 1,2,\ldots,n$, mais il n'y a pas la rénormalisation $\sum\alpha_i = 1$.

2.2.3. Statistique exhaustive correspondante de la famille exponentielle

Nous allons supposer que la distribution $f(x|\theta)$ appartient à la famille exponentielle :

$$f(x|\theta) = a(x)b(\theta)\exp\left[\sum_l c_l(x)d_l(\theta)\right] \tag{2.26}$$

Nous savons alors construire une statistique exhaustive $\tilde{T}$ pour un échantillon $\mathbf{x} = (x_1, x_1, \ldots, x_k)$:

$$\tilde{T}(\mathbf{x}) = \left\{ \sum_j c_1(x_j),\ \sum_j c_2(x_j),\ \ldots,\ \sum_j c_p(x_j);\ k \right\}$$

d'où la statistique exhaustive "normalisée" est (*cf.* formule (1.14)) :

$$T(\mathbf{x}) = \left\{ \frac{1}{k}\sum_j c_1(x_j),\ \frac{1}{k}\sum_j c_2(x_j),\ \ldots,\ \frac{1}{k}\sum_j c_p(x_j);\ k \right\}$$

$$= \left\{ \overline{c_1(\mathbf{x})},\ \overline{c_2(\mathbf{x})},\ \ldots,\ \overline{c_p(\mathbf{x})};\ k \right\}$$

Nous considérons maintenant n échantillons $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ tels que $\mathbf{x}_i$ ait pour taille k_i, et nous notons α_i le score associé à cet échantillon $\mathbf{x}_i$. La fonction de vraisemblance s'écrit donc :

$$f'(\mathbf{x}|\theta) = \prod_i f(\mathbf{x}_i|\theta)^{\alpha_i} \propto b(\theta)\exp\left[\sum_l d_l(\theta)\sum_i \alpha_i \sum_j c_l(x_j^i)\right] \tag{2.27}$$

où $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n$ est l'échantillon total. Une statistique exhaustive $\tilde{T}(\mathbf{x})$ est définie par :

$$\tilde{T}(\mathbf{x}) = \sum \alpha_i \tilde{T}(\mathbf{x}_i) \tag{2.28}$$

d'où la statistique exhaustive normalisée $T(\mathbf{x})$:

$$T(\mathbf{x}) = \left\{ \frac{\sum \alpha_i k_i \overline{c_i(\mathbf{x}_i)}}{\sum \alpha_i k_i} \; ; \; \sum \alpha_i k_i \right\} \tag{2.29}$$

Nous pourrons combiner cette méthode des scores d'expert aux diverses méthodes développées ci-dessous pour pouvoir imposer des poids aux échantillons et aux lots correspondants.

Dans la combinaison avec la méthode des échantillons équilibrés, il existera deux séries de scores. Les premiers scores $\{\alpha_i\}$, dits "externes", représentent les poids accordés aux lots de production; les deuxièmes scores "internes" $\{K/k_i\}$ dans l'analyse des échantillons équilibrés seront pris en compte pour éliminer les effets dûs aux tailles différentes des échantillons (cf. paragraphe 2.3).

Pour la méthode des familles paramétriques (cf. paragraphe 2.4), la situation est différente, parce que cette méthode utilise directement tous les résultats mélangés (dit aussi non-classés) $\{x_j^i\}$. Nous devons multiplier les résultats de chaque échantillon $\mathbf{x}_i = \{x_j^i;\ j = 1,2,\ldots,k_i\}$ par le score correspondant α_i. L'analyse bayésienne est menée avec la nouvelle série d'échantillons $\alpha_1\mathbf{x}_1,\ \alpha_2\mathbf{x}_2,\ \ldots,\ \alpha_n\mathbf{x}_n$.

Cependant, avec l'ensemble $\{\alpha_i\mathbf{x}_i\}$, les autres méthodes ne donnent pas le même résultat que celui ci-dessus (cf. formule (2.27)) avec l'ensemble $\{(\mathbf{x}_i,\alpha_i)\}$, parce que les fonctions $\{c_i\}$ ne sont pas toutes linéaires et donc on n'a pas l'égalité : $\tilde{T}(\alpha_i\mathbf{x}_i) = \alpha_i\tilde{T}(\mathbf{x}_i)$ (cf. paragraphe 2.5).

2.3. METHODES DES ECHANTILLONS EQUILIBRES

Nous supposons maintenant que les techniques de production sont homogènes et que chaque lot répondant aux mêmes conditions de spécification doit peser un poids égal dans l'interpétation de l'information a priori. Par conséquent, dans la fonction de vraisemblance, un poids identique est accordé à chacun des n échantillons, constituant cette information a priori.

Plus précisément, nous avons un échantillon $\mathbf{x}_i$ constitué par k_i valeurs

(ou expériences). Sa fonction de vraisemblance est le produit de k_i densités des expériences $\{x_j^i \; ; \; j = 1,2,\ldots,k_i\}$ réalisées dans le lot (L_i) :

$$f(\mathbf{x}_i|\theta) = \prod_i f(x_j^i|\theta)$$

Nous pouvons dire que le lot (L_i) est représenté k_i fois dans la fonction de vraisemblance au sens usuel. Pour éliminer cet effet des tailles des échantillons, il est naturel de choisir, pour chaque échantillon $\mathbf{x}_i$, un score α_i proportionnel à l'inverse de la taille k_i. Nous avons alors

$$\alpha_i = K/k_i \tag{2.30}$$

avec $K = \left(\sum k_i^{-1}\right)^{-1}$ constante de régularisation telle que $\sum \alpha_i = 1$. Donc, nous avons un problème d'échantillons avec des poids différents. D'après les formules (2.25) et (2.26), nous obtenons la fonction de vraisemblance approchée et la distribution a priori correspondante :

$$\pi(\theta) \propto f'(\mathbf{x}|\theta) = \prod_i f(\mathbf{x}_i|\theta)^{K/k_i} \tag{2.31}$$

Cette méthode spécifique s'appelée <u>méthode des échantillons équilibrés</u>.

Dans le cas d'une distribution exponentielle, le calcul du noyau de la distribution a priori est tout simplement réduit à une opération de conjuguaison linéaire des statistiques exhaustives des échantillons :

$$\tilde{T}(\mathbf{x}) = \frac{\sum \tilde{T}(\mathbf{x}_i)/k_i}{\sum 1/k_i} \tag{2.32}$$

d'où la statistique exhaustive normalisée :

$$T(\mathbf{x}) = \left\{ \frac{\sum c_i(\mathbf{x}_i)}{n} \; ; \; \frac{1}{\frac{1}{n}\sum 1/k_i} = KH \right\} \tag{2.33}$$

La statistique $KH = n \cdot K$ est la moyenne harmonique des tailles $\{k_i\}$.

Si nous la comparons avec la statistique exhaustive normalisée obtenue

ECOLE NATIONALE DES PONTS & CHAUSSÉES
CENTRE PÉDAGOGIQUE DE DOCUMENTATION ET DE COMMUNICATION

par la méthode séquentielle, soit :

$$T(\mathbf{x}) = \left\{ \frac{\sum k_i \overline{c_i(\mathbf{x}_i)}}{\sum k_i} \; ; \; \sum k_i = KT \right\} \tag{2.34}$$

on s'aperçoit que le paramètre de "taille" joue un rôle différent dans les deux méthodes.

2.3.1. Exemple du modèle binomial-bêta (2)

Pour illustrer cette méthode, nous reprenons l'exemple de la distribution binomiale. Avec n résultats d'expériences $\mathbf{x}_1 = (x_1, k_1)$, $\mathbf{x}_2 = (x_2, k_2), \ldots,$ $\mathbf{x}_n = (x_k, k_n)$, la distribution a priori du taux θ est :

$$\pi(\theta) \propto \prod_i \left(\theta^{x_i} (1-\theta)^{k_i - x_i}\right)^{K/k_i} = \left(\theta^{\Sigma x_i/k_i} (1-\theta)^{n - \Sigma x_i/k_i}\right)^K \tag{2.35}$$

c'est une distribution bêta de paramètres :

$$\begin{cases} \alpha = K\sum(x_i/k_i) + 1 \\ \beta = Kn - K\sum(x_i/k_i) + 1 \end{cases} \tag{2.36}$$

Désignons par $\overline{\theta} = \frac{1}{n}\sum\theta_i$ la moyenne des taux empiriques $\{\theta_i = x_i/k_i\}$, l'estimateur (α, β) ci-dessus devient :

$$\begin{cases} \alpha = KH\cdot\overline{\theta} + 1 \\ \beta = KH\cdot(1-\overline{\theta}) + 1 \end{cases} \tag{2.37}$$

Le maximum de vraisemblance et la moyenne (a priori) du taux θ sont les suivants :

$$\hat{\theta}_\pi = \frac{\alpha - 1}{\alpha + \beta - 2} = \overline{\theta} \tag{2.38}$$

$$\overline{\theta}_\pi = \mathbb{E}^\pi(\theta) = \frac{\alpha}{\alpha + \beta} = \frac{KH\cdot\overline{\theta} + 1}{KH + 2} \tag{2.39}$$

Si la moyenne empirique $\overline{\theta}$ est petite, $\overline{\theta} < 1/2$ par exemple, la distribution π

de θ est asymétrique et nous avons $\bar{\theta} = \hat{\theta}_\pi < \bar{\theta}_\pi < 1/2$. Dans le cas contraire, $\bar{\theta} > 1/2$, nous avons $1/2 < \bar{\theta}_\pi < \hat{\theta}_\pi = \bar{\theta}$. Il est clair que si $\bar{\theta} = 1/2$, alors on a l'égalité : $\bar{\theta}_\pi = \hat{\theta}_\pi = \bar{\theta}$.

La méthode séquentielle conduit à la même conclusion avec la valeur de critère $\hat{\theta} = \sum x_i / \sum k_i = \sum k_i \theta_i / \sum k_i$ au lieu de la moyenne $\bar{\theta}$ des taux empiriques.

Si les deux taux empiriques sont presque égaux et inférieurs à 1/2 : $\bar{\theta} \approx \hat{\theta} < 1/2$, alors nous avons toujours $\frac{KH \cdot \bar{\theta} + 1}{KH + 2} < \frac{KT \cdot \hat{\theta} + 1}{KT + 2}$, c'est à dire que l'estimateur a priori du taux θ par la méthode des échantillons équilibrés est inférieur à celui par la méthode séquentielle, parce que le nombre total KT est supérieur à la moyenne harmonique KH des nombres.

Si l'on prend une statistique exhaustive $\tilde{T}(\mathbf{x}_i) = \mathbf{x}_i = (x_i, k_i)$, on a :

$$\begin{aligned} \tilde{T}(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n) &= K \sum \tilde{T}(\mathbf{x}_i)/k_i \\ &= K\left(\sum x_i/k_i, n\right) \\ &= Kn\left(\frac{1}{n}\sum \theta_i, 1\right) \end{aligned}$$

et la statistique exhautive normalisée est définie par :

$$T(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n) = \left(\frac{1}{n}\sum \theta_i,\ Kn\right) = (\bar{\theta},\ KH)$$

2.4. METHODES DES FAMILLES PARAMETRIQUES

Nous rappelons que nous sommes en présence d'une série d'échantillons, prélevés dans différents lots de production, chacun portant une information sur l'état de la nature θ.

Le principe de ces méthodes est de supposer que la loi de distribution a priori π possède une forme paramétrique ; on calcule ce paramètre à partir de la distribution marginale p de la variable aléatoire X. Quand le choix de la famille paramétrique est fait, il existe plusieurs méthodes pour déterminer le paramètre en tenant compte de la relation suivante :

$$p(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta \tag{2.40}$$

ou

$$p(x;\xi) = \int_{\Theta} f(x|\theta)\pi(\theta;\xi)d\theta$$

où ξ est le paramètre de la distribution a priori π. D'où nous déduisons des égalités entre les valeurs caractéristiques (moments, fractiles, ordres ...) d'un élément π de la famille considérée et de la distribution marginale p.

Dans ce cas, la détermination des paramètres de π est basée sur le fait qu'il faut au moins autant de relations que de paramètres à estimer ; ainsi le choix d'une distribution a priori se ramène éventuellement à la détermination de la distribution p.

Nous pouvons utiliser également la méthode du maximum de vraisemblance, mais la solution est parfois difficile à trouver.

Nous précisons ici que pour cette famille de méthodes, les échantillons sont regroupés dans un échantillon global suivant la distribution p au lieu de la distribution conditionnelle $f(x|\theta)$. A partir de ces résultats, nous estimons de façon empirique des valeurs caractéristiques de p, et puis nous en déduisons celles de π.

2.4.1. Méthode des moments

En fonction du nombre de paramètres de π, la méthode la plus facile est d'établir les égalités entre les moments de p et les moments de π à l'aide du lemme suivant :

Lemme : Désignons respectivement les n-ième moments des distributions marginale p et conditionnelle f de la variable aléatoire X, (ceux-ci étant donnés par rapport à la densité $f(x|\theta)$) par :

$$\upsilon_n = \mathbb{E}(X^n) \qquad \mu_n = \mathbb{E}[(X-\mathbb{E}(X))^n]$$

$$\upsilon_n^{\theta} = \mathbb{E}_{\theta}(X^n) \qquad \mu_n^{\theta} = \mathbb{E}_{\theta}[(X-\mathbb{E}_{\theta}(X)^n] \quad .$$

Supposons que les intégrales existent, alors nous avons :

$$\upsilon_n = \mathbb{E}^{\pi}\left[\upsilon_n^{\theta}\right] \qquad \mu_n = \sum_k C_n^k \mathbb{E}^{\pi}\left[\mu_k^{\theta}\left(\upsilon_1^{\theta} - \upsilon_1\right)^{n-k}\right] \tag{2.41}$$

Démonstration : Soit μ la mesure image de l'espace probabilisé Ω par X. Ce lemme est démontré par les calculs suivants dans le cas du modèle dominé :

$$\begin{aligned}
\mathbb{E}(X^n) &= \int_{\Omega} x^n p(x) d\mu \\
&= \iint_{\Omega\times\Theta} x^n f(x|\theta)\pi(\theta) d\mu d\theta \\
&= \int_{\Theta} \mathbb{E}_{\theta}(X^n)\pi(\theta) d\theta = \mathbb{E}^{\pi}\left[\mathbb{E}_{\theta}(X^n)\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[(X-\mathbb{E}(X))^n] &= \int_{\Omega} [x-\mathbb{E}(X)]^n p(x) d\mu \\
&= \iint_{\Omega\times\Theta} [x-\mathbb{E}(X)]^n f(x|\theta)\pi(\theta) d\mu d\theta \\
&= \int_{\Theta} \sum_k C_n^k \cdot \mathbb{E}_{\theta}[(X-\mathbb{E}_{\theta}(X))^k] \cdot [\mathbb{E}_{\theta}(X)-\mathbb{E}(X)]^{n-k} \pi(\theta) d\theta \\
&= \sum_k C_n^k \cdot \mathbb{E}^{\pi}\left[\mathbb{E}_{\theta}[(X-\mathbb{E}_{\theta}(X))^k] \cdot [\mathbb{E}_{\theta}(X)-\mathbb{E}(X)]^{n-k}\right] .
\end{aligned}$$

Les séries $\{\upsilon_n\}$ et $\{\mu_n\}$ sont bien sûr liées. Il suffit de calculer une série pour déterminer les paramètres de la distribution a priori.

La plupart des distributions ont au plus deux paramètres, il suffit donc de calculer les relations pour la moyenne et la variance.

Notation: $\mu = \mathbb{E}^P(X)$ — moyenne marginale

$\sigma^2 = \mathbb{E}^P[(X-\mu)^2] = \mathbb{V}^P(X)$ — variance marginale

$\mu_\theta = \mathbb{E}_\theta(X)$ — moyenne conditionnelle

$\sigma_\theta^2 = \mathbb{E}_\theta[(X-\mu_\theta)^2] = \mathbb{V}_\theta(X)$ — variance conditionnelle

$\mu_\pi = \mathbb{E}^\pi(\theta)$ — moyenne a priori du paramètre θ

$\sigma_\pi^2 = \mathbb{E}^\pi[(\theta-\mu_\pi)^2] = \mathbb{V}^\pi(\theta)$ — variance a priori du paramètre θ

Dans ce cas particulier, le lemme suivant donne :

$$\begin{cases} \mu = \mathbb{E}^\pi[\mu_\theta] \\ \sigma^2 = \mathbb{E}^\pi[\sigma_\theta^2] + \mathbb{E}^\pi[(\mu_\theta-\mu)^2] \end{cases} \tag{2.42}$$

Conséquence : Soit c une constante. Si $\mu_\theta = c\theta$, alors on a $\mu = c\mu_\pi$ et $\sigma^2 = \mathbb{E}^\pi[\sigma_\theta^2] + c^2\sigma_\pi^2$.

Remarquons que la méthode des moments mène souvent à des difficultés pratiques et en particulier à des valeurs "impossibles" des paramètres de la distribution a priori π. Au chapitre IV, nous montrerons plusieurs exemples.

2.4.2. Exemple du modèle binomial-bêta (3)

Nous continuons à étudier l'exemple de la variable binomiale $X(k,\theta)$.

Nous considérons que la distribution a priori est conjuguée de la fonction de vraisemblance, et donc est une distribution bêta :

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{2.43}$$

dont la moyenne a priori et la variance a priori du taux θ sont :

$$\begin{cases} \mu_\pi = \dfrac{\alpha}{\alpha+\beta} \\ \sigma_\pi^2 = \dfrac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} \end{cases} \tag{2.44}$$

Si nous supposons que le nombre de tirages k est fixé a priori, nous avons la moyenne et la variance conditionnelles de la variable $X(k,\theta)$:

$$\begin{cases} \mu_\theta = k\theta \\ \sigma_\theta^2 = k\theta(1-\theta) \end{cases} \tag{2.45}$$

D'après le lemme ci-dessus, nous obtenons la moyenne et la variance de la variable X en fonction du couple (α,β) et du nombre k :

$$\begin{cases} \mu = k\mu_\pi = \dfrac{k\alpha}{\alpha+\beta} \\ \sigma^2 = E^\pi[\sigma_\theta^2] + k^2\sigma_\pi^2 = \dfrac{k\alpha\beta(\alpha+\beta+k)}{(\alpha+\beta+1)(\alpha+\beta)^2} \end{cases}$$

d'où les paramètres α et β sont en fonction de μ et σ^2 :

$$\begin{cases} \alpha = \dfrac{\mu^2(1-\mu/k) - \sigma^2\mu/k}{\sigma^2 - \mu(1-\mu/k)} \\ \beta = \alpha\left(\dfrac{k}{\mu} - 1\right) \end{cases} \tag{2.46}$$

Il reste donc à calculer la moyenne et la variance (μ,σ^2) de tous les échantillons regroupés; ceux-ci sont simplement estimées de façon empirique soit :

$$\begin{cases} \mu = m = \dfrac{1}{n}\sum_i x_i \\ \sigma^2 = v = \dfrac{1}{n}\sum_i (x_i-\mu)^2 \end{cases}$$

Le paramètre k peut être considéré normalement comme la moyenne des nombres $\{k_i\}$: $k = KM = \frac{1}{n}\sum k_i$.

Avec la définition de la moyenne empirique totale du taux $\hat{\theta} = \Sigma x_i/\Sigma k_i = m/KM$, la formule (2.46) devient :

$$\begin{cases} \alpha = \dfrac{(KM\cdot\hat{\theta})^2(1-\hat{\theta}) - v\hat{\theta}}{v - KM\cdot\hat{\theta}(1-\hat{\theta})} = \dfrac{m^2(1-\hat{\theta}) - v\hat{\theta}}{v - m(1-\hat{\theta})} \\ \beta = \alpha(1/\hat{\theta} - 1) \end{cases} \tag{2.47}$$

Pour que α et β soient positifs, il faut satisfaire la condition :

soit $\begin{cases} 0 < \hat{\theta} < 1 \\ m^2(1-\hat{\theta})/\hat{\theta} < v < m(1-\hat{\theta}) \end{cases}$ soit $\begin{cases} 0 < \hat{\theta} < 1 \\ m^2(1-\hat{\theta})/\hat{\theta} > v > m(1-\hat{\theta}) \end{cases}$

Dans le cas particulier où chaque expérience est faite de manière à tirer des objets jusqu'à l'obtention de x unités défectueuses, la variance empirique est nulle et cette méthode n'est plus adaptée à ce problème.

2.4.3. Méthode des fractiles

Dans la méthode précédente, nous n'avons présenté que la moyenne et la variance. Naturellement, nous pouvons choisir d'autres moments, mais un tel choix changera éventuellement l'estimateur des paramètres α et β. De plus, il arrive souvent que les moments d'une distribution π n'existent pas (cf. paragraphe 4.3). En général, l'estimation de la probabilité d'une région est relativement facile et sensible par rapport à celle de moments, parce que l'estimation par des moments imposent des restrictions sur les queues des distributions p et π (cf. [7] pour l'analyse de sensibilité ou de robustesse). Cette approche est un prolongement de la méthode des vraisemblances relatives (cf.paragraphes 1.5.1 et 1.5.2).

Nous pouvons établir des égalités entre des fractiles de $\pi(\theta)$ et ceux de $p(x)$ et choisir le paramètre tel que la densité obtenue ait ses fractiles les plus proches possible de ceux de $\pi(\theta)$. Mais, l'expression analytique n'est pas si simple que pour les autres estimateurs.

Un α-fractile d'une distribution de probabilité est un point x_α tel que la variable X ait la probabilité α d'être inférieure ou égale à x_α; et α est appelé l'ordre de ce fractile.

Désignons par $\alpha^\theta(x_\alpha) = \mathbb{P}^\theta\{ X < x_\alpha\}$ l'ordre du fractile de la distribution $f(x|\theta)$ associé au point x_α. Soit en général $\Omega = [a,+\infty)$ avec a un nombre

fini ou $-\infty$. Alors on obtient :

$$\alpha = \mathbb{P}\{X < x_\alpha\} = \int_a^{x_\alpha} p(x)dx$$

$$= \int_a^{x_\alpha} \int_\Theta f(x|\theta)\pi(\theta)d\theta dx$$

$$= \int_\Theta \alpha^\theta(x_\alpha)\pi(\theta)d\theta$$

$$= \mathbb{E}^\pi\left[\alpha^\theta(x_\alpha)\right] \tag{2.48}$$

2.4.4. Méthode du maximum de vraisemblance et Exemple du modèle binomial-bêta (4)

Comme nous nous plaçons dans le cas où les résultats $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ sont supposés des réalisations de la variable X suivant la loi de distribution p, nous pouvons construire la quantité de vraisemblance comme la fonction de vraisemblance de ces résultats :

$$L(\xi;\mathbf{x}) = p(\mathbf{x};\xi) = \prod_i p(x_i;\xi) = \prod_i \int_\Theta f(x_i|\theta)\pi(\theta;\xi)d\theta \tag{2.49}$$

où le paramètre ξ de la distribution π est à déterminer.

Le principe du maximum de vraisemblance est de chercher des valeurs qui rendent maximum la quantité de vrasemblance, dont l'équation s'écrit :

$$\frac{\partial L(\xi;\mathbf{x})}{\partial \xi_i} = 0 \qquad \text{pour } i = 1, 2, \ldots, q$$

Le paramètre est un vecteur $\xi = (\xi_1, \xi_2, \ldots, \xi_q)$ lorsque $q > 1$.

Cette méthode est très répandue dans l'estimation statistique, à cause de nombreuses propriétés intéressantes comme son efficacité lorsque le nombre q grand, son exhaustivité s'il existe un estimateur exhaustif pour le problème étudié, son invariance et sa convergence (cf. [41] et [49] pour les

définitions). Mais l'équation de vraisemblance est souvent moins aisée à résoudre, surtout pour obtenir une solution analytique.

Par exemple, nous reprenons la distribution binomiale $X(k,\theta)$ et nous avons montré précédemment que $f(x|\theta) = B(k,\theta)(x)$ et $\pi(\theta) = Be(\alpha,\beta)(\theta)$. Dans ce cas, nous avons $\xi = (\alpha,\beta)$, et la distribution p est obtenue par le calcul suivant :

$$p\big(x;(\alpha,\beta)\big) = \int_{[0,1]} B(k,\theta)(x)\cdot Be(\alpha,\beta)(\theta)d\theta$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(k+1)}{\Gamma(\alpha+\beta+k)} \times \frac{\Gamma(\alpha+x)\Gamma(\beta+k-x)}{\Gamma(x+1)\Gamma(k-x+1)}$$

où $C_k^x = \frac{k!}{x!\,(k-x)!} = \frac{\Gamma(k+1)}{\Gamma(x+1)\Gamma(k-x+1)}$ (cf. formule (2.15)). Donc, la fonction de vraisemblance a pour forme :

$$L\big((\alpha,\beta);\mathbf{x}\big) \propto \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+k)}\right)^n \prod_i \big(\Gamma(\alpha+x_i)\Gamma(\beta+k-x_i)\big)$$

Désignons $\psi(x) = \frac{d\log\Gamma(x)}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}$. Cette fonction est appelée digamma. Comme les valeurs rendant la fonction $L\big((\alpha,\beta);\mathbf{x}\big)$ maximum sont équivalentes à celles rendant son logarithme $\log\big[L\big((\alpha,\beta);\mathbf{x}\big)\big]$ maximum, nous déduisons donc l'équation de vraisemblance :

$$\begin{cases} \psi(\alpha+\beta+k) - \psi(\alpha+\beta) = \frac{1}{n}\sum_i \psi(\alpha+x_i) - \psi(\alpha) \\ \psi(\alpha+\beta+k) - \psi(\alpha+\beta) = \frac{1}{n}\sum_i \psi(\beta+k-x_i) - \psi(\beta) \end{cases}$$

Une solution de cette équation doit être obtenue par itérations, ce qui rend le calcul plus lourd que pour d'autres estimateurs.

2.5. METHODES DU "QUASI-ECHANTILLON"

L'interprétation de l'information a priori est ici tout à fait différente. Les échantillons ne sont plus regroupés en un seul échantillon.

Les expériences conçues sont supposées issues du même modèle formel, mais les valeurs du paramètres θ sont différentes. En effet, à partir d'une série d'échantillons, les valeurs de θ sont obtenues par l'une des estimations classiques :

$$\begin{array}{cccc} x_1, & x_2, & \ldots, & x_n \\ \downarrow & \downarrow & & \downarrow \\ \theta_1, & \theta_2, & \ldots, & \theta_n \end{array}$$

Nous pouvons dire que ces valeurs $\{\theta_i\}$ constituent un "<u>quasi-échantillon</u>" de la variable θ.

Il est clair que l'on peut utiliser ces valeurs pour construire une loi de distribution de la variable θ. C'est ainsi que nous retrouvons le problème fondamental de l'analyse statistique : estimer une distribution à partir d'un échantillon mais dans le cadre bayésien, ceci peut se faire à partir des vraisemblances relatives, surtout dans le cas d'un espace Θ discret.

Au paragraphe 1.4, nous avons proposé une méthode d'estimation de la distribution par une méthode des vraisemblances relatives. La distribution construite est discrète. Pour obtenir une distribution continue, une méthode possible est de construire un histrogramme et puis d'ajuster une densité convenable. Plus généralement dans ce cadre non paramétrique, nous pouvons utiliser des méthodes de noyaux. Parzen (1962) ([32]) a proposé la formule suivante :

$$\pi(\theta) \approx \pi_n(\theta) = \frac{1}{nh}\sum_i K\left(\frac{\theta - \theta_i}{h}\right) \tag{2.50}$$

où $h = h(n)$ satisfait certaines conditions de régularité et où le noyau K peut être choisi parmi l'une des densités standards, par exemple $N(0,1)$.

Si la forme de la distribution $\pi(\theta)$ est supposée connue, nous pouvons utiliser à nouveau l'estimation des familles paramétriques pour le paramètre de cette distribution a priori à partir du quasi-échantillon $(\theta_1, \theta_2, \ldots, \theta_n)$, d'où le nom de la méthode du quasi-échantillon.

Nous nous limitons dans la suite au cas paramétrique. La méthode du quasi-échantillon comprend deux étapes: l'estimation de $(\theta_1, \theta_2, \ldots, \theta_n)$, puis

des paramètres de la distribution a priori $\pi(\theta)$. Ces estimations seront menées par des méthodes classiques : soit moments, soit maximum de vraisemblance, et nous proposons donc pour ce quasi-échantillon quatre méthodes, selon l'estimation utilisée à chaque étape :

— méthode des doubles moments (ou "moments-moments"), notée par MM,
— méthode des "maxima-moments",
— méthode des "moments-maxima",
— méthode des doubles maxima de vraisemblance (ou "maxima-maxima"), notée par VV.

Pour illustrer le principe du quasi-échantillon, nous prenons dans le paragraphe suivant l'estimateur du maximum de vraisemblance pour les deux étapes, soit la méthode des doubles maxima de vraisemblance.

2.5.1. Construction d'un quasi-échantillon du paramètre θ

Supposons que l'information a priori soit une série d'échantillons $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$ associés à une série de fonctions de vraisemblance $L(\theta;\ \mathbf{x}_1)$, $L(\theta;\ \mathbf{x}_2)$, ..., $L(\theta;\ \mathbf{x}_n)$ avec $L(\theta;\ \mathbf{x}_i) = f(\mathbf{x}_i|\theta)$, ainsi qu'une série de distributions a priori $\pi(\theta|\mathbf{x}_1)$, $\pi(\theta|\mathbf{x}_2)$, ..., $\pi(\theta|\mathbf{x}_n)$, obtenues grâce à la méthode de la conjuguée naturelle $\pi(\theta|\mathbf{x}_i) \propto L(\theta;\mathbf{x}_i)$.

Alors pour chaque échantillon $\mathbf{x}_i$, l'estimateur du maximum de vraisemblance est une valeur de θ, qui rend $L(\theta;\mathbf{x}_i)$ maximale ; ainsi $\pi(\theta|\mathbf{x}_i)$ est obtenue comme une solution de l'équation de vraisemblance :

$$\frac{dL(\theta;\mathbf{x}_i)}{d\theta} = \frac{\partial\pi(\theta|\mathbf{x}_i)}{\partial\theta} = 0 \tag{2.51}$$

Cet estimateur est noté par θ_i, $i = 1,2,\ldots,n$. L'ensemble $\{\theta_i\}$ constitue un échantillon de taille n de la variable θ.

S'il existe une statistique exhaustive $T(\mathbf{x})$ par rapport à θ, à chaque échantillon $\mathbf{x}_i$ correspond une valeur $\theta_i \in \Theta$ du paramètre qui ne dépend que de $T(\mathbf{x}_i)$. Remarquons que l'échantillon $\{\theta_i\}$ ne dépend donc que des valeurs de la statistique exhaustive $T(\mathbf{x}_1)$, $T(\mathbf{x}_2)$, ..., $T(\mathbf{x}_n)$.

2.5.2. Estimateur des paramètres de la loi a priori

Soit $\pi(\theta|\xi)$ la densité de distribution a priori. A partir de l'échantillon $(\theta_1,\theta_2,\ldots,\theta_n)$ de la variable θ, la fonction de vraisemblance de cet échantillon est une fonction de la variable ξ, définie par :

$$L(\xi;\ \theta_1,\theta_2,\ldots,\theta_n) = \prod_i \pi(\theta_i|\xi) \tag{2.52}$$

Le maximum de vraisemblance du paramètre $\xi = (\xi_1,\xi_2,\ldots,\xi_q)$ est une valeur maximisant $L(\xi = (\xi_1,\xi_2,\ldots,\xi_q);\ \theta_1,\theta_2,\ldots,\theta_n)$, c'est à dire qu'en ce point-là, les dérivées partielles de cette fonction s'annulent,

$$\frac{\partial L}{\partial \xi_1} = \frac{\partial L}{\partial \xi_2} = \ldots = \frac{\partial L}{\partial \xi_q} = 0 \tag{2.53}$$

Nous montrons qu'éventuellement, chaque paramètre est représenté sous une forme analytique, induite de celle de la distribution a priori $\pi(\theta|\xi)$ et déterminée par les valeurs de l'échantillon de θ. Nous arrivons à choisir la distribution a priori.

Si la distribution a priori est conjuguée de la fonction de vraisemblance, $\pi(\theta|\xi) \propto f(\mathbf{x}|\theta)$ et que la distribution sous-jacente $f(x|\theta)$ admet une statistique exhaustive, on a $f(\mathbf{x}|\theta) \propto g(T(\mathbf{x})|\theta)$ et le paramètre $\xi = (\xi_1,\xi_2,\ldots,\xi_q)$ est fonction de la statistique exhaustive $T(\mathbf{x})$.

<u>Remarque</u>

Dans le cas d'échantillons avec des scores différents $(\mathbf{x}_i,\alpha_i)$, la fonction de vraisemblance approchée de $(\mathbf{x}_i,\alpha_i)$ est $f(\mathbf{x}_i|\theta)^{\alpha_i}$, mais l'estimateur du maximum de vraisemblance θ_i est le même que dans le cas usuel $f(\mathbf{x}_i|\theta)$ dans la première étape de la méthode du quasi-échantillon (cf. paragraphes 2.2 et 2.5.1); l'estimateur des moments de la distribution $f(x|\theta)$ ne change par non plus. Cependant, le quasi-échantillon de la variable θ prend en compte les poids accordés aux échantillons, et donc aux lots correspondants. Comme l'information provenant de l'échantillon $\mathbf{x}_i$ est affectée d'un poids α_i, ce poids α_i est appliqué également à l'estimateur θ_i. Alors le quasi-échantillon devient $(\alpha_1\theta_1,\ \alpha_2\theta_2,\ldots,\ \alpha_n\theta_n)$, avec lequel nous entrons dans la

deuxième étape de la méthode. Si nous notons : $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ et $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ et "^{t}" l'opération de transposition, alors nous avons :

$$\underline{\alpha}^t\underline{\theta} = (\alpha_1\theta_1, \alpha_2\theta_2, \ldots, \alpha_n\theta_n) \tag{2.54}$$

2.5.3. Exemple du modèle binomial-bêta (5) et Divers estimateurs du quasi-échantillon

Revenons à l'exemple de la variable binomiale $X(k,\theta)$. Soit $\mathbf{x}_i = (x_i, k_i)$ un résultat d'expérience. Les estimateurs des moments et du maximum de vraisemblance de θ sont

$$\begin{cases} \theta_i = (x_i+1)/(k_i+2) \\ \theta_i = x_i/k_i \end{cases} \tag{2.55}$$

Ces valeurs $\{\theta_i\}$ constituent un quasi-échantillon de la distribution a priori bêta : $\pi(\theta) = Be(\alpha,\beta)(\theta)$.

Des équations des moments de la distribution bêta (cf. formule (2.44)), nous déduisons l'estimateur suivant :

$$\begin{cases} \alpha = \delta_\pi^{-2}(1 - \mu_\pi) - \mu_\pi \\ \beta = \alpha(\mu_\pi^{-1} - 1) \end{cases} \tag{2.56}$$

où $\delta_\pi = \sigma_\pi/\mu_\pi$ est le coefficient de variation de la variable θ. Les paramètres μ_π et δ_π sont estimés de façon empirique à partir du quasi-échantillon $\{\theta_i\}$:

$$\begin{cases} \mu_\pi = \bar{\theta} = \frac{1}{n}\sum_i \theta_i \\ \delta_\pi = d_\pi \end{cases} \tag{2.57}$$

Il est clair que $d_\pi \neq 0$, équivalent à $v_\pi \neq 0$ et revient à dire que les θ_i sont non égaux entre eux.

L'estimation par la méthode du maximum de vraisemblance est relativement compliquée parce que la fonction gamma $\Gamma(x)$ intervient dans la fonction de vraisemblance du quasi-échantillon $(\theta_1, \theta_2, \ldots, \theta_n)$:

$$L(\alpha,\beta;\ \theta_1,\theta_2,\ldots,\theta_n) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^n \left(\prod_i \theta_i\right)^{\alpha-1} \left(\prod_i (1-\theta_i)\right)^{\beta-1} \tag{2.58}$$

d'où l'équation de vraisemblance :

$$\begin{cases} \psi(\alpha) - \psi(\alpha+\beta) = \overline{\log\theta} = \frac{1}{n}\sum_i \log\theta_i \\ \psi(\beta) - \psi(\alpha+\beta) = \overline{\log(1-\theta)} = \frac{1}{n}\sum_i \log(1-\theta_i) \end{cases} \tag{2.59}$$

Comme pour l'estimateur du maximum de vraisemblance (2.49), le calcul doit être fait à l'aide de la méthode d'itération.

2.6. EXEMPLE NUMERIQUE ET GRAPHIQUE

Nous continuons à étudier l'exemple de la distribution binomiale $B(k,\theta)$ avec des données et comparons les estimateurs développés ci-dessus.

Dans une expérience d'un nombre k d'objets, nous trouvons x unités (ou nombres) présentant le caractère appelé par convention "défectueux". Dans ce cas, la variable X suit une distribution binomiale $B(k,\theta)$ avec un taux θ inconnu.

Nous avons montré aux paragraphes précédents que la distribution a priori de θ est une distribution bêta :

$$\pi(\theta) = Be(\alpha,\beta)(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{2.60}$$

2.6.1. Résumé des estimateurs

Supposons $\mathbf{x}_1 = (x_1,k_1)$, $\mathbf{x}_2 = (x_2,k_2)$, ..., $\mathbf{x}_n = (x_n,k_n)$ n échantillons (ou événements) de la variable binomiale X avec x_i nombre entier non négatif tel que $0 \le x_i \le k_i$. La probabilté de chacun des événements :

$$f(\mathbf{x}_i|\theta) = C_{k_i}^{x_i}\ \theta^{x_i}\ (1-\theta)^{k_i-x_i} \quad .$$

Résumons 6 estimateurs des paramètres de la distribution a priori :

(1) Méthode séquentielle SS :

$$\begin{cases} \alpha = KT\cdot\hat{\theta} + 1 & = n\cdot m + 1 \\ \beta = KT(1 - \hat{\theta}) + 1 & = n(KM - m) + 1 \end{cases}$$

(2) Méthode des échantillons équilibrés EE :

$$\begin{cases} \alpha = KH\cdot\bar{\theta} + 1 \\ \beta = KH\cdot(1 - \bar{\theta}) + 1 \end{cases}$$

(3) Méthode des moments OO :

$$\begin{cases} \alpha = \dfrac{m^2(1-\hat{\theta}) - v\hat{\theta}}{v - m(1-\hat{\theta})} = \dfrac{m(KM-m) - v}{v\cdot KM - m(KM-m)}m \\ \beta = \alpha(1/\hat{\theta} - 1) \end{cases}$$

(4) Méthode du maximum de vraisemblance :

$$\begin{cases} \psi(\alpha+\beta+k) - \psi(\alpha+\beta) = \dfrac{1}{n}\sum\psi(\alpha+x_i) - \psi(\alpha) \\ \psi(\alpha+\beta+k) - \psi(\alpha+\beta) = \dfrac{1}{n}\sum\psi(\beta+k-x_i) - \psi(\beta) \end{cases}$$

(5) Méthode des doubles moments MM :

$$\begin{cases} \alpha = d_\pi^{-2} - (d_\pi^{-2} + 1)\bar{\theta} \\ \beta = \alpha(1/\bar{\theta} - 1) \end{cases}$$

(6) Méthode des doubles maxima :

$$\begin{cases} \psi(\alpha) - \psi(\alpha+\beta) = \overline{\log\theta} \\ \psi(\beta) - \psi(\alpha+\beta) = \overline{\log(1-\theta)} \end{cases}$$

avec les notations suivantes :

n — nombre d'expériences (d'échantillons)

$KT = \sum k_i$ — nombre total des nombres (ou unités)

$KM = \frac{1}{n}\sum k_i$ — moyenne des nombres $\{k_i\}$

$KH = \left(\frac{1}{n}\sum k_i^{-1}\right)^{-1}$ — moyenne harmonique des nombres $\{k_i\}$

$m = \frac{1}{n}\sum x_i$ — moyenne des unités défectueuses $\{x_i\}$

$v = \frac{1}{n}\sum (x_i-\mu)^2$ — variance des unités défectueuses $\{x_i\}$

$\hat{\theta} = \sum x_i / \sum k_i = \sum k_i\theta_i / \sum k_i$ — moyenne empirique total de l'échantillon composé $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$

$\bar{\theta} = \frac{1}{n}\sum \theta_i$ — moyenne des taux empiriques $\{\theta_i = x_i/k_i\}$

$d_\pi = \sqrt{v_\pi}/\mu_\pi$ — coefficient de variation empirique de $\{\theta_i\}$ avec $\mu_\pi = \theta_\pi = \bar{\theta}$ et v_π la variance empirique.

2.6.2. Caractéristiques a priori et a posteriori du taux

Supposons maintenant une expérience réalisée $\mathbf{x} = (x,k)$ dans un lot (L). Nous avons à estimer le taux de défectueux θ dans ce lot en tenant compte de l'information a priori fondée sur n expériences.

Nous avons vu comment utiliser cette information a priori pour établir un préavis sur le taux. Ce préavis est exprimé en une distribution $\pi(\theta) = \pi_0(\theta)$, distribution bêta qui sera ensuite corrigée par l'expérience aquise $\mathbf{x}$.

D'après le théorème de Bayes, la distribution a posteriori du taux θ est aussi une distribution bêta :

$$\pi_1(\theta) \propto f(\mathbf{x}|\theta)\pi(\theta) = C_k^x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha+x-1}(1-\theta)^{\beta+(k-x)-1} \tag{2.61}$$

d'où les paramètres indexés par "$_1$" :

$$\begin{cases} \alpha_1 = \alpha + x \\ \beta_1 = \beta + k - x \end{cases} \tag{2.62}$$

Nous indiquons commme caractéristiques des distributions a priori et a posteriori le mode (valeur qui rend la densité maximum, donc appelée aussi l'estimateur du maximum de vraisemblance) et la moyenne du taux θ, étant l'estimateur de Bayes minimisant le coût quadratique (cf. paragraphe 1.4.5):

$$\begin{cases} \hat{\theta}_{\pi 0} = \dfrac{\alpha - 1}{\alpha + \beta - 2} & \text{si } \alpha > 1 \text{ et } \alpha+\beta > 2 \\ \bar{\theta}_{\pi 0} = \mu_{\pi 0} = \dfrac{\alpha}{\alpha + \beta} \end{cases} \tag{2.63}$$

$$\begin{cases} \hat{\theta}_{\pi 1} = \dfrac{\alpha_1 - 1}{\alpha_1 + \beta_1 - 2} = \dfrac{\alpha + x - 1}{\alpha + \beta + k - 2} \\ \bar{\theta}_{\pi 1} = \mu_{\pi 1} = \dfrac{\alpha_1}{\alpha_1 + \beta_1} = \dfrac{\alpha + x}{\alpha + \beta + k} \end{cases} \tag{2.64}$$

Nous pouvons aussi associer à l'estimation du taux θ un intervalle de confiance pour une probabilité donnée p, soit (a,b), défini par :

$$p = \mathbb{P}^{\pi}\{a < \theta < b | \mathbf{x} = (x,k)\}$$

$$= \int_b^a Be(a+x, \beta+k-x)(\theta)d\theta$$

$$= \frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+x)\Gamma(\beta+k-x)} \int_b^a \theta^{\alpha+x-1}(1-\theta)^{\beta+k-x-1} d\theta$$

Les résultats déterminés respectivement par les quatres méthodes (1), (2), (3) et (5) sont donnés dans le tableau suivant, où le paramètre α est déterminé par l'estimateur correspondant décrit dans le paragraphe 2.6.1.

Notons $\tilde{\theta} = x/k$ le taux empirique pour le lot. Nous pouvons constater que le résultat bayésien est un peu près une pondération entre l'information a priori et l'information provenant du lot présent représentés respectivement par (θ_0, k_0) et $(\tilde{\theta}, k)$. En effet, pour chaque méthode, nous pouvons exprimer le résultats d'une manière formelle :

$$\theta_1 \approx \frac{k_0\theta_0 + k\tilde{\theta}}{k_0 + k} \tag{2.65}$$

avec le taux a posteriori $\theta_1 = \bar{\theta}_{\pi 1}$ ou $\hat{\theta}_{m1}$ et le taux a priori $\theta_0 = \bar{\theta}_{\pi 0}$ ou $\hat{\theta}_{\pi 0}$.

	l'a priori		l'a posteriori		Caractéristique de l'a priori
	$\bar{\theta}_{\pi 0}$	$\hat{\theta}_{\pi 0}$	$\bar{\theta}_{\pi 1}$	$\hat{\theta}_{\pi 1}$	
SS	$\frac{\hat{\theta}+1/KT}{1+2/KT}$	$\hat{\theta}$	$\frac{KT\hat{\theta} + k\tilde{\theta}+1}{KT + k+2}$	$\frac{KT\hat{\theta} + k\tilde{\theta}}{KT + k}$	$k_0 = KT$ $\theta_0 = \hat{\theta}$
EE	$\frac{\bar{\theta}+1/KH}{1+2/KH}$	$\bar{\theta}$	$\frac{KH\bar{\theta} + k\tilde{\theta}+1}{KH + k+2}$	$\frac{KH\bar{\theta} + k\tilde{\theta}}{KH + k}$	$k_0 = KH$ $\theta_0 = \bar{\theta}$
OO	$\hat{\theta}$	$\hat{\theta}\times\frac{\alpha-1}{\alpha-2\hat{\theta}}$	$\frac{k_0\hat{\theta} + k\tilde{\theta}}{k_0 + k}$	$\frac{k_0\hat{\theta} + k\tilde{\theta}-1}{k_0 + k-2}$	$k_0 = \alpha/\hat{\theta}$ $\theta_0 = \hat{\theta}$
MM	$\bar{\theta}$	$\bar{\theta}\times\frac{\alpha-1}{\alpha-2\bar{\theta}}$	$\frac{k_0\bar{\theta} + k\tilde{\theta}}{k_0 + k}$	$\frac{k_0\bar{\theta} + k\tilde{\theta}-1}{k_0 + k-2}$	$k_0 = \alpha/\bar{\theta}$ $\theta_0 = \bar{\theta}$

Tableau 2.2 Moyennes et modes a priori et a posteriori du taux θ

Nous considérons ici la variable θ comme le taux de défectueux dans le cas du contrôle de la qualité. Dans ce cas, le taux empirique total $\hat{\theta}$ et la moyenne des taux empiriques $\bar{\theta}$ sont inférieurs à 1/2. Donc, il est clair que la distribution a priori de la variable θ est toujours asymétrique, et que nous avons $\bar{\theta}_{\pi 0} < \hat{\theta}_{\pi 0}$, c'est à dire que la moyenne a priori est supérieure au maximum de vraisemblance, pour chaque méthode utilisée :

SS : $$\bar{\theta}_{\pi 0} = \frac{\hat{\theta} + 1/(n\cdot KM)}{1 + 2/(n\cdot KM)} > \hat{\theta}_{\pi 0} = \hat{\theta}$$

EE : $$\bar{\theta}_{\pi 0} = \frac{\bar{\theta} + 1/KH}{1 + 2/KH} > \hat{\theta}_{\pi 0} = \bar{\theta}$$

OO : $$\bar{\theta}_{\pi 0} = \hat{\theta} > \hat{\theta}_{\pi 0} = \hat{\theta}\frac{\alpha - 1}{\alpha - 2\hat{\theta}}$$

MM : $$\bar{\theta}_{\pi 0} = \bar{\theta} > \hat{\theta}_{\pi 0} = \bar{\theta}\frac{\alpha - 1}{\alpha - 2\bar{\theta}}$$

Bien entendu, pour les méthodes des moments OO et des doubles moments MM, il

faut que la condition $\alpha > 1$ soit satisfaite.

Parmi ces estimateurs de la distribution a priori, seul l'estimateur séquentiel dépend du nombre n des échantillons. La moyenne a priori tend vers le maximum de vraisemblance quand n est assez grand : $\bar{\theta}_{\pi 0} \longrightarrow \hat{\theta}_{\pi 0}$.

2.6.3. Résultats numériques et Comparaison des méthodes

Nous envisagerons deux cas différents : dans le premier cas, les expériences ont le même nombre de tirages k et des unités défectueuses différentes x = 1, 3, 5, 7 ; dans le second, les expériences ont le même nombre d'unités défectueuses x et des nombres de tirages différents k = 30, 40, 60, 70. Chacun des cas a ici quatre échantillons.

Pour comparer les méthodes de détermination de la distribution a priori, nous prenons, à titre d'exemple, k = 40 et x = 5 respectivement. Nous calculons également, pour k = 10, 20, 30,... et pour x = 1, 2, 3, ..., dans les deux cas, les valeurs (α,β), la moyenne $\bar{\theta}_{\pi 0}$, le maximum de vraisemblance $\hat{\theta}_{\pi 0}$ et les α-fractiles correspondant aux probabilités $\alpha = 0.90$ et 0.95. Le taux emprique total correspondant est $\hat{\theta} = \bar{\theta} = 0.4$, 0.2, 0.133, ... dans le premier cas et $\hat{\theta} = 0.02$, 0.04, 0.6, ... dans le second, mais la moyenne des taux empiriques n'est pas la même : $\bar{\theta} = 0.022$, 0.045, 0.067, ... Les résultats sont donnés à l'annexe 1.

Supposons qu'une expérience soit réalisée dans un lot : $\mathbf{x} = (4,50)$. La distribution a posteriori du taux θ est évaluée à partir de l'information (a posteriori) résultant de deux informations du passé et du présent, elle est constituée par n+1 = 5 expériences.

Information a priori 4 expériences	+	Expérience **x** du lot $\tilde{\theta} = 0.08$	$\Longrightarrow$	Information résultante 5 expériences

Nous donnons les résultats dans les tableaux 2.3 et 2.4. Les statistiques des quatre échantillons sont respectivement donnés ci-dessous.

Cas 1. Nombre de tirages constant k = 40

KT = 160 KM = KH = 40 m = 4 v = 5

$\hat{\theta} = \bar{\theta} = 0.10$ $d_{\pi}^{-2} = 16/5$

Cas 2. Nombre unités défectueuses constant x = 5

KT = 200 KM = 50 KH = 44.80 m = 5 v = 0

$\hat{\theta} = 0.10$ $\bar{\theta} = 0.1282$ $d_{\pi}^{-2} = 8.8583$

	Résultats a priori		Résultats a posteriori		Ecart
	(α,β)	$\bar{\theta}_{\pi 0}$	(α_1,β_1)	$\bar{\theta}_{\pi 1}$	$\Delta = \lvert\bar{\theta}_{\pi 0} - \bar{\theta}_{\pi 1}\rvert$
SS	(17, 145)	.1049	(21, 191)	.0991	.00588
EE	(5, 37)	.1190	(9, 83)	.0978	.02122
OO	(9.93,89.36)	.1	(13.93,135.36)	.0933	.00670
MM	(2.78,25.02)	.1	(6.78,71.02)	.0871	.01285

Tableau 2.3 Taux a priori, empirique et a posteriori pour le 1er cas

	Résultats a priori		Résultats a posteriori		Ecart
	(α,β)	$\bar{\theta}_{\pi 0}$	(α_1,β_1)	$\bar{\theta}_{\pi 1}$	$\Delta = \lvert\bar{\theta}_{\pi 0} - \bar{\theta}_{\pi 1}\rvert$
SS	(21, 181)	.1040	(25, 227)	.0992	.00475
EE	(6, 40.8)	.1282	(10, 86.8)	.1033	.02490
OO	non valable à cause de v = 0 (cf. paragraphe 2.4.2)				
MM	(7.76,61.75)	.1116	(11.76,107.75)	.0984	.01322

Tableau 2.4 Taux a priori, empirique et a posteriori pour le 2ème cas

Dans cet exemple, le taux empirique est $\tilde{\theta} = 0.08$ et le lot (L) présente une diminution du taux θ par rapport au taux a priori θ_0 voisin de 0.10. Il est naturel que le taux a posteriori soit aussi diminué.

Nous définissons l'écart entre deux estimateurs a priori $\bar{\theta}_{\pi 0}$ et a posteriori $\bar{\theta}_{\pi 1}$ par $\Delta = \bar{\theta}_{\pi 0} - \bar{\theta}_{\pi 1} > 0$. La valeur Δ est l'une des caractéristiques décrivant la différence entre les quatres méthodes. Cette différence est en effet dûe à l'interprétation de l'information a priori par $(\bar{\theta}_{\pi 0}, k_0)$.

Rappelons que le taux a posteriori $\theta_1 = \bar{\theta}_{\pi 1}$ est un peu près une pondération entre l'information a priori (θ_0, k_0) et l'information du présent $(\tilde{\theta}, k)$. Comme le taux a priori et le taux empirique sont relativement voisins par rapport aux paramètres k_0 et k, la représentation de ces deux informations dans le taux θ_1 dépend essentiellement du rapport k/k_0, ainsi l'écart Δ :

$$\Delta = \bar{\theta}_{\pi 0} - \bar{\theta}_{\pi 1} \approx \frac{k}{k_0 + k}(\bar{\theta}_{\pi 0} - \tilde{\theta}) \qquad (2.66)$$

Au vu de ces tableaux, nous remarquons que les méthodes séquentielle SS et des moments OO donnent des corrections par l'expérience moins grandes que par les méthodes des échantillons équilibrés EE et des doubles moments MM.

Dans le premier cas du nombre k constant, nous avons les estimateurs de la distribution a priori :

$$\text{SS} : \begin{cases} \alpha = n \cdot m + 1 = 17 \\ \beta = n(KM-m) + 1 = 4k - 15 \end{cases} \qquad k_0 = KT = 4k = 160$$

$$\text{EE} : \begin{cases} \alpha = m + 1 = 5 \\ \beta = KH - m + 1 = k - 3 \end{cases} \qquad k_0 = k = 40$$

$$\text{OO} : \begin{cases} \alpha = m \times \dfrac{m(KM-m) - v}{KM \cdot v - m(KM-m)} = 4\dfrac{4k-21}{k+16} \overset{<}{\longrightarrow} 16 \\ \beta = \alpha(KM/m - 1) = (k-4)\dfrac{4k-21}{k+16} \end{cases} \qquad k_0 = \alpha/\hat{\theta} = k\frac{4k-21}{k+16} = 198.56$$

$$\text{MM} : \begin{cases} \alpha = d_\pi^{-2} - (d_\pi^{-2}+1)\dfrac{m}{k} = 4\dfrac{4k-21}{5k} \overset{<}{\longrightarrow} 3.2 \\ \beta = \alpha(k/m - 1) = (k-4)\dfrac{4k-21}{5k} \end{cases} \qquad k_0 = \alpha/\bar{\theta} = \frac{4k-21}{5} = 27.8$$

Nous en déduisons que pour les méthodes SS et OO, le paramètre k_0 est de l'ordre de grandeur de KT = 4k ou près de 4k, tandis que pour les méthodes EE et MM, k_0 n'est que de l'ordre de grandeur de KH = k ou près de $\frac{4}{5}$k. En effet, les méthodes EE et MM éliminent de façon directe ou indirecte les effets dûs aux nombres $\{k_i\}$: la méthode EE prend en compte les scores $\{K/k_i\}$ et la méthode MM considère les taux empiriques $\{\theta_i = x_i/k_i\}$. C'est pour cela que les rapports k/k_0 sont plus grands ici que dans les méthodes SS et OO et les corrections pour une expérience quelconque seront plus importantes.

Du fait que $\alpha \approx k_0\theta_0$ (cf. paragraphe 2.6.1), si le paramètre k_0 est grand, alors le paramètre α l'est aussi. Cependant pour les méthodes OO et MM le paramètre α est limité respectivement à α = 16 et 3.2.

Quant à la méthode EE, le nombre k_0 = KH ne dépend pas du nombre n des échantillons et il est inférieur à KM la moyenne des nombres $\{k_i\}$, KH ≤ KM, grâce à l'inégalité de Cauchy :

$$\left(\sum a_i b_i\right)^2 \leq \left(\sum a_i^2\right)\left(\sum b_i^2\right) \tag{2.67}$$

pour tous réels a_i et b_i. Dans notre cas, il suffit de prendre $a_i^2 = k_i$ et $b_i^2 = 1/k_i$.

Par contre, pour la méthode séquentielle SS, nous avons k_0 = KT = 4·KM, et $\alpha = n \cdot m + 1$. Néanmoins, pour utiliser avec succès cette méthode, il faut limiter le nombre n pour que l'information a priori ne pèse pas trop.

Cette remarque est aussi vraie pour d'autres distributions.

CHAPITRE III FAMILLES PARTICULIÈRES DE DISTRIBUTIONS CONJUGUÉES ET MODÈLES BAYÉSIENS CORRESPONDANTS

3.0. OBJECTIF ET COMMENTAIRES

Dans l'approche paramétrique, nous avons vu que l'utilisation de familles de distributions conjuguées est essentielle dans l'étude des modèles continus car elle simplifie le passage de la distribution a priori à la distribution a posteriori. Il s'agit en effet d'un changement des paramètres ξ (ou hyperparamètres) des distributions du paramètre θ (cf. paragraphe 1.6).

Nous rappelons que les distributions bayésiennes ou prédictives sont obtenues à partir des probabilités des observations réalisées antérieurement d'une variable X, pondérées des distributions a priori et a posteriori. Dans un contrôle statistique de la qualité, on étudie souvent le rapport entre la probabilité d'acceptation et le niveau réel de qualité (proportion de défectueux, valeur moyenne du caractère mesuré, valeur correspondant à un certain fractile ...), afin de pouvoir décider d'accepter ou refuser un lot examiné selon les règles de conformité.

Ces distributions interviennent aussi dans les études de fiabilité : estimation, à partir d'un échantillon recueilli pendant un temps spécifié, des paramètres d'une loi de survie et des fonctions qui en découlent (durée moyenne de vie, ...) ; et fixation des règles de décision (acceptation ou rejet ou autre) d'un lot d'éléments dont la loi de survie est continue en fonction de l'information fournie par un échantillon, ...

L'objectif du présent chapitre est de développer et de découvrir des familles de distributions conjuguées unidimensionnelles et les distributions bayésiennes correspondantes, à partir des distributions particulières (sous-jacentes) $f(x|\theta)$.

Le lecteur peut en première lecture passer directement au résumé à la

fin du chapitre, considérer ce chapitre comme une annexe.

3.0.1. Deux familles générales et Leurs modèles bayésiens

L'hypothèse de normalité est très fréquente dans l'analyse statistique. Elle est essentiellement due au théorème limite central : si "une variable" est la résultante d'un grand nombre de causes, petites, à effet additif, cette variable suit une loi normale. En termes plus formels, la somme d'un grand nombre de variables aléatoires indépendantes, à variances uniformément bornées, a, après une normalisation adéquate, une loi asymptotiquement normale (gaussienne). Ce fameux résultat garantit alors que la distribution normale $N(\mu,\sigma^2)$ est une bonne approximation, dans bien nombreux cas, de la réalité. Cette distribution présente l'avantage d'offrir une grande facilité pour exploiter les statistiques, ceci l'a rendu extrêment populaire.

Nous allons développer les modèles probabilistes issus de la distribution normale. Dans le chapitre IV, nous préciserons et comparerons différentes formes du modèle normal-gamma : soit empirique $NG(m,v,k)$ ou "théorique" $NG(m,v,k,n)$, cette dernière est utilisée dans certains travaux actuels (cf. [36]).

Comme nous l'avons expliqué au début de ce travail, il arrive souvent que la normalité des distributions soit mal vérifiée pour diverses raisons : non-symétrie, aplatissement trop lent ou trop rapide, troncatures ; nous essayerons donc d'étendre cette étude aux divers cas des distributions normale inverse, log-normale, gamma, gamma inverse, de Pareto et aux distributions de valeurs extrêmes comme Weibull et Fréchet.

Nous verrons que ces distributions peuvent en effet être regroupées en deux familles: distributions normale, log-normale et normale inverse constituant la <u>famille normale</u>; et distributions gamma, gamma inverse, de Weibull et de Fréchet constituant une autre famille, dite <u>famille gamma</u>.

Toutes ces distributions dépendent de deux paramètres θ_1 et θ_2.

Pour la famille normale, le modèle du couple $\underline{\theta} = (\theta_1,\theta_2)$ sera du type normal-gamma ; si θ_2 est une dispersion connue, la variable θ_1, représentant

une position de la distribution, sera adaptée au modèle normal ; par contre, si θ_1 est fixé, la variable θ_2 suivra alors une distribution gamma.

Quant à la famille gamma, nous trouvons une représentation de la forme générale: $G(\upsilon,\nu,x_0,\alpha,\beta)$, dite <u>distribution gamma généralisée</u>, avec des distributions particulières: (1) gamma $G(\alpha,\beta) = G(1,1,0,\alpha,\beta)$; (2) gamma inverse $GI(\alpha,\beta) = G(1,-1,0,\alpha,\beta)$; (3) Weibull $W(\alpha,\beta) = G(-\alpha,\alpha,\varepsilon,\alpha,\beta)$ et (4) Fréchet $Fr(\alpha,\beta) = G(\alpha,-\alpha,\varepsilon,\alpha,\beta)$. Si l'on ne considère que la variable $\theta = \theta_2 = \beta^{\upsilon}$ et que α est fixé, alors quelle que soit une distribution de telle forme, le modèle bayésien sera gamma.

3.0.2. Estimation des paramètres connus

Rappelons que dans le cas dominé, $f(x|\theta)$ est la densité de probabilité de X relative à la mesure de Lebesgue υ sur l'espace euclidien $\mathbb{R}^k$, θ étant un paramètre. Nous notons aussi $f(\mathbf{x}|\theta) = L(\theta|\mathbf{x}) = L_{\mathbf{x}}(\theta)$ la fonction de vraisemblance, fonction de θ seul si $\mathbf{x} = (x,k)$ est un échantillon avec $k > 1$.

Si l'on veut étudier un des paramètres soit par exemple $\theta = \theta_2$ dans ce cas, on suppose θ_1 et θ_3 connus, constants et estimés par l'une des méthodes classiques comme celle du maximum de vraisemblance, notés par $\hat{\theta}_1$ et $\hat{\theta}_3$. Les distributions bayésiennes (ou prédictives a priori et a posteriori) dépendront bien sûr de ces valeurs $\hat{\theta}_1$ et $\hat{\theta}_3$.

Une autre approche consisterait à approximer les distributions de paramètres θ_1 et θ_3 par des distributions normales :

$$\theta_1 \text{ --- } N\left(\hat{\theta}_1,\sigma_{11}^2\right) \qquad \text{et} \qquad \theta_3 \text{ --- } N\left(\hat{\theta}_3,\sigma_{33}^2\right) \tag{3.00}$$

avec σ_{11}^2 et σ_{33}^2 définies par $\sigma_{11}^2 = (1,0,0)^t\Sigma(1,0,0)$ et $\sigma_{33}^2 = (0,0,1)^t\Sigma(0,0,1)$, où

$$\Sigma^{-1} = \left\{ -\frac{\partial^2 L\left(\theta_1,\theta_2,\theta_3;x\right)}{\partial\theta_i\partial\theta_j} \; ; \; i,\ j = 1,2,3 \right\}$$

est l'inverse de la matrice de variances-covariances. Nous nous limiterons dans ce chapitre au premier cas.

En général, l'indice inférieur "$_0$" est utilisée pour le paramètre de la distribution a priori et l'indice inférieur "$_1$" l'est pour celui de la distribution a posteriori. Les lettres grecques désignent souvent les variables des paramètres des distributions considérées. Signalons que nous omettrons "$_0$" dans le texte s'il n'y a pas de confusion possible et nous chercherons les distributions bayésiennes a priori et a posteriori de même forme ; seule changera l'indexation du paramètre par "$_0$" ou par "$_1$".

Dans ce chapitre, nous supposons avoir à notre disposition un échantillon de taille k, que nous désignons par $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. Les valeurs $\{x_i\}$ sont prises par le processus $X = (X_i)_{1 \le i \le k}$ où les variables $\{X_i\}$ sont indépendantes de même loi, loi que nous cherchons à déterminer.

3.1. DISTRIBUTION NORMALE (LAPLACE-GAUSS)

La distribution normale est entièrement caractérisée par sa moyenne μ et son écart-type $\sigma > 0$, puisque sa fonction de densité s'écrit :

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{3.10}$$

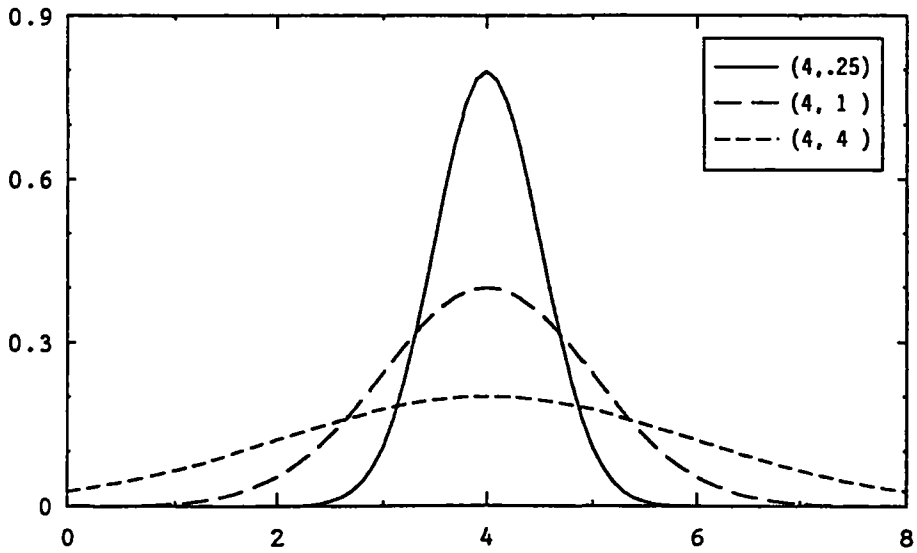


Figure 3.1 Distribution normale

Le graphique montre les courbes des densités des distributions normales correspondant à la même moyenne et aux différentes valeurs de variance, indiquées sur le graphe 3.1.

Rappelons que par définition, le mode noté M_o est la valeur où la densité atteint son maximum dans le cas de monomode. Pour une distribution normale, le mode coïncide avec la moyenne μ, c'est à dire que la courbe est toujours symétrique autour de cette valeur. Lorsque l'écart-type σ augmente, la distribution devient plus dispersée.

Trois cas se présentent suivant que (1) la moyenne μ et l'écart-type σ sont inconnus ; (2) μ est inconnu avec σ connu ; et (3) σ est inconnu avec μ connu.

3.1.1. μ et σ inconnus

Fonction de vraisemblance

Soit $\mathbf{x}$ un échantillon d'une distribution normale de moyenne μ et écart-type σ inconnus. Pour simplifier les formules, posons $\eta = 1/(2\sigma^2)$, η est appelée précision de la distribution normale $N(\mu,\sigma^2)$. La fonction de densité de $\mathbf{x}$ est:

$$f(\mathbf{x}|\mu,\eta) \propto \sqrt{\frac{\eta}{\pi}}^{k} \exp\left[-\eta\sum_i (x_i - \mu)^2\right] \tag{3.110}$$

Désignons la moyenne et la variance empiriques de l'échantillon $\mathbf{x}$ par :

$$m = \frac{1}{k}\sum_i x_i \qquad \text{et} \qquad v = \frac{1}{k}\sum_i (x_i - m)^2 \quad .$$

Grâce à l'égalité $\sum(x_i-\mu)^2 = k(m-\mu) + \sum(x_i-m)^2$, nous obtenons la fonction de vraisemblance en fonction de m, v et k :

$$f(\mathbf{x}|\mu,\eta) \propto \eta^{k/2}\exp\left[-\eta k(\mu-m)^2\right]\exp(-\eta k v) \tag{3.111}$$

ceci implique qu'une statistique exhaustive $T(\mathbf{x})$ par rapport au couple (μ,η) des variables est le vecteur (m,v,k), défini par extension sur $\mathbb{R}\times\mathbb{R}^+$.

Distribution a priori

Le principe de la conjuguée de la fonction de vraisemblance nous permet de faire la démarche essentielle :

— attribuer au couple (μ,η), une distribution de probabilité :

$$\pi(\mu,\eta) \propto \eta^{k_0/2}\exp\left[-\eta k_0(\mu-m_0)^2\right]\exp\left(-\eta k_0 v_0\right) \tag{3.112}$$

avec (m_0, v_0, k_0), $v_0 > 0$ et $k_0 > 0$. Après normalisation, elle devient une distribution appelée distribution normale-gamma et notée par :

$$NG\left(\mu,\eta;\ m_0, v_0, k_0\right) = \sqrt{\frac{\eta k_0}{\pi}}\cdot\exp\left[-\eta k_0(\mu-m_0)^2\right]\cdot\frac{(k_0 v)^{(k_0+1)/2}}{\Gamma\left(\frac{k_0+1}{2}\right)}\eta^{(k_0-1)/2}e^{-\eta k_0 v_0} \tag{3.113}$$

Il n'est pas difficile de montrer la propriété suivante :

$$\pi(\mu,\eta) = \pi(\mu|\eta)\pi(\eta) = N\left(\mu;\ m_0, (2k_0\eta)^{-1}\right)G\left(\eta;\ \frac{k_0+1}{2}, k_0 v_0\right) \tag{3.114}$$

avec $\pi(\mu|\eta)$ distribution normale de moyenne m_0 et de précision $k_0\eta$ et $\pi(\eta)$ distribution gamma de paramètres $(k_0+1)/2$ et $k_0 v_0$.

Distribution a posteriori

Ayant à notre disposition un échantillon $\mathbf{x}$, ainsi que sa statistique exhaustive (m, v, k) , la distribution de la variable (μ,η) est modifiée selon le théorème de Bayes :

$$\pi(\mu,\eta|\mathbf{x}) \propto f(\mathbf{x}|\mu,\eta)\pi(\mu,\eta)$$

$$\propto \eta^{(k+k_0)/2}\exp\left[-\eta\left(k(\mu-m)^2+k_0(\mu-m_0)^2\right)\right]\exp\left[-\eta\left(kv+k_0 v_0\right)\right]$$

Il en résulte que la distribution a posteriori de (μ,η) est aussi normale-gamma avec le paramètre (m_1, v_1, k_1) défini ci-dessous :

$$\pi(\mu,\eta|\mathbf{x}) = NG\left(\mu,\eta;\ m_1, v_1, k_1\right) \tag{3.115}$$

$$\begin{cases} k_1 = k_0 + k \\ k_1 m_1 = k_0 m_0 + km \\ k_1 v_1 = k_0 v_0 + kv + (m-m_0)^2 kk_0/k_1 \end{cases} \tag{3.116}$$

Ainsi, les distribution normale-gamma constituent une famille de distributions conjuguées par rapport à la famille de distributions normales.

Opération de conjuguaison et Stabilité de "(m,v,k)"

Il n'est pas étonnant que la statistique exhaustive "(m,v,k)" satisfasse une stabilité au sens suivant :

Soit $\mathbf{x}_0 = (x_1^0, x_2^0, \ldots, x_{k0}^0)$ un échantillon de la distribution normale. Si le paramètre de la distribution a priori (m_0, v_0, k_0) est estimé à partir de cet échantillon, donc la statistique exhaustive $T(\mathbf{x}_0)$. Après l'obtention de $\mathbf{x} = (x_1, x_2, \ldots, x_k)$, le paramètre de la distribution a posteriori (m_1, v_1, k_1) est la statistique exhaustive de l'échantillon composé $\mathbf{x}_1 = (x_1^1, x_2^1, \ldots, x_{k1}^1)$ $= \mathbf{x} \cup \mathbf{x}_0 = (x_1, x_2, \ldots, x_k, x_1^0, x_2^0, \ldots, x_{k0}^0)$ avec $k_1 = k + k_0$, parce que nous avons

$$k_1 m_1 = \sum_{i=1}^{k+k_0} x_i^1 = \sum_{i=1}^{k} x_i + \sum_{i=1}^{k_0} x_i^0 = km + k_0 m_0$$

$$k_1 v_1 = \sum_{i=1}^{k+k_0} (x_i^1 - m_1)^2 = \sum_{i=1}^{k} (x_i - m_1)^2 + \sum_{i=1}^{k_0} (x_i^0 - m_1)^2$$

$$= kv + k(m - m_1)^2 + k_0 v_0 + k_0 (m_0 - m_1)^2$$

$$= kv + k_0 v_0 + kk_0 (m - m_0)^2 / k_1 \tag{3.117}$$

Soit $T(\mathbf{x}) = (m, v, k)$ la statistique exhaustive de l'échantillon $\mathbf{x}$. Comme on a $f(\mathbf{x}|\theta) \propto g(\theta|T(\mathbf{x}))$, la stabilité précédente entraine que :

$$g(\theta|T(\mathbf{x} \cup \mathbf{x}_0)) \propto g(\theta|T(\mathbf{x})) \times g(\theta|T(\mathbf{x}_0)) \quad .$$

C'est la propriété développée aux paragraphes 1.6.2 et 2.1.2.

Distribution bayésienne

La distribution bayésienne de la variable X est obtenue à l'aide de la distribution conditionnelle de X sachant (μ,η), pondérée par la distribution a priori de (μ,η). Ceci nous amène à intégrer leurs fonctions de densité :

$$
\begin{aligned}
p(x) &= \iint_{\mathbb{R}\times\mathbb{R}^+} f(x|\mu,\eta)\pi(\mu,\eta)d\mu d\eta \\
&= \int_{\mathbb{R}^+}\left(\int_{\mathbb{R}} f(x|\mu,\eta)\pi(\mu|\eta)d\mu\right)\pi(\eta)d\eta \\
&= \int_{\mathbb{R}^+}\left(\int_{\mathbb{R}} \frac{\sqrt{k}}{\pi}\eta\exp\left[-\eta(x-\mu)^2-k\eta(\mu-m)^2\right]d\mu\right)\pi(\eta)d\eta \\
&\propto \int_{\mathbb{R}^+} \exp\left[-\eta\frac{k}{k+1}(x-m)^2\right]\eta^{k/2}\exp\left(-\eta kv\right)d\eta \\
&\propto \left[v + \frac{(x-m)^2}{k+1}\right]^{-(k/2+1)}
\end{aligned}
\tag{3.118}
$$

où le paramètre (m,v,k) peut être soit (m_0,v_0,k_0) s'il s'agit d'une distribution bayésienne a priori, soit (m_1,v_1,k_1) s'il s'agit d'une distribution a posteriori.

Nous obtenons le noyau d'une distribution de Student non centrée. Alors la variable $Y = (X-m)/\sqrt{v}$ suit la distribution (centrée) de Student à (k+1) degrés de liberté et Y^2 suit la distribution de Fisher-Snedecor de paramètre 1 et (k+1), ce que nous symbolisons par $(X-m)/\sqrt{v} \approx\approx\approx T(k+1)$ et $(X-m)^2/v \approx\approx\approx F(1,k+1)$. La moyenne et la variance bayésiennes sont:

$$
\begin{cases}
M = \mathbb{E}(X) = m \\
V = \mathbb{V}(X) = \dfrac{k+1}{k-1}\cdot v \longrightarrow v \quad \text{si } k > 1 \longrightarrow \infty
\end{cases}
\tag{3.119}
$$

Puisque la distribution de Student est asymptotiquement approximée par la distribution normale réduite, $T(k+1) \approx\approx\approx N(0,1)$, cela revient à dire que la variable X est asymptotiquement normale de moyenne m et variance v. Ce

résultat confirme que ce modèle bayésien est "équivalent" au modèle classique si le paramètre (m,v,k) est la statistique exhaustive de l'échantillon $\mathbf{x}$. Tous les résultats sont illustrés ci-dessous pour les deux approches, avec $X|_{(\mu,\eta)} \approx\approx\approx N(\mu,\sigma^2)$:

bayésienne $(\mu,\eta) \approx\approx\approx NG(m,v,k)$ $\quad X \approx\approx\approx m+\sqrt{v}T(k+1) \longrightarrow N(m,v)$ pour k grand

classique $\quad \mu = m \quad \sigma^2 = v \quad X \approx\approx\approx N(m,v)$.

Remarque sur l'a priori non-informatif

Comme une distribution normale de moyenne μ et variance σ^2 inconnues a une densité de la forme $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$, nous avons vu dans 2.1.4 qu'un principe d'invariance permettait de choisir une distribution a priori: $\pi(\mu,\sigma^2) = \sigma^{-1}$. En terme de précision η, nous prendrons plutôt $\pi(\mu,\eta) = \sqrt{\eta}$ que $\pi(\mu,\eta) = \sqrt{2\eta}$.

En introduisant cette distribution a priori dans la formule séquentielle (2.11), nous obtenons finalement la distribution du paramètre $\theta = (\mu,\eta)$:

$$\pi(\mu,\eta) = N\left(\mu;\ m,\ \frac{1}{2k\eta}\right)G\left(\eta;\ \frac{k}{2}+1,\ kv\right) \quad .$$

Il existe seulement 1/2 degré de différence avec le modèle (3.114), lorsque le paramètre (m,v,k) est la statistique exhaustive de l'échantillon $\mathbf{x}$. Remarquons que l'on reste dans le cadre des distributions conjuguées par rapport à la fonction de vraisemblance.

3.1.2. μ inconnu et σ connu

Soit $\mathbf{x}$ un échantillon tiré au sort à partir d'une distribution normale de moyenne μ inconnue et d'écart-type σ connu. Par le même calcul que précédemment, la fonction de vraisemblance s'écrit :

$$f(\mathbf{x}|\mu) \propto \sigma^{-k}\exp\left[-\frac{k(\mu-m)^2}{2\sigma^2}\right] \tag{3.120}$$

donc le couple (m,k) est une statistique exhaustive par rapport à la variable μ définie sur $\mathbb{R}$.

Le principe de la conjuguée naturelle permet cette fois-ci d'attribuer à la variable μ, une distribution de probabilité <u>normale</u> avec $v_0 > 0$ fixé :

$$\pi(\mu) = N\left(\mu;\ m_0, v_0\right) \tag{3.121}$$

Nous pouvons considérons également le forme empirique de la distribution π de manière à ce que

$$\pi(\mu) = \sqrt{\frac{\eta k}{\pi}}\exp\left(-\ k_0\eta(\mu - m_0)^2\right) = N\left(\mu;\ m_0, \frac{1}{2k_0\eta} = v_0\right) \tag{3.122}$$

Une telle forme fait apparaître le rôle du paramètre "k" dans le modèle considéré, lorsque l'estimation de la distribution a priori π est réalisée à partir des échantillons antérieurs comme nous le ferons dans cette étude. La précision (ou la variance) peut être considérée soit comme constante, soit comme estimée de manière empirique à l'aide des échantillons antérieurs. Pour ces deux cas, nous verrons au chapitre IV que les estimateurs sont différents.

Il n'est pas étonnant dans ce cas que les distributions a posteriori et bayésienne soient aussi normales en tenant compte des calculs :

$$\pi(\mu|\mathbf{x}) \propto f(\mathbf{x}|\mu)\pi(\mu) \propto \exp\left[-\left(\frac{k(\mu-m)^2}{\sigma^2} + \frac{(\mu-m_0)^2}{v_0}\right)\right] \tag{3.123}$$

$$\begin{aligned} p(x) &= \int_{\mathbb{R}} \frac{1}{2\pi\sigma\sqrt{v}}\exp\left[-\ \frac{(x-\mu)^2}{\sigma^2} - \frac{(\mu-m)^2}{v}\right]d\mu \\ &= \frac{1}{\sqrt{2\pi\left(\sigma^2+v\right)}}\exp\left[-\ \frac{1}{2\left(\sigma^2+v\right)}\left(x-m\right)^2\right] \end{aligned} \tag{3.124}$$

où les paramètres m_1 et v_1 de l'a posteriori sont définis par :

$$\begin{cases} m_1 = \left(k\eta m + h_0 m_0\right)/h_1 \\ v_1 = \dfrac{1}{2h_1} \qquad \text{avec } h_1 = k\eta + h_0 \end{cases} \tag{3.125}$$

où v_1 dépend de la taille k de l'échantillon $\mathbf{x}$. La moyenne et la variance

bayésiennes sont définies par

$$\begin{cases} M = \mathbb{E}(X) = m \\ V = \mathbb{V}(X) = \sigma^2 + v \end{cases} \tag{3.126}$$

où $(m,v) = (m_0,v_0)$ ou (m_1,v_1).

Dans ce cas, nous pouvons dire que la famille de distributions normales est conjuguée par rapport à la famille de distributions normales d'écart-type σ connu.

loi $f(x\|\mu,\sigma^2)$	loi a priori $\pi(\mu)$	loi bayésienne $p(x)$
$\mathbb{E}(X\|\mu) = \mu$	$\mathbb{E}(\mu) = m_0 = m$	$\mathbb{E}(X) = m$
$\mathbb{V}(X\|\mu) = \sigma^2 = v$	$\mathbb{V}(\mu) = v_0 = v/k$	$\mathbb{V}(X) = v(1+\frac{1}{k})$

Dans le cas où le paramètre (m_0,v_0) de la distribution a priori est déterminé par la statistique exhaustive (m,v,k) d'un échantillon (x,k), les moyennes et les variances des distributions correspondantes sont indiquées dans le tableau ci-dessus. Nous constatons que le modèle est "équivalent" au modèle classique, que $V = \mathbb{V}(X)$ converge vers la variance empirique v quand k tend vers l'infini, et que le paramètre μ est souvent estimé par m, $M = \mathbb{E}(X)$ étant égale aussi à cette moyenne d'après la formule (3.126).

D'autre part, comme la densité d'une distribution normale d'écart-type σ connu peut être exprimée sous la forme $f(x-\mu)$, en utilisant un principe d'invariance nous pouvons choisir comme distribution a priori la mesure uniforme: $\pi(\mu) = 1_{\mathbb{R}}(\mu)$. A partir d'un échantillon $\mathbf{x}$, nous appliquons la formule séquentielle (2.11) et nous trouvons la même distribution du paramètre μ que précédemment (cf. formules (3.120) et (3.121)).

3.1.3. μ connu et σ inconnu

Soit $\mathbf{x}$ un échantillon d'une distribution normale de moyenne μ connue et d'écart-type σ inconnu et donc de précision η inconnue. Sa fonction de vrai-

semblance s'écrit :

$$f(\mathbf{x}|\eta) \propto \eta^{k/2}\exp\left[-\eta\left(kv+k(\mu-m)^2\right)\right] \tag{3.130}$$

donc le triplet (m,v,k) est une statistique exhaustive de $\mathbf{x}$ par rapport à la variable η, définie sur $\mathbb{R}^{+*}$.

Le principe de la conjuguée naturelle attribue à η une <u>distribution gamma</u> :

$$\pi(\eta) = G\left(\eta;\ a_0, b_0\right) \tag{3.131}$$

Pour la même raison que précédemment, nous gardons aussi la forme empirique de ce modèle comme celle conjuguée à la fonction de vraisemblance :

$$\pi(\eta) = \frac{\left(kv+k(\mu-m)^2\right)^{k/2+1}}{\Gamma\left(k/2+1\right)}\eta^{k/2}\exp\left[-\ \eta k\left(v\ +\ (\mu-m)^2\right)\right] \tag{3.132}$$

Cette forme permettra de traiter les cas où la moyenne est considérée comme constante ou comme estimée de manière empirique à l'aide des échantillons antérieurs.

Il est clair que la distribution a posteriori est aussi gamma, puisque nous avons

$$\pi(\eta|\mathbf{x}) \propto f(\mathbf{x}|\eta)\pi(\eta) \propto \eta^{k/2+a_0-1}\exp\left[-\eta\left(k(\mu-m)^2+kv+b_0\right)\right] \tag{3.133}$$

d'où le paramètre : $\begin{cases} a_1 = a_0 + \dfrac{k}{2} \\ b_1 = b_0 + k\left(v\ +\ (\mu-m)^2\right) \end{cases}$ (3.134)

Nous montrons alors que la famille de distributions gamma est conjuguée par rapport à la famille des distributions normales de moyenne connue.

Remarquons que les statistiques (a,b) satisfont aussi la stabilité au sens défini au paragraphe 3.1.1. Si le paramètre (a_0, b_0) est déterminé par la statistique exhaustive d'un échantillon $\mathbf{x}_0$ de manière à ce que $a_0 = \frac{k_0}{2} + 1$ et $b_0 = k_0\left(v\ +\ (\mu-m_0)^2\right)$ (cf. formule (3.131)), alors étant donné un échantillon $\mathbf{x}$, nous pouvons calculer :

$$k_1\left(v_1 + (\mu-m_1)^2\right) = \sum \left(x_i^1 - m_1\right)^2 + k_1\left(\mu - m_1\right)^2$$

$$= kv + k_0 v_0 + \frac{kk_0}{k_1}(m_0 - m)^2 + k_1\left(\mu - \frac{k_0 m_0 + km}{k_0 + k}\right)^2$$

$$= k\left(v + (\mu-m)^2\right) + k_0\left(v + (\mu-m_0)^2\right)$$

avec $k_1 m_1 = km + k_0 m_0$, d'où les paramètres de la distribution a posteriori : $(a_1-1) = (a_0-1) + \frac{k}{2}$ et $b_1 = k_1\left(v_1 + (\mu-m_1)^2\right)$, estimés de la même manière que précédemment mais à partir de l'échantillon composé $\mathbf{x}_1 = \mathbf{x} \cup \mathbf{x}_0$ (cf. notations au paragraphe 3.1.1). Finalement, nous avons l'opérateur de conjuguaison linéaire :

$$g\left(\theta \mid T(\mathbf{x} \cup \mathbf{x}_0)\right) = g\left(\theta \mid T(\mathbf{x})\right) g\left(\theta \mid T(\mathbf{x}_0)\right)$$

avec μ fixé.

La distribution bayésienne de la variable aléatoire X devient une distribution de Student non centrée avec $(a,b) = (a_0, b_0)$ ou (a_1, b_1) :

$$p(x) = \int_{\mathbb{R}^{+*}} \sqrt{\frac{\eta}{\pi}} \exp\left(-\eta(x-\mu)^2\right) \frac{b^a}{\Gamma(a)} \eta^{a-1} \exp\left(-\eta b\right) d\eta$$

$$\propto \left[b + (x-\mu)^2\right]^{-a-1/2} \tag{3.135}$$

Notons $c^2 = \frac{b}{2a}$, la variable $Y = \frac{X-\mu}{c}$ suit <u>une distribution de Student</u> à 2a degrés de liberté, c'est à dire $\frac{X-\mu}{c} \approx\approx\approx T(2a)$ ou $\left(\frac{X-\mu}{c}\right)^2 \approx\approx\approx F(1,2a)$. La moyenne et la variance bayésiennes sont :

$$\begin{cases} M = \mathbb{E}(X) = \mu \\ V = \mathbb{V}(X) = \dfrac{b}{2(a-1)} \qquad \text{si } a > 1 \end{cases} \tag{3.136}$$

Pour la même raison que dans le cas où μ et η sont inconnus, la distribution est asymptotiquement normale, $N\left(\mu, c^2\right)$. Lorsque le paramètre (a,b) est déterminé par la statistique exhaustive (m,v,k) de manière à ce que $a = \frac{k}{2}+1$ et $b = k\left(v + (\mu-m)^2\right)$, le modèle est "équivalent" au modèle classique pour k grand si μ est estimé par m ; dans ce cas, $c^2 = \frac{b}{2a} = \frac{k}{k-1}v = \frac{1}{k-1}\sum\left(x_i-m\right)^2$ est

l'estimateur sans biais de la variance et V = v.

D'autre part, comme une distribution normale de paramètre σ inconnu a une densité de la forme $\frac{1}{\sigma}f\left(\frac{x}{\sigma}\right)$, un principe d'invariance nous conduirait à une distribution a priori proportionnelle à σ^{-1} donc à $\sqrt{\eta}$; soit $\pi(\eta) = \sqrt{\eta}$.

Etant donné un échantillon **x**, la formule séquentielle nous conduit à une distribution a priori de type gamma : $\pi(\eta) = G\left(\eta;\ \frac{k+1}{2} + 1,\ k[v+(\mu-m)^2]\right)$. Si, dans le modèle précédent, le paramètre (a,b) de la distribution a priori est déterminé par la statistique exhaustive (m,v,k) de l'échantillon **x** de manière à ce que $a = \frac{k}{2}+1$ et $b = k\left(v+(\mu-m)^2\right)$, alors nous pouvons dire qu'il n'y a que 1/2 degré de différence sur le premier paramètre entre ces deux modèles considérés (cf. formules (3.132) et (3.131)).

3.2. DISTRIBUTION LOG-NORMALE

Cette loi de distribution, notée $LN\left(\alpha,\beta^2\right)$, de la variable X est obtenue lorsque son logarithme Y = lnX suit une distribution normale $N\left(\alpha,\beta^2\right)$. La densité de cette distribution pour tout $x \in \mathbb{R}^{+*}$ s'écrit :

$$f(x|\alpha,\beta) = \frac{1}{x\sqrt{2\pi}\beta}\exp\left[-\frac{1}{2}\left(\frac{\ln x - \alpha}{\beta}\right)^2\right] \tag{3.20}$$

Les paramètres α et β sont la moyenne et l'écart-type de Y = lnX. Cependant, les relations entre la moyenne μ et la variance σ^2 de X et celles α et β^2 de lnX sont :

$$\begin{cases} \mu = \exp\left(\alpha+\beta^2/2\right) \\ \sigma^2 = \exp\left(2\alpha+\beta^2\right)\left(e^{\beta^2}-1\right) \end{cases} \quad \text{et} \quad \begin{cases} \alpha = \ln(\mu) - \frac{1}{2}\ln\left(1+\delta^2\right) \\ \beta = \sqrt{\ln(1+\delta^2)} \end{cases} \tag{3.21}$$

avec $\delta = \frac{\sigma}{\mu}$ le coefficient de variation de la variable X.

Le mode $M_o = \exp\left(\alpha-\beta^2\right) = \mu\left(1+\delta^2\right)^{-3/2}$ est toujours inférieur à la moyenne et à la médiane $M_e = \exp(\alpha)$. Si l'on étudie la courbe $f(M_o) = \sup_{x>0}\{f(x)\}$ en fonction de σ^2, μ étant fixé , on voit qu'elle a un seul maximum $\frac{1}{\mu}\sqrt{e/\pi}$ et l'atteint au point $M_o = \mu e^{-3/4}$. Dans ce cas, $\delta^2 = \sqrt{e} - 1$ et $\beta^2 = \frac{1}{2}$. La figure

3.2 illustre ce résultat, pour les densités log-normales correspondant à la même moyenne $\mu = 4$ et à différentes variances. Le cas $\sigma^2 = 10.38$ correspond à la distribution pour laquelle $f(M_0)$ est minimale, $\mathrm{Min}\{f(M_0)\} = 0.2325$. Elles sont asymétriques et ceci d'autant plus que δ et σ augmentent. Au contraire, si δ devient petit (<0.25 par exemple), la distribution log-normale approche la distribution normale et devient très concentrée.

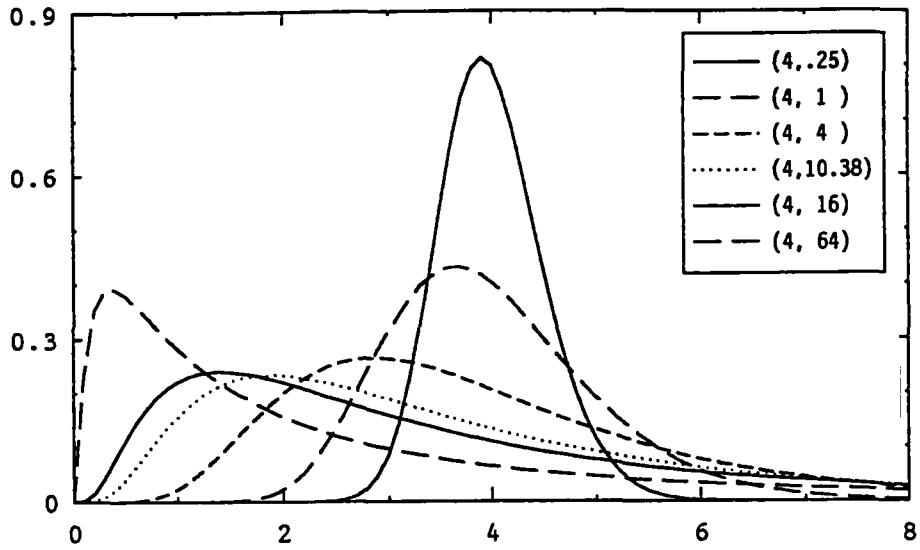


Figure 3.2 Distribution log-normale

La spécificité de cette distribution, outre son asymétrie, est que le support admet une borne inférieure $x_0 = 0$; et elle est donc portée par $\mathbb{R}^{+*}$, c'est à dire que la probabilité d'apparition de valeurs négatives est nulle. Il faut indiquer que cette distribution sera souvent adaptée à des problèmes réels de modélisation de variables (cf. par exemple [43] et [12]).

Comme la variable lnX suit une distribution normale, nous utilisons les modèles précédents en considérant la variable lnX et non X elle-même. Nous indiquons les formules principales en distinguant les trois cas suivants : (1) α et β inconnus; (2) α inconnu et β connu et (3) α connu et β inconnu.

Soit $\mathbf{x}$ un échantillon d'une distribution log-normale de paramètres α et β. Nous notons $\lambda = 1/(2\beta^2)$ la précision de la distribution log-normale. Nous utilisons les statistiques m_{ln}, v_{ln} et k au lieu de m, v et k dans le cas de

distributions normales. Nous obtenons les trois modèles normal-gamma, normal et gamma correspondant aux trois possibilités de (α,β). Les formules de passage de l'a priori à l'a posteriori sont données par les suivantes :

$$\begin{cases} k_1 m_1 = k m_{ln} + k_0 m_0 \\ k_1 = k + k_0 \\ k_1 v_1 = k v_{ln} + k_0 v_0 + (m_{ln} - m_0)^2 k_0 k / k_1 \end{cases} \tag{3.22}$$

$$\begin{cases} h_1 m_1 = m_{ln} k\lambda + m_0 h_0 \\ v_1 = \dfrac{1}{2h_1} \qquad \text{avec } h_1 = k/(2\beta)^2 + 1/(2v_0) = k\lambda + h_0 \end{cases} \tag{3.23}$$

$$\begin{cases} a_1 = a_0 + \dfrac{k}{2} \\ b_1 = b_0 + k[v_{ln} + (\alpha - m_{ln})^2] \end{cases} \tag{3.24}$$

Nous avons les propriétés suivantes :

— la famille de distributions normale-gamma est conjuguée par rapport à la famille de distributions log-normales;

— la famille de distributions normales est conjuguée par rapport à la famille de distributions log-normales de deuxième paramètre β connu;

— la famille de distributions gamma est conjuguée par rapport à la famille de distributions log-normales de premier paramètre α connu.

Les distributions bayésiennes de la variable lnX sont de Student dans le cas où α et β sont inconnus, ou dans le cas où α est connu et β inconnu. Par contre dans le cas où α est inconnu et β connu, la distribution a priori du paramètre α est normale, $N(m,v)$, et la distribution bayésienne de X est log-normale $LN(m, v+\beta^2)$, avec la moyenne et la variance calculées par :

$$\begin{cases} M = \mathbb{E}(X) = \exp\left(m + \dfrac{v+\beta^2}{2}\right) \\ V = \mathbb{V}(X) = \exp\left(2m+v+\beta^2\right)\left[\exp\left(v+\beta^2\right) - 1\right] \end{cases}$$

où $(m,v) = (m_0, v_0)$ ou (m_1, v_1).

3.3. DISTRIBUTION NORMALE INVERSE (WALD)

La distribution normale inverse est aussi entièrement caractérisée par deux paramètres μ et σ, puisque sa fonction de densité s'écrit :

$$f(x|\mu,\beta) = \frac{1}{\sqrt{2\pi}\beta x^{3/2}}\exp\left[-\frac{1}{2x}\left(\frac{x-\mu}{\mu\beta}\right)^2\right] \tag{3.30}$$

Le paramètre μ est la moyenne théorique d'une telle variable X et doit être positif parce que sa variance est définie par $\sigma^2 = \mu^3\beta^2$ et $\delta = \sqrt{\mu}\beta$ est le coefficient de variation. Cette distribution intervient en particulier dans les études de fiabilité avec des données de survie (cf. [11] et [26]).

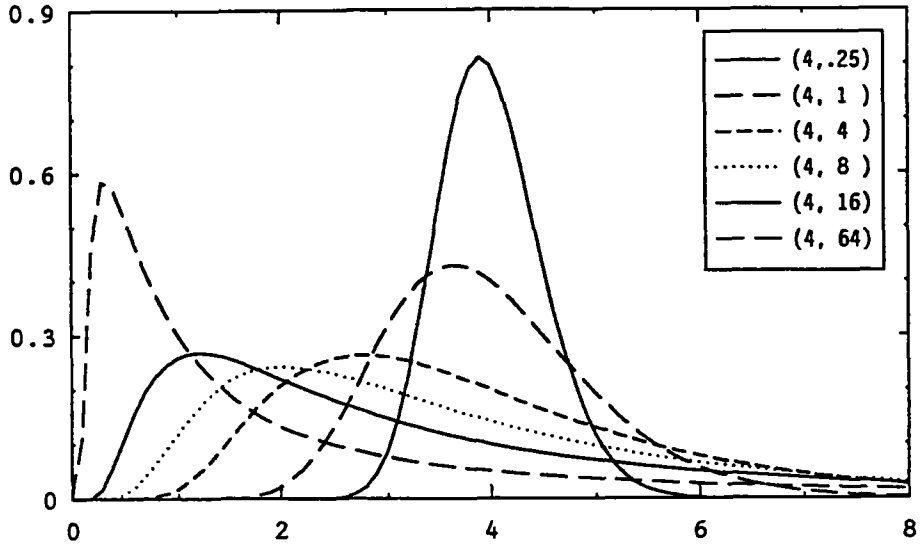


Figure 3.3 Distribution normale inverse

Le mode défini par $M_o = \sqrt{\mu^2 + \left(\frac{3}{2}\mu^2\beta^2\right)^2} - \frac{3}{2}\mu^2\beta^2$ est strictement inférieur à la moyenne. Etant donné μ, la courbe $f(M_o) = \sup_{x>0}\{f(x)\}$ en fonction de σ^2 a un seul minimum $\frac{1}{\mu}\sqrt{8/(e\pi)}$ qui est atteint au point $M_o = \frac{\mu}{2}$. Dans ce cas, $\delta^2 = \mu\beta^2 = 1/2$. La figure 3.3 ci-desous montre les densités des distributions normales inverses correspondant à la même moyenne et à différentes variances. Les valeurs du paramètre β^2 associées sont respectivement 0.0039, 0.0156, 0.0625, 0.125, 0.25 et 1. Le cas $\sigma^2 = 8$ ($\beta^2 = 0.125$) correspond à la distribution pour laquelle $f(M_o)$ est minimale: $\mathrm{Min}\{f(M_o)\} = 0.2420$.

L'asymétrie mesurée par exemple par la distance entre M et M_0 $d = M-M_0$, augmente avec la variance σ^2. Comme la distribution log-normale, la distribution normale inverse approche une distribution normale lorsque δ est petit.

Soit $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ un échantillon tiré au sort d'une distribution normale inverse. Pour chacun x_i, la densité de probabilité s'écrit :

$$f(x_i|\mu,\beta) \propto \beta^{-1}\exp\left[-\frac{1}{2}\beta^{-2}x_i\left(\frac{1}{\mu}-\frac{1}{x_i}\right)^2\right] .$$

En notant $m_H = \left(\frac{1}{k}\sum x_i^{-1}\right)^{-1}$ la moyenne harmonique et $m_{-1} = m_H^{-1}$ le moment réciproque d'ordre -1, la fonction de vraisemblance de $\mathbf{x}$ s'écrit :

$$f(\mathbf{x}|\mu,\beta) \propto \beta^{-k}\exp\left[-\frac{1}{2}\beta^{-2}km\left(\frac{1}{\mu}-\frac{1}{m}\right)^2-\frac{1}{2}\beta^{-2}k\left(\frac{1}{m_H}-\frac{1}{m}\right)\right] \qquad (3.31)$$

Dans ce cas, la statistique exhaustive associée est $T(\mathbf{x}) = (m, m_H, k)$.

3.3.1. Couple des variables (μ,β^2)

Considérons les deux variables $\alpha = \mu^{-1}$ et $\lambda = \frac{1}{2}\beta^{-2}$; alors nous trouvons aussi pour ce couple une autre <u>forme normale-gamma</u> :

$$\pi(\alpha,\lambda) \propto \lambda^{k/2}\exp\left[-\lambda km\left(\alpha-m^{-1}\right)^2\right]\exp\left[-\lambda k\left(m_H^{-1}-m^{-1}\right)\right] \qquad (3.310)$$

d'où les égalités :

$$\pi(\alpha,\lambda) = \pi(\alpha|\lambda)\pi(\lambda) = N\left(\alpha;\ \frac{1}{m},\frac{1}{2km\lambda}\right)G\left(\lambda;\ \frac{k+1}{2},k\left(m_H^{-1}-m^{-1}\right)\right) \qquad (3.311)$$

En introduisant les paramètres $a_0 = m^{-1}$, $b_0 = m_H^{-1} - m^{-1}$, $k_0 = k$, nous avons :

$$\pi(\alpha,\lambda) = N\left(\alpha;\ a_0,\frac{a_0}{2k_0\lambda}\right)G\left(\lambda;\ \frac{k_0+1}{2},k_0b_0\right) \qquad (3.312)$$

Nous avons toujours $b_0 \geq 0$ puisque la moyenne harmonique n'est pas supérieure à la moyenne arithmétique $m_H \leq m$, d'après l'inégalité de Cauchy (cf. paragraphe 2.6.3) Cette différence b_0 représente aussi la dispersion des données.

En comparant les deux formes de la distribution normale-gamma (3.113) et (3.312), nous voyons que cette dennière a un facteur de plus $a = m^{-1}$ dans le paramètre de variance de la partie normale, qui n'influence en effet que la variance de α :

$$\pi(\alpha,\lambda) : \quad \begin{cases} \mathbb{E}(\alpha) = a_0 = \dfrac{1}{m} & \Longrightarrow \quad \mathbb{E}(\mu) = m \\ \mathbb{V}(\alpha) = \dfrac{1}{k-1}a_0 b_0 = \dfrac{1}{(k-1)m}\left(m_H^{-1} - m^{-1}\right) & \\ \mathbb{E}(\lambda) = \dfrac{k+1}{2kb} & \Longrightarrow \quad \mathbb{E}(\beta^2) = \dfrac{k}{k+1}\left(m_H^{-1} - m^{-1}\right) \end{cases} ;$$

$$\pi(\mu,\eta) : \quad \begin{cases} \mathbb{E}(\mu) = m \\ \mathbb{V}(\mu) = \dfrac{1}{k-1}v \\ \mathbb{E}(\eta) = \dfrac{k+1}{2kv} & \Longrightarrow \quad \mathbb{E}(\sigma^2) = \dfrac{k}{k+1}v \end{cases} \tag{3.313}$$

Posons $a = m^{-1}$ et $b = m_H^{-1} - m^{-1}$. La distribution a posteriori après une observation $\mathbf{x}$ est aussi normale-gamma de la forme (3.313) avec les paramètres définis par :

$$\begin{cases} k_1/a_1 = k_0/a_0 + k/a \\ k_1 = k_0 + k \\ k_1 b_1 = k_0 b_0 + kb + \dfrac{a}{k_1} \times \dfrac{k_0}{a_0} \times \dfrac{k}{a}\left(a - a_0\right)^2 \end{cases} \tag{3.314}$$

La famille de distributions normale-gamma de la forme (3.313) constitue une famille de distributions conjuguées par rapport à la famille de distributions normale inverses.

Il n'est pas difficile de vérifier que la définition des statistique $a = m^{-1}$ et $b = m_H^{-1} - m^{-1}$ satisfait aussi une stabilité au sens suivant: si (a_0, b_0) est déterminé par la statistique exhaustive (m_0, m_{H0}, k_0) d'un échantillon $\mathbf{x}_0$, alors, étant donné un échantillon $\mathbf{x}$, (a_1, b_1) est déterminé par la statistique exhaustive (m_1, m_{H1}, k_1) de l'échantillon composé $\mathbf{x}_1 = \mathbf{x} \cup \mathbf{x}_0$.

Notons $(a,b,k) = (a_0,b_0,k_0)$ ou (a_1,b_1,k_1). La distribution prédictive est obtenue à l'aide des calculs suivants :

$$p(x) = \iint_{\mathbb{R}\times\mathbb{R}^+} f(x|\alpha,\lambda)\pi(\alpha,\lambda)d\alpha d\lambda$$

$$= \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}} f(x|\alpha,\lambda)\pi(\alpha|\lambda)d\alpha \right) \pi(\lambda)d\lambda$$

$$= \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}} \frac{\sqrt{k/a}}{\pi x^{3/2}} \lambda \exp\left[- \lambda x \left(x^{-1}-\alpha\right)^2 - \lambda \frac{k}{a} \left(\alpha - a\right)^2 \right] d\alpha \right) \pi(\lambda)d\lambda$$

$$= \int_{\mathbb{R}^+} \sqrt{\frac{\lambda k/a}{\pi(x+k/a)x^3}} \exp\left[- \lambda \frac{a/x}{x+k/a} \left(x - a^{-1}\right)^2 \right] G\left(\frac{k+1}{2}, kb\right)(\lambda)d\lambda$$

$$\propto \left(x^3(x+k/a)\right)^{-1/2} \left[b + \frac{a/x}{x+k/a} \left(x - a^{-1}\right)^2 \right]^{-(k/2+1)} \qquad (3.315)$$

Mais pour calculer la moyenne et la variance, nous ne pouvons plus utiliser la technique de changement de l'ordre des intégrales comme précédemment, puisque nous avons $\int_{\mathbb{R}^{+*}} x f(x|\alpha,\lambda)dx = \frac{1}{\alpha} = \mu$; comme la distribution $\pi(\alpha|\lambda)$ est normale, il est clair que l'intégrale est infinie : $\int_{\mathbb{R}} \frac{1}{\alpha} \pi(\alpha|\lambda)d\alpha = \infty$. En effet, cette distribution n'a aucun moment. En effet pour x suffisament grand, le facteur suivant est asymptotiquement indépendant de x :

$$\left[b + \frac{a/x}{x+k/a} \left(x - a^{-1}\right)^2 \right]^{-(k/2+1)} \approx O(1)$$

Le facteur est dit de l'ordre 1 à l'infini. Nous en déduisons donc :

$$x^n p(x) = x^n \left(x^3(x+k/a)\right)^{-1/2} \left[b + \frac{a/x}{x+k/a} \left(x - a^{-1}\right)^2 \right]^{-(k/2+1)} \approx O(x^{n-2})$$

L'intégrale du n-ième moment $\mu_n = \int_{\mathbb{R}^{+*}} x^n p(x)dx$ est infinie, même pour n = 1.

3.3.2. Variable μ

Comme dans le cas de la distribution normale, λ étant fixé, la variable $\alpha = \mu^{-1}$ est adaptée au <u>modèle normal</u> :

$$\pi(\alpha) = N\left(\alpha;\ \frac{1}{m}, \frac{1}{2km\lambda}\right) = N\left(\alpha;\ m_0, v_0\right) \qquad (3.320)$$

La propriété de conjuguaison est vérifiée pour la famille de lois de distributions normales inverses de paramètre β connu, et le passage de l'a

priori à l'a posteriori est réalisé à l'aide de la formule :

$$\begin{cases} h_1 m_1 = h_0 m_0 + \lambda k \\ v_1 = \dfrac{1}{2h_1} \end{cases} \quad \text{avec } h_1 = \lambda k m + h_0 \text{ et } h_0 = \frac{1}{2v_0} \tag{3.321}$$

La distribution bayésienne est de la forme particulière :

$$p(x) \propto \frac{1}{\sqrt{(h+\lambda x)x^3}} \exp\left[- \lambda \frac{xh}{h+\lambda x}\left(\frac{1}{x} - m\right)^2\right] \tag{3.322}$$

avec $(m,v) = (m_0, v_0)$ ou (m_1, v_1) et $h = \frac{1}{2v}$. Elle n'a aucun moment comme précédemment parce que la fonction à intégrer est aussi de l'ordre (n-2) pour x suffisament grand : $x^n p(x) \approx O(x^{n-2})$.

3.3.3. Variable β^2

Pour la variable $\lambda = 1/(2\beta^2)$, μ étant fixé, le modèle est gamma :

$$\pi(\lambda) = G\left(\lambda;\ \frac{k+1}{2}, k\left(m_H^{-1} - m^{-1} + m(\mu^{-1} - m^{-1})^2\right)\right) = G\left(\lambda;\ a_0, b_0\right) \tag{3.330}$$

La famille de distributions gamma est conjugée par rapport à la famille de distributions normales inverses du paramètre μ connu. La formule suivante permet de réaliser le passage :

$$\begin{cases} a_1 = a_0 + \dfrac{k}{2} \\ b_1 = b_0 + k\left(b + m(\mu^{-1} - m^{-1})^2\right) \end{cases} \tag{3.331}$$

La distribution prédictive avec $(a,b) = (a_0, b_0)$ ou (a_1, b_1) est

$$p(x) \propto \frac{1}{x^{3/2}\left(x(x^{-1} - \mu^{-1})^2 + b\right)^{a/2+1}} \tag{3.332}$$

La moyenne et la variance prédictives sont calculées en changeant l'ordre des intégrales :

$$\begin{cases} M = \mathbb{E}(X) = \mu \\ V = \mathbb{V}(X) = \dfrac{b}{2(a-1)}\mu^3 \end{cases} \tag{3.333}$$

3.4. DISTRIBUTION GAMMA

La distribution gamma est une autre distribution particulièrement fréquente. Nous verrons dans la suite que les modèles étudiés non issus de la famille normale (distributions normale, log-normale et normale inverse) contiennent souvent des distributions gamma.

La densité de cette distribution de paramètres α et β pour tout $x \in \mathbb{R}^{+*}$ s'écrit :

$$f(x|\alpha,\beta) = \frac{\beta}{\Gamma(\alpha)}(x\beta)^{\alpha-1}e^{-x\beta} \tag{3.40}$$

Si $\alpha = 1$, c'est une loi exponentielle. Si α est un entier positif, elle est une loi d'Erlang. Si $\alpha = \upsilon/2$ et $\beta = 1/2$, nous avons une loi du chi-deux à ν degrés de liberté, $\chi^2(\nu)$.

Les relations entre les moyenne μ et variance σ^2 et les deux paramètres sont données par :

$$\begin{cases} \mu = \alpha\cdot\beta^{-1} \\ \sigma^2 = \alpha\cdot\beta^{-2} \end{cases} \quad \text{et} \quad \begin{cases} \alpha = \mu^2/\sigma^2 \\ \beta = \mu/\sigma^2 \end{cases} \tag{3.41}$$

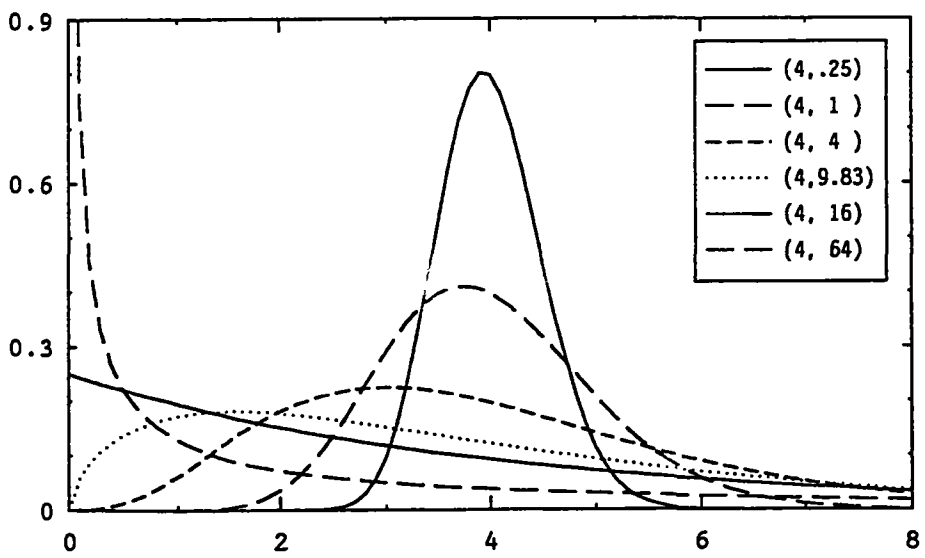


Figure 3.4 Distribution gamma

Le mode est $M_o = (\alpha-1)/\beta = \mu-\sigma^2/\mu$, toujours situé à gauche de la moyenne — ces deux valeurs se rapprochent lorsque σ décroissant. La figure 3.4 montre les densités des distributions gamma correspondant à la même moyenne $\mu = 4$ et à différentes variances. Les couples (α,β) sont respectivement (64,16.0), (16,4), (4,1), (1.63,0.41), (1,0.25) et (0.25,0.06). A $\sigma^2 = 9.83$ ou $\beta = 1.63$, la courbe $f(M_o)$ atteint son minimum $\mathrm{Min}\{f(M_o)\} = 0.18$. L'asymétrie, mesurée par exemple par la distance entre M et M_o: $d = M-M_o$, augmente avec la variance.

Comme pour la distribution log-normale, le support de la distribution gamma prossède aussi une borne inférieure $x_o = 0$. Par translation, cette borne peut être amenée en un point spécifique. Le problème qui se pose immédiatement en pratique dans les cas bornés, est celui de choisir cette borne: le minimum des données, ou une valeur plus petite que ce minimum.

Nous étudions deux cas : (1) α et β sont inconnus, et (2) β est inconnu avec α connu.

3.4.1. α et β inconnus

Fonction de vraisemblance

Soit $\mathbf{x}$ un échantillon d'une distribution gamma de paramètres α et β inconnus. La fonction de vraisemblance est le produit des probabilités pour chaque x_i, $1 \le i \le k$:

$$f(\mathbf{x}|\alpha,\beta) = \prod_i \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-x_i\beta} = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^k m_G^{k(\alpha-1)} e^{-km\beta} \qquad (3.410)$$

il en résulte donc qu'une statistique exhaustive de l'échantillon $\mathbf{x}$ est $T(\mathbf{x}) = (m, m_G, k)$ avec $m_G = \left(\prod x_i\right)^{1/k}$ la moyenne géométrique.

Distributions a priori

En tenant compte du principe de la conjuguée naturelle, nous avons envie de choisir la distribution a priori du couple (α,β) des variables,

défini par extension sur $\mathbb{R}^+\times\mathbb{R}^+$ sous forme de la fonction de vraisemblance.

Nous faisons d'abord la normalisation pour la variable β et pour α fixé et nous avons donc :

$$\pi(\alpha,\beta) \propto \frac{(km)^{k\alpha+1}}{\Gamma(k\alpha+1)}\beta^{k\alpha}e^{-km\beta}\cdot\frac{\alpha\Gamma(k\alpha)}{[k^{\alpha}\Gamma(\alpha)]^{k}}\exp\left(-\alpha k\log(m/m_G)\right) \tag{3.411}$$

Ensuite, nous faisons une approximation exponentielle de la fonction Γ par la formule exponentielle: $\Gamma(\alpha) \approx \sqrt{2\pi}\, e^{-\alpha}\alpha^{\alpha-1/2}$ (cf. par exemple [1]). En posant $r = \log m - \log m_G = \log\bar{x} - \overline{\log x}$, nous obtenons une "approximation" de π par $\bar{\pi}$, où

$$\pi(\alpha,\beta) \approx \bar{\pi}(\alpha,\beta) = \frac{(km)^{k\alpha+1}}{\Gamma(k\alpha+1)}\beta^{k\alpha}e^{-km\beta}\cdot\frac{(kr)^{(k+3)/2}}{\Gamma\left(\frac{k+3}{2}\right)}\alpha^{\frac{k+1}{2}}e^{-kr\alpha} \tag{3.412}$$

Cette distribution $\bar{\pi}$ est appelée <u>distribution gamma-gamma</u>, notée par $GG(m,r,k)$ (l'indice "$_0$" est omis), et vérifie la propriété suivante :

$$\bar{\pi}(\alpha,\beta) = \bar{\pi}(\beta|\alpha)\bar{\pi}(\alpha) \tag{3.413}$$

où $\pi(\beta|\alpha)$ est la distribution gamma $G(k\alpha+1,km)$, et $\pi(\alpha)$ une fonction de α proche de la distribution gamma $G\left(\frac{k+3}{2},kr\right)$. Nous avons $\pi(\beta|\alpha) = \bar{\pi}(\beta|\alpha)$ et $\pi(\alpha) \approx \bar{\pi}(\alpha)$, approximation valable pour α grand.

<u>Distribution a posteriori</u>

Avec un échantillon $\mathbf{x}$, et sa statistique exhaustive $T(\mathbf{x}) = (m,m_G,k)$, comme précédemment, la distribution a posteriori de la variable (α,β) est approximée par une distribution gamma-gamma dont les paramètres (m_1,r_1,k_1) sont définis par :

$$\begin{cases} k_1 = k_0 + k \\ k_1m_1 = k_0m_0 + km \\ k_1r_1 = k_0r_0 - k_0\log m_0 + k_1\log m_1 - k\log m_G \end{cases} \tag{3.414}$$

$$= k_0r_0 + kr + k_1\log m_1 - (k_0\log m_0 + k\log m)$$

Cette relation implique que la famille de distributions gamma-gamma est à l'approximation près, conjuguée par rapport à la famille de distributions gamma.

Opération de conjuguaison et Stabilité de l'estimateur r

La définition de la statistique r satisfait une stabilité au sens suivant : si le paramètre de la loi a priori r_0 est estimé par $\left(\log m_0 - \log m_{G0}\right)$, où $T(\mathbf{x}_0) = (m_0, m_{G0}, k_0)$ est une statistique exhaustive d'un échantillon $\mathbf{x}_0$, alors étant donné un échantillon $\mathbf{x}$, la formule (3.415) permet de calculer r_1 de l'a posteriori sous la même forme $\left(\log m_1 - \log m_{G1}\right)$, où $T(\mathbf{x}_1) = (m_1, m_{G1}, k_1)$ est la statistique exhaustive de l'échantillon composé $\mathbf{x}_1 = \mathbf{x} \cup \mathbf{x}_0$. Finalement, nous avons l'opérateur linéaire:

$$g\left(\theta \mid T(\mathbf{x} \cup \mathbf{x}_0)\right) = g\left(\theta \mid T(\mathbf{x})\right) + g\left(\theta \mid T(\mathbf{x}_0)\right) \quad .$$

Cette statistique représente aussi une dispersion à l'intérieur du lot. La concavité de la fonction logx assure que r est toujours positif. D'après cette remarque, nous pouvons dire que le triplet (m,r,k) est une statistique exhaustive pour ce modèle.

Distribution bayésienne

La distribution bayésienne de la variable X est obtenue en intégrant la distribution conditionnelle de X connaissant le paramètre (α, β), suivant toutes les probabilités d'apparition du couple (α, β) :

$$\begin{aligned} p(x) &= \iint_{\mathbb{R}^+ \times \mathbb{R}^+} f(x \mid \alpha, \beta)\pi(\alpha, \beta)\, d\alpha\, d\beta \\ &= \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}^+} f(x \mid \alpha, \beta)\pi(\beta \mid \alpha)\, d\beta \right) \pi(\alpha)\, d\alpha \\ &= \int_{\mathbb{R}^+} \frac{\Gamma(k\alpha+\alpha+1)}{\Gamma(k\alpha+1)\Gamma(\alpha)} \cdot \frac{(km)^{k\alpha+1} x^{\alpha-1}}{(km+x)^{k\alpha+\alpha+1}} \cdot G\left(\frac{k+3}{2}, kr\right)(\alpha)\, d\alpha \qquad (3.415) \end{aligned}$$

C'est une distribution que l'on ne se calcule que de façon numérique. Mais, il n'est pas difficile d'obtenir sa moyenne : $M = \mathbb{E}(X) = m_0$ ou m_1, mais les variances $\mathbb{V}(X)$ et $\mathbb{V}(X|\mathbf{x})$ n'existent pas puisque

$$\int_{\mathbb{R}^+} (x-m)^2 p(x)dx = \int_{\mathbb{R}^+}\left[\int_{\mathbb{R}^+}\left(\int_{\mathbb{R}^+} x^2 f(x|\alpha,\beta)dx\right)\pi(\beta|\alpha)d\beta\right]\pi(\alpha)d\alpha - m^2$$

$$= \int_{\mathbb{R}^+}\left[\int_{\mathbb{R}^+} \alpha(\alpha+1)\beta^{-2}\pi(\beta|\alpha)d\beta\right]\pi(\alpha)d\alpha - m^2$$

$$= m^2(1+\frac{1}{k})\int_{\mathbb{R}^+} \frac{1}{\alpha-1/k}\pi(\alpha)d\alpha$$

d'où $\alpha = 1/k$ est un point singulier.

Par contre, si l'on modifie légèrement la forme "approximée" de la loi de distribution a priori, nous définissons ici une <u>forme "théorique" de la distribution gamma-gamma</u>: $\pi'(\alpha,\beta) = G(\beta;\ k\alpha, km)G(\alpha;\ kt, kr)$ avec un quatrième paramètre $t > 0$. La formule du passage de la distribution a priori à la distribution a posteriori reste la même que précédemment avec $t_1 = t + t_0$. La distribution bayésienne devient alors une combinaison d'une distribution gamma et d'une distribution de Fisher-Snedecor :

$$p(x) = \int_{\mathbb{R}^+} \frac{1}{m}F\left(\frac{x}{m};\ 2k\alpha, 2\alpha\right)\cdot G(\alpha;\ kt, kr)d\alpha \qquad (3.416)$$

3.4.2. α connu et β inconnu

Soit $\mathbf{x}$ un échantillon d'une distribution gamma de paramètres α connu et β inconnu. La fonction de vraisemblance de cet échantillon est :

$$f(x|\beta) \propto \beta^{k\alpha}e^{-\Sigma x_i\beta} \qquad (3.420)$$

donc (m,k) est une statistique exhaustive par rapport à β.

Le principe de la conjuguée naturelle donne pour distribution a priori

de la variable β une distribution gamma, et la distribution a posteriori reste dans la famille de distributions gamma :

$$\pi(\beta) = G(\beta;\ a_0, b_0) \quad \text{et} \quad \pi(\beta|x) = G(\beta;\ a_1, b_1) \tag{3.421}$$

La formule de passage de la distribution a priori à l'a posteriori est :

$$\begin{cases} a_1 = a_0 + k\alpha \\ b_1 = b_0 + km \end{cases} \tag{3.422}$$

La famille de distributions gamma est conjuguée par rapport à la famille de distributions gamma de premier paramètre α connu. L'opérateur associé est aussi linéaire :

$$g(\theta|T(\mathbf{x}\cup\mathbf{x}_0)) = g(\theta|T(\mathbf{x})) + g(\theta|T(\mathbf{x}_0)) \quad .$$

La distribution bayésienne résulte des calculs suivants :

$$p(x) = \int_{\mathbb{R}^{+*}} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta} \cdot \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-b\beta} d\beta = \frac{\Gamma(a+\alpha)}{\Gamma(a)\Gamma(\alpha)} \cdot \frac{b^a\ x^{\alpha-1}}{(b+x)^{a+\alpha}} \tag{3.423}$$

avec $(a,b) = (a_0, b_0)$ ou (a_1, b_1). Donc la variable $Y = \frac{a}{\alpha b} X$ admet une loi de distribution de Fisher-Snedecor de paramètres 2α et $2a$, d'où la moyenne et la variance bayésiennes :

$$\begin{cases} M = \mathbb{E}(X) = \dfrac{\alpha b}{a-1} \\ V = \mathbb{V}(X) = \left(\dfrac{\alpha b}{a-1}\right)^2 \dfrac{a+\alpha-1}{\alpha(a-2)} \qquad \text{si } a > 2 \end{cases} \tag{3.424}$$

Si le paramètre (a,b) est déterminé par la statistique exhaustive (m,k) d'un échantillon $\mathbf{x}$, où la valeur α est considérée comme constante, de sorte que $a = k\alpha+1$ et $b = km$, alors la moyenne bayésienne $M = \mathbb{E}(X)$ est égale à la moyenne empirique m, mais la variance bayésienne n'est pas reliée simplement avec la variance empirique v, parce que nous avons $V = \mathbb{V}(X) = m^2 \frac{k+1}{k\alpha-1}$. Par contre si α est estimé par $m^2/v = d^{-2}$, d étant le coefficient de variation empirique, dans ce cas $V = v\frac{k+1}{k-d^2}$ est supérieure à la variance empirique v.

3.5. DISTRIBUTION GAMMA GENERALISEE

Remarquons que, lorsque X admet une loi de distribution gamma $G(\alpha,\beta)$, la variable $Y = \beta X$ a une distribution gamma $G(\alpha,1)$. De plus, si α est fixé, nous venons de montrer que la distribution a priori de β est gamma.

De manière plus générale, si l'on se donne une variable X dont la distribution dépend de deux paramètres α et β telle que $Y = \beta^{\upsilon}|X-x_{\circ}|^{\nu}$ suit une distribution gamma $G(\alpha,1)$, alors la distribution a priori de la variable β^{υ} sera encore gamma, et la statistique exhaustive générale d'un échantillon $\mathbf{x}$ est alors

$$T(\mathbf{x}) = \left(m_{\nu} = \frac{1}{k}\sum|x_i-x_{\circ}|^{\nu},\ k\right) \tag{3.50}$$

au lieu de $T(\mathbf{x}) = (m,k)$.

La distribution de X est appelée la <u>distribution gamma généralisée</u> et définie soit sur $\{x;\ x > x_{\circ} = \varepsilon\}$, soit sur $\{x;\ x < x_{\circ} = \omega\}$ par :

$$f(x|\upsilon,\nu,x_{\circ},\alpha,\beta) = |\nu|\frac{\beta^{\upsilon}}{\Gamma(\alpha)}|x-x_{\circ}|^{\nu-1}\left(\beta^{\upsilon}|x-x_{\circ}|^{\nu}\right)^{\alpha-1}\exp\left(-\beta^{\upsilon}|x-x_{\circ}|^{\nu}\right) \tag{3.51}$$

($|\cdot|$ est la fonction de la valeur absolue) avec des cas particuliers :

— avec $(1,1,0,\alpha,\beta)$, c'est une loi gamma $G(\alpha,\beta)$ (du premier espèce);
— avec $(1,-1,0,\alpha,\beta)$, c'est une loi gamma inverse $GI(\alpha,\beta)$;
— avec $(-\alpha,\alpha,\varepsilon,1,\beta)$, c'est une loi de Weibull de type minimum $W(\alpha,\beta)$;
— avec $(\alpha,-\alpha,\varepsilon,1,\beta)$, c'est une loi de Fréchet de type maximum $Fr(\alpha,\beta)$;
— avec $(-\alpha,\alpha,\omega,1,\beta)$, c'est une loi de Weibull de type maximum;
— avec $(\alpha,-\alpha,\omega,1,\beta)$, c'est une loi de Fréchet de type minimum;

où α et β sont des réels positifs (cf. paragraphes 3.6 et 3.7). Quant à $G(-1,1,0,\alpha,\beta)$, la distribution est appelée gamma de seconde espèce :

$$f(x|\alpha,\beta) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}e^{-x/\beta} \tag{3.52}$$

Il suffit de considérer les variables α et $\lambda = \beta^{-1}$ dans le cas de la distribution gamma de permière espèce (cf. paragraphe 3.4).

La moyenne et la variance de la variable $Y = \beta^{\upsilon}|X-x_{\circ}|^{\nu}$ sont égales à α, d'où

$$\begin{cases} \mathbb{E}(|X-x_{\circ}|^{\nu}) = \alpha/\beta^{\upsilon} \\ \mathbb{V}(|X-x_{\circ}|^{\nu}) = \alpha/\beta^{2\upsilon} \end{cases} \tag{5.53}$$

La formule du passage de la distribution a priori de la variable $\lambda = \beta^{\upsilon}$ à la distribution a posteriori est la suivante :

$$\begin{cases} a_1 = a_0 + k\alpha \\ b_1 = b_0 + km_{\upsilon} \end{cases} \tag{3.54}$$

Nous avons la propriété de sa conjuguaison suivante: la famille de distributions gamma est conjuguée par rapport à la famille de distributions gamma généralisées de paramètre β inconnu. Il n'est pas difficile de vérifier que l'opérateur associé est linéaire.

La distribution bayésienne est calculée par

$$p(x) = \int_{\mathbb{R}^{+*}} |\nu|\frac{\lambda}{\Gamma(\alpha)}|x-x_{\circ}|^{\nu-1}\left(\lambda|x-x_{\circ}|^{\nu}\right)^{\alpha-1}\exp\left(-\lambda|x-x_{\circ}|^{\nu}\right)\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}d\lambda$$

$$= |\nu|\frac{\Gamma(a+\alpha)}{\Gamma(a)\Gamma(\alpha)}\cdot\frac{b^a|x-x_{\circ}|^{\nu(\alpha-1)}}{\left(b+|x-x_{\circ}|^{\nu}\right)^{a+\alpha}}|x-x_{\circ}|^{\nu-1} \tag{3.55}$$

avec $(a,b) = (a_0,b_0)$ ou (a_1,b_1), donc la variable $Y = \frac{a}{\alpha b}|X-x_{\circ}|^{\nu}$ admet une de <u>distribution de Fisher-Snedecor</u>, $F(2\alpha,2a)$. Sa moyenne est :

$$\begin{cases} \mathbb{E}(|X-x_{\circ}|) = \left(\frac{\alpha b}{a-1}\right)^{1/\nu} & \text{si } \nu > 0 \\ \mathbb{E}(|X-x_{\circ}|) = \left(\frac{\alpha b}{a-1}\right)^{-1/|\nu|} & \text{si } \nu < 0 \end{cases} \tag{3.56}$$

3.6. DISTRIBUTION GAMMA INVERSE

Etant donnés α et β, la distribution gamma inverse a pour densité une fonction définie sur $\mathbb{R}^+$:

ECOLE NATIONALE DES PONTS & CHAUSSÉES
CENTRE PÉDAGOGIQUE DE DOCUMENTATION ET DE COMMUNICATION

$$f(x|\alpha,\beta) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{\beta}{x}\right)^{\alpha+1} e^{-\beta/x} \tag{3.60}$$

La variable $Y = \beta/X$ suit une loi exponentielle, cas particulier cité au paragraphe 3.5, et nous adoptons un modèle gamma pour β. Les relations entre la moyenne et la variance et les deux paramètres sont données par :

$$\begin{cases} \mu = \dfrac{\beta}{\alpha-1} \\ \sigma^2 = \mu^2 \dfrac{1}{\alpha-2} \quad \text{si } \alpha > 2 \end{cases} \quad \text{et} \quad \begin{cases} \alpha = \mu^2/\sigma^2 + 2 \\ \beta = \mu(\mu^2/\sigma^2+1) \end{cases} \tag{3.61}$$

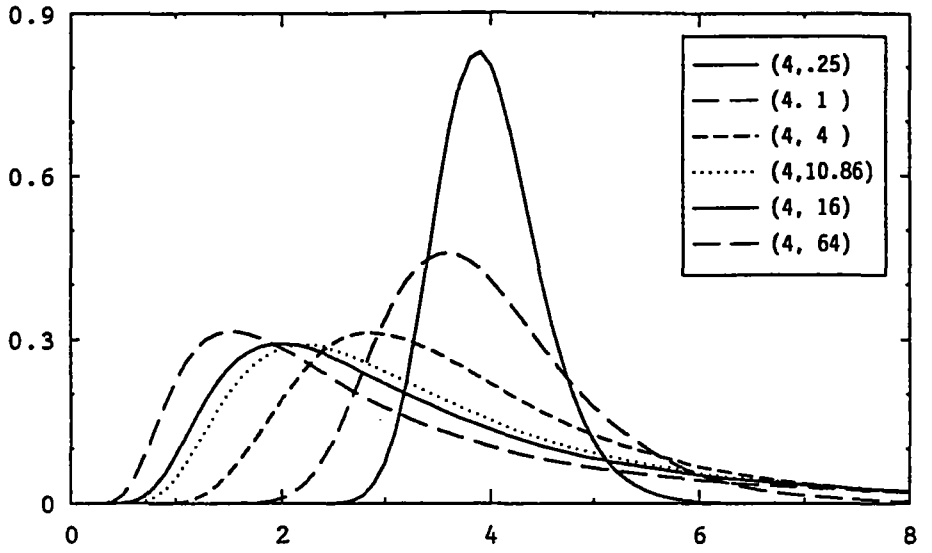


Figure 3.5 Distribution gamma inverse

Le mode est $M_o = \frac{\beta}{\alpha+1} = \mu\frac{\alpha-1}{\alpha+1}$, toujours situé à gauche de la moyenne — ces deux valeurs se rapprochent lorsque δ décroit. La figure 3.6 montre les densités des distributions gamma correspondant à la même moyenne et à différentes variances. Les couples de paramètres (α,β) sont respectivement (66, 260), (18,68), (6,20), (3.47,9.89), (3,80) et (2.25,50). Pour $\sigma^2 = 10.86$ ($\beta = 3.47$), la courbe $f(M_o)$ atteint son seul minimum $\text{Min}\{f(M_o)\} = 0.29$. Elles sont asymétriques et ceci augmente avec la variance σ^2 avec la limite de $\alpha = 2$ où $\sigma^2 = \infty$.

Si l'on considère la variable $Y = X^{-1}$, alors la distribution de Y est gamma et la distribution de $\theta = (\alpha,\beta)$ est gamma-gamma. La transformation de

la statistique exhaustive de Y en celle de X est la suivante :

$$T(y) = (m^y, m_G^y, k) \longrightarrow (m_H^{-1}, m_G^{-1}, k) = T(x)$$

et le paramètre de la distribution gamma-gamma devient:

$$\left(m_o = m^y,\ r_o = \log m^y - \log m_G^y,\ k_o\right) \longrightarrow \left(n_o = m_H^{-1},\ s_o = \log m_H^{-1} - \log m_G^{-1},\ k_o\right)$$
$$= \log m_G - \log m_H$$

Quant au problème de la variable β, α étant fixé, c'est un cas particulier de la distribution gamma généralisée (cf. paragraphe 3.5). Le modèle gamma est adopté et la distribution prédictive de la variable $Y = X^{-1}$ est de type de Ficher-Snedecor.

3.7. DISTRIBUTIONS DE VALEURS EXTREMES : WEIBULL ET FRECHET

Les distributions de Weibull et de Fréchet sont obtenues, lorsque les variables $Y = (X/\beta)^\alpha$ ou $Y = (\beta/X)^\alpha$ suivent respectivement la loi exponentielle réduite e^{-y} avec α et β deux réels positifs. Elles peuvent cependant être considérées comme distributions asymptotiques des valeurs maximales et minimales pour des distributions initiales à la densité respectivement à décroissance exponentielle telle que Cauchy, Pareto, ... ou bornée inférieurement telle que uniforme, Bêta, ... (cf. [21] et [42]). Plus généralement, si $Y = \left(\frac{X-\varepsilon}{\upsilon-\varepsilon}\right)^{\pm\alpha}$ suit la loi exponentielle, alors X suit une loi de Weibull ou de Fréchet à trois paramètres, dont ε est la borne inférieure ε (x, $\upsilon > \varepsilon$).

Les densités de ces distributions sont définies sur $\{x;\ x > \varepsilon\}$ par :

$$f(x|\alpha,\beta,\varepsilon) = \frac{\alpha}{\upsilon-\varepsilon}\left(\frac{x-\varepsilon}{\upsilon-\varepsilon}\right)^{\alpha-1} \exp\left[-\left(\frac{x-\varepsilon}{\upsilon-\varepsilon}\right)^{\alpha}\right] \tag{3.70a}$$

$$f(x|\alpha,\beta,\varepsilon) = \frac{\alpha}{\upsilon-\varepsilon}\left(\frac{\upsilon-\varepsilon}{x-\varepsilon}\right)^{\alpha+1} \exp\left[-\left(\frac{\upsilon-\varepsilon}{x-\varepsilon}\right)^{\alpha}\right] \tag{3.70b}$$

Les statistiques de ces deux distributions sont données dans le tableau suivant.

	Weibull	Fréchet
mode M_o	$\varepsilon + \beta\left(1-\frac{1}{\alpha}\right)^{1/\alpha}$ $\alpha > 1$	$\varepsilon + \beta\left(1+\frac{1}{\alpha}\right)^{1/\alpha}$
médiane M_e	$\varepsilon + \beta(\ln 2)^{1/\alpha}$ $\beta = \upsilon-\varepsilon$	$\varepsilon + \beta(\ln 2)^{-1/\alpha}$ $\beta = \upsilon-\varepsilon$
moyenne M	$\varepsilon + \beta\Gamma\left(1+\frac{1}{\alpha}\right)$	$\varepsilon + \beta\Gamma\left(1-\frac{1}{\alpha}\right)$ $\alpha > 1$
variance V	$\beta^2\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2\right]$	$\beta^2\left[\Gamma\left(1-\frac{2}{\alpha}\right)-\Gamma\left(1-\frac{1}{\alpha}\right)^2\right]$ $\alpha > 2$
moments	$\overline{(x-\varepsilon)^n} = \beta^n\Gamma\left(1+\frac{n}{\alpha}\right)$ $n > -\alpha$	$\overline{(x-\varepsilon)^n} = \beta^n\Gamma\left(1-\frac{n}{\alpha}\right)$ $\alpha > n$

Tableau 3.1 Statistiques des lois de Weibull et de Fréchet

Si l'on ne considère que deux paramètres α et $\beta = \upsilon$ (avec $\varepsilon = 0$), parmi les différents estimateurs, on utilise souvent celui des deux premiers moments :

$$\begin{aligned} \mu &= \beta\Gamma\left(1\pm\frac{1}{\alpha}\right) \\ \sigma^2 &= \beta^2\left[\Gamma\left(1\pm\frac{2}{\alpha}\right) - \Gamma\left(1\pm\frac{1}{\alpha}\right)^2\right] \end{aligned} \tag{3.71}$$

On en déduit que α est la solution de l'équation suivante :

$$\Gamma\left(1\pm\frac{2}{\alpha}\right) - (\delta^2+1)\Gamma\left(1\pm\frac{1}{\alpha}\right)^2 = 0 \tag{3.72}$$

Quant aux trois paramètres $(\alpha, \upsilon, \varepsilon)$, ils admettent plusieurs estimateurs tels que les trois premiers moments, le maximum de vraisemblance, ... (cf. [15] et [48]).

Les figures 3.6 et 3.7 illustrent les densités de Weibull et de Fréchet associées à la même moyenne et à différentes variances. Les cas $\sigma^2 = 8.13$ et 57.73 correspondent aux distributions pour lesquelles $f(M_o)$ est minimale, respectivement $\mathrm{Min}\{f(M_o)\} = 0.17$ et 0.36. Elles sont asymétriques et ceci augmente avec la variance σ^2 avec une limite de $\alpha = 2$ où $M_o = 1.84$ et $f(M_o) = 0.36$ pour celle de Fréchet.

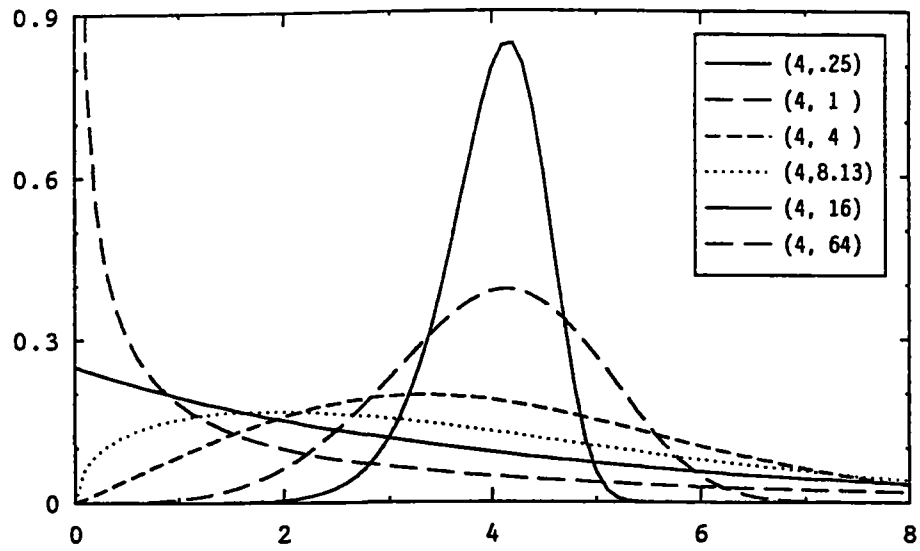


Figure 3.6 Distribution de Weibull (de minimum)

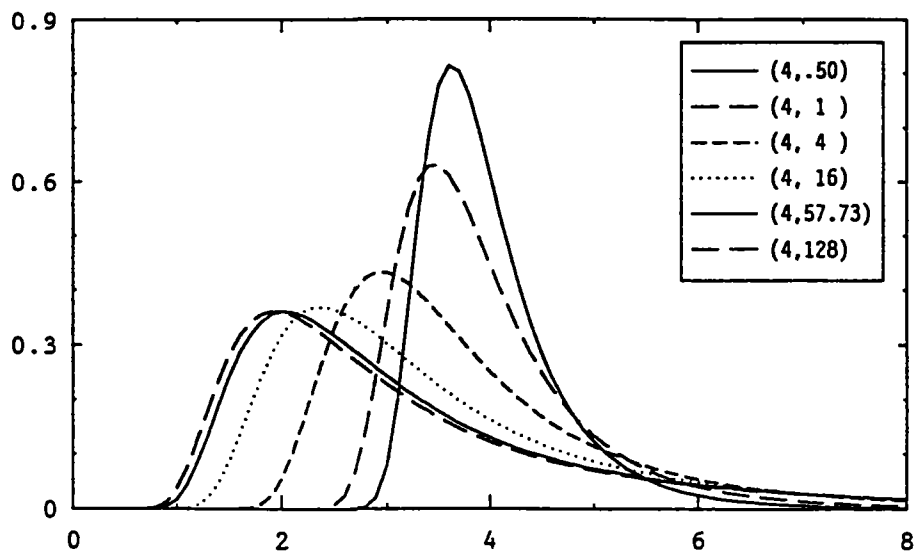


Figure 3.7 Distribution de Fréchet (de maximum)

Par le principe de symétrie $\mathrm{Min}\{X_i\} = -\mathrm{Max}\{-X_i\}$, nous obtenons les deux autres lois de valeurs minimale et maximale, appelées respectivement Weibull de maximum et Fréchet de minimum dont les densités sont défines sur l'espace $\{x;\ x < \omega\}$ par :

$$f(x|\alpha,\beta,\omega) = \frac{\alpha}{\omega-\upsilon}\cdot\left(\frac{\omega-x}{\omega-\upsilon}\right)^{\alpha-1}\exp\left[-\left(\frac{\omega-x}{\omega-\upsilon}\right)^{\alpha}\right] \tag{3.73a}$$

$$f(x|\alpha,\beta,\omega) = \frac{\alpha}{\omega-\upsilon}\cdot\left(\frac{\omega-\upsilon}{\omega-x}\right)^{\alpha+1}\exp\left[-\left(\frac{\omega-\upsilon}{\omega-x}\right)^{\alpha}\right] \tag{3.73b}$$

En utilisant le changement de variables $Y = \omega - \frac{(\upsilon-\varepsilon)(\omega-\upsilon)}{(x-\varepsilon)}$, la loi de Fréchet de minimum est transformée en celle de Weibull de même type ; la loi de Weibull de maximum correspond à celle de Fréchet de même type.

Nous nous limitons au cas où β est inconnu, donc $\lambda = \beta^{\mp\alpha}$ est inconnu et nous allons voir que la famille de distributions conjuguées est la famille gamma pour la variable λ.

Les fonctions de vraisemblance d'un échantillon $\mathbf{x} = (x,k)$ sont

$$f(\mathbf{x}|\beta) \propto \beta^{\mp k\alpha}\exp\left[-\beta^{\mp\alpha}\sum_i (x_i-\varepsilon)^{\pm\alpha}\right] \tag{3.74}$$

Si l'on désigne les moments empiriques d'ordre $\pm\alpha$, avec la valeur ε par $m_{\pm\alpha} = \frac{1}{n}\sum(x_i-\varepsilon)^{\pm\alpha}$, le couple $(m_{\pm\alpha},k)$ est donc une statistique exhaustive par rapport à la variable λ définie sur $\mathbb{R}^{+*}$.

Ce sont des cas particuliers de la distribution gamma généralisée. Donc nous avons le modèle gamma pour la variable λ avec la formule de passage :

$$\begin{cases} a_1 = a_0 + k \\ b_1 = b_0 + km_{\pm\alpha} \end{cases} \tag{3.75}$$

Pour la variable β, les distributions deviennent :

$$\pi(\beta) = \frac{\alpha b_0^{a_0}}{\Gamma(a_0)}\beta^{\mp\alpha a_0-1}\exp\left(-b_0\beta^{\mp\alpha}\right) \tag{3.76}$$

La propriété de conjuguaison de la famille de distributions gamma est valable pour les deux familles de distributions de Weibull et de Fréchet de premier paramètre α et de borne ε fixés. L'opérateur de conjuguaison est linéaire.

Les distributions bayésiennes sont calculées par :

$$p(x) = \int_{\mathbb{R}^{+*}} f(x|\lambda)\pi(\lambda)d\lambda = \alpha a\frac{b^a(x-\varepsilon)^{\pm\alpha-1}}{\left(b+(x-\varepsilon)^{\pm\alpha}\right)^{a+1}} \tag{3.77}$$

où (a,b) peut être (a_0,b_0) ou (a_1,b_1). Si l'on introduit les changements de variables $Y = (X-\varepsilon)^{\pm\alpha}/b + 1$ respectivement dans les deux formules , alors la variable Y admet la distribution de Pareto P(1,a) (cf. formule (3.70)). Or, la variable $Y = \frac{a}{b}(X-\varepsilon)^{\pm\alpha}$ admet une distribution de Fisher-Snedecor F(2,2a). Ainsi, nous obtenons la moyenne et la variance bayésiennes :

$$\begin{cases} \mathbb{E}[(X-\varepsilon)^{\pm\alpha}] = \dfrac{ab}{a-1} \\ \mathbb{V}[(X-\varepsilon)^{\pm\alpha}] = \dfrac{ab^2}{(a-1)^2(a-2)} \end{cases} \tag{3.78}$$

La moyenne et la variance bayésiennes de X sont calculées directement, nous obtenons respectivement les résultats pour la distribution de Weibull et pour la distribution de Fréchet :

$$\begin{cases} \mathbb{E}(X) = \varepsilon + \dfrac{b^{1/\alpha}}{\Gamma(a)}\cdot\Gamma\left(1+\dfrac{1}{\alpha}\right)\Gamma\left(a-\dfrac{1}{\alpha}\right) \\ \mathbb{V}(X) = b^{2/\alpha}\left[\Gamma\left(1+\dfrac{2}{\alpha}\right)\dfrac{\Gamma\left(a-\frac{1}{\alpha}\right)}{\Gamma(a)} - \Gamma\left(1+\dfrac{1}{\alpha}\right)^2\dfrac{\Gamma\left(a-\frac{1}{\alpha}\right)^2}{\Gamma(a)^2}\right] \end{cases}$$

$$\begin{cases} \mathbb{E}(X) = \varepsilon + \dfrac{b^{-1/\alpha}}{\Gamma(a)}\Gamma\left(1-\dfrac{1}{\alpha}\right)\Gamma\left(a+\dfrac{1}{\alpha}\right) \\ \mathbb{V}(X) = b^{-2/\alpha}\left[\Gamma\left(1-\dfrac{2}{\alpha}\right)\dfrac{\Gamma\left(a+\frac{2}{\alpha}\right)}{\Gamma(a)} - \Gamma\left(1-\dfrac{1}{\alpha}\right)^2\dfrac{\Gamma\left(a+\frac{1}{\alpha}\right)^2}{\Gamma(a)^2}\right] \end{cases}$$

En utilisant une approximation de la fonction Γ (cf. paragraphe 3.4) pour a assez grand, les résultats se simplifient pour les distributions de Weibull et de Fréchet :

$$\begin{cases} \mathbb{E}(X) \approx \varepsilon + (b/a)^{1/\alpha}\Gamma\left(1+\dfrac{1}{\alpha}\right) \\ \mathbb{V}(X) \approx (b/a)^{2/\alpha}\left[\Gamma\left(1+\dfrac{2}{\alpha}\right) - \Gamma\left(1+\dfrac{1}{\alpha}\right)^2\right] \end{cases} \tag{3.79a}$$

$$\begin{cases} \mathbb{E}(X) \approx \varepsilon + (a/b)^{1/\alpha}\Gamma\left(1-\dfrac{1}{\alpha}\right) \\ \mathbb{V}(X) \approx (a/b)^{2/\alpha}\left[\Gamma\left(1-\dfrac{2}{\alpha}\right) - \Gamma\left(1-\dfrac{1}{\alpha}\right)^2\right] \end{cases} \tag{3.79b}$$

Toutes ces formules sont similaires à celles de la moyenne et la variance

théoriques en fonction de $(\alpha,\beta,\varepsilon)$ ci-dessus.

L'approximation de la fonction Γ est valable pour a assez grand (cf. paragraphe 3.4.1). Si le paramètre (a_0,b_0) de la distribution a priori est estimé par la méthode séquentielle, on a : $a_0 = k + 1$ et $b_0 = km_{\pm\alpha}$, où $(m_{\pm\alpha},k)$ est la statistique exhaustive d'un échantillon $\mathbf{x}$. En ce qui concerne la distribution bayésienne a posteriori, d'après la formule (3.73), nous avons : $a_1 = a_0 + k$. Donc si $a = a_0$ ou a_1 est grand, cela revient à dire que la taille d'échantillon est grande. De plus, la moyenne et la variance bayésiennes ne dépendent pas des estimateurs empiriques m et v, mais de $m_{\pm\alpha}$ les moments empiriques (non centrés) d'ordre $\pm\alpha$. La valeur de α est souvent calculée à l'aide de la formule (3.72) ou les trois premiers moments si $\varepsilon \neq 0$. Il enrésulte que les analyses classique et bayésienne ne donnent toujours pas les mêmes résultats, même si la taille de l'échantillon est grande. Ce qui semble surprenant.

Dans le cas où la variation du paramètre α est assez grande, la variable $\lambda = \beta^{\mp\alpha}$ est très dispersée et donc la procédure ne sera plus adaptée. On doit considérer la variable $\lambda = \beta^{\mp 1}$ directement, puis utiliser le modèle gamma pour λ, mais le modèle prédictif n'est pas simple.

3.8. DISTRIBUTION DE PARETO

C'est la distribution de la variable $X = \alpha\cdot e^{Y}$ définie sur l'ensemble $\{x;\ x > \alpha\} \subset \mathbb{R}$, lorsque la variable $Y = \log(X/\alpha)$ suit une loi exponentielle de paramètre $\beta > 0$ qui n'est autre que la loi gamma généralisée (cf. paragraphe 3.5). Sa densité est :

$$f(x|\beta) = \beta\alpha^{\beta}x^{-\beta-1} \tag{3.80}$$

Les relations entre (μ,σ^2) et (α,β) sont données par :

$$\begin{cases} \mu = \dfrac{\beta\alpha}{\beta-1} & \beta > 1 \\ \sigma^2 = \dfrac{\beta\alpha^2}{(\beta-1)^2(\beta-2)} & \beta > 2 \end{cases} \qquad \begin{cases} \beta = 1+\sqrt{1+\mu^2/\sigma^2} = 1+\dfrac{1}{\delta}\sqrt{1+\delta^2} \\ \alpha = \mu\left(1 - \dfrac{\delta}{\delta+\sqrt{1+\delta^2}}\right) \end{cases} \tag{3.81}$$

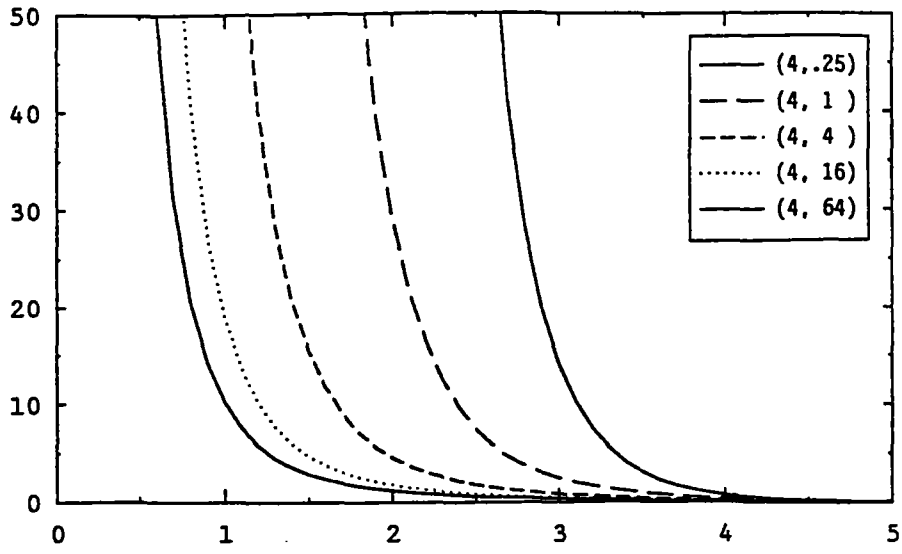


Figure 3.8 Distribution de Pareto

La distribution est décroissante en x. La figure 3.8 montre différentes densités de Pareto associées aux différents couples des valeurs (α,β).

L'inconnu du problème est le paramètre β. Dans ce cas, désignons par **x** un échantillon d'une distribution de Pareto avec un même multiplicateur α, alors la fonction de vraisemblance s'écrit :

$$f(x|\beta) \propto \beta^{k}\exp\left[-\beta k\left(\log m_G - \log\alpha\right)\right] \tag{3.82}$$

Nous trouvons que le couple $(t = \log m_G - \log\alpha,\ k)$ est une statistique exhaustive par rapport à la variable β définie sur $\mathbb{R}^{+*}$, où m_G est la moyenne géométrique de l'échantillon. Et le noyau de la fonction de vraisemblance est proportionnel à la densité d'une distribution gamma. Nous choisissons donc pour distribution a priori de la variable β , une distribution gamma :

$$\pi(\beta) = G\left(\beta;\ a_0, b_0\right) \tag{3.83}$$

La distribution a posteriori de la variable β est gamma : $G\left(a_1, b_1\right)$ avec les paramètres définis par :

$$\begin{cases} a_1 = a_0 + k \\ b_1 = b_0 + kt \end{cases} \tag{3.84}$$

Ainsi, la famille de distributions gamma est conjuguée par rapport à la famille de distributions de Pareto de paramètre α connu.

La définition de la norme statistique "t" est aussi stable comme celles des normes dans les cas précédents. Si le paramètre t_0 de la distribution a priori est déterminé par $(\log m_{G0} - \log\alpha_0)$, où (m_{G0}, k_0) est la statistique exhaustive d'un échantillon $\mathbf{x}_0$ avec le multiplicateur α_0 associé, alors $\mathbf{x}$ étant un échantillon avec le multiplicateur α, nous déduisons de la formule (3.84) le paramètre t_1 de l'a posteriori sous la même forme $(\log m_{G1} - \log\alpha_1)$, où (m_{G1}, k_1) est la statistique exhaustive de l'échantillon composé $\mathbf{x}_1 = \mathbf{x} \cup \mathbf{x}_0$ avec le multiplicateur α_1 défini comme la moyenne géométrique de (α_0, k_0) et (α, k) : $\alpha_1 = \left(\alpha^k \times \alpha_0^{k_0}\right)^{1/k_1}$ avec $k_1 = k + k_0$. Finalement, l'opérateur de conjuguaison est linéaire :

$$g\left(\theta \mid T(\mathbf{x} \cup \mathbf{x}_0)\right) = g\left(\theta \mid T(\mathbf{x})\right) + g\left(\theta \mid T(\mathbf{x}_0)\right) \quad .$$

Comme la variable X est définie sur l'ensemble $\{x > \alpha\}$, la statistique $t = \log m_G - \log\alpha$ est toujours positive.

Maintenant, nous pouvons effectuer le calcul des intégrales suivantes :

$$p(x) = \int_{\mathbb{R}^{+*}} \beta\alpha^{\beta} x^{-\beta-1} \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-\beta b} d\beta = \frac{ab\, x^{a-1}}{\left(b + \log x - \log\alpha\right)^{a+1}} \qquad (3.85)$$

Il en résulte que la variable $Y = \frac{a}{b}(\log X - \log\alpha)$ suit une loi de Fisher-Snedecor de paramètres 2 et 2a avec $(a,b) = (a_0, b_0)$ ou (a_1, b_1).

Comme dans le cas de lois log-normales, la moyenne et la variance bayésiennes n'existent pas , et nous ne pouvons utiliser ce modèle qu'après une transformation des variables de X en logX avec α choisi a priori.

Si le paramètre (a,b) de la loi a priori est défini par la statistique exhaustive $(t = \log m_G - \log\alpha,\ k)$ d'un échantillon $\mathbf{x}$ et par le paramètre α, de sorte que $a = k+1$ et $b = kt$, alors la variable $Y = \frac{k+1}{k} \cdot \frac{\log X - \log\alpha}{\log m_G - \log\alpha}$ suit une loi de F de paramètres 2 et 2(k+1). Donc la moyenne et la variance bayésiennes de la variable $Y = \log X$ sont définies par : $\mathbb{E}(\log X) = \log m_G$, pour $k > 1$, $\mathbb{V}(\log X) = \frac{k+1}{k-1}(\log m_G - \log\alpha)^2$.

3.9. RESUME

Dans les deux tableaux ci-dessous, nous donnons une liste des lois de distribution continues étudiés dans ce chapitre. Dans ces tableaux, nous indiquons :

— distribution sous-jacente $f(x|\theta)$
— statistique exhaustive
— distributions a priori et a posteriori (conjuguées)
— formule de passage de l'a priori à l'a posteriori

— distribution sous-jacente $f(x|\theta)$
— distribution du paramètre θ
— distribution bayésienne $p(x)$

Distributions Conjuguées et Détermination De Distribution A Posteriori

Loi Apportée $f(x\|\theta)$		Loi A Priori $\pi(\theta)$	Echantillon $\mathbf{x} = (x, k)$	Loi A Posteriori $\pi(\theta\|x)$
Normale $N(\mu, \sigma^2)$	$\mu,\ \eta$	$\pi(\mu,\eta) = NG(\mu,\eta;\ m_0, v_0, k_0)$ $= N(\mu;\ m_0, \frac{1}{2k_0\eta})G(\eta;\ \frac{k_0+1}{2}, k_0 v_0)$	(m, v, k)	$k_1 = k_0 + k$ $k_1 m_1 = k_0 m_0 + km$ $k_1^2 v_1 = k_1(k_0 v_0 + kv) + k_0 k(m-m_0)^2$
	μ	$\pi(\mu) = N(\mu;\ m_0, v_0)$	$(m, k),\ \eta$ $[\eta = 1/(2v)]$	$h_1 m_1 = h_0 m_0 + k\eta m$ $v_1 = 1/(2h_1)$ avec $h_1 = h_0 + k\eta$
	$\eta = \frac{1}{2}\sigma^{-2}$	$\pi(\eta) = G(\eta;\ a_0, b_0)$	$(m, v, k),\ \mu$	$a_1 = a_0 + k/2$ $b_1 = b_0 + k(v + (\mu-m)^2)$
$X \approx\approx\approx LN(\alpha, \beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha, \beta^2)$			(m_{ln}, v_{ln}, k)	idem
Normale inverse $NI(\mu, \beta^2)$	$\alpha,\ \lambda$	$\pi(\alpha,\lambda;\ a_0 = m_0^{-1}, b_0 = m_{H0}^{-1} - m_0^{-1}, k_0)$ $= N(\alpha;\ a_0, \frac{a_0}{2k_0\lambda})G(\lambda;\ \frac{k_0+1}{2}, k_0 b_0)$	(m, m_H, k) $a = m^{-1}$ $b = m_H^{-1} - m^{-1}$	$k_1 = k_0 + k$ $k_1/a_1 = k_0/a_0 + k/a$ $k_1 b_1 = k_0 b_0 + kb + \frac{a_1}{k_1}\frac{k_0}{a_0}\frac{k}{a}(a-a_0)^2$
	$\alpha = \mu^{-1}$	$\pi(\alpha) = N(\alpha;\ m_0, v_0)$	$(m, k),\ \lambda$ $[\lambda = m^3/(2v)]$	$h_1 m_1 = h_0 m_0 + \lambda k$ $v_1 = 1/(2h_1)$ avec $h_1 = h_0 + \lambda km$
	$\lambda = \frac{1}{2}\beta^{-2}$	$\pi(\lambda) = G(\lambda;\ a_0, b_0)$	$(m, m_H, k),\ \mu$ $b = m_H^{-1} - m^{-1}$	$a_1 = a_0 + \frac{k}{2}$ $b_1 = b_0 + k(b + m(\mu^{-1} - m^{-1})^2)$

Distributions Conjuguées et Détermination De Distribution A Posteriori (suite)

Loi Apportée $f(x\|\theta)$		Loi A Priori $\pi(\theta)$	Echantillon $\mathbf{x} = (x, k)$	Loi A Posteriori $\pi(\theta\|x)$
Gamma $G(\alpha,\beta)$	$\alpha,\ \beta$	$\pi(\alpha,\beta) = GG(\alpha,\beta;\ m_0, r_0, k_0)$ $= G(\beta;\ k_0\alpha, k_0 m_0)G(\alpha;\ \frac{k_0+3}{2}, k_0 r_0)$	(m, m_G, k) $m_G = (\prod x_i)^{1/k}$ $r = \log(m/m_G)$	$k_1 = k_0 + k$ $k_1 m_1 = k_0 m_0 + km$ $k_1 r_1 = k_0 r_0 + kr + k_1 \log m_1 - k_0 \log m_0 - k\log m$
Gamma inverse $GI(\alpha,\beta)$	$\alpha,\ \beta$	$\pi(\alpha,\beta) = GG(\alpha,\beta;\ n_0, s_0, k_0)$ $= G(\beta;\ k_0\alpha, k_0 n_0)G(\alpha;\ \frac{k_0+3}{2}, k_0 s_0)$	(m_H, m_G, k) $m_H^{-1} = \frac{1}{k}\sum x_i^{-1}$ $s = \log(m_G/m_H)$	$k_1 = k_0 + k$ $k_1 n_1 = k_0 n_0 + km_H^{-1}$ $k_1 s_1 = k_0 s_0 + ks + k_1 \log n_1 - k_0 \log n_0 - k\log m_H^{-1}$
Gamma généralisée $G(\upsilon,\nu,x_\circ,\alpha,\beta)$ $\lambda = \beta^{\upsilon}$		$\pi(\lambda) = G(\lambda;\ a_0, b_0)$	$(m_\nu, k),\ \alpha$ $\alpha = m_\nu^2/V(m_\nu)$	$a_1 = a_0 + k\alpha$ $b_1 = b_0 + km_\nu$
Gamma Gamma inverse	$\lambda = \beta$	$G(\alpha,\beta) = G(1,1,0,\alpha,\beta)$ $GI(\alpha,\beta) = G(1,-1,0,\alpha,\beta)$	$(m_{\pm 1}, k),\ \alpha$ $m_{-1} = m_H^{-1}$	$a_1 = a_0 + k\alpha$ $b_1 = b_0 + km_{\pm 1}$
Weibull Fréchet	$\lambda = \beta^{\mp\alpha}$	$W(\alpha,\beta) = G(-\alpha,\alpha,\varepsilon,1,\beta)$ $Fr(\alpha,\beta) = G(\alpha,-\alpha,\varepsilon,1,\beta)$	$(m_{\pm\alpha}, k),\ \alpha$	$a_1 = a_0 + k$ $b_1 = b_0 + km_{\pm\alpha}$
Pareto $P(\alpha,\beta)$	β	$\pi(\beta) = G(\beta;\ a_0, b_0)$	$(t, k),\ \alpha$ $t = \log(m_G/\alpha)$	$a_1 = a_0 + k$ $b_1 = b_0 + kt$

DISTRIBUTIONS USUELLES ET DISTRIBUTIONS BAYÉSIENNES

Distribution au sens usuel $f(x\|\theta)$	Distribution a priori de θ $\pi(\theta)$	Distribution au sens bayésien $p(x) = \int_\Theta f(x\|\theta)\pi(\theta)d\theta$
$X \approx\approx\approx N(\mu,\sigma^2)$	$(\mu,\eta) \approx\approx\approx NG(m,v,k)$ $\eta = 1/(2\sigma^2)$	$X \approx\approx\approx m + \sqrt{v}T(k+1)$ $\cong N(m,v)$ pour k grand
$X \approx\approx\approx N(\mu,\sigma^2)$ σ^2 connue	$\mu \approx\approx\approx N(m,v)$	$X \approx\approx\approx N(m,v+\sigma^2)$
$X \approx\approx\approx N(\mu,\sigma^2)$ μ connue	$\eta \approx\approx\approx G(a,b)$	$X \approx\approx\approx \mu+cT(2a)$ avec $c^2=\frac{b}{2a}$ $\cong N(\mu,c^2)$ pour a grand
$X \approx\approx\approx LN(\alpha,\beta) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha,\beta^2)$		
$X \approx\approx\approx NI(\mu,\beta^2)$ voir la section 3.3		
$X \approx\approx\approx G(\alpha,\beta)$	$(\alpha,\beta) \approx\approx\approx GG(m,r,k)$	$mX \approx\approx\approx \int F(2k\alpha,2\alpha)(x)\pi(\alpha)d\alpha$ $\pi(\alpha) = G(\frac{k+3}{2},kr)(\alpha)$
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow X^{-1} \approx\approx\approx G(\alpha,\beta)$		
$X \approx\approx\approx G(\upsilon,\nu,x_\circ\alpha,\beta)$ Gamma généralisée	$\lambda = \beta^\upsilon \approx\approx\approx G(a,b)$	$\|X-x_\circ\|^\nu \approx\approx\approx \frac{\alpha b}{a}F(2\alpha,2a)$
$X \approx\approx\approx G(\alpha,\beta) = G(1,1,0,\alpha,\beta)$ $X \approx\approx\approx GI(\alpha,\beta) = G(1,-1,0,\alpha,\beta)$		$X^{\pm 1} \approx\approx\approx \frac{\alpha b}{a}F(2\alpha,2a)$
$X \approx\approx\approx W(\alpha,\beta) = G(-\alpha,\alpha,0,1,\beta)$ $X \approx\approx\approx Fr(\alpha,\beta) = G(\alpha,-\alpha,0,1,\beta)$		$X^{\pm\alpha} \approx\approx\approx \frac{b}{a}F(2,2a)$
$X \approx\approx\approx P(\alpha,\beta)$ α connu	$\beta \approx\approx\approx G(a,b)$	$\ln\frac{X}{\alpha} \approx\approx\approx \frac{b}{a}F(2,2a)$

CHAPITRE IV ESTIMATIONS DES PARAMÈTRES DE LA LOI A PRIORI

4.0. INTRODUCTION

D'après ce qui précède, pour exploiter une distribution bayésienne (ou prédictive a posteriori) d'une variable X, la méthodologie bayésienne, bien que relativement simple en théorie, entraîne une certaine complexité dans les processus d'analyse, même si l'on connaît la forme de la distribution des observations réalisées par le passé. Lorsque la forme de la loi de distribution est choisie, développer un tel modèle consiste à choisir un estimateur convenable de son paramètre.

Nous avons vu comment construire la distribution a posteriori et les distributions bayésiennes a priori et a posteriori à partir d'un échantillon et d'une distribution a priori. Il reste à évaluer la distribution a priori. Selon les principes exposés au chapitre II, nous allons établir les différents estimateurs de ses paramètres à partir de la réalisation d'expériences identiques que l'on peut obtenir. Toutes les formules sont listées dans les tableaux à l'annexe 2.

Pour un modèle quelconque appliqué, le résultat prédictif a posteriori est toujours une pondération entre deux informations : information a priori et information provenant du lot à tester, mais les poids représentant ces informations dépendent des méthodes d'exploitation. Nous prendrons le modèle normal-gamma pour comparer ces méthodes et nous étudierons également l'influence des paramètres de la distribution a priori et de l'échantillon $\mathbf{x}$ sur les modèles prédictifs.

Ensuite, nous présenterons le modèle normal-gamma à quatre paramètres $NG(m,v,k,n)$ pour le couple des variables (μ,η), représentant les paramètres inconnus d'une distribution normale. Ce modèle ne peut être validé que par la méthode du maximum de vraisemblance dans le cas où les résultats sont regroupés (cf. paragraphe 2.4) et par les méthodes du quasi-échantillon :

parmi celles-ci, la méthode des doubles maxima de vraisemblance a été développée dans les travaux antérieurs, référencés presque tous par le livre de H. Raiffa & R. Schlaifer (1961) ([36]), mais l'estimateur obtenu est aussi appelé l'estimateur du maximum de vraisemblance. Par contre, notre modèle normal-gamma à trois paramètres $NG(m,v,k)$ présenté au chapitre III peut être combiné avec différentes méthodes que nous avons développées au chapitre II.

Nous ferons dans l'application numérique ci-après quelques commentaires sur ces estimateurs de paramètres pour montrer les conditions d'utilisation et leur influence sur le comportement de la distribution prédictive, lors de l'expérience réalisée sur un lot à tester.

Enfin, nous montrons à la fin du chapitre les modèles issus de la normalité de la distribution sous-jacente, dans un tableau comparatif général des méthodes retenues, ainsi que la modification due à l'estimateur v' de la variance sans biais au lieu de celui de variance minimale. Mais nous verrons que cette modification est très faible.

Supposons dans ce chapitre que l'information a priori forme un ensemble de n échantillons $\mathbf{x}_1 = (x_1,k_1)$, $\mathbf{x}_2 = (x_2,k_2)$, ..., $\mathbf{x}_n = (x_n,k_n)$. Comme tout le modèles considérés ici possèdent une statistique exhaustive, cet ensemble est tranformé en un ensemble plus petit $T(\mathbf{x}_1)$, $T(\mathbf{x}_2)$,..., $T(\mathbf{x}_n)$ sans réduire la connaissance a priori.

Dans ce chapitre, nous supprimerons tous les indices "$_0$" puisque nous traiterons uniquement la distribution a priori.

4.1. ESTIMATION SEQUENTIELLE

Nous savons que l'obtention de la distribution a priori du paramètre inconnu θ est fondée sur le principe de la conjuguée naturelle. Nous avons montré au chapitre I que pour chaque distribution exponentielle, l'opération de conjuguaison exprimée en sa statistique exhaustive $\tilde{T}(\mathbf{x})$ est linéaire, en passant la distribution a priori à la distribution a posteriori.

La méthode séquentielle repose sur le principe : considérer une loi de

distribution a posteriori comme loi a priori pour un échantillon suivant. A partir des n échantillons et en partant d'une mesure uniforme du paramètre inconnu θ, la distribution a priori de θ obtenue vérifie la propriété :

$$\tilde{T}(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n) = \tilde{T}(\mathbf{x}_1) + \tilde{T}(\mathbf{x}_2) + \ldots + \tilde{T}(\mathbf{x}_n) \tag{4.10}$$

En terme de noyaux de fonction de vraisemblance $f(\mathbf{x}|\theta) \propto g\big(\tilde{T}(\mathbf{x})|\theta\big)$ d'un échantillon quelconque $\mathbf{x}$, cet opération de conjuguaison correspondant aux n échantillons s'écrit :

$$\begin{aligned} \prod_i g\big(\tilde{T}(\mathbf{x}_i)|\theta\big) &\propto g\big(\sum_i \tilde{T}(\mathbf{x}_i)|\theta\big) \\ &\propto g\big(\tilde{T}(\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n)|\theta\big) \end{aligned} \tag{4.11}$$

puisque toutes les distributions usuelles appartiennent à la famille exponentielle (cf. paragraphe 1.1). Nous avons alors

$$\pi(\theta) \propto \prod_i f\big(\mathbf{x}_i|\theta\big) \propto \prod_i g\big(\tilde{T}(\mathbf{x}_i)|\theta\big) \propto g\big(\tilde{T}(\mathbf{x})|\theta\big) \tag{4.12}$$

où $\tilde{T}(\mathbf{x})$ est la statistique exhaustive de l'échantillon regroupé $\mathbf{x}$:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_1 \cup \mathbf{x}_1 \cup \ldots \cup \mathbf{x}_n \\ &= \{x_j^i;\ j = 1,2,\ldots,n_i \text{ et } i = 1,2,\ldots,n\} \end{aligned} \tag{4.13}$$

Ainsi, nous montrons que la distribution a priori de θ est déterminée par la statistique exhaustive $\tilde{T}(\mathbf{x})$ de l'échantillon regroupé $\mathbf{x}$.

Nous précisons ici les trois cas correspondant la distribution normale $N\big(\mu,\sigma^2\big)$, les autres étant donnés dans le tableau à l'annexe 2.

Nous définissons d'abord la moyenne des variances pondérées ou non par les tailles :

$$VT = \frac{1}{KT}\sum_i k_i v_i \qquad \text{et} \qquad VM = \frac{1}{n}\sum_i v_i$$

Nous désignons par MT, VMT et KT la moyenne totale, la variance totale et la taille totale de l'échantillon $\mathbf{x}$ définis par :

$$\begin{cases} MT = \frac{1}{KT}\sum_i k_i m_i = \frac{1}{KT}\sum_i \sum_j x_j^i \\ KT = \sum k_i \\ VMT = \frac{1}{KT}\sum_i k_i \left(v_i + (m_i - MT)^2\right) = \frac{1}{KT}\sum_i \sum_j \left(x_j^i - MT\right)^2 \end{cases} \tag{4.14}$$

<u>Cas 1. μ et $\eta = 1/(2\sigma^2)$ inconnus et modèle normal-gamma NG(m,v,k)</u>

L'estimateur du paramètre de $\pi(\mu,\eta)$ est :

$$(m,v,k) = (MT, VMT, KT)$$

d'où la moyenne de la variable μ et celle de la varaible σ^2 :

$$\mathbb{E}(\mu) = MT \qquad \mathbb{E}(\eta) = \frac{k+1}{2kv}$$

$$\Longrightarrow \quad \mathbb{E}(\sigma^2) = \frac{KT}{KT+1} VMT \approx VMT \quad .$$

<u>Cas 2. μ inconnu et η connu et modèle normal N(m,v)</u>

Dans le cas où la variance σ^2 est considérée constante, l'estimateur du paramètre du modèle normal est :

$$(m,v) = (MT,\ VT/KT)$$

Par contre, si nous considérons les variances v_i (ou les précisions h_i) empiriques des échantillons dans ce modèle, l'estimateur est

$$(m,v) = \left(MHT,\ \frac{1}{2KT \cdot HT}\right)$$

où $MHT = \sum_i k_i h_i m_i / \sum_i k_i h_i$ et $HT = \frac{1}{KT}\sum_i k_i h_i$.

La statistique m = MHT est la moyenne des moyennes pondérées par les

précisions et les tailles. Si l'on désigne par VHT la moyenne harmonique des variances pondérées par les tailles, on a VHT = 1/(2HT) et

$$(m,v) = (MHT,\ VHT/KT)$$

Cas 3. μ connu et η inconnu et modèle gamma G(a,b)

L'estimateur du paramètre de la distribtuion gamma est

$$(a,b) = (KT/2+1,\ KT\cdot VMT)$$

lorsque la moyenne est constante, μ = MT par exemple. Si l'on prend $\mu_i = m_i$, alors nous obtenons :

$$(a,b) = (KT/2+1,\ KT\cdot VT).$$

La moyenne a priori de la variance est : $\mathbb{E}(\sigma^2) = \frac{1}{2\mathbb{E}(\eta)} = VMT\frac{KT}{KT+2} \approx VMT$ ou $VT\frac{KT}{KT+2} \approx VT$ dans le deuxième cas.

4.2. ESTIMATION DES ECHANTILLONS EQUILIBRES

L'idée de cette méthode est de construire la fonction de vraisemblance bayésienne de n échantillons de manière suivante :

$$L'(\theta;\mathbf{x}) = L'(\theta;\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n) = \prod_i f(\mathbf{x}_i|\theta)^{K/k_i} \tag{4.20}$$

et d'appliquer le principe de la conjuguée naturelle :

$$\pi(\theta) \propto L'(\theta;\mathbf{x}) = \prod_i f(\mathbf{x}_i|\theta)^{K/k_i} \tag{4.21}$$

avec $K = \left(\sum k_i^{-1}\right)^{-1}$. Dans ces formules, les scores $\{K/k_i\}$ sont utilisés pour éliminer les effets dus aux tailles différentes. Donc tous les échantillons, ainsi que les lots correspondants, présentent réellement des poids identiques dans l'interpètation de l'information a priori.

Dans le cas de la distribution normale de moyenne μ et de variance σ^2 inconnues, la distribution normale-gamma $\pi(\theta)$ est directement déterminée par la statistique exhaustive (cf. paragraphes 2.3 et 3.1.1), dont le calcul est le suivant :

$$\begin{aligned}\pi(\mu,\eta) &\propto \prod_i f(x_i|\mu,\eta)^{K/k_i}\\ &\propto \prod_i \left(\eta^{k_i/2}\exp\left(-k_i\eta(m_i-\mu)^2\right)\exp\left(-k_i v_i\eta\right)\right)^{K/k_i}\\ &\propto \left(\prod_i \eta^{1/2}\exp\left(-\eta(m_i-\mu)^2\right)\exp\left(-v_i\eta\right)\right)^{K} \qquad (4.22)\end{aligned}$$

d'où la statistique exhaustive de cette fonction de vraisemblance approchée est définie par

$$\begin{cases} MM = \frac{1}{n}\sum_i m_i \\ VMM = \frac{1}{n}\sum_i \left(v_i + (m_i - MM)^2\right) \\ KH = \left(\frac{1}{n}\sum_i k_i^{-1}\right)^{-1} \end{cases} \qquad (4.23)$$

où KH = nK est la moyenne harmonique des tailles $\{k_i\}$.

De la même manière, nous obtenons l'estimateur du modèle normal de la variable μ dans les cas où la variance σ^2 est soit constante soit estimée de façon empirique pour chacun des échantillons :

$$\begin{cases} m = MM \\ v = VM/KH \end{cases} \qquad (4.24)$$

$$\begin{cases} m = MHM \\ v = VH/KH \end{cases} \qquad (4.25)$$

où $MHM = \sum h_i m_i / \sum h_i$ et $HM = \frac{1}{n}\sum h_i$ sont la moyenne des moyennes pondérées par les précisions et la moyenne des précisions ; et VH = 1/(2HM) est la moyenne harmonique des variances.

L'estimateur du paramètre du modèle gamma de la variable η dans les cas

où la moyenne μ est soit constante μ = MM par exemple, soit estimée de façon empirique pour chacun des échantillons vaut :

$$\begin{cases} a = KH/2 + 1 \\ b = KH \cdot VMM \end{cases} \tag{4.46}$$

$$\begin{cases} a = KH/2 + 1 \\ b = KH \cdot VM \end{cases} \tag{4.47}$$

Donc, la moyenne a priori de la variance est $\mathbb{E}(\sigma^2) = VMM\frac{KH}{KH+2} \approx VMM$ ou $VM\frac{KH}{KH+2} \approx VM$ dans les deuxième cas.

4.3. ESTIMATIONS DES FAMILLES PARAMETRIQUES

Nous rappelons que ces méthodes de construction d'une distribution a priori $\pi(\theta)$ du paramètre θ se fait à partir de plusieurs séries de résultats non classés. Ces résultats sont donc considérés comme répartis suivant une distribution marginale qui est ici la distribution bayésienne p(x) résultant d'une moyenne des fonctions de densité $f(x|\theta)$ pondérées par les probabilités des valeurs possibles du paramètre θ.

La formule $p(x) = \int_\Theta f(x|\theta)\pi(\theta)d\theta$ permet d'établir les relations entre les moments de p(x) et ceux de $\pi(\theta)$, d'où le paramètre de la distribution a priori en fonction des moments de p(x), estimés de façon empirique :

$$k = KT = \sum_i k_i \tag{4.30}$$

$$\upsilon_1 = MT = \frac{1}{k}\sum_i \sum_j x_{ij} \qquad \mu_2 = VMT = \frac{1}{k}\sum_i \sum_j (x_j^i - MT)^2$$

$$\mu_3 = V3 = \frac{1}{k}\sum_i \sum_j (x_j^i - MT)^3 \qquad \mu_4 = V4 = \frac{1}{k}\sum_i \sum_j (x_j^i - MT)^4$$

A partir des résultats non-classés, nous pouvons construire la fonction de vraisemblance d'après la même formule ci-dessus, d'où l'estimateur du maximum de vraisemblance, mais les calculs sont souvent difficiles ; nous ne le calculons que pour les trois modèles issus de la distribution normale.

4.3.1. Distributions normale $N(\mu,\sigma^2)$ et log-normale $LN(\alpha,\beta^2)$

Cas 1. μ, $\eta = 1/(2\sigma^2)$ inconnus

Loi a priori : $\pi(\mu,\eta) = NG(\mu,\eta;\ m,v,k) = N(\mu;\ m,\frac{1}{2k\eta})G(\eta;\ \frac{k+1}{2},kv)$

Loi bayésienne : $p(y = \frac{x-m}{\sqrt{v}}) = T(y;\ k+1)$

(a) Moments : $\upsilon_1 = \mathbb{E}^p[X] = m \qquad \mu_{2n+1} = \mathbb{E}^p[(X-m)^{2n+1}] = 0$

$$\mu_{2n} = \mathbb{E}^p[(X-m)^{2n}] = v^n(2n-1)!!\frac{(k+1)^n}{(k-1)(k-3)\ldots(k-2n+1)}$$

Paramètres : $m = MT \qquad v = \frac{C4+3}{2C4+3}VMT \qquad k = 3 + \frac{6}{C4}$ (4.311a)

avec $C4 = V4/VMT^2 - 3$ coefficient d'aplatissement empirique

Remarque : Pour que les intégrales μ_2 et μ_4 existent, il faut que la condition $k > 3$ soit vérifiée, c'est à dire $C4 > 0$. Alors la distribution bayésienne est plus pointue que la distribution normale, qui a 0 pour coefficient d'aplatissement.

(b) Fonction de vraisemblance : $\left[\frac{\Gamma(\frac{k}{2}+1)}{\sqrt{(k+1)v}\,\Gamma(\frac{k+1}{2})}\right]^{KT}\prod_i\left(1 + \frac{(x_i - m)^2}{(k+1)\cdot v}\right)^{-(k/2+1)}$

Désignons $c_i = (x_i - m)^2 + (k+1)v$. L'équation de vraisemblance est :

$$m = \sum x_i c_i^{-1} / \sum c_i^{-1} \qquad v = \frac{1}{k+2}\left(\frac{1}{KT}\sum c_i^{-1}\right)^{-1} \tag{4.311b}$$

$$\psi(\frac{k}{2}+1) - \psi(\frac{k+1}{2}) + \log(\frac{k+1}{k+2}) = \log(\frac{1}{KT}\sum c_i^{-1}) - \frac{1}{KT}\sum \log c_i^{-1}$$

En utilisant une approximation de la fonction digamma pour x suffisamment grand : $\psi(x) \approx \log(x) - \frac{1}{2x}$ (cf. [1]), nous avons

$$\frac{1}{(k+1)(k+2)} = \log(\frac{1}{KT}\sum c_i^{-1}) - \frac{1}{KT}\sum \log c_i^{-1}$$

Nous en concluons que le paramètre "k" est une grandeur pour mesurer la

dispersion des résultats.

Cas 2. μ inconnu (σ^2 connu)

Loi a priori : $\pi(\mu) = N(\mu;\ m, v)$

Loi bayésienne : $p(x) = N(x;\ m, \sigma^2+v)$

Paramètres : $m = MT \qquad v = VMT - \sigma^2$ (4.312)

Remarque : L'estimateur du maximum de vraisemblance est égal à celui des moments.

Cas 3. η inconnu (μ connu)

Loi a priori : $\pi(\eta) = G(\eta;\ a, b)$

Loi bayésienne : $p\left(y = \sqrt{\frac{2a}{b}}(x-\mu) = \frac{x-\mu}{\sqrt{v}}\right) = T(y;\ 2a)$

(a) Moments : $\upsilon_1 = \mu \qquad \mu_{2n+1} = 0$

$$\mu_{2n} = (2n-1)!!\frac{a^n}{(a-1)(a-2)\ldots(a-n)}\left(\frac{b}{2a}\right)^n \qquad \text{si } n < a$$

Paramètres : $a = 2 + \frac{3}{C4} \qquad b = 2\cdot VMT\left(1 + \frac{3}{C4}\right)$ (4.313a)

Remarque : Pour que μ_2 et μ_4 existent, la condition $a > 2$ doit être vérifiée, c'est à dire que $C4 > 0$. Si la distribution a priori gamma $G(a,b)$ est paramétrée par (v,k), de sorte que $a = \frac{k}{2}+1$ et $b = kv$, alors nous avons : $k = 2 + \frac{6}{C4}$ et $v = VMT$.

On a souvent $\mu = MT$.

(b) Fonction de vraisemblance : $\left[\frac{\Gamma\left(a+\frac{1}{2}\right)}{\sqrt{2av}\,\Gamma(a)}\right]^{KT} \prod_i \left(1 + \frac{(x_i - m)^2}{2a\cdot v}\right)^{-(a+1/2)}$

Notons $c_i = (x_i - m)^2 + 2a\cdot v$. L'équation de vraisemblance est :

$$m = \sum x_i c_i^{-1} / \sum c_i^{-1} \qquad v = \frac{1}{2a+1}\left(\frac{1}{KT}\sum c_i^{-1}\right)^{-1} \tag{4.313b}$$

$$\frac{1}{2a(2a+1)} = \log\left(\frac{1}{KT}\sum c_i^{-1}\right) - \frac{1}{KT}\sum \log c_i^{-1} \qquad \text{(pour a grand).}$$

Cas 4. Distribution log-normale

Nous pouvons considérer la variable lnX qui suit une loi normale, ainsi que les échantillons de X transformés en échantillons de lnX sur lesquels on applique les procédures précédentes.

Nous remarquons que dans le cas où α est inconnu et β connu et pour une distribution log-normale, la variable X admet aussi une distribution log-normale au sens bayésien. Lorsque l'on choisit comme distribution a priori une distribution normale $N(m,v)$, les deux procédures, considérant la variable lnX ou la variable X elle-même, donnent deux estimateurs différents du paramètre de la distribution a priori normale-gamma.

Loi bayésienne	$p(x) = LN(x;\ m, v+\beta^2)$	$p(y = \ln x) = N(y;\ m, v+\beta^2)$
Moments :	$\upsilon_n = \exp\left(nm + \frac{n^2}{2}(v+\beta^2)\right)$	$\upsilon_1(\ln X) = m$
	$\mu_2 + \upsilon_1^2 = \upsilon_2$	$\mu_{2n}(\ln X) = (2n-1)!!(v+\beta^2)$
Paramètres :	$m = \ln MT - \ln\sqrt{1+D2}$	$m = \mathbb{E}(\ln X) = \frac{1}{KT}\sum_i \sum_j \log x_j^i$
	$v = \ln(D2+1) - \beta^2$	$v = \mathbb{V}(\ln X) - \beta^2$
		$v = = \frac{\sum\sum\left(\ln x_j^i - \mathbb{E}(\ln X)\right)^2}{\sum k_i} - \beta^2$

où $D2 = VMT/MT^2$ est le carré du coefficient de variation empirique de tous les données regroupées. La fonction de moyenne et la fonction de variance sont ici désignées par les notations :

$$\mathbb{E}(X) = \frac{1}{k}\sum_i x_i \qquad \text{et} \qquad \mathbb{V}(X) = \frac{1}{k}\sum_i \left(x_i - \mathbb{E}(X)\right)^2 \quad .$$

4.3.2. Distribution normale inverse $NI(\mu,\beta^2)$

Cas 1. et 2. μ et β inconnus et μ inconnu (β connu) : non valables !

En effet, aucun moments d'ordre ≥ 1 n'existe (cf. paragraphe 3.3) et la méthode du maximum de vraisemblance est très compliquées.

Cas 3. $\lambda = 1/(2\beta^2)$ inconnu (μ connu)

Loi a priori : $\pi(\lambda) = G(\lambda;\ a,b)$

Loi bayésienne : $p(x) \propto \left(x^{3/2}[x(x^{-1}-\mu^{-1})^2+b]\right)^{-(a/2+1)}$

Moments : $\upsilon_1 = \mu \qquad \mu_2 = \mu^3\dfrac{b}{2(a-1)} \qquad \mu_3 = \dfrac{3}{4}\mu^5\dfrac{b^2}{(a-1)(a-2)}$

Paramètres : $a = 2 + \dfrac{3\sqrt{VMT}}{B1-3\sqrt{VMT}} \qquad b = 2VMT(a-1)/\mu^3 \qquad (4.32)$

avec $C3 = \pm\sqrt{B1} = V3/VMT^{3/2}$ coefficient d'asymétrie empirique

Remarque : Pour que μ_2 et μ_3 existent, il faut que la condition "a > 2" soit vérifiée, c'est à dire $B1 > 3\sqrt{VMT}$, autrement dit, $V3 > 3\cdot VMT^2$.

4.3.3. Distributions gamma $G(\alpha,\beta)$ et gamma inverse $GI(\alpha,\beta)$

Cas 1. $G(\alpha,\beta)$ avec α, β inconnus : non valables !

En effet, aucun moments d'ordre > 1 n'existe (cf. paragraphe 3.4.1) et la méthode du maximum de vraisemblance est très compliquées.

Cas 2. $G(\alpha,\beta)$ avec β inconnu (α connu)

Loi a priori : $\pi(\beta) = G(\beta;\ a,b)$

Loi bayésienne : $p\left(y = \dfrac{a}{\alpha b}x\right) = F(y;\ 2\alpha, 2a)$

Moments : $\upsilon_n = b^n \frac{\alpha(\alpha+1)\ldots(\alpha+n-1)}{(a-1)(a-2)\ldots(a-n)}$

Paramètres : $a = \frac{2\cdot D2+1-1/\alpha}{D2 - 1/\alpha}$ $\qquad b = \frac{D2+1}{\alpha\cdot D2-1}MT$ (4.33)

Valeur α : (a) Moments : $\alpha = D2^{-1}$

Dans ce cas, la distribution a priori est Dirac en 1/MT.

(b) Maximum de vraisemblance :

$$\log\alpha - \psi(\alpha) - \log(MT/MGT) = 0$$

d'où $\alpha \approx \left(2\log(MT/MGT)\right)^{-1}$

avec $MGT = \left(\prod x_j^i\right)^{1/KT}$ la moyenne géométrique totale. Mais, il faut vérifier la condition : $\alpha \geq D2^{-1}$.

Quant à la distribution gamma inverse, comme $Y = X^{-1}$ suit une distribution gamma, il suffit de procéder de la même manière que précédemment sur la variable Y au lieu de X.

4.3.4. Distributions de Weibull $W(\alpha,\beta)$ et de Fréchet $Fr(\alpha,\beta)$

Loi a priori : $\pi(\lambda = \beta^{\mp\alpha}) = G\left(\lambda;\ a,b\right)$

Loi bayésienne : $p\left(y = \frac{a}{b}x^{\pm\alpha}\right) = F\left(y;\ 2,2a\right)$

Moments : $\upsilon_n = b^{\pm n/\alpha}\Gamma\left(1 \pm \frac{n}{\alpha}\right)\frac{\Gamma\left(a \mp \frac{n}{\alpha}\right)}{\Gamma(a)}$

Paramètres : $b = \left[\frac{MT\cdot\Gamma(a)}{\Gamma\left(1 \pm \frac{1}{\alpha}\right)\Gamma\left(a \mp \frac{1}{\alpha}\right)}\right]^{\alpha}$

a est solution de l'équation suivante :

$$\frac{\Gamma\left(a \mp \frac{2}{\alpha}\right)\Gamma(a)}{\Gamma^2\left(a \mp \frac{1}{\alpha}\right)} = (D2+1)\frac{\Gamma^2\left(1 \pm \frac{1}{\alpha}\right)}{\Gamma\left(1 \pm \frac{2}{\alpha}\right)} \qquad (4.34)$$

Valeur α : (a) Moments : α est solution de l'équation suivante :

$$\Gamma\left(1 \pm \frac{2}{\alpha}\right) = (D2 + 1)\Gamma^2\left(1 \pm \frac{1}{\alpha}\right)$$

L'équation (4.34) devient : $\Gamma\left(a \mp \frac{2}{\alpha}\right)\Gamma(a) = \Gamma^2\left(a \mp \frac{1}{\alpha}\right)$.

(b) Maximum de vraisemblance : α est solution de l'équation suivante :

$$\frac{1}{\alpha}\left(1+\left(m_{\pm\alpha} - 1\right)\log\left(m_{\pm\alpha}\right)\right) = \frac{1}{m_{\pm\alpha}}\left(\overline{x^{\pm\alpha}\log\left(x^{\pm 1}\right)} - \overline{x^{\pm\alpha}}\cdot\overline{\log\left(x^{\pm 1}\right)}\right)$$

(c) Moyenne et Médiane : $$\frac{MT}{Me} = \frac{\Gamma\left(1 \pm \frac{1}{\alpha}\right)}{(\ln 2)^{\pm\alpha}}$$

(d) Moyenne et Mode : $$\frac{MT}{Mo} = \frac{\Gamma\left(1 \pm \frac{1}{\alpha}\right)}{\left(1 \mp \frac{1}{\alpha}\right)^{1/\alpha}}$$

4.3.5. Distribution de Pareto $P(\alpha,\beta)$

Nous avons montré au paragraphe 3.8 qu'il n'existe aucun moment de la distribution bayésienne de la variable X puisque $p\left(y=\frac{a}{b}\ln\frac{x}{\alpha}\right) = F\left(y;\ 2,2a\right)$. Nous considérons directement la variable $Y = \ln\frac{x}{\alpha}$ qui suit une distribution gamma $G\left(1,\beta\right)$. Suivant la procédure du paragraphe 4.3.3, nous avons la distribution a priori : $\pi(\beta) = G\left(a,b\right)(\beta)$ avec les paramètres définis par

$$\begin{cases} a = \dfrac{2\cdot DA2}{DA2 - 1} \\[2ex] b = \dfrac{DA2 + 1}{DA2 - 1}\times MAT \end{cases} \tag{4.35}$$

où MAT, VAT et $DA2 = VAT/MAT^2$ sont respectivement la moyenne, la variance et le carré du coefficient de variation empirique de $\ln(X/\alpha)$. Et l'échantillon "total" est $\{y_j = \log(x_j/\alpha);\ j = 1,2,\ldots,k_i \text{ et } i = 1,2,\ldots,n\}$.

La valeur α peut être estimée par la méthode des moments :

$$\alpha = MT\left(1 - \frac{1}{1+\sqrt{1+D2^{-1}}}\right)$$

4.4. ESTIMATIONS DU QUASI-ECHANTILLON

Rappelons que la démarche est réalisée en deux étapes : en premier lieu il s'agit de construire un quasi-échantillon du paramètre θ à partir de la distribution conditionnelle de la variable aléatoire X, $f(x|\theta)$, avec les n échantillons disponibles ; dans le second cas, supposons que la forme de la distribution a priori de θ soit connue, $\pi(\theta)$, alors la fonction de vraisemblance de ce quasi-échantillon donnera un estimateur du paramètre de $\pi(\theta)$ par une des méthodes classiques.

Dans le présent, nous ne citons dans les deux étapes que la méthode des moments et la méthode du maximum de vraisemblance. Nous avons ainsi les quatre combinaisons :

— méthode des doubles moments, notée par MM;

— méthode des maxima-moments, notée par VM;

— méthode des moments-maxima, notée par MV;

— méthode des doubles maxima de vraisemblance, notée par VV.

4.4.1. Distributions Normale $N(\mu,\sigma^2)$ et log-normale $LN(\alpha,\beta^2)$

Pour le paramètre (μ,σ^2) de la distribution normale, l'estimateur des moments coïncide avec celui du maximum de vraisemblance. Dans la première étape, les deux estimateurs se réduient en un seul, et donc c'est la deuxième étape qui donne deux estimateurs différents, citons MM et VV. Quant à la distribution log-normale, il suffit d'appliquer les mêmes procédures sur la variable lnX au lieu de la variable X.

Considérons d'abord les n échantillons $\mathbf{x}_i$, i =1,2,...,n, chacun suivant une distribution normale $N(\mu_i,\sigma_i^2)$. Pour chaque échantillon $\mathbf{x}_i$, l'estimateur du paramètre $[\mu_i,\eta_i=1/(2\sigma_i^2)]$ est

$$\begin{cases} \mu_i = m_i \\ \eta_i = h_i = (2v_i)^{-1} \end{cases} \tag{4.41}$$

où h_i est appelée la précision empirique.

Posons préalablement

$$\begin{cases} MM = \frac{1}{n}\sum_i m_i \\ MV = \frac{1}{n}\sum_i (m_i - MM)^2 \\ ML = \frac{1}{n}\sum_i \log m_i \end{cases} \qquad \begin{cases} HM = \frac{1}{n}\sum_i h_i \\ HV = \frac{1}{n}\sum_i (h_i - HM)^2 \\ HL = \frac{1}{n}\sum_i \log h_i \end{cases}$$

<u>Cas 1. μ, $\eta = 1/(2\sigma^2)$ inconnus</u>

Le paramètre inconnu est $\theta = (\mu, \eta)$. Ainsi, $\{\theta_i = (\mu_i, \eta_i);\ i = 1,2,\ldots,n\}$ constitue un quasi-échantillon de ce paramètre.

Nous avons montré que la variable (μ, η) suit une distribution normale-gamma paramétrée par le vecteur (m, v, k), nous avons $\pi(\mu, \eta) = NG(\mu, \eta;\ m, v, k)$. Les moments des deux premiers ordres de la distribution normale-gamma sont :

$$\begin{cases} \mathbb{E}(\mu) = m & (11) \\ \mathbb{V}(\mu) = \frac{v}{k-1} & (12) \end{cases} \qquad \begin{cases} \mathbb{E}(\eta) = \frac{k+1}{2kv} & (13) \\ \mathbb{V}(\eta) = \frac{k+1}{2}(kv)^{-2} & (14) \end{cases}$$

Pour résoudre trois paramètres, il suffit de prendre trois parmi les quatre relations. Nous en donnons deux comme estimateur des doubles moments MM, la première solution prend en compte la variation entre les moyennes tant que la deuxième ne considère que la moyenne et la variance des précisions :

$$\begin{array}{ccc} (11) + (12) + (13) & \qquad & (11) + (13) + (14) \\ \Downarrow & & \Downarrow \\ \begin{cases} m = MM \\ v = (k-1)MV \\ k = \frac{1+c}{2c}\left[1+\sqrt{1+\frac{4c}{(1+c)^2}}\right] \end{cases} & & \begin{cases} m = MM \\ v = VH\frac{k+1}{k} \\ k = 2HM^2/HV - 1 \end{cases} \end{array} \tag{4.411a}$$

avec c = MV/VH et VH = 1/(2HM).

Quant à l'estimateur du maximum de vraisemblance, en admettant la notation $\pi(\mu,\eta) = \pi(\mu,\eta|m,v,k)$, la fonction de vraisemblance de l'échantillon $\{\theta_i = (\mu_i,\eta_i)\}$ s'écrit donc :

$$L(m,v,k;\ \theta_1,\theta_2,\ldots,\theta_n) = \prod_i \pi(\theta_i|m,v,k)$$

$$= \left(\frac{k}{\pi}\right)^{n/2} \Gamma\left(\frac{k+1}{2}\right)^{-n} \left(\prod_i \eta_i\right)^{k/2} (kv)^{n\frac{k+1}{2}} \exp\left[-kv\sum_i \eta_i\right] \exp\left[-k\sum_i \eta_i(\mu_i-m)^2\right] .$$

D'autre part, le maximum de vraisemblance (m,v,k) du quasi-échnatillon est obtenu en déterminant pour quelles valeurs les dérivées partielles s'annulent : $\frac{\partial L}{\partial m} = 0$, $\frac{\partial L}{\partial v} = 0$, $\frac{\partial L}{\partial k} = 0$. Avec la fonction $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$, nous en déduisons le résultat :

$$m = \frac{\sum \mu_i \eta_i}{\sum \eta_i} \qquad\qquad v = \frac{k+1}{2k}\left(\frac{1}{n}\sum_i \eta_i\right)^{-1}$$

$$k^{-1} + \log\left(\frac{k+1}{2}\right) - \psi\left(\frac{k+1}{2}\right) = \log\frac{1}{n}\sum_i \eta_i - \frac{1}{n}\sum_i \log\eta_i + 2\frac{1}{n}\sum_i \eta_i(\mu_i-m)^2$$

En utilisant une approximation de la fonction ψ pour k suffisamment grand :

$$\psi\left(\frac{k+1}{2}\right) \approx \log\frac{k+1}{2} - \frac{1}{k+1}$$

et posant

$$MHM = \frac{1}{n}\sum_i h_i m_i \Big/ \frac{1}{n}\sum_i h_i$$

$$MH4 = \frac{1}{n}\sum_i h_i m_i^2$$

nous avons l'estimateur des doubles maxima de vraisemblance VV :

$$\begin{cases} m = MHM \\ v = \frac{1}{2HM} \times \frac{k+1}{k} \\ k \approx \left(\log HM - HL + 2(MH4 - MHM^2 \cdot HM)\right)^{-1} \end{cases} \qquad (4.411b)$$

où $\mathbb{E}(\eta) = \frac{k+1}{2kv}$ = HM est la moyenne des précisions empiriques $\{h_i = 1/(2v_i)\}$. Donc, nous avons :

$$\mathbb{E}(\sigma^2) = \frac{1}{2\mathbb{E}(\eta)} = VH$$

c'est à dire que la moyenne a priori de la variance est égale à la moyenne harmonique des variances.

La concavité stricte de la fonction $\log x - x^2$ assure que k est toujours positif et devient assez grand lorsque les précisions h_i sont toutes voisines les unes des autres et les moyennes m_i sont aussi ; k est infini si les h_i sont égaux. Par contre, il n'est pas évident d'avoir k grand lorsque la moyenne est grande, même si les écarts entre les h_i sont petits.

Cas 2. μ inconnu ($\eta = 1/(2\sigma^2)$ connu)

Comme nous ne considérons que la variabilité du paramètre μ, c'est donc l'ensemble $\{\theta_i = \mu_i;\ i = 1,2,\ldots,n\}$ qui constitue un quasi-échantillon de la variable $\theta = \mu$.

La distribution de la variable μ est normale $N(m,v)$. Si la variance est considérée comme constante, nous avons les estimateurs MM et MV égaux à :

$$\begin{cases} m = \mathbb{E}(\mu) = MM \\ v = \mathbb{V}(\mu) = MV \end{cases} \tag{4.412a}$$

Notons que ce cas, noté par N10-MM, ne prend pas en compte les variances des n échantillons, tant qu'elles sont faibles ou constantes, ni leur écart, mais il étudie la dispersion entre les moyennes.

Par ailleurs, si l'on suppose que la variable μ admet aussi une distribution normale $N\left(m, v = \frac{1}{2k\eta}\right)$ paramétrée par le vecteur (m,k) avec $\eta = 1/(2\sigma^2)$ estimée de façon empirique pour chacun des échantillons, elle est conjuguée de la fonction de vraisemblance (cf. paragraphe 3.1.2) de manière à ce que :

$$\pi(\mu) = \sqrt{\frac{k\eta}{\pi}} \cdot \exp\left(-\ \eta k(\mu - m)^2\right)$$

Dans ce cas, le quasi-échantillon est l'ensemble $\{(\mu_i, \eta_i);\ i = 1,2,\ldots,n\}$ et admettant $\pi(\mu) = \pi(\theta=(\mu,\eta)|m,k)$. L'estimateur des moments MM est

$$\begin{cases} m = \mathbb{E}(\mu) = MM \\ v = \dfrac{1}{2k\cdot HM} = \dfrac{VH}{KM} \end{cases} \tag{4.412b}$$

lorsque nous prenons k = KM et η = HM les moyennes des tailles $\{k_i\}$ et des précisions $\{h_i\}$.

La fonction de vraisemblance de cet échantillon devient :

$$L(m,k;\ \theta_1,\theta_2,\ldots,\theta_n) = \left(\frac{k}{\pi}\right)^{n/2}\left(\prod_i \eta_i\right)^{1/2}\exp\left(-\ k\sum_i \eta_i(\mu_i-m)^2\right)$$

Son maximum de vraisemblance (m,k) est obtenu losrque les dérivées partielles $\frac{\partial L}{\partial m}$, $\frac{\partial L}{\partial k}$ s'annulent. En utilisant $\eta_i = h_i$ et les notations précédentes, nous avons l'estimateur des doubles maxima de vraisemblance VV :

$$\begin{cases} m = MHM \\ \eta = HM \\ k = \dfrac{1}{2}\left(MH4 - MHM^2\cdot HM\right)^{-1} \end{cases} \tag{4.412c}$$

d'où $v = \frac{1}{2k\eta} = MH4/HM - MHM^2$.

Il est clair que si l'on suppose $\sigma_i^2 = \sigma^2$, donc $h_i = h$, cette formule donc devient la précédente que l'on appliquera au cas où la variation des variances est négligeable.

Cas 3. η inconnu (μ connu)

Pour le paramètre de la distribution normale $N(\mu_i, \sigma_i^2)$, les estimateurs des moments et du maximum de vraisemblance de la variable $\eta = 1/(2\sigma^2)$ sont différents et donnés respectivement par :

$$\begin{aligned} &\text{soit} \quad \eta_i = h_i = 1/(2v_i) \\ &\text{soit} \quad \eta_i' = h_i' = \frac{1}{2}\left(v_i + (\mu-m_i)^2\right)^{-1} \end{aligned} \tag{4.413}$$

Si $\mu_i = m_i$, les deux estimateurs coïncident. $\{\theta_i = \eta_i$ ou η'_i; $i = 1, 2, \ldots, n\}$ constitue alors un quasi-échantillon du paramètre η.

La distribution de la variable η est gamma $G(a,b)$, d'où l'estimateur de moments :

$$\begin{cases} a = \mathbb{E}(\eta)^2/\mathbb{V}(\eta) = HM^2/HV \\ b = \mathbb{E}(\eta)/\mathbb{V}(\eta) = a/HM \end{cases} \tag{4.413a}$$

La fonction de vraisemblance de l'échantillon de η s'écrit :

$$L(a,b;\ \eta_1,\eta_2,\ldots,\eta_n) = \left(\frac{b^a}{\Gamma(a)}\right)^n \prod_i \eta_i^{a-1} \exp\left(-\ b\sum_i \eta_i\right)$$

d'où l'estimateur du maximum de vraisemblance avec $\mathbb{E}(\log\eta) = HL$:

$$\begin{cases} a/b = \mathbb{E}(\eta) = HM \\ \log b - \psi(a) + \mathbb{E}(\log\eta) = 0 \end{cases}$$

$$\Longrightarrow \quad \begin{cases} a \approx \frac{1}{2}\left(\log HM - HL\right)^{-1} \\ b = a/HM \end{cases} \tag{4.431b}$$

Si l'on suppose les μ_i constantes : $\mu_i = \mu = MM$ ou MT par exemple, les deux autres estimateurs sont calculés avec $\{\eta'_i;\ i = 1,2,\ldots,n\}$. Dans ce cas, nous étudions la dispersion à l'intérieur des lots et la dispersion des lots avec une position moyenne μ (globale) ou valeur cible.

4.4.2. Distribution normale inverse $NI(\mu,\beta^2)$

Considérons d'abord les n échantillons $\mathbf{x}_i$, $i = 1,2,\ldots,n$, chacun suivant la distribution normale inverse $NI(\mu_i,\beta_i^2)$. L'estimateur des moments et l'estimateur du maximum de vraisemblance du paramètre $(\alpha_i = \mu_i^{-1}, \lambda_i = 1/(2\beta_i^2))$ sont différents et donnés par

$$\begin{cases} \alpha_i = \mu_i^{-1} = m_i^{-1} \\ \lambda_i = \mu_i^3/(2\sigma_i^2) = m_i^3 h_i \end{cases} \tag{4.42}$$

$$\begin{cases} \alpha_i = \mu_i^{-1} = m_i^{-1} \\ \lambda_i = l_i = \frac{1}{2}\left(m_{Hi}^{-1} - m_i^{-1}\right)^{-1} \end{cases}$$

où m_{Hi} est la moyenne harmonique empirique de l'échantillon $\mathbf{x}_i$.

Posons préalablement :

$$MIH = \frac{1}{n}\sum_i h_i/m_i$$

$$\begin{cases} MIM = \frac{1}{n}\sum_i m_i^{-1} \\ MIV = \frac{1}{n}\sum_i \left(m_i^{-1} - MIM\right)^2 \end{cases} \qquad \begin{cases} MLM = \frac{1}{n}\sum_i l_i/m_i \\ MLV = \frac{1}{n}\sum_i l_i/m_i^2 \end{cases}$$

$$\begin{cases} LM = \frac{1}{n}\sum_i l_i \\ LV = \frac{1}{n}\sum_i \left(l_i - LM\right)^2 \\ LL = \frac{1}{n}\sum_i \log l_i \end{cases} \qquad \begin{cases} MH7 = \frac{1}{n}\sum_i h_i m_i^3 \\ MH8 = \frac{1}{n}\sum_i \left(h_i m_i^3 - MH7\right)^2 \\ MH9 = \frac{1}{n}\sum_i \log(h_i m_i^3) = 3ML + HL \end{cases}$$

<u>Cas 1. $\alpha = 1/\mu$, $\lambda = 1/(2\beta^2)$ inconnus</u>

Le paramètre inconnu est $\theta = (\alpha, \lambda)$. Chaque ensemble $\{\theta_i = (\alpha_i, \lambda_i)$; $i = 1, 2, \ldots, n\}$ constitue un quasi-échantillon du paramètre θ. Nous avons montré que la variable (α, λ) suit une distribution normale- gamma paramétrée par le vecteur (m, m_H, k) et aussi par $(a = m^{-1}, b, k)$:

$$\pi(\alpha, \lambda) = NG(\alpha, \lambda;\ a, b, k)$$

$$= N\left(\alpha;\ a, \frac{a}{2k\lambda}\right) G\left(\lambda;\ \frac{k+1}{2}, kb\right)$$

d'où les moments de deux premiers ordres de la distribution normale-gamma :

$$\begin{cases} \mathbb{E}(\alpha) = a & (21) \\ \mathbb{V}(\alpha) = \frac{ab}{k-1} & (22) \end{cases} \qquad \begin{cases} \mathbb{E}(\lambda) = \frac{k+1}{2kb} & (23) \\ \mathbb{V}(\lambda) = \frac{k+1}{2}(kb)^{-2} & (24) \end{cases}$$

Parmi les différentes solutions, nous en donnons deux

$$(21) + (22) + (23) \qquad\qquad (21) + (23) + (24)$$

$$\Downarrow \qquad\qquad\qquad\qquad \Downarrow$$

$$\begin{cases} a = MIM = MH^{-1} \\ b = (k-1)\dfrac{MIV}{MIM} \\ k = \dfrac{1+c}{2c}\left[1+\sqrt{1+\dfrac{4c}{(1+c)^2}}\right] \end{cases} \qquad \begin{cases} a = MIM = MH^{-1} \\ b = \dfrac{1}{2\mathbb{E}(\lambda)}\times\dfrac{k+1}{k} \\ k = 2\mathbb{E}(\lambda)^2/\mathbb{V}(\lambda) - 1 \end{cases} \qquad (4.421a)$$

avec $c = 2\dfrac{\mathbb{E}(\lambda)\cdot MIV}{MIM}$ et $[\mathbb{E}(\lambda),\mathbb{V}(\lambda)] = \begin{cases} [MH7, MH8] \\ [LM, LV] \end{cases}$.

On obtient les deux estimateurs des doubles moments MM et des maxima-moments MM et VM. Comme nous avons $\alpha = \mu^{-1}$ et $a = m^{-1}$, la moyenne de μ est égale à la moyenne harmonique des moyennes :

$$\mathbb{E}(\mu) = MH = \left(\frac{1}{n}\sum m_i^{-1}\right)^{-1} .$$

Quant à l'estimateur du maximum de vraisemblance, en admettant la notation $\pi(\alpha,\lambda) = \pi(\alpha,\lambda|a,b,k)$, la fonction de vraisemblance de l'échantillon $\{\theta_i = (\alpha_i,\lambda_i)\}$ s'écrit donc comme :

$$L(a,b,k;\ \theta_1,\theta_2,\ldots,\theta_n) = \prod_i \pi(\theta_i|a,b,k)$$

$$= \left(\frac{k}{\pi a}\right)^{n/2}\Gamma\left(\frac{k+1}{2}\right)^{-n}\left(\prod_i\lambda_i\right)^{k/2}(kb)^{n\frac{k+1}{2}}\exp\left[-kb\sum_i\lambda_i\right]\exp\left[-\frac{k}{a}\sum_i\lambda_i(\alpha_i-a)^2\right]$$

D'autre part, le maximum de vraisemblance (a,b,k) est obtenu en déterminant pour quelles valeurs les dérivées partielles s'annulent : $\dfrac{\partial L}{\partial a} = 0$, $\dfrac{\partial L}{\partial b} = 0$, $\dfrac{\partial L}{\partial k} = 0$, d'où un système d'équations :

$$b = \frac{k+1}{2k}\left(\frac{1}{n}\sum_i\lambda_i\right)^{-1}$$

$$k^{-1} = \frac{2}{a}\times\frac{1}{n}\sum_i\lambda_i(\alpha_i-a)^2 + 4\frac{1}{n}\sum_i\lambda_i(\alpha_i-a)$$

$$k^{-1} + \log\left(\frac{k+1}{2}\right) - \psi\left(\frac{k+1}{2}\right) = \log\frac{1}{n}\sum_i\lambda_i - \frac{1}{n}\sum_i\log\lambda_i + \frac{2}{a}\times\frac{1}{n}\sum_i\lambda_i(\alpha_i-a)^2$$

Nous en déduisons l'estimateur du maximum de vraisemblance :

$$\begin{cases} a = \dfrac{1}{2k\mathbb{E}(\lambda)}\left[\sqrt{1+16k^2\mathbb{E}(\lambda)\mathbb{E}(\lambda\alpha)} - 1\right] \\ b = \dfrac{k+1}{2k\mathbb{E}(\lambda)} \\ \dfrac{1}{k+1} - \dfrac{2}{k}\left[\sqrt{1+16k^2\mathbb{E}(\lambda)\mathbb{E}(\lambda\alpha)} - 1\right] = \log\mathbb{E}(\lambda) - \mathbb{E}(\log\lambda) - 4\mathbb{E}(\lambda\alpha) \end{cases} \tag{4.421b}$$

Nous obtenons en effet les deux estimateurs des moments-maxima MV et des doubles maxima de vraisemblance VV, avec les définitions :

$$[\mathbb{E}(\lambda), \mathbb{E}(\lambda\alpha), \mathbb{E}(\log\lambda)] = \begin{cases} [MH7, MH4, MH9] \\ [LM, MLM, LL] \end{cases}$$

<u>Cas 2. $\alpha = \mu^{-1}$ inconnu ($\lambda = 1/(2\beta^2)$ connu)</u>

Nous ne considérons que la variabilité du paramètre $\alpha = \mu^{-1}$. L'ensemble $\{\alpha_i = m_i^{-1};\ i = 1,2,\ldots,n\}$ constitue un quasi-échantillon de α.

La distribution de la variable α est normale $N(m,v)$, d'où l'estimateur des doubles moments MM :

$$\begin{cases} m = \mathbb{E}(\alpha) = MIM \\ v = \mathbb{V}(\alpha) = MIV \end{cases} \tag{4.422a}$$

Par ailleurs, si l'on suppose que la variable α admet aussi une distribution normale $N\left(m=a, v=\frac{a}{2k\lambda}\right)$ paramétrée par le vecteur (a,k) avec $\lambda = 1/(2\beta^2)$ connu, elle est ici conjuguée de la fonction de vraisemblance (cf. paragraphe 3.3.2) de manière à ce que :

$$\pi(\alpha) = \sqrt{\frac{k\lambda}{\pi a}}\cdot\exp\left(-\frac{\lambda k}{a}(\alpha - a)^2\right) .$$

Nous considérons plutôt le quasi-échantillon comme l'ensemble $\{\theta_i = (\alpha_i, \lambda_i)\ ;\ i = 1,2,\ldots,n\}$, et admettant $\pi(\alpha) = \pi(\theta=(\alpha,\lambda)|a,k)$. La fonction de vraisemblance de cet échantillon devient :

$$L(a,k;\ \theta_1,\theta_2,\ldots,\theta_n) = \prod_i \pi(\theta_i|a,k)$$

$$= \left(\frac{k}{\pi a}\right)^{n/2}\left(\prod_i \lambda_i\right)^{1/2}\exp\left(-\frac{k}{a}\sum_i \lambda_i(\alpha_i - a)^2\right)$$

Son maximum de vraisemblance (a,k) est obtenu lorsque les dérivées partielles $\frac{\partial L}{\partial a}$, $\frac{\partial L}{\partial k}$ s'annulent, d'où nous avons :

$$\begin{cases} a = \dfrac{\mathbb{E}(\lambda\alpha)}{\mathbb{E}(\lambda)} \\ k^{-1} = \dfrac{2}{a} \times \dfrac{1}{n} \sum_i \lambda_i (\alpha_i - a)^2 \end{cases} \tag{4.422b}$$

$$= \frac{2}{a}\left(\mathbb{E}(\lambda\alpha^2) - \mathbb{E}(\lambda\alpha)^2/\mathbb{E}(\lambda)\right)$$

Pour les méthodes MV et VV, nous obtenons $\left(m = a,\ v = \frac{a}{2k\lambda}\right)$ avec λ = MH7 et λ = LM :

$$\begin{cases} m = MH4/MH7 \\ v = MH1/MH7 - m^2 \end{cases}$$

$$\begin{cases} m = MLM/LM \\ v = MLV/LM - m^2 \end{cases}$$

Si l'on suppose que les variances des échantillons sont égales $\beta_i^2 = \beta^2$, donc $\lambda_i = \lambda$, alors la formule (4.422b) donnera le même résultat que (4.422a), ainsi nous avons MM = VM = MV = VV.

Cas 3. $\lambda = 1/(2\beta^2)$ inconnu (μ connu)

Pour le paramètre de la distribution normale $N(\mu_i, \beta_i^2)$, les estimateurs des moments et du maximum de vraisemblance de la variable $\lambda = 1/(2\beta^2)$ sont différents :

$$\text{soit} \qquad \lambda_i = h_i m_i^3$$

$$\text{soit} \qquad \lambda'_i = l'_i = \frac{1}{2}\left(m_i(\mu_i^{-1} - m_i^{-1})^2 + m_{Hi}^{-1} - m_i^{-1}\right)^{-1} \tag{4.423}$$

Si $\mu_i = m_i$, les deux estimateurs coïncident. $\{\theta_i = \lambda_i$ ou $\lambda'_i;\ i = 1, 2, \ldots, n\}$ constitue alors un quasi-échantillon du paramètre λ.

Comme la distribution de la variable λ est gamma, suivant la procédure du cas de la distribution normale de précision inconnue η (cf. formules (4.413a et b)), nous avons les estimateurs :

$$\begin{cases} a = \mathbb{E}(\lambda)^2/\mathbb{V}(\lambda) \\ b = a/\mathbb{E}(\lambda) \end{cases} \tag{4.423a}$$

$$\begin{cases} a \approx \frac{1}{2}\left(\log\mathbb{E}(\lambda) - \mathbb{E}(\log\lambda)\right)^{-1} \\ b = a/LM \end{cases} \tag{4.423b}$$

Si l'on suppose les μ_i constantes : $\mu_i = \mu$ = MIM ou MIT = $\left(\frac{1}{KT}\sum x_i^{-1}\right)^{-1}$, il suffit d'utiliser les formules (4.413a et b) avec $\{\theta_i = \lambda'_i;\ i = 1,2,\ldots,n\}$ et nous trouvons les estimateurs MV et VV.

4.4.3. Distributions gamma $G(\alpha,\beta)$ et gamma inverse $GI(\alpha,\beta)$

Nous donnerons les estimateurs pour la distribution gamma. Pour la loi gamma inverse, il suffit de considérer la variable $Y = X^{-1}$ au lieu de X.

Considérons les n échantillons $\mathbf{x}_i$, i = 1, 2,..., n, chacun suivant une distribution gamma $G(\alpha_i,\beta_i)$ et les statistiques exhaustives (m_i, m_{Gi}, k_i). Les estimateurs des moments et du maximum de vraisemblance sont donnés par :

$$\begin{cases} \alpha_i = m_i^2/v_i^2 = 2h_i m_i^2 \\ \beta_i = m_i/v_i^2 = 2h_i m_i \end{cases}$$

$$\begin{cases} \alpha_i = \frac{1}{2}\left(\log m_i - \log m_{Gi}\right)^{-1} = \frac{1}{2r_i} \\ \beta_i = \frac{1}{2r_i m_i} \end{cases} \tag{4.43}$$

Posons d'abord :

$$\begin{cases} MH0 = \frac{1}{n}\sum_i h_i m_i^2 \log(h_i m_i) \\ MR0 = \frac{1}{n}\sum_i r_i^{-1}\log\left(m_i r_i\right)^{-1} \end{cases}$$

$$\begin{cases} MH1 = \frac{1}{n}\sum_i h_i m_i \\ MH2 = \frac{1}{n}\sum_i (h_i m_i - MH1)^2 \\ MH3 = \frac{1}{n}\sum_i \log(h_i m_i) \end{cases} \qquad \begin{cases} MH4 = \frac{1}{n}\sum_i h_i m_i^2 \\ MH5 = \frac{1}{n}\sum_i (h_i m_i^2 - MH4)^2 \\ MH6 = \frac{1}{n}\sum_i \log(h_i m_i^2) \end{cases}$$

$$\begin{cases} RI1 = \frac{1}{n}\sum_i r_i^{-1} \\ RI2 = \frac{1}{n}\sum_i \left(r_i^{-1} - RI1\right)^2 \\ RI3 = \frac{1}{n}\sum_i \log r_i^{-1} \end{cases} \qquad \begin{cases} MR1 = \frac{1}{n}\sum_i \left(r_i m_i\right)^{-1} \\ MR2 = \frac{1}{n}\sum_i \left[\left(r_i m_i\right)^{-1} - MR1\right]^2 \\ MR3 = \frac{1}{n}\sum_i \log\left(r_i m_i\right)^{-1} \end{cases}$$

<u>Cas 1. α, β inconnus</u>

Les deux ensemble $\{\theta_i = (\alpha_i, \beta_i)\ ;\ i = 1,2,\ldots,n\}$ obtenus ci-dessus constituent deux quasi- échantillons du paramètre $\theta = (\alpha,\beta)$.

Nous avons montré que cette variable suit une distribution gamma-gamma $\pi(\alpha,\beta) = GG(m,r,k)(\alpha,\beta)$. Les moments d'ordre 1 et d'ordre 2 sont

$$\begin{cases} \mathbb{E}(\alpha) = \frac{k+3}{2kr} & (31) \\ \mathbb{V}(\alpha) = \frac{k+3}{2}(kr)^{-2} & (32) \end{cases} \qquad \begin{cases} \mathbb{E}(\beta) = (1+\frac{k+3}{2r})(km)^{-1} & (33) \\ \mathbb{V}(\beta) = (1+\frac{k+3}{2r})(km)^{-2} & (34) \end{cases}$$

Pour les mêmes raisons que précédemment, nous choisissons deux estimateurs :

$$\begin{array}{ccc} (31) + (32) + (33) & & (31) + (33) + (34) \\ \Downarrow & & \Downarrow \\ \begin{cases} m = \frac{\mathbb{E}(\alpha)+1/k}{\mathbb{E}(\beta)} \\ r = \frac{\mathbb{E}(\alpha)}{k\mathbb{V}(\alpha)} \\ k = 2\mathbb{E}(\alpha)^2/\mathbb{V}(\alpha) - 3 \end{cases} & & \begin{cases} m = \frac{\mathbb{E}(\beta)}{k\mathbb{V}(\beta)} \\ r = \frac{1+3/k}{2\mathbb{E}(\alpha)} \\ k = \frac{1}{\mathbb{E}(\alpha)}\left[\frac{\mathbb{E}(\alpha)^2}{\mathbb{V}(\beta)} - 1\right] \end{cases} \end{array} \qquad (4.431a)$$

Notons $\pi(\alpha,\beta) = \pi(\alpha,\beta|m,r,k)$, la fonction de vraisemblance de $\{\theta_i\}$ est

$$L(m,r,k;\ \theta_1,\theta_2,\ldots,\theta_n) = \prod_i \pi(\theta_i|m,r,k)$$

$$= \frac{(km)^{k\Sigma\alpha_i+n}}{\prod\Gamma(k\alpha_i+1)}\left(\prod_i \beta_i^{\alpha_i}\right)^{k/2} \frac{(kr)^{n\frac{k+3}{2}}}{\Gamma\left(\frac{k+3}{2}\right)}\left(\prod_i \alpha_i\right)^{\frac{k+1}{2}} \exp\left[-\ k\left(m\sum_i \beta_i - r\sum_i \alpha_i\right)\right] \quad .$$

d'où l'estimateur du maximum de vraisemblance (m,r,k) :

$$\begin{cases} m = \dfrac{\mathbb{E}(\alpha)+1/k}{\mathbb{E}(\beta)} \\ r = \dfrac{1+3/k}{2\mathbb{E}(\alpha)} \\ \dfrac{1}{2}\left(\log(\frac{k+3}{2}) - \psi(\frac{k+3}{2})\right) + \mathbb{E}(\alpha)\log[k\mathbb{E}(\alpha)+1] - \mathbb{E}[\alpha\log(k\alpha+1)] \\ \qquad = \mathbb{E}(\alpha)\log\mathbb{E}(\beta) - \mathbb{E}(\alpha\log\beta) + \dfrac{1}{2}[\log\mathbb{E}(\alpha) - \mathbb{E}(\log\alpha)] \end{cases} \tag{4.431b}$$

En utilisant l'approximation de la fonction ψ pour k assez grand, nous avons

$$k \approx 2\left[\log\mathbb{E}(\alpha) - \mathbb{E}(\log\alpha) + 2\left(\mathbb{E}(\alpha)\log\mathbb{E}(\beta) - \mathbb{E}(\alpha\log\beta)\right)\right]^{-1} \quad .$$

La concavité stricte de la fonction $x\log x - \frac{1}{2}\log x$ assure que k est toujours positif, et devient assez grand lorsque les (α_i, β_i) sont tous voisins les uns des autres; k est infini s'ils sont égaux.

Les estimateurs donnés par les formules (4.431a et b) doivent être combinés avec des statistiques de $\{\alpha_i, \beta_i\}$ données dans le tableau ci-dessous (cf. formule (4.43)).

Méthode	$\mathbb{E}(\alpha)$	$\mathbb{V}(\alpha)$	$\mathbb{E}(\beta)$	$\mathbb{V}(\beta)$	$\mathbb{E}(\log\alpha)$	$\mathbb{E}(\log\beta)$	$\mathbb{E}(\alpha\log\beta)$
Moments	2MH4	4MH5	2MH1	4MH2	MH6+log2	MH3+log2	MH0
Maximum	$\frac{1}{2}$RI1	$\frac{1}{4}$RI2	$\frac{1}{2}$MR1	$\frac{1}{4}$MR2	RI3-log2	MR3-log2	MR0

Tableau 4.1 Statistique des paramètres de la distribution gamma

Cas 2. β inconnu (α connu)

Si l'on suppose que le premier paramètre α de la distribution gamma $G(\alpha,\beta)$ est connu, la distribution de la variable β est aussi gamma $G(a,b)$.

De la même manière qu'au paragraphe 4.413, les estimateurs sont :

$$\begin{cases} a = \mathbb{E}(\beta)^2/\mathbb{V}(\beta) \\ b = \mathbb{E}(\beta)/\mathbb{V}(\beta) \end{cases} \tag{4.432}$$

$$\begin{cases} a \approx \frac{1}{2}\left(\log\mathbb{E}(\beta)-\mathbb{E}(\log\beta)\right)^{-1} \\ b = a/\mathbb{E}(\beta) \end{cases}$$

avec les statistiques de $\{\beta_i\}$ définies dans le tableau ci-dessus.

4.4.4. Distributions de Weibull $W(\alpha,\beta)$ et de Fréchet $Fr(\alpha,\beta)$

Considérons n échantillons x_i, $i = 1,2,\ldots,n$, chacun suivant une loi de Weibull $W(\alpha_i,\beta_i)$ ou de Fréchet $Fr(\alpha_i,\beta_i)$ avec α_i connu.

Il existe deux quasi-échantillons de la variable $\lambda = \beta^{\mp\alpha}$, le premier constitué de n estimateurs des moments et le second de n maxima de vraisemblance :

$$\text{soit} \qquad \lambda_i = \left(m_i^{-1}\Gamma\left(1 \pm \frac{1}{\alpha_i}\right)\right)^{\mp\alpha_i} \tag{4.44}$$

$$\text{soit} \qquad \lambda_i = m_{\pm\alpha i}^{-1}$$

tous avec α_i solution de l'équation :

$$\Gamma\left(1 \pm \frac{2}{\alpha}\right) - (d^2+1)\Gamma\left(1 \pm \frac{1}{\alpha}\right) = 0$$

où $d^2 = v/m^2 = (2hm^2)^{-1}$ est le carré du coefficient de variation.

Posons :

$$\begin{cases} \mathbb{E}(\lambda) = MGM = \frac{1}{n}\sum_i\left(\Gamma\left(1+\frac{1}{\alpha_i}\right)/m_i\right)^{\mp\alpha_i} \\ \mathbb{V}(\lambda) = MGV = \frac{1}{n}\sum_i\left[\left(\Gamma\left(1+\frac{1}{\alpha_i}\right)/m_i\right)^{\mp\alpha_i} - MGM\right]^2 \\ \mathbb{E}(\log\lambda) = MGL = \frac{1}{n_i}\sum\log\left[\Gamma\left(1+\frac{1}{\alpha_i}\right)/m_i\right]^{\mp\alpha_i} \end{cases} \tag{4.44a}$$

$$\begin{cases} \mathbb{E}(\lambda) = MAM = \frac{1}{n}\sum_i m_{\pm\alpha i}^{-1} \\ \mathbb{V}(\lambda) = MAV = \frac{1}{n}\sum_i\left(m_{\pm\alpha i}^{-1} - MAM\right)^2 \\ \mathbb{E}(\log\lambda) = MAL = \frac{1}{n}\sum_i\log\left(m_{\pm\alpha i}\right)^{-1} \end{cases} \tag{4.44b}$$

Comme la variable λ suit une loi de distribution gamma, les estimateurs du paramètre (a,b) sont les mêmes que dans les formules précédentes (4.423a, b), mais avec les définitions ci-dessus des statistiques de $\{\lambda_i\}$.

4.4.5. Distribution de Pareto $P(\alpha,\beta)$

Considérons n échantillons x_i, $i = 1,2,\ldots,n$, chacun suivant une loi de Pareto $P(\alpha_i,\beta_i)$ avec α_i connu.

Il existe deux quasi-échantillons de la variable β, le premier constitué de n estimateurs des moments et le second de n maxima de vraisemblance :

$$\begin{aligned} &\text{soit} \qquad \beta_i = 1 + \sqrt{1+2h_i m_i^2} = t'_i \\ &\text{soit} \qquad \beta_i = \left(\log(m_{Gi}/\alpha_i)\right)^{-1} = t_i^{-1} \end{aligned} \tag{4.45}$$

tous avec l'estimateur des moments :

$$\alpha_i = m_i\left(1 - \frac{1}{1+\sqrt{1+2h_i m_i^2}}\right) .$$

Comme la variable β suit aussi une distribution gamma, pour établir les estimateurs, il suffit d'utiliser les formules (4.432a et b) avec les statistiques de $\{\beta_i\}$ suivantes :

$$\begin{cases} \mathbb{E}(\beta) = TM = \frac{1}{n}\sum_i\left(1+\sqrt{1+2h_i m_i^2}\right) \\ \mathbb{V}(\beta) = TV = \frac{1}{n}\sum_i\left(1+\sqrt{1+2h_i m_i^2} - TM\right)^2 \\ \mathbb{E}(\log\beta) = TL = \frac{1}{n}\sum_i\log\left(1+\sqrt{1+2h_i m_i^2}\right) \end{cases} \tag{4.45a}$$

$$\begin{cases} \mathbb{E}(\beta) = TIM = \frac{1}{n}\sum_i t_i^{-1} \\ \mathbb{V}(\beta) = TIV = \frac{1}{n}\sum_i\left(t_i^{-1} - TIL\right)^2 \\ \mathbb{E}(\log\beta) = TIL = \frac{1}{n}\sum_i\log t_i^{-1} \end{cases} \tag{4.45b}$$

4.5. MODELE NORMAL-GAMMA A QUATRE PARAMETRES NG(m,v,k,n)

4.5.1. Forme mathématique et Estimateur sans biais

Rappelons que dans le chapitre III, nous avons étudié le modèle normal-gamma $NG(m,v,k)$ du couple $\theta = (\mu,\eta)$, considéré comme paramètres inconnus de la distribution normale.

Dans les travaux actuels (cf. [36], [17], [24] et [35]), on utilise plutôt le modèle normal-gamma à quatre paramètres $NG(m,v,k,n)$ sous la forme suivante :

$$NG(m,v,k,n)(\mu,\eta) = \sqrt{\frac{k\eta}{2\pi}}\exp\left(-\ k\eta(m-\mu)^2\right)\frac{(nv)^{n/2}}{\Gamma\left(\frac{n}{2}\right)}\eta^{n/2-1}\exp\left(-\ nv\eta\right) \tag{4.51}$$

Désignons $\mathbf{x} = (x_1,x_2,\ldots,x_k)$ un échantillon de la distribution normale $N(\mu,\sigma^2)$. L'évaluation de la distribution a priori est basée aussi sur la méthode de la conjuguée de la fonction de vraisemblance (cf. paragraphe 1.7). Cependant, dans le développement de cette fonction (similaire à ce que est fait au paragraphe 3.1.1), la variance retenue est l'estimateur sans biais : $v' = \frac{1}{k-1}\sum(x_i-m)^2$ au lieu de celui de variance minimale $v = \frac{1}{k}\sum(x_i-m)^2$, et la distribution a priori devient la suivante :

$$\bar{\pi}(\mu,\eta) \propto \eta^{k/2}\exp\left(-\ k\eta(m-\mu)^2\right)\exp\left(-\ (k-1)v'\eta\right) \quad .$$

donc cette distribution a pour forme : $\bar{\pi}(\mu,\eta) = \bar{\pi}(\mu|\eta)\bar{\pi}(\eta)$ avec $\bar{\pi}(\mu|\eta)$ la distribution normale $N\left(m,\frac{1}{2k\eta}\right)$ et $\bar{\pi}(\eta)$ la distribution gamma $G\left(\frac{k+1}{2},(k-1)v'\right)$.

Dans ce modèle, on introduit un quatrième paramètre $n = k-1$ pour la loi gamma $\bar{\pi}(\eta) = G\left(\eta;\ \frac{n+2}{2},\ nv'\right)$, mais on conserve la distribution normale de $\bar{\pi}(\mu|\eta)$ pour le paramètre k soit $N\left(m,\frac{1}{2k\eta}\right)$. On a donc

$$\bar{\pi}(\mu,\eta) \propto \sqrt{\frac{k\eta}{2\pi}}\exp\left(-\ k\eta(m-\mu)^2\right)\eta^{n/2}\exp\left(-\ nv'\eta\right) \quad .$$

Enfin pour permettre au paramètre "a" de la distribution gamma $G(a,b)$ d'être aussi proche que l'on veut de 0, on suppose que $\bar{\pi}(\eta)$ est une loi de

distribution $G\left(\frac{n}{2}, nv'\right)(\eta)$. La forme ainsi retenue pour la distribution normale-gamma à quatre paramètres NG(m,v,k,n) est la suivante :

$$\pi'(\mu,\eta) = N\left(m, \frac{1}{2k\eta}\right)(\mu) G\left(\frac{n}{2}, nv'\right)(\eta)$$

au lieu de la forme "empirique" :

$$\bar{\pi}(\mu,\eta) = N\left(m, \frac{1}{2k\eta}\right)(\mu) G\left(\frac{n+2}{2}, nv'\right)(\eta)$$

ou

$$\pi(\mu,\eta) = N\left(m, \frac{1}{2k\eta}\right)(\mu) G\left(\frac{k+1}{2}, kv\right)(\eta)$$

A partir de cette distribution normale-gamma, suivant la même procédure qu'au paragraphe 3.1.1, la distribution bayésienne de la variable X, étant la distribution marginale, est aussi de Student:

$$p(x) = T\left(m, v'\frac{k+1}{k}, n\right)(x) \tag{4.52}$$

Pour valider un modèle bayésien, il restera à choisir une méthode de détermination de la distribution a priori. Nous allons montrer en premier lieu que les trois premières méthodes développées au chapitre II ne peuvent fonctionnner pour permettre de différencier k et n dans le modèle à quatre paramètres $NG(m,v,k,n)$. On ne peut donc utiliser que les méthodes du quasi-échantillon.

Usuellement, dans le cas de la distribution normale, on applique le modèle $NG(m,v,k,n)$ associé à une méthode de détermination de la distribution a priori, dite méthode du maximum de vraisemblance. Ici, nous utilisons les méthodes de détermination de la distribution a priori que nous avons développées au chapitre II. La méthode du maximum de vraisemblance correspond à une méthode du quasi-échantillon que nous appelons méthode des doubles maxima de vraisemblance.

Nous allons faire une comparaison entre les estimateurs des paramètres des deux modèles $NG(m,v,k,n)$ et $NG(m,v,k)$, et étudier leurs influences sur les résultats prédictifs.

Dans l'étude du présent, nous prenons le modèle normal-gamma à trois paramètres $NG(m,v,k)$, qui permet d'utiliser de façon natuelle les méthodes de détermination de la distribution a priori que nous avons développées au chapitre II sans y ajouter un paramètre "n" supplémentaire.

4.5.2. Non validation du modèle pour certaines méthodes

Supposons l'information a priori constituée d'une série des échantillons $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$ de même loi de distribution normale.

Méthode séquentielle

La méthode séquentielle avec au départ une mesure uniforme, donne une forme "directe" pour la distribution a priori qui est la suivante (cf. formule (2.13)) :

$$\pi(\mu,\eta) \propto \prod_i f(\mathbf{x}_i|\mu,\eta) = \prod_i \prod_j f(x_j^i|\mu,\eta) \quad .$$

La statistique exhaustive est le triplet (MT,VMT,KT) formé de la moyenne, la variance et la taille de l'échantillon total (cf. paragraphe 4.1), et donc la forme de distribution est "empirique" $NG(m,v,k)$ avec m = MT, v = VMT et k = KT, non pas "théorique" $NG(m,v,k,n)$ à cause de la modification de la forme ci-dessus. (Bien entendu, nous pouvons toujours définir artificiellement le paramètre n = k-1).

Quant aux estimateurs de la variance, on utilise plutôt l'estimateur de variance minimale VMT, parce qu'il est calculé à partir "d'un échantillon cumulé" : $\mathbf{x} = (\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$; la taille de cet échantillon est généralement élevée, soit $KT = \sum k_i$.

Même si l'on utilise l'estimateur sans biais v' et n = k-1, le résultat bayésien ne change pas. En effet, suivant la procédure de calcul du paragraphe 3.3.1, la distribution bayésienne de X est aussi une loi de Student : $T(m,v'\frac{k-1}{k},k+1)$. Grâce à l'égalité : $v = v'\frac{k-1}{k}$, nous retouvons la même distribution de Student T(m,v,k+1) que pour l'estimateur de variance minimale.

Méthode des échantillons équilibrés

Si nous appliquons cette méthode pour le modèle $NG(m,v,k,n)$, nous avons la même conclusion que précédemment, puisque c'est toujours le même principe de la conjuguée naturelle. Donc on ne peut pas trouver la forme modifiée à quatre paramètres. Mais la fonction de vraisemblance est ici approchée et la statistique exhaustive (MM,VMM,KH) détermine directement le paramètre du modèle normal-gamma (m,v,k).

Si l'on veut utiliser l'estimateur de la variance sans biais v' au lieu de v, la forme de la distribution a priori sera :

$$\pi'(\mu,\eta) \propto \prod_i f(x_i|\mu,\eta)^{K/k_i}$$

$$\propto \prod_i \left(\eta^{k_i/2}\exp\left(-k_i\eta(m_i-\mu)^2\right)\exp\left(-k_i v'_i\eta\right)\right)^{K/k_i}$$

$$\propto \left(\prod_i \eta^{1/2}\exp\left(-\eta(m_i-\mu)^2\right)\exp\left(-\frac{k_i-1}{k_i}v'_i\eta\right)\right)^{K} \quad .$$

Grâce à l'égalité $v_i = \frac{k_i-1}{k_i}\cdot v'_i$, les facteurs $\frac{k_i-1}{k_i}$ ajustés sur les variances v'_i donneront le même résultat que dans le cas de l'estimateur de variance minimale v et nous avons : $\pi'(\mu,\eta|m,v',k,n=k-1) = \pi(\mu,\eta|m,v,k)$. La distribution bayésienne ou prédictive reste alors le même pour les deux estimateurs de la variance.

Méthodes des moments

Rappelons que pour le modèle $NG(m,v,k,n)$, la distribution bayésienne est aussi une loi de Student: $p(x) = T\left(m,v\frac{k+1}{k},n\right)$. Dans cette méthode, toutes les données mélangées sont supposées comme réparties suivant cette distribution non conditionnelle p(x).

La distribution a priori peut être déterminée par les relations entre des moments de p et ses paramètres :

$$\upsilon_1 = m \qquad\qquad \mu_{2m-1} = 0$$

$$\mu_{2m} = (2m-1)!!\frac{n^m}{(n-2)(n-4)\ldots(n-2m)}\left(\frac{k-1}{k}v'\right)^2$$

Il est clair que les deux paramètres v et k ne sont pas indépendants. Donc il n'y pas de solution pour cette méthode. Pour la même raison, la méthode du maximum de vraisemblance n'est pas valable, sauf si l'on impose la relation n = k-1.

4.5.3. Comparaison des estimateurs du quasi-échantillon

Rappelons que sous l'hypothèse de la normalité, nous appliquons deux méthodes pour déterminer la distribution a priori, qui sont dérivées de la méthode du quasi-échantillon : méthode des doubles moments MM et méthode des doubles maxima de vraisemblance VV.

La première étape de cette démarche consiste à chercher les valeurs les plus probables du paramètre (μ,η) de la distribution normale, à partir des n échantillons. Les résultats sont les mêmes par la méthode des moments ou par la méthode du maximum de vraisemblance :

$$x_i \longrightarrow \begin{cases} \mu_i = m_i \\ \eta_i = h_i = 1/(2v_i) \text{ ou } h'_i = 1/(2v'_i) \end{cases}$$

Ces résultats ne dépendent pas du modèle appliqué.

A partir de l'échantillon constitué des n valeurs de (μ,η), les paramètres du modèle $NG(m',v',k',n')$ sont obtenus encore une fois par la méthode des moments ou par la méthode du maximum de vraisemblance.

Nous utilisons le symbole "'" pour distinguer les paramètres du modèle à quatre paramètres de ceux du modèle à trois paramètres.

<u>Méthode des doubles moments MM</u>

Les moments de deux premiers ordres de la distribution $NG(m',v',k',n')$ sont estimés de façon empirique à partir du quasi-échantillon. Les équations de ces moments sont fonctions de (m',v',k',n') :

$$MM = \mathbb{E}(\mu) = m' \qquad MV = \mathbb{V}(\mu) = \frac{n'}{n'-2}\times\frac{v'}{k'}$$

$$HM' = \mathbb{E}(\eta) = \frac{1}{2v'} \qquad HV' = \mathbb{V}(\eta) = \frac{1}{2n'v'^2} \tag{*}$$

d'où

$$m' = MM \qquad v' = \frac{1}{2HM'} = VH'$$

$$n' = \frac{2HM'^2}{HV'} \qquad k' = \frac{VH'}{MV}(1-2/n')^{-1} \tag{4.53}$$

Notons $d' = \frac{\sqrt{HV'}}{HM'}$ le coefficient de variation empirique des $\{h'_i = (2v')^{-1}\}$. Pour que $k' = VH'\left(MV\cdot(1 - d'^2)\right)^{-1}$ soit positif, il faut que la variation des variances soit faible : $d'^2 < 1$. Quant au modèle $NG(m,v,k)$, le champ d'application est plus large, puisqu'il exige que le carré du coefficient de variation des précisions $\{h_i\}$ soit inférieur à 2 : $d^2 < 2$ avec $d = \frac{\sqrt{HV}}{HM}$ (cf. formule (4.411a)). Pour comparer ces deux modèles, nous pouvons supposer que $HM' \approx HM$ et $HV' \approx HV$, donc $d' \approx d$.

Nous savons que pour les deux modèles normal-gamma : $NG(m',v',k',n')$ et $NG(m,v,k)$, les distributions bayésiennes sont de Student: $T(m',v'\frac{k'+1}{k'},n')$ et $T(m,v,k+1)$. Les moyennes bayésiennes (ou prédictives) sont toutes égales à la moyenne des moyennes : $\mathbb{E}(X) = m = m' = MM$, mais les variances ne sont pas les mêmes, cette différence est due aux paramètres k' et k :

$$V' = v'\frac{k'+1}{k'}\times\frac{n'}{n'-2} = (k'+1)MV$$

$$V = v\frac{k+1}{k-1} = (k+1)MV$$

Remarquons qu'il y a deux possibilités pour le choix des paramètres du modèle $NG(m,v,k)$ (cf. 4.411a).

Si nous prenons $k = 2\frac{HM^2}{HV} - 1 = 2d^{-2} - 1$, alors nous avons :

$$k/k' = \frac{MV}{VH}(1-d^2)\times(2d^{-2} - 1) = \frac{MV}{VH}(2d^{-2} - 3 + d^2)$$

Comparer k et k' revient à comparer $x = \frac{VH}{MV}$ et $y = 2d^{-2} + d^2 - 3$. Dans la pratique, on considère souvent la situation où la valeur MV est assez grande et alors x est inférieur à y. Par suite, on a $k' < k$ et la variance prédictive a priori V' obtenue par le modèle $NG(m',v',k',n')$ est inférieure à celle V

obtenue par le modèle $NG(m,v,k)$. Si $x \geq y$, ce qui sera le cas lorsque d^2 est proche de 1, au contraire on a $V' \geq V$.

Si nous choisissons $k = \frac{c+1}{2c}\left(1 + \sqrt{1+\frac{4c}{(1+c)^2}}\right)$ pour le modèle $NG(m,v,k)$, nous avons la même conclusion $V' < V$, si $d^2 \leq \frac{c}{c+1}$, où $c = MV/VH$, puisque nous avons :

$$\frac{V'}{V} \approx \frac{c + \frac{1}{1-d^2}}{c+\frac{1}{2}(c+1)\left(1+\sqrt{1+\frac{4c}{(1+c)^2}}\right)} \leq \frac{c + 1/(1-d^2)}{1 + 2c}$$

<u>Méthode des doubles maxima de vraisemblance</u>

Comme nous l'avons développé au paragraphe 4.4.1, l'estimateur du maximum de vraisemblance du modèle $NG(m',v',k',n')$ est :

$$m' = MHM' \qquad v' = VH'$$

$$\frac{1}{k'} = 2\left(MH4' - MHM'^2 HM'\right) \qquad \frac{1}{n'} = \log(HM') - HL' \tag{4.54}$$

Les moyennes et les variances de (μ,η) sont :

$$\mathbb{E}(\mu) = m' = MHM' \qquad V'_\mu = \mathbb{V}(\mu) = \frac{n'}{n'-2}\times\frac{VH'}{k'}$$

$$\mathbb{E}(\eta) = \frac{1}{2v'} = HM' \qquad V'_\eta = \mathbb{V}(\eta) = \frac{1}{2n'v'^2} = \frac{2}{n'}HM'^2$$

Nous donnons les résultats correspondant au modèle $NG(m,v,k)$:

$$\mathbb{E}(\mu) = m = MHM \qquad V_\mu = \mathbb{V}(\mu) = \frac{v}{k-1} = \frac{k+1}{k-1}\times\frac{VH}{k}$$

$$\mathbb{E}(\eta) = \frac{k+1}{2kv} = HM \qquad V_\eta = \mathbb{V}(\eta) = \frac{k+1}{2}(kv)^{-2} = \frac{2}{k+1}HM^2$$

Pour comparer les modèles, nous supposons que $v'_i \approx v_i$, alors $MHM' \approx MHM$, $HM' \approx HM$, ...etc. Les moyennes de μ et η sont presque égales ; ce sont aussi les variances qui font la différence entre deux modèles normal-gamma.

Pour que les variances V_μ et V'_μ existent, il faut satisfaire les conditions : $n' > 2$ et $k > 1$.

Désignons $\begin{cases} x = 2\left(MH4 - MHM^2 \cdot HM\right) \\ y = \log(HM) - HL \end{cases}$

Nous avons alors $k = \frac{1}{x+y}$; et les conditions "$n' > 2$ et $k > 1$" sont équivalentes aux conditions $y < 1/2$ et $x+y < 1$. Pour comparer la différence entre les variances de la moyenne μ, nous considérons la fonction suivante :

$$(V'_\mu - V_\mu)/VH = \frac{1+x}{1-2y} - \frac{1+x+y}{1-x-y}$$

Dans le cas de l'égalité, nous obtenons la condition : $x^2 + x = 2y^2 + xy$ avec les contraintes ci-dessus. La relation entre les deux variances V'_μ et V_μ dépend de la relation entre les deux statistiques x et y :

— Si $x^2 + x \geq 2y^2 + xy$, alors $V'_\mu \leq V_\mu$;

— Si $x^2 + x < 2y^2 + xy$, alors $V'_\mu > V_\mu$.

Nous illustrons les zones des contraintes de (x,y). La courbe au milieu représente la limite : $x^2 + x = 2y^2 + xy$ telle que $V'_\mu = V_\mu$. Dans la zone à droite, $V'_\mu < V_\mu$; dans la zone à gauche, $V'_\mu > V_\mu$.

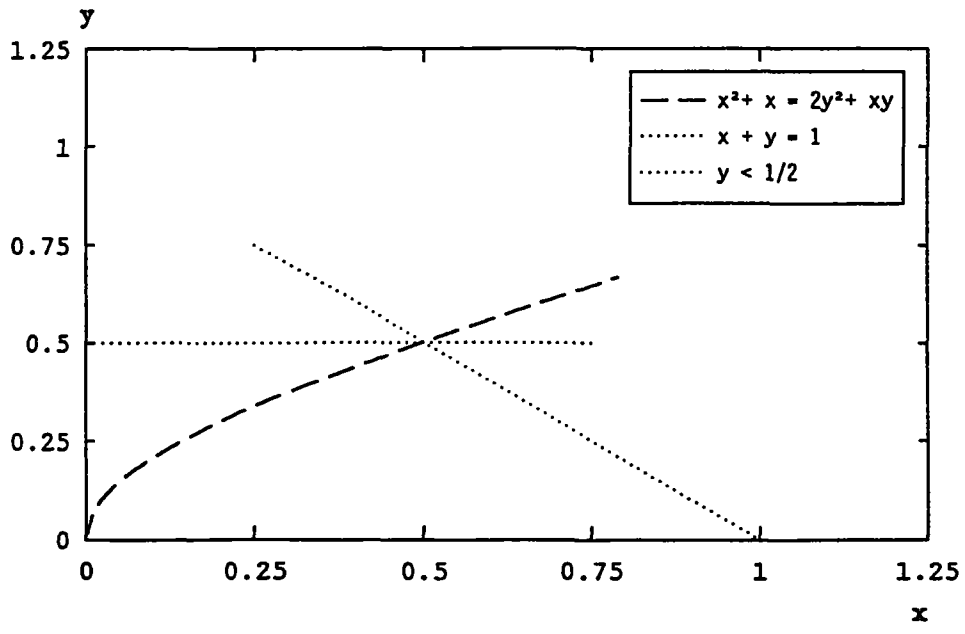


Figure 4.1 Zone de définition

Signification des paramètres et leurs Contraintes

Considérons d'abord la statistique y :

$$y = \log(HM) - HL = \log\left(\frac{1}{n}\sum h_i\right) - \frac{1}{n}\sum \log h_i \tag{4.55}$$

Etant la différence du logarithme des précisions, cette statistique y mesure une dispersion sur la précision η, ainsi que sur la variance $\sigma^2 = 1/(2\eta)$, qui est $\frac{1}{n'} = \mathbb{V}(\eta)/[2\mathbb{E}(\eta)^2]$.

Dans le cas où la dispersion y est petite, c'est à dire lorsque les variances sont relativement grandes et assez proches les unes des autres, le paramètre n' est grand. Pour ces deux modèles normal-gamma, on a utilisé une approximation de la fonction digamma $\psi\left(\frac{n'}{2}\right)$ et $\psi\left(\frac{k+1}{2}\right)$, à condition que n' et k soient suffisamment grands.

Ensuite, nous étudions la statistique x :

$$x = 2\left(MH4 - MHM^2HM\right) = \frac{2}{n}\left[\sum h_i m_i^2 - \frac{\left(\sum h_i m_i\right)^2}{\sum h_i}\right]$$

$$= \frac{2}{n}\sum h_i\left(m_i - \frac{\sum h_i m_i}{\sum h_i}\right)^2 \tag{4.56}$$

Nous pouvons imaginer que la statistique $Vm = \frac{x}{2HM}$ est en quelque sorte une variance des moyennes $\{m_i\}$ pondérées par les précisions $\{h_i\}$, ou appelée variance empirique de la moyenne pondérée par la précision. La statistique $Hm = \frac{1}{2Vm}$ est appelée précision empirique de la moyenne.

Nous pouvons conclure que le paramètre $k' = \frac{1}{x} = \frac{Hm}{HM}$ est un rapport entre deux grandeurs :

— la précision Hm empirique de la moyenne μ pondérée par la précision η. Dans cette statistique, les précisions $\{h_i\}$ représentent un poids de chacune des moyennes $\{m_i\}$ dans l'information a priori. Si toutes les moyennes $\{m_i\}$ sont très voisines, Hm est proche de zéro même si les $\{h_i\}$ sont différentes l'une de l'autre ; par contre, si les $\{h_i\}$

sont voisines ou constantes, Hm est tout simplement la précision de la moyenne μ. Cette propriété pour caractériser la dispersion de la moyenne μ s'interprète par le fait que la variance de la distribution normale-gamma s'écrit :

$$V(\mu) = \frac{n'}{n'-2} \times \frac{v'}{k'} = \frac{n'}{n'-2} \times \frac{v'}{k'} \approx Vm$$

pour n' assez grand.

— la précision moyenne HM des résultats aléatoires de la variable X dans les lots considérés. Pour avoir une bonne précision, il nécessite des moyens de mise en oeuvre, donc le coût va augmenter.

La précision Hm est une mesure au sens global de la production ; et la précision HM est une mesure au sens local vis à vis d'un lot précis. Il est clair que plus la précision est grande, plus le résultat est précis.

Désignons $VH = \frac{1}{2HM}$ la moyenne harmonique des variances, nous trouvons la relation $k' = \frac{VH}{Vm}$ en terme de variances.

4.5.4. Modèle NG(m,v,k) avec l'estimateur sans biais

Pour le modèle $NG(m,v,k)$, nous pouvons également utiliser l'estimateur de la variance sans biais. Dans ce cas, le modèle normal-gamma à trois paramètres devient :

$$\pi(\mu,\eta) = N\left(\mu;\ m,\ \frac{1}{2k\eta}\right) G\left(\eta;\ \frac{k+1}{2},\ (k-1)v'\right) \quad .$$

Nous venons de montrer que pour les trois premières méthodes séquentielle SS, des échantillons équilibrés EE et des moments OO, l'utilisation de l'estimateur sans biais n'entraine pas d'influence sur le résultat prédictif a priori. Par contre, pour les méthodes du quasi-échantillon, elle induit peu des changements sur les estimateurs des paramètres.

Les estimateurs des doubles moments et des doubles maxima de vraisemblance sont respectivement donnés par

$$\begin{cases} m = MM \\ v = k \cdot MV \\ k = \frac{1+c'}{2c'}\left[1+\sqrt{1+\frac{4c'}{(1+c')^2}}\right] \end{cases} \qquad \begin{cases} m = MM \\ v = VH'\frac{k+1}{k} \\ k = 2HM'^2/HV' - 1 \end{cases} \tag{4.57}$$

$$\begin{cases} m = MHM' \\ v = VH'\frac{k+1}{k-1} \\ k \approx \left(\log HM' - HL' + 2(MH4' - MHM'^2 HM')\right)^{-1} \end{cases} \tag{4.58}$$

où c' = MV/VH'. Les statistiques avec le symbole "'" sont calculés avec les précisions $\{h'_i = 1/(2v'_i)\}$ au lieu des précisions $\{h_i = 1/(2v_i)\}$.

Comparant ces estimateurs avec les formules (4.411a et b) dans le cas de l'estimateur de variance minimale, le seul paramètre v change légèrement. Nous savons que la distribution bayésienne de la variable X est une loi de Student $T(m, v\frac{k-1}{k}, k+1)$, il est clair que le paramètre de la variance $v\frac{k-1}{k}$ est égal de manière formelle à celui de l'estimateur de variance minimale.

Finalement, la différence entre les modèles estimés par les deux estimateurs différents de la variance dépendra de l'écart entre les statistiques calculées avec les présicions $\{h'_i\}$ ou $\{h_i\}$. Dans le cas où la taille est relativement élevée, nous avons $h'_i = h_i$, donc les deux modèles sont quasiment les mêmes.

4.6. ETUDE DE LA SENSIBILITE ET COMPARAISON DES METHODES

Au vu de la facilité de calculs pour des lois normales et de l'interpétation des statistiques associées, l'étude de ce paragraphe sur le modèle normal-gamma $NG(m, v, k)$ a pour but de montrer l'influence de l'information a priori et de l'information provenant d'un lot sur la distribution a posteriori et la distribution prédictive. Pour le modèle normal-gamma à quatre paramètres, nous ferons auusi quelques commentaires sur le problème de la sensibilité.

Nous allons comparer les cinq méthodes à titre d'exemple pour le modèle normal-gamma $NG(m, v, k)$, méthodes que nous venons de développer.

4.6.1. Etude de la sensibilté du modèle normal-gamma

Nous savons que dans l'approche bayésienne, les relations entre les trois informations a priori (du passé), du présent et a posteriori (résultant des deux premières) sont exprimées par la formule de passage de la distribution a priori à la distribution posteriori :

$$\begin{cases} m_1 = \dfrac{k_0 m_0 + km}{k_0 + k} \\ v_1 = \dfrac{k_0}{k_1} v_0 + \dfrac{k}{k_1} v + \dfrac{k_0 k}{k_1 k_1}(m - m_0)^2 \\ k_1 = k_0 + k \end{cases} \tag{4.61}$$

Au vu de cette relation, nous constatons que le résultat a posteriori est une pondération entre les deux informations du passé et du présent, représentés respectivement par (m_1, v_1, k_1), (m_0, v_0, k_0) et (m, v, k).

Nous savons que les résultats prédictifs a priori i = 0 et a posteriori i = 1 sont exprimés en moyenne et variance de la façon suivante :

$$M_i = m_i \qquad V_i = v_i \frac{k_i + 1}{k_i - 1} \tag{4.62}$$

Le résultat a posteriori (M_1, V_1) est "à peu près" une pondération entre les résultats a priori (M_0, V_0) et empirique (m, v), parce que nous ne pouvons pas calculer directement la distribution prédictive a posteriori à partir de la distribution prédictive a priori (cf. paragraphe 1.8).

Le paramètre a posteriori v_1 est une pondération entre les paramètres v_0, v et le carré de l'écart entre les moyennes a priori m_0 et empirique m. La moyenne m_0 représente en effet une <u>valeur cible</u> dans la procédure. Mais cette valeur est flottante lorsque l'on prendra un processus glissant dans le temps, puisque elle dépend directement des n échantillons antérieurs pris dans l'information a priori pendant une période convenable.

Cette remarque permet de conclure qu'à partir du résultat prédictif (M_1, V_1), est prise la décision, décision d'autant plus fiable que les deux informations sont concordantes, puisque la variance v_1 diminue, ainsi que

V_1, quand l'écart entre la moyenne empirique m et la moyenne a priori $M_0 = m_0$ tend vers 0.

Dans ce modèle, les paramètres k_0 et k jouent un rôle important, car ce sont eux qui déterminent les poids des deux informations dans la pondération et permettent également le calcul du résultat prédictif. Il est évident que plus le rapport k/k_0 augmente, plus la moyenne prédictive M_1 est proche de la moyenne empirique m ; inversement si ce rapport k/k_0 diminue, $M_1 = m_1$ est plus proche de $M_0 = m_0$.

Pour la variance prédictive, nous arrivons à la même conclusion, puique le paramètre v_1 est en effet fonction de k et k_0 :

$$v_1 = h(k,k_0) = \frac{1}{k_0 + k}\left(k_0 v_0 + kv + k_0 m_0^2 + km^2 - (k_0 + k)m_1^2\right)$$

d'où
$$\begin{cases} \dfrac{\partial}{\partial k} h(k,k_0) = \dfrac{k_0}{(k_0+k)^2}\left(v - v_0 + m^2 - m_0^2\right) \\ \dfrac{\partial}{\partial k_0} h(k,k_0) = \dfrac{k}{(k_0+k)^2}\left(v_0 - v + m_0^2 - m^2\right) \end{cases}$$

Il existe trois cas possibles selon la relation entre les deux valeurs $v + m^2$ et $v_0 + m_0^2$:

— Si $v + m^2 > v_0 + m_0^2$, v_1 est croissante en k et décroissante en k_0.

— Si $v + m^2 < v_0 + m_0^2$, v_1 est décroissante en k et croissante en k_0.

— Si $v + m^2 = v_0 + m_0^2$, v_1 est constante pour les k et k_0 quelconques.

Les comportements asymptotiques de la moyenne M_1 et de la variance V_1 sont

$$\begin{cases} \lim\limits_{k_0 \to \infty} M_1 = M_0 \\ \lim\limits_{k_0 \to \infty} V_1 = V_0 \end{cases} \qquad \begin{cases} \lim\limits_{k \to \infty} M_1 = m \\ \lim\limits_{k \to \infty} V_1 = v \end{cases} \tag{4.63}$$

Choisir un modèle bayésien convenable conduit à définir à la fois une information a priori, un plan d'échantillonnage et une méthode de détermination de la distribution a priori, de manière à obtenir le rapport $\rho = k_0/k$

raisonnable pour chaque problème posé. Un tel choix permettra de résoudre des problèmes d'un petit nombre k d'essais, parce que l'information a priori ne pèsera pas trop par rapport à un petit échantillon prélevé dans un lot.

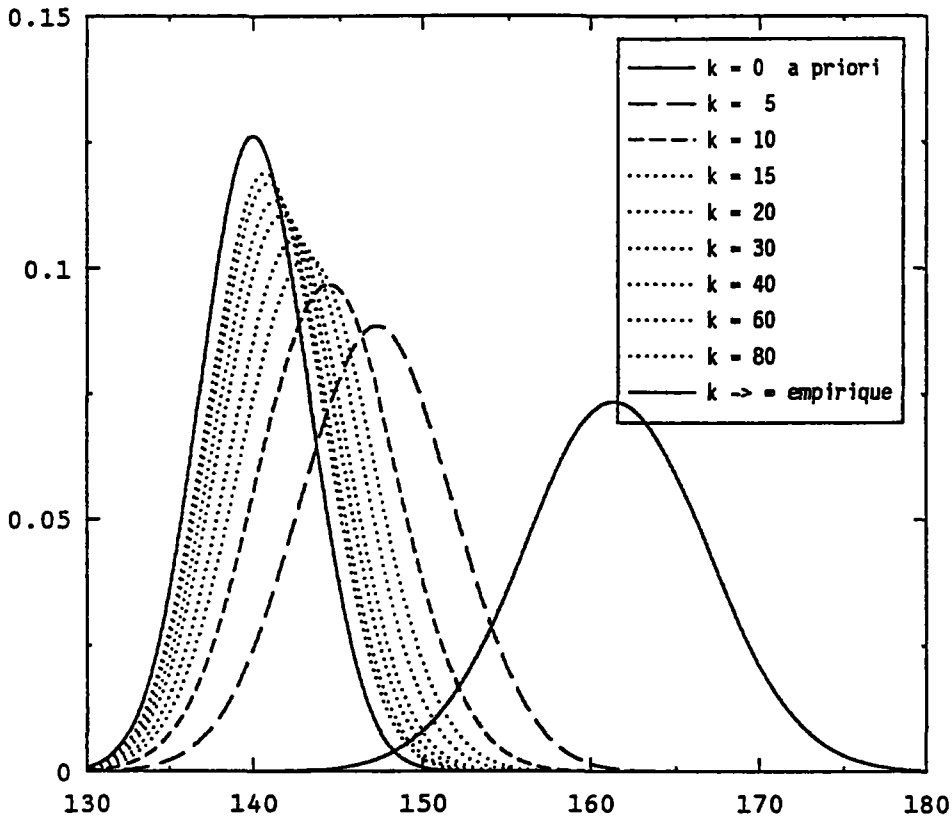


Figure 4.2 Comportement asymptotique

Nous montrons cet effet de la pondération et ce comportement asymptotique, en traçant les distributions prédictives a priori p(x) et empirique f(x|m,v) considérées comme les distributions limites, et une série de distributions a posteriori p(x|x) correspondant aux différentes tailles k avec la moyenne m et la variance v empiriques fixées, dont les résultats sont donnés dans le tableau 5.18 pour le lot n°18.

Quant au modèle normal-gamma à quatre paramètres $NG(m,v,k,n)$, la statistique exhaustive d'un échantillon est $(m,v',k,n = k-1)$. Si nous reprenons les notations (m'_0,v'_0,k'_0,n'_0) pour les paramètres de la distribution a priori, alors la formule du passage de l'a priori à l'a posteriori est la suivante :

$$\begin{cases} m'_1 = \dfrac{k'_0 m'_0 + km}{k'_0 + k} \\ v'_1 = \dfrac{n'_0}{n'_1} v'_0 + \dfrac{n}{n'_1} v' + \dfrac{k'_0 k}{n'_1 k_1}(m - m'_0)^2 \\ k'_1 = k'_0 + k \\ n'_1 = n'_0 + n + 1 = n'_0 + k \end{cases} \tag{4.64}$$

Les résultats prédictifs a priori i = 0 et a posteriori i = 1 sont de la forme suivante :

$$\begin{cases} M'_i = m'_i \\ V'_i = v'_{i-1} \dfrac{k'_i + 1}{k'_i} \times \dfrac{n'_{i-1}}{2} \end{cases} \tag{4.65}$$

Il est clair que le résultat a posteriori (M'_1, V'_1) est aussi à peu près une pondération entre les résultats a priori (M'_0, V'_0) et empirique (m, v). Mais dans cette pondération, les poids représentant les informations du passé et du présent sont respectivement par (k'_0, n'_0) et $(k, n = k-1)$, donc nous avons deux rapports k'_0/k et n'_0/n.

Le modèle $NG(m, v, k, n)$ aura le même comportement symptotique que précédemment :

$$\begin{cases} \lim\limits_{k'_0 \to \infty} M'_1 = M'_0 \\ \lim\limits_{k'_0 \to \infty} V'_1 = V'_0 \end{cases} \qquad \begin{cases} \lim\limits_{k \to \infty} M'_1 = m \\ \lim\limits_{k \to \infty} V'_1 = v \end{cases} \tag{4.66}$$

Pour les deux modèles normal-gamma, les moyennes a priori sont égales : $M_1 = m_1 = m'_1 = M'_1$, égale à la moyenne des moyennes MM pour la méthode des doubles moments et égale à la moyenne empirique de la moyenne pondérée par la précision MHM pour la méthode des doubles maxima de vraisemblance.

Si nous appliquons la méthode des doubles moments MM, le modèle à quatre paramètres $NG(m, v, k, n)$ favorise légèrement l'information provenant du lot à tester par rapport au modèle à trois paramètres $NG(m, v, k)$, dans le cas où la moyenne μ de la variable X est relativement grande ou dans le cas où la variance de la moyenne μ est grande (cf. paragraphe 4.5.3), parce que nous avons $k'_0 < k_0$.

Au contraire, si nous appliquons la méthode des doubles maxima de vraisemblance VV, le modèle $NG(m,v,k,n)$ favorise toujours plus l'information a priori que le modèle $NG(m,v,k)$. En effet, nous avons montré au paragraphe 4.5.3 que la relation entre le paramètre k_0 du modèle $NG(m,v,k)$ et le couple (k'_0,n'_0) du modèle $NG(m,v,k,n)$ déterminée par la formule suivante :

$$\frac{1}{k_0} = \frac{1}{k'_0} + \frac{1}{n'_0}$$

Les paramètres k'_0 et n'_0 sont alors tous supérieurs au paramètre k_0.

Quant à la comparaison entre les variances prédictives (a posteriori) V'_1 et V_1, le résultat sera un peu plus compliqué, car leur rapport dépend de plusieurs paramètres : (m'_0,v'_0,k'_0,n'_0), v', $n = k-1$, (m_0,v_0,k_0) et (m,v,k). Nous n'entrerons pas dans le détail.

4.6.2. Comparaison des méthodes

Résumons ici les cinq estimateurs du modèle normal-gamma $NG(m,v,k)$ dans le tableau 4.2. Nous notons ce modèle par N11.

Rappelons que l'estimateur séquentiel est la statistique exhaustive (MT,VMT,KT) de l'échantillon composé $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n = \{x^i_j \ ; \ i = 1,2,\ldots,n$ et $j = 1,2,\ldots,k_i\}$ avec les définitions suivantes :

MT — la moyenne des moyennes $\{m_i\}$ pondérées par les tailles $\{k_i\}$,

VMT —la moyennes des $\{v_i + (m_i - MT)^2\}$ pondérées par les tailles,

KT — la taille totale $\sum k_i$.

Dans ce cas, plus la taille de l'échantillon est élevée, plus cet échantillon aura du poids dans l'information a priori, ainsi que dans les distributions du paramètre $\theta = (\mu,\eta)$ et de la variable X.

Quant à l'estimateur des moments, il est naturel que nous trouvions la même conclusion parce que les statistiques sont calculées à partir de toutes les données mélangées.

En général, pour un problème de poids non indentiques, une solution qui

satisfait sans doute plus l'esprit serait d'avoir des tailles constantes d'échantillons formant l'information a priori. Dans ce cas, les deux estimateurs de la moyenne m_0 et de la "variance" v_0 deviennent respectivement :

MT = MM — la moyenne des moyennes $\{m_i\}$

VMT = VMM — la moyenne des $\{v_i + (m_i - MM)^2\}$

mais les deux estimateurs du paramètre k_0 ne changent pas.

N11	Méthode	Paramètres m_0	v_0	k_0
$NG(m, v, k)$ μ , η $\eta = \frac{1}{2}\sigma^{-2}$	SS	MT	VMT	KT
	EE	MM	VMM	KH
	OO	MT	$VMT\frac{k_0-1}{k_0+1}$	$3+\frac{6}{C4}$
	MM	MM	$(k_0-1)MV$	$\frac{1+c}{2c}\left[1+\sqrt{1+\frac{4c}{(1+c)^2}}\right]$ c=MV/VH
		MM	$VH\frac{k_0+1}{k_0}$	$2\frac{HM^2}{HV} - 1$
	VV	MHM	$VH\frac{k_0+1}{k_0}$	$\log(HM) - HL + \frac{Vm}{VH}$

Tableau 4.2 Estimateur du modèle normal-gamma

Pour la méthode séquentielle, k_0= KT = n×KM est la totalité des essais; l'information a priori donnera généralement un poids "n" plus fois important qu'un échantillon dont la taille est moyenne k ≈ KM. Pour la méthode des moments, le paramètre k_0 dépend du coefficient d'aplatissement C4.

D'après ces remarques, les deux méthodes seront adaptées plutôt dans les cas des échantillons de taille faible, par exemple k < 5. Au lieu de chercher des significations des statistiques de chaque échantillon, nous cumulerons toutes les informations en un seul échantillon si la situation le permet, et puis nous l'introduirons dans le processus séquentiel (SS), nous

calculerons les moments empiriques, qui détermineront les paramètres de la distribution a priori (OO).

Cependant, les méthodes des échantillons équilibrés EE et du quasi-échantillon MM et VV permettent de traiter les moyennes des statistiques $\{m_i\}$, $\{h_i\}$, ..., indépendamment des tailles $\{k_i\}$ des échantillons.

Pour la méthode des doubles maxima de vraisemblance, le paramètre de la moyenne est la moyenne des moyennes $\{m_i\}$ pondérées par les précisions $\{h_i\}$: m_0 = MHM. Ainsi plus la précision obtenue pour un échantillon est élevée, plus cet échantillon prendra un poids important dans l'information a priori, et donc aussi dans les distributions prédictives a priori et a posteriori. Quant à la variance, le paramètre v_0 est à une constante près la moyenne harmonique des variances $\{v_i\}$: VH = $\left(\frac{1}{n}\sum v_i^{-1}\right)^{-1}$.

Les tailles $\{k_i\}$ et les variances ou précisions $\{h_i\}$ n'interviennent pas dans la détermination de la moyenne a priori m_0= MM, par les méthodes EE et MM. Le paramètre v_0 déterminé par la méthode MM dépend essentiellement de la variance de la moyenne tandis que la méthode EE donne v_0= VMM la moyenne des $\{v_i + (m_i - MM)^2\}$.

L'utilisation de l'une des trois méthodes EE, MM et VV dépendra des relations entre les facteurs suivants :

— les variances des échantillons : grandeur et dispersion ;

— l'effet de ces variances sur les moyennes ;

— la dispersion des moyennes.

Si nous prenons un score identique pour chacun des n échantillons 1/n, l'estimateur des scores d'expert du modèle normal-gamma est : (MT,VMT,KT/n), parce que la fonction de vraisemblance bayésienne est : $L'_{\mathbf{x}}(\theta) = \prod f(\mathbf{x}_i|\theta)^{1/n}$ et le facteur 1/n ne donne qu'un effet sur le paramètre k_0 = KT/n = KM, par rapport à l'estimateur séquentiel. Nous l'appelons la méthode séquentielle modifiée, notée aussi par SS.

Dans le cas où les tailles sont identiques, il est clair que cet estimateur donne le même résultat que l'estimateur des échantillons équilibrés EE puisque nous avons (MT = MM, VMT = VMM, KM = KH).

4.6.3. RESUME DES ESTIMATEURS DES MODELES NORMAL ET GAMMA

Nous nous plaçons toujours dans le cadre de la normalité de la distribution conditionnelle de la variable X étudiée. A travers ce chapitre et le chapitre III, nous avons vu qu'il existe plusieurs associations de modèles et de méthodes. Désignons respectivement par N11, N10 et N01 les modèles normal-gamma, normal et gamma, le premier étudie la moyenne et la précision ensemble (μ,η), le deuxième étude la moyenne μ et la dernier étude la précision η (équivalente à la variance σ^2).

Pour la moyenne μ, nous pouvons distinguer les cas suivants :

— variance connue et constante;

— variance connue, mais variable d'un lot à l'autre;

— variance inconnue et variable d'un lot à l'autre.

Dans le dernier cas, l'analyse est réalisée par le modèle normal-gamma N11 ; dans les deux permiers cas, nous prenons le modèle normal N10, dont le modèle prédictif de la variable X étudiée est aussi normal (cf. paragraphe 3.1.2) ; mais les paramètres seront différents. De plus, la pondération des informations du passé (a priori) et du présent (provenant de l'échantillon du lot à tester) va influer sur ces paramètres.

Nous remarquons que dans le deuxième cas, la méthode des familles paratriques n'est pas valable (cf. paragraphe 2.4).

Nous résumons dans les tableaux 4.3 et 4.4 tous les estimateurs (m_0,v_0) des modèles normaux, ainsi que les paramètres (m_1,v_1) des distributions normales a posteriori.

Rappelons que la moyenne prédictive de la variable X est soit $M_0= m_0$ a priori, soit $M_1= m_1$ a posteriori ; et que la variance prédictive de X est la somme de la variance de μ et la variance empirique: $V_1= v_1+ v$. Au vu de ces tableaux, nous constatons que la variance a priori de μ peut être exprimée par $v_0= V/k_0$, où V est une sorte de moyenne des variances. Donc nous pouvons prendre comme variance prédictive a priori $V_0 = v_0 + V = V(1 + 1/k_0)$. Nous montrons que la moyenne a posteriori M_1 est une pondération entre la moyenne

a priori M_0 et la moyenne empirique m, mais leurs poids dépendent de la méthode utilisée.

N10 $\sigma_1^2 = v_1$	Méthode	Paramètres m_0	Paramètres v_0	Pondération m_1	Pondération v_1
$N(m,v)$	SS	MHT	$\frac{VHT}{KT}$	$\frac{KT \cdot MHT/VHT + k \cdot m/v}{KT/VHT + k/v}$	$\frac{1}{KT/VHT + k/v}$
	EE	MHM	$\frac{VH}{KH}$	$\frac{KH \cdot MHM/VH + k \cdot m/v}{KH/VH + k/v}$	$\frac{1}{KH/VH + k/v}$
	MM	MM	$\frac{VH}{KM}$	$\frac{KM \cdot MM/VH + k \cdot m/v}{KM/VH + k/v}$	$\frac{1}{KM/VH + k/v}$
	VV	MHM	$\frac{VH}{k_0} = Vm$	$\frac{k_0 \cdot MHM/VH + k \cdot m/v}{k_0/VH + k/v}$	$\frac{1}{k_0/VH + k/v}$

Tableau 4.3 Estimateur et Pondération du modèle normal avec $\sigma_1^2 = v_1$

N10 σ^2=cte	Méthode	Paramètres m_0	Paramètres v_0	Pondération m_1	Pondération v_1
$N(m,v)$	SS	MT	$\frac{VT}{KT}$	$\frac{KT \cdot MT/VT + k \cdot m/v}{KT/VT + k/v}$	$\frac{1}{KT/VT + k/v}$
	EE	MM	$\frac{VM}{KH}$	$\frac{KH \cdot MM/VM + k \cdot m/v}{KH/VM + k/v}$	$\frac{1}{KH/VM + k/v}$
	OO	MT	0	MT	0
	MM=VV	MM	MV	$\frac{MT/MV + k \cdot m/v}{1/MV + k/v}$	$\frac{1}{1/MV + k/v}$

Tableau 4.4 Estimateur et Pondération du modèle normal avec σ^2 constante

---

—— k_0 est le paramètre défini par la formule (4.412b) et $Vm = VH/k_0$ est défini au paragraphe 4.5.3 (cf. formule 4.56)).

De façon similaire, pour la variance σ^2, nous distinguons également les trois cas suivants :

— moyenne connue et constante;

— moyenne connue, mais variable d'un lot à l'autre;

— moyenne inconnue et variable d'un lot à l'autre.

NO1	Méthode	Paramètres a_0	Paramètres b_0	Prédictions V_0	Prédictions V_1
$\mu_i = m_i$	SS	KT/2+1	KT·VT	VT	$\frac{KT \cdot VT + kv}{KT + k}$
$G(a,b)$	EE	KH/2+1	KH·VM	VM	$\frac{KH \cdot VM + kv}{KH + k}$
$\eta = \frac{1}{2}\sigma^{-2}$	MM	HM^2/HV	$2a_0VH$	$VH\frac{a_0}{a_0-1}$	$\frac{2a_0VH + kv}{2a_0-2+k}$
	VV	$\frac{1}{2}(\log HM - HL)^{-1}$	$2a_0VH$	$VH\frac{a_0}{a_0-1}$	$\frac{2a_0VH + kv}{2a_0-2+k}$

Tableau 4.5 Estimateur et Prédiction du modèle gamma avec $\mu_i = m_i$

NO1	Méthode	Paramètres a_0	Paramètres b_0	Préictions V_0	Préictions V_1
μ =cte	SS	KT/2+1	KT·VMT (μ=MT)	VMT	$\frac{KT \cdot VMT + kv}{KT + k}$
$G(a,b)$	EE	KH/2+1	KH·VMM (μ=MM)	VMM	$\frac{KH \cdot VMM + kv}{KH + k}$
$\eta = \frac{1}{2}\sigma^{-2}$	OO	2 + 3/C4	$2VMT(a_0-1)$ (μ=MT)	VMT	$\frac{2(a_0-1)VMT+kv}{2(a_0-1)+k}$
	MM	HMb^2/HVb	a_0/HMb (μ=MM)	$VHb\frac{a_0}{a_0-1}$	$\frac{2a_0VHb+kv}{2a_0-2+k}$
	VV	$\frac{1}{2}(\log HMb - HLb)^{-1}$	a_0/HMb (μ=MM)	$VHb\frac{a_0}{a_0-1}$	$\frac{2a_0VHb+kv}{2a_0-2+k}$

Tableau 4.6 Estimateur et Prédiction du modèle gamma avec μ = cte

Dans le dernier cas, le modèle du couple (μ, η) est normal-gamma N11; et dans les deux premier cas, nous choisissons le modèle gamma NO1, dont le modèle prédictif de la variable X étudiée est une distribution de Student (cf. paragraphe 3.1.3). Les moyennes a priori et a posteriori de la variance σ^2 sont définies par $V_i = \frac{b_i}{2(a_i-1)}$, $i = 0,1$.

Dans le tableau 4.6, les définitions des statistiques avec le suffixe "b" sont les mêmes que sans ce suffixe, mais nous utilisons les symboles des précision $\{h'_i = \frac{1}{2}[v_i + (\mu - m_i)^2]^{-1}\}$ au lieu des $\{h_i = 1/(2v_i)\}$.

CHAPITRE V ETUDE DU CAS DES BOULONS HR

5.0. EXPOSE DE L'ENVIRONNEMENT

Ce chapitre traite du contrôle des boulons à haute résistance. Nous ne nous intéressons pas ici au contrôle des assemblages. La vérification de conformité est effectuée dans le cadre de l'élaboration des normes nationales par le contrôleur et les laboratoires de la marque N.F. (CTICM, CEBTP, LPC).

Les boulons, constitués de vis, écrous et rondelles, sont fabriqués par la société "Forges et Boulonnerie d'Ars-sur-Moselle", qui possède la marque N.F. et est présenté sur les grands chantiers de la construction métallique: TGV, Tunnel sous la Manche, ... etc. Ces boulons sont repartis en quatres catégories de qualité :

— HR1 : boulons de classe 10.9 noirs (non revêtus ou non protégés),
— HR1Z: boulons de classe 10.9 galvanisés,
— HR2 : boulons de classe 8.8 noirs,
— HR2Z: boulons de classe 8.8 galvanisés.

Dans le contrôle en usine ou de réception, les lots sont bien définis et parfaitement identifiables par des paramètres géométriques (longueur, diamètre, ...), le numéro de coulée d'acier et le traitement thermique.

Le contrôle est réalisé par échantillonnage. Un échantillon est prélevé en usine dans chaque lot de fabrication. La vérification comporte les essais principaux suivants :

— essai d'aptitude à l'emploi (allongements rémanents);

CTICM : Centre Technique et Industriel de la Construction Métallique
CEBTP : Centre d'Etude du Bâtiment et des Travaux Publics
LPC : Laboratoire des Ponts et Chaussées

— essai de traction (résistance à la traction, limite d'élasticité...)
— essai de résilience (KCU);
— essai de traction avec cale biais;
— essai de dureté HV/30 (mesure sous une charge de 30daN).

Outre les caractéristiques de l'acier comme sa limite d'élasticité, sa résistance à la traction, sa résilience, etc ..., le contrôle porte principalement sur le coefficient de frottement K, qui détermine le couple de serrage appliqué aux assemblages des ouvrages métalliques, donc sur la sécurité obtenue. La valeur moyenne $\overline{K}$ de chaque lot de fabrication doit être déterminée en usine à l'état de livraison.

Pour distinguer la taille k d'échantillon et le coefficient K, nous utilisons ici la notation K majuscule au lieu de k en convention habituelle.

Cette étude comporte deux problèmes :

— l'estimation dans un lot de la moyenne $\mu = \overline{K}$ et de l'écart-type σ du coefficient K ;

— le suivi de l'évolution de la qualité dans le temps, permettant de faire un contrôle de la production globale.

Après un bref rappel du contrôle et de ses exigences, nous montrerons l'apport de l'approche bayésienne, puis nous effectuerons des applications numériques dans l'hypothèse d'une distribution normale du coefficient K, et enfin nous proposerons quelques modèles spécifiques adaptés au cas concret du contrôle des boulons.

5.1. COEFFICIENT K

5.1.1. Définition et Principe du contrôle

Les valeurs du coefficient K obtenues sur banc d'essai automatique sont expérimentales et dépendent :

— de l'état de lubrifiation des vis et des écrous,
— de l'état de finition des filets des vis et écrous,

— de l'état de finition des surfaces de contact écrou-rondelle.

Le coefficient K a une influence importante sur le couple C de règlage des clés dans le cas de "serrage par couple imposé" (cf. norme NF P 22-466), et donc sur la précontrainte P. La formule fondamentale, qui lie le couple de serrage par la clef dynamométrique à la précontrainte réelle dans les boulons, est de la forme suivante :

$$C = k_o P_{nom} \cdot K \cdot d \cdot 10^{-3} \qquad (5.00)$$

avec C : couple moyen de serrage (en newton-mètre),

d : diamètre nominal de la vis (en millimètre),

P_{nom}: précontrainte nominale garantie (en newton),

k_o: coefficient minorateur de sécurité pris généralement entre 0.72 et 0.90, la norme a fixé ce coefficient à 0.80.

Pour tenir compte des dispersions du coefficient K et du fonctionnement des clés, on majore le couple C de serrage de manière à avoir le maximum de boulons au-dessus de $P_U = 0.8 \cdot P_{nom}$, P_U valeur de la précontrainte prise en compte dans le calcul de dimensionnement (cf. norme NF E 27-701 et [56]).

Le contrôle de conformité actuel ne prend en considération que la moyenne $\bar{K}$, adoptée dans la formule (5.00). L'écart-type de K n'interviendra que de façon indirecte pour calculer le coefficient minorateur k_o, que J. Jacquet (1978) ([24]) a proposé et fixé une fois pour toutes dans les cas de règlage par le couple imposé. Cet écart-type apportera parfois une information supplémentaire pour rejeter des lots de boulons trop dispersés.

Lors de la mise en oeuvre des boulons, on ne peut pas toujours échapper à la présence de plusieurs lots sur un même chantier. Pour un diamètre donné, il y a en général 1, 2 ou 3 lots pour un chantier moyen (cf. [54]). Toutefois, le mélange des lots n'est pas pris en compte dans le contrôle actuel des assemblages (cf. norme NF P 22-466). Donc, on ne peut espérer que les lots utilisés dans un chantier proviennent d'un même fabricant et, aient les caractéristiques de dimensionnement.

C'est ainsi que le contrôle des boulons HR impose une exigence sur la moyenne du coefficient K. En même temps, il est souhaitable d'avoir une

faible dispersion à l'intérieur d'un lot et des productions de lots les plus homogènes possibles, donc une faible dispersion entre les lots successifs.

5.1.2. Précision de l'analyse statistique

Les tableaux 5.1 et 5.2 dans la page suivante donnent les répartitions de la production et des contrôles de l'année 1988, en nombre de pièces des boulons ; ils sont tirés du "Procès Verbal des Essais de Conformité, 1er Semestre 1989", relatif à la vérification de la marque N.F. des boulons HR à serrage contrôlé de l'AFNOR.

Nous constatons que la proportion de pièces testées dans l'ensemble de la production d'une même qualité, tous diamètres confondus, est faible : de 5 à 8 pour 10000 environ. En effet, généralement un lot peut contenir de 500 à 6000 boulons d'une même qualité et d'un diamètre donné, et les mesures du coefficient K sont faites sur 5 boulons par lot. Le rapport entre la taille d'un échantillon et le nombre de pièces dans le lot est donc faible.

> L'objectif est d'affiner une bonne fiabilité sur l'estimation de $\bar{K}$ sans augmenter les tailles d'échantillons testés pour des raisons de coût d'essais.

Pour les boulons HR d'une même qualité (répondant aux mêmes conditions de spécifications), les techniques de fabrication sont, sinon identiques, du moins très proches entre les lots successifs pendant une certaine période. Dans cette condition, nous pouvons utiliser les techniques bayésiennes, qui permettent de prendre en compte des informations antérieures de manière à avoir des informations supplémentaires, et rendre ainsi plus significatifs les résultats de l'analyse statistique par échantillonnage.

5.1.3. Prospection sur l'ensemble des données

Les mesures du coefficient K sont obtenues lors des essais d'aptitude à l'emploi (cf. norme NF E 27-702), et fournies par la société "Forges et Boulonnerie d'Ars-sur-Moselle".

Résultats pour l'année 1988

trimestre	1er	2ème	3ème	4ème	Annuel 1988
HR1	358 085	488 355	377 128	393 763	1 617 331
HR1Z	328 678	187 531	167 729	158 326	842 264
HR2	228 447	494 582	433 923	358 802	1 542 754
HR2Z	275 672	173 842	153 622	141 070	774 206

Tableau 5.1 Répartition de la production en nombre de pièces, tous diamètres confondus

Diamètre	12	14	16	18	20	22	24	27	30	Total
HR1	115	75	225	190	225	135	220	95	70	1350
HR1Z	65	40	45	85	155	90	100	60	55	695
HR2	55	30	185	110	185	70	90	0	40	765
HR2Z	60	60	100	80	135	40	70	35	20	600

Tableau 5.2 Répartition de nombres de pièces testées

Classe	Etat de surface	Symbole abrégé	Lot de fabrication n°	Nombre d'essais k	Moyenne $m \times 10^3$	Ecart-type $s \times 10^3$	Coeff. de variation $d = \frac{s}{m}$
10.9	non protégé	HR1	8085	5	162.00	6.325	0.0390
	galvanisé	HR1Z	8090	5	147.20	3.655	0.0248
8.8	non protégé	HR2	8088	5	159.00	5.831	0.0367
	galvanisé	HR2Z	8082	5	149.00	2.236	0.0600

Tableau 5.3 Valeurs de K dans les lots de différentes qualités

Lot n°	Nombre d'essais k	Moyenne $m \times 10^3$	Ecart-type $s \times 10^3$	Précision $h \times 10^{-3}$	Coef. de variation $d = \frac{s}{m}$
1	10	156.00	3.000	55.556	0.0192
2	5	162.00	6.782	10.870	0.0419
3	5	162.00	4.000	31.250	0.0247
4	5	169.00	7.348	9.259	0.0435
5	10	162.00	6.000	13.889	0.0370
6	5	173.00	4.000	31.250	0.0231
7	5	187.00	4.000	31.250	0.0214
8	5	165.00	3.162	50.000	0.0192
9	5	175.00	5.477	16.667	0.0313
10	5	176.00	6.633	11.364	0.0377
11	10	161.50	3.202	48.780	0.0198
12	5	149.00	4.899	20.833	0.0329
13	5	157.00	4.000	31.250	0.0255
14	5	159.00	6.633	11.364	0.0417
15	10	153.30	6.558	11.625	0.0428
16	5	159.00	5.831	14.706	0.0367
17	5	160.00	6.325	12.500	0.0395
18	5	140.00	3.162	50.000	0.0226
Moyenne	6.1	162.54	5.056	25.690	0.0311
Ecart-type	2.1	10.48	2.115	15.577	0.0089

Tableau 5.4 Valeurs de K dans les lots de la qualité HR2 (1988)

Le tableau 5.3 issu des cahiers de mesures concerne des boulons de différentes qualités, et le tableau 5.4 est relatif aux boulons de la même qualité HR2, d'un même diamètre ϕ = 18mm, et tous de l'année 1988. Ces deux tableaux montrent que le coefficient K varie non seulement entre des lots de différentes qualités, mais également entre des lots d'une même qualité et à l'intérieur d'un même lot.

Dans le tableau 5.4, nous donnons également les moyennes et les écart-types des statistiques des échantillons $\mathbf{x}_i$: tailles k_i, moyennes m_i, écart-types $s_i = \sqrt{v_i}$, précisions $h_i = 1/(2v_i)$ et coefficients de variation $d_i = s_i/m_i$ définis par :

$$KM = \mathbb{E}(k) = \frac{1}{n}\sum k_i \qquad KS = \mathbb{S}(k) = \sqrt{\frac{1}{n}\sum(k_i - KM)^2}$$

$$MM = \mathbb{E}(m) = \frac{1}{n}\sum m_i \qquad MS = \mathbb{S}(m) = \sqrt{\frac{1}{n}\sum(m_i - MM)^2}$$

$$SM = \mathbb{E}(s) = \frac{1}{n}\sum s_i \qquad SS = \mathbb{S}(s) = \sqrt{\frac{1}{n}\sum(s_i - SM)^2}$$

$$HM = \mathbb{E}(h) = \frac{1}{n}\sum h_i \qquad HS = \mathbb{S}(h) = \sqrt{\frac{1}{n}\sum(h_i - HM)^2}$$

$$DM = \mathbb{E}(d) = \frac{1}{n}\sum d_i \qquad DS = \mathbb{S}(d) = \sqrt{\frac{1}{n}\sum(d_i - DM)^2} \quad .$$

L'analyse statistique des valeurs du coefficient K montre que la dispersion à l'intérieur d'un lot est relativement faible (le coefficient de variation est de l'ordre de 10^{-2}) par rapport à celle entre les lots ; ceci se voit en comparant l'écart-type des moyennes empiriques, $MS = 10.48\times10^{-3}$, et la moyenne des écart-types empiriques, $SM = 5.056\times10^{-3}$.

Nous avons précisé ci-dessus que pour faciliter le contrôle des assemblages, il est préférable que la valeur moyenne $\mu = \bar{K}$ varie peu d'un lot à l'autre. Si la dispersion entre les lots successifs paraissait très grande, il conviendrait de faire une analyse spécifique sur l'ensemble des lots pris en compte, de manière à pouvoir effectuer un contrôle par échantillonnage. C'est un sujet d'étude qui reste ouvert.

Notre analyse statistique portera donc sur la dispersion entre les lots successifs et celle à l'intérieur des lots. La séquence des moyennes prédictives (ou bayésiennes a posteriori) de chaque lot montrera l'évolution du niveau de la qualité, sur lequel nous pourrons éventuellement fixer un critère pour déclencher une alarme en cas de changement brutal. Ceci conduira à la notion d'un <u>contrôle au sens global de la production</u>.

5.2. APPROCHE BAYESIENNE

Dans le contrôle classique, pour un lot à tester, la moyenne $\mu = \bar{K}$ et l'écart-type σ du coefficient K sont <u>estimés</u> par la moyenne et l'écart-type empiriques ($m, s = \sqrt{v}$) d'un échantillon **x** prélevé dans ce lot. Or, nous avons remarqué ci-dessus que dans le cas des boulons, les tailles d'échantillons sont faibles ; il peut donc arriver avec une probabilité assez forte que les valeurs "estimées" (m, s) soient éloignées des <u>valeurs réelles (μ, σ)</u> :

phénomène de biais statistique; ceci serait grave surtout si la moyenne m était très éloignée de $\mu = \bar{K}$.

Dans l'approche bayésienne, on estime dans un premier temps la distribution du coefficient K à partir des n échantillons précédents $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$; cette distribution est appelée distribution prédictive a priori $\mathbb{P}_0(\cdot) = \mathbb{P}(\cdot)$ avec une moyenne et un écart-type a priori (M_0, S_0). Ensuite, l'information fournie par l'échantillon $\mathbf{x}$ du lot considéré vient corriger cette première estimation, la distribution du coefficient K est réajustée pour ce lot, et devient une distribution prédictive a posteriori $\mathbb{P}_1(\cdot) = \mathbb{P}_{\mathbf{x}}(\cdot) = \mathbb{P}(\cdot|\mathbf{x})$ avec une moyenne et un écart-type a posteriori (M_1, S_1).

Dans ce modèle bayésien, la prédiction a posteriori M_1 est située entre la prédiction a priori M_0 et l'estimation empirique m. Il s'agit en effet d'une pondération entre l'information a priori (du passé) $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ et l'information contenue dans l'échantillon (du présent) $\{\mathbf{x}\}$, que nous allons préciser dans la suite. Or le plus souvent et surtout dans le cas de petits échantillons à risque statistique élevé, les deux moyennes M_0 et μ se trouvent "du même coté" de m. Donc, corriger m par un certain poids du passé permettra d'augmenter la fiabilité de l'analyse sans prélever un nombre plus important de pièces dans le lot considéré.

Pour compléter, l'écart-type prédictif S_1 sera aussi une moyenne pondérée des deux écart-types S_0 et s lorsque M_0 et m seront très voisines. Dans le cas contraire, S_1 sera égale à cette moyenne augmentée d'une quantité croissante avec l'écart entre M_0 et m, qui rend compte de l'incertitude résultant de deux estimations contradictoires.

La figure ci-dessous montre la variation des moyennes pour les 18 lots de la qualité HR2, de diamètre 18mm, tous de l'année 1988, ainsi que l'effet de la pondération par le passé représenté par la courbe en trait continu. Chacun des échantillons a une moyenne m_i et un écart-type s_i empiriques et l'intervalle représente la position de m_i avec les écars $\pm s_i$, i = 1, 2, ..., 18. Nous prenons les 10 premiers échantillons pour constituer l'information a priori, la prédiction par la méthode des échantillons équilibrés commence pour le lot n°11: $M_1 = M_{11}$. Ensuite, les lots n°2 à n°11 constituent l'information a priori pour l'étude du lot n°12 et nous calculons $M_1 = M_{12}$, et ainsi de suite jusqu'au lot n°18. Le processus est donc glissant.

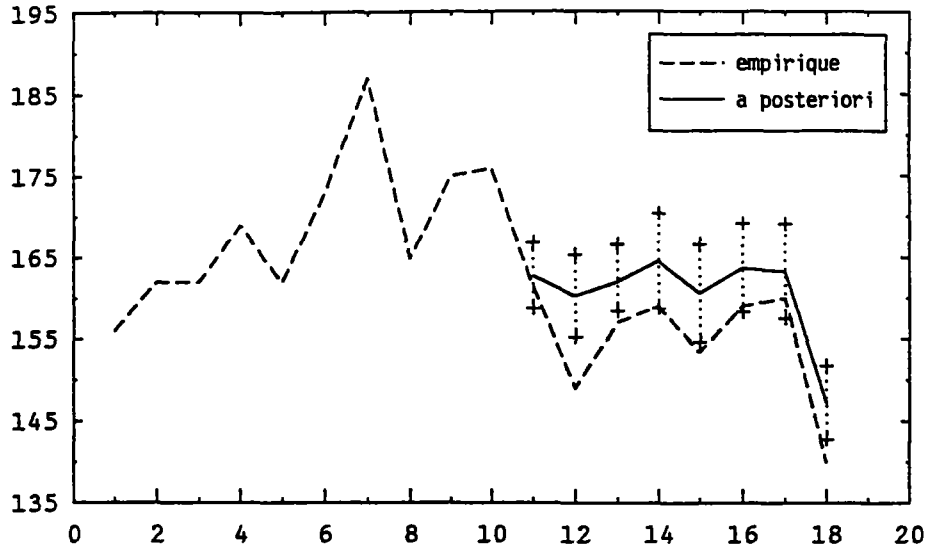


Figure 5.1 Prédictions empiriques et bayériennes

Outre cette "pondération par le passé", l'approche bayésienne permet, en cas d'absence d'échantillon prélevé dans un lot, d'utiliser, avec précautions, la distribution a priori évaluée à partir des n échantillons précédents comme une distribution prédictive $\mathbb{P}_0$, et d'adopter (M_0, S_0) comme une estimation du couple (μ,σ). Cette propriété peut permettre éventuellement de choisir un espacement convenable des séquences d'échantillonnage et une fréquence minimum des contrôles, en vue de réduire le volume global d'essais.

Ajoutons qu'à partir d'une distribution a priori $\mathbb{P}_0(\cdot) = \mathbb{P}(\cdot)$ donnée, nous pouvons calculer la distribution a posteriori $\mathbb{P}_1(\cdot) = \mathbb{P}_{\mathbf{x}}(\cdot)$ ou $= \mathbb{P}(\cdot|\mathbf{x})$ avec des échantillons de différentes tailles prélevés dans le lot. En général, si le résultat n'est pas satisfaisant avec un échantillon $\mathbf{x} = \mathbf{x}'$, nous reprendrons le calcul avec un échantillon "renforcé" $\mathbf{x} = \mathbf{x}' \cup \mathbf{x}''$, et ainsi de suite jusqu'à ce que le résultat semble suffisamment vraisemblable. Nous allons montrer au paragraphe 5.6.3 que pour obtenir une confirmation du résultat, il suffit d'un léger enrichissement. Inversement, si le résultat est assez bon et l'évolution du niveau de la qualité est lente, nous nous contenterons d'échantillons de taille initiale. La procédure bayésienne permettra ainsi d'économiser des essais en modulant leur nombre en fonction des résultats obtenus.

De manière imagée, nous voyons que lorsque la présomption est favorable

(a priori satisfaisant), et qu'elle est confirmée, même par un échantillon limité, l'approche bayésienne conduira à "faire confiance" au lot à tester. Si l'information provenant du lot contredit la présomption favorable, cette approche proposera d'enrichir la connaissance de ce lot par rééchantillonnage avant de le déclarer non-conforme. Inversement dans le cas d'un a priori défavorable, c'est à dire de valeur a priori proche de la limite d'acceptation, des valeurs empiriques défavorables conduisent à un rejet immédiat du lot, tandis que des valeurs empiriques favorables entraineront, le cas échéant, un rééchantillonnage de confirmation.

5.3. MISE EN OEUVRE DES MODELES BAYESIENS

5.3.1. Quelques modèles pour la distribution normale de K

Dans le cas des boulons, les coefficients de variation empiriques sont très faibles, pour les échantillons $\mathrm{Max}\{d_i = s_i/m_i\} < 0.050$ et pour l'ensemble des données regroupées DT = 0.068. Il en est de même du coefficient d'aplatissement C4 = 0.431 et du coefficient d'asymétrie C3 = 0.340 (cf. tableaux 5.4. et 5.5). Le paramètre d'aplatissement γ_2 rend compte de la concentration de la distribution autour du mode, ou de la moyenne dans le cas symétrique; la distribution normale étant prise comme référence : $\gamma_2 = 0$ et $\gamma_1 = 0$ (où les lettres greques désignent les valeurs théoriques).

Nous supposons donc, dans notre analyse, la normalité de la distribution des valeurs du coefficient K. La distribution normale est caractérisée par sa moyenne μ et son écart-type σ. Dans ce cas, il existe trois modèles de base, notés par N10, N01 et N11, qui prennent en compte respectivement la variablité de la moyenne $\mu = \bar{K}$, de l'écart-type σ (équivalent à la variance σ^2 ou à la précision $\eta = 1/(2\sigma^2)$), ou des deux ensemble (μ, σ). Dans chacun de ces modèles, nous étudions respectivement la moyenne $\theta = \mu$ avec un écart-type σ fixé, l'écart-type $\theta = \sigma$ avec une moyenne fixée μ, ou à la fois la moyenne et l'écart-type $\theta = (\mu, \sigma)$.

Prendre en compte la variation du coefficient K entre les lots, K étant considéré comme une variable aléatoire dont la valeur pour chaque boulon représente un tirage, revient ici à étudier les variations des paramètres μ

et σ. Dans l'approche classique, pour chaque lot, le couple (μ, σ) est considéré comme inconnu, mais déterministe, et estimé de manière empirique par la statistique (m, s) d'un échantillon **x**. Dans l'approche bayésienne au contraire, le couple (μ, σ) est considéré comme un vecteur aléatoire dont la connaissance dans chaque lot est exprimée par une distribution de probabilité. Si le lot était totalement connu (toutes les valeurs mesurées), cette distribution serait réduite au couple (m, s) (masse de Dirac). Dans le cas contraire, elle est d'autant plus "étalée" que la connaissance du lot est plus faible. Toutefois avant de commencer l'échantillonnage du lot, nous adoptons pour (μ, σ) une distribution a priori établie comme il suit. Précisons que dans le cas d'un écart-type σ fixé, étudier $\theta = (\mu, \sigma)$ revient à étudier $\theta = \mu$; et dans le cas d'une moyenne μ fixée, étudier $\theta = (\mu, \sigma)$ est équivalente à étudier $\theta = \sigma$.

Nous pouvons procéder à l'analyse statistique du coefficient K de deux manières : la première est de construire la distribution prédictive a posteriori à partir de laquelle nous prévoyons la moyenne et l'écart-type du lot; la deuxième est de faire une estimation bayésienne du paramètre θ considéré à partir de la distribution a posteriori $\pi_1(\theta) = \pi(\theta|\mathbf{x})$, avec $\theta = \mu$, σ, ou $\theta = (\mu, \sigma)$. Nous remarquons que si nous étudions la moyenne $\theta = \mu$, alors la moyenne prédictive du coefficient K est égale à la moyenne de la moyenne μ : $M_1 = m_1$ pour le modèle normal. Nous obtenons le même résultat pour le modèle normal-gamma où $\theta = (\mu, \sigma)$. Par contre pour l'écart-type σ, il y aura une petite différence entre ces deux approches puisque la distribution prédictive du coefficient K est une loi de Student $T(m_1, v_1, t_1)$ pour le modèle gamma ou normal-gamma (cf. paragraphes 3.1 et 5.5.3).

Nous n'étudions que trois modèles prédictifs du coefficient K puisque qu'ils permettent d'appliquer les critères du contrôle, qui sont liés généralement à un seuil inférieur ou supérieur ou à un ensemble correspondant aux fractiles de la distribution prédictive. Ils comprennent également des critères de type $M_1 \pm \lambda S_1$, plus généralement des intervalles de tolérance $[M_1 - \lambda_1 S_1, M_1 + {}_2 S_1]$ avec les risques de premiers espèces α_1 et α_2, tandis que l'estimation à partir des modèles $\pi_1(\theta)$ ne concerne que la moyenne de θ et son écart-type : $\mathbb{E}^{\pi}(\theta)$ et $\mathbb{S}^{\pi}(\theta)$.

Après les études des modèles descriptifs, nous proposerons au paragraphe 5.6. des modèles spécifiques pour résoudre le problème des boulons HR.

5.3.2. Prise en compte du passé

Rappelons que dans la procédure bayésienne, le premier point important est d'évaluer les variabilités des paramètres μ et σ par une distribution de probabilité $\pi(\theta)$ avec $\theta = \mu$, σ ou (μ,σ) à partir de l'information a priori. Les restes sont les conséquences de la méthodologie bayésienne, en utilisant la théorie de Bayes et la théorie de probabilité conditionnelle.

Cette évaluation ne pourra être réalisée qu'à partir de l'expérience professionnelle, d'opinions d'experts, et d'observations acquises etc. Grâce à la continuité de la production des boulons HR, l'information a priori est constituée ici des "n" échantillons antérieurs de même qualité pendant une certaine période convenable, liée à une fréquence d'échantillonnage, c'est à dire à la date d'échantillonnage et à la taille des échantillons.

La population de ces échantillons varie peu au-dessus de 5. Nous avons précisé aux chapitres précédents que, afin que l'information a priori basée sur les n échantillons antérieurs ne pèse pas trop par rapport à celle du lot considéré, il faut limiter le nombre n, donc la durée de la période correspondante. Nous prenons en pratique n constant, ce qui rend le processus glissant.

La construction de la distribution a priori se fait à l'aide d'une des méthodes développées au chapitre II :

— méthode séquentielle, notée par SS,

— méthode des échantillons équilibrés, notée par EE,

— méthode des moments, notée par OO,

— méthode des doubles moments, notée par MM,

— méthode des doubles maxima de vraisemblance, notée par VV.

Au stade actuel de notre analyse, nous accordons un poids identique aux n échantillons antérieurs, donc aux lots correspondants. Toutefois dans le cas où cette information a priori mélange des lots de la production en cours et des lots des productions passées, nous pourrons introduire des coefficients et la méthode des scores d'expert permet d'obtenir une pondération spécifique des lots pris en compte (cf. paragraphe 2.2).

L'application numérique est faite dans le cas des boulons de la qualité HR2, diamètre 18mm pour l'année 1988 (cf. tableau 5.4). L'histrogramme 5.2 contient la totalité de pièces contrôlées réparties sur l'intervalle [135,190]. L'annexe A.3 du rapport donne les histrogrammes pour les qualités HR1, HR1Z et HR2Z, diamètre 18mm pour la même année.

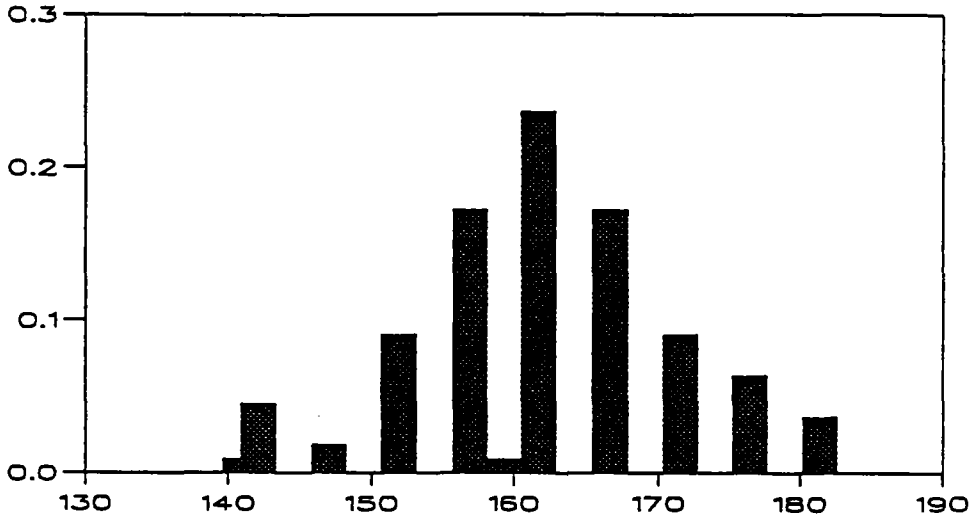


Figure 5.2 Histrogramme des valeurs K de la qualité HR2 (1988)

A titre d'illustration, les statistiques de toutes les mesures mélangées sont données dans le tableau 5.5 et elles sont définies par :

$KT = \sum_i k_i$ — taille totale

$MT = \frac{1}{KT}\sum_i \sum_j x_j^i$ — moyenne totale

$VMT = \frac{1}{KT}\sum_i \sum_j (x_j^i - MT)^2$ — variance totale

$DT = \frac{\sqrt{VMT}}{MT}$ — coefficient de variation total

$C3 = V3/VMT^{3/2}$ — coefficient d'asymétrie total

$C4 = V4/VMT^2 - 3$ — coefficient d'aplatissement

$Max = Max\{x_j^i\}$ — valeur maximale

$Min = Min\{x_j^i\}$ — valeur minimale

avec $V3 = \frac{1}{KT}\sum_i \sum_j (x_j^i - MT)^3$ et $V4 = \frac{1}{KT}\sum_i \sum_j (x_j^i - MT)^4$.

KT	MT×10³	VMT×10⁶	ET×10³	DT	C3	C4	Max×10³	Min×10³
110	161.755	122.185	11.054	.068	.340	.431	135.000	190.000

Tableau 5.5 Information sur l'ensemble des données de HR2 (1988)

Pour étudier l'influence de la distribution a priori $\pi_0(\theta) = \pi(\theta)$ et de l'échantillon $\mathbf{x}$ sur la distribution posteriori $\pi_1(\theta) = \pi(\cdot|\mathbf{x})$ et sur la distribution prédictive a posteriori de densité $p_1(\cdot) = p_{\mathbf{x}}(\cdot)$, nous allons étudier deux échantillons n°17 et n°18 ; le premier constitue un lot "moyen" conforme aux prédictions a priori et le second un lot "extrême" avec des valeurs faibles (m,s). Les 10 échantillons précédents constituent l'information a priori du lot à étudier.

Désignons par (MM,SM) les moyennes des moyennes $\{m_i\}$ et des écart-types $\{s_i\}$ des n = 10 échantillons précédents. Nous avons ici :

— pour le lot n°17 : (MM, SM) = (164.18, 5.040)
(m, s) = (160.00, 6.325) (lot "moyen") ;

— pour le lot n°18 : (MM, SM) = (161.48, 5.272)
(m, s) = (140.00, 3.162) (lot "extrême").

Les valeurs mesurées du coefficient K sont multipliées par 10^3, ainsi qu'elles sont relevées et consignées dans les cahiers de mesures. Ceci ne modifie pas nos conclusions grâce à la linéarité des opérations dans l'hypothèse de la normalité de la population faite ici, à l'exception de l'estimateur des doubles maxima de vraisemblance VV. Mais dans le cas des boulons HR, nous verrons dans la suite que nous choisirons plutôt les méthodes des échantillons équilibrés et des doubles moments, qui affectent des poids équilibres aux informations du passé et du présent, tandis que la méthode des doubles maxima de vraisemblance donnera un poids relativement faible au passé par rapport au présent.

	Lot n°	Nombre d'essais k	Moyenne $m \times 10^3$	Ecart-type $s \times 10^3$	Précision $h \times 10^{-3}$	Coef. de variation $d = \frac{s}{m}$
	7	5	187.00	4.000	31.250	0.0214
	8	5	165.00	3.162	50.000	0.0192
	9	5	175.00	5.477	16.667	0.0313
	10	5	176.00	6.633	11.364	0.0377
	11	10	161.50	3.202	48.780	0.0198
	12	5	149.00	4.899	20.833	0.0329
	13	5	157.00	4.000	31.250	0.0255
	14	5	159.00	6.633	11.364	0.0417
	15	10	153.30	6.558	11.625	0.0428
	16	5	159.00	5.831	14.706	0.0367
Moyenne		6.1	164.18	5.040	24.784	0.0309
Ecart-type		2.1	11.13	1.315	14.184	0.0085
Lot à tester	17	5	160.00	6.325	12.500	0.0395

Tableau 5.6.1 Valeurs K dans les 10 lots n°7 à n°16 de l'année 1988 (HR2)

	Lot n°	Nombre d'essais k	Moyenne $m \times 10^3$	Ecart-type $s \times 10^3$	Précision $h \times 10^{-3}$	Coef. de variation $d = \frac{s}{m}$
	8	5	165.00	3.162	50.000	0.0192
	9	5	175.00	5.477	16.667	0.0313
	10	5	176.00	6.633	11.364	0.0377
	11	10	161.50	3.202	48.780	0.0198
	12	5	149.00	4.899	20.833	0.0329
	13	5	157.00	4.000	31.250	0.0255
	14	5	159.00	6.633	11.364	0.0417
	15	10	153.30	6.558	11.625	0.0428
	16	5	159.00	5.831	14.706	0.0367
	17	5	160.00	6.325	12.500	0.0395
Moyenne		6.1	161.48	5.272	22.909	0.0327
Ecart-type		2.1	8.14	1.316	14.442	0.0082
Lot à tester	18	5	140.00	3.162	50.000	0.0226

Tableau 5.6.2 Valeurs K dans les 10 lots n°8 à n°17 de l'année 1988 (HR2)

5.4. RESULTATS PREDICTIFS ET COMPARAISON DE MODELES

Nous allons étudier les résultats prédictifs a posteriori des lots n°17 et n°18 en utilisant trois modèles N10, N01 et N11 avec les cinq différentes méthodes de détermination de la distribution a priori. Cette étude permet de visualiser la sensibilité des estimateurs, de définir leur usage selon les circonstances de l'analyse.

Nous précisons d'abord que l'information contenue dans un échantillon $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ peut être remplacée par un triplet $(m, v=s^2, k)$, appelé statistique exhaustive de l'échantillon $\mathbf{x}$ (cf. paragraphes 1.1 et 3.1). Ceci signifie vis à vis de l'estimation de la distribution normale de la variable étudiée, ce triplet contient toute l'information (nécessaire et suffisante) apportée par un échantillon complet. Donc, l'ensemble des échantillons $\{\mathbf{x}_i;\ i = 1,2,\ldots n,\}$ peut être réduit à l'ensemble des triplets $\{(m_i, v_i, k_i)\ ;\ i = 1,2,\ldots,n\}$. Nous prenons dans la suite comme estimateur de la variance celui de variance minimale, soit $v = \frac{1}{k}\sum(x_j-m)^2$, éventuellement biaisé. Nous remarquons que dans notre analyse, il donne un résultat a posteriori proche de celui à l'estimateur sans biais $v' = \frac{1}{k-1}\sum(x_j-m)^2$ et fournit des formules plus légères (cf. paragraphes 4.5 et 4.6).

Modèle	Distribution π_i	Distribution p_i	M_i	S_i
N11	Normale -Gamma $NG(m_i, v_i, k_i)$	Student $T(m_i, v_i, k_i+1)$	m_i	$\sqrt{\frac{k_i+1}{k_i-1}v_i}$
N10	Normale $N(m_i, v_i)$	Normale $N(m_i, v_i+v)$	m_i	$\sqrt{v_i+v}$
N01	Gamma $G(a_i, b_i)$	Student $T(\mu, \frac{b_i}{2a_i-}, 2a_i)$	m	$\sqrt{\frac{b_i}{2(a_i-1)}}$

Tableau 5.7 Modèles bayésiens

Nous résumons ensuite dans le tableau 5.7 les distributions a priori π_0 et a posteriori π_i, prédictives a priori p_0 et a posteriori p_i associées aux

trois modèles, et les formules pour calculer la moyenne M_i et l'écart-type S_i prédictifs en fonction des paramètres m_i, v_i et k_i des distributions soit a priori i = 0, soit a posteriori i = 1. Dans ce tableau, les paramètres a_i et b_i peuvent être exprimés en fonction de (m_i, v_i, k_i) de manière à ce que : $a_i = k_i/2 + 1$ et $b_i = k_i\left(v_i + (\mu - m_i)^2\right)$ (cf. paragraphe 3.1.3).

Le résultat prédictif a posteriori exprimés à l'aide de (M_1, S_1) sont donnés dans les deux tableaux 5.8.1 et 5.8.2, dans lesquels nous indiquons :

— les trois modèles N10, N01 et N11 pour la distribution normale de K, correspondant respectievement aux cas de la prise en compte de la variabilité de μ, σ ou (μ, σ);

— les méthodes appliquées pour déterminer la distribution a priori;

— la moyenne et l'écart-type empiriques du lot (m, s), sur lesquels est basée l'approche classique, afin de comparer avec l'approche bayésienne;

— les moyennes des moyennes et des écart-types des 10 échantillons précédents.

Au vu de ces deux tableaux, nous donnons quelques remarques générales suivantes.

5.4.1. Ecart-type prédictif

Pour le modèle N11, les méthodes de détermination de la distribution a priori donnent en général un écart-type prédictif S_1 nettement supérieur à l'écart-type empirique s, pour le lot moyen ou le lot extrême. En effet, nous avons :

$$v_1 = \frac{k_0}{k_1} v_0 + \frac{k}{k_1} v + \frac{k_0 k}{k_1 k_1} (m - m_0)^2 \qquad (5.41)$$

D'après les relations entre v_i et V_i, nous pouvons dire que la variance prédictive (a posteriori) $V_1 = S_1^2$ est à peu près une pondération entre la variance prédictive a priori $V_0 = S_0^2$ et la variance empirique v, augmentée d'une quantité croissante avec le carré de l'écart entre la moyenne a priori

(MM,SM) = (164.180 , 5.040) et (m,s) = (160.000 , 6.325)						
Moyennes et Ecart-types Prédictifs (M_1, S_1)						
	N11 — (μ, σ^2)	N10 — μ		N01 — σ^2		
		$\sigma_1^2 = v_1$	$\sigma^2 =$ cte		$\mu_1 = m_1$	μ = MT
SS	162.815 11.69	163.639 6.35	162.887 6.36		5.30	11.51
EE	162.200 11.15	163.252 6.52	162.600 6.60		5.76	10.17
OO	—	—	163.050 0	160.000	—	—
MM	162.112 6.66	162.943 6.51	160.253 6.89		5.96	8.34
VV	160.156 8.07	160.307 6.89	idem		5.92	8.32

Tableau 5.8.1 Prédiction après la réception du lot n°17

(MM,SM) = (161.480 , 5.272) et (m,s) = (140.000 , 3.162)						
Moyennes et Ecart-types Prédictifs (M_1, S_1)						
	N11 — (μ, σ^2)	N10 — μ		N01 — σ^2		
		$\sigma_1^2 = v_1$	$\sigma^2 =$ cte		$\mu_1 = m_1$	μ = MT
SS	159.200 10.84	158.007 3.21	156.743 3.22		5.25	9.46
EE	151.305 14.35	147.230 3.37	145.873 3.38		4.50	16.14
OO	157.138 11.91	—	160.800 0	140.000	—	12.41
MM	149.590 12.83	147.620 3.36	140.629 3.46		4.46	21.16
VV	141.681 8.36	140.865 3.45	idem		4.48	20.99

Tableau 5.8.2 Prédiction après la réception du lot n°18

— ce symbole signifie que la méthode n'est pas valable à cause de la négativité de C4 (cf. chapitre IV).

— ce symbole signifie que la méthode n'existe pas pour le modèle conserné.

$M_0 = m_0$ et la moyenne empirique m. Donc, nous en déduisons que ce modèle N11 sera adapté au cas où le contrôle demande une position de référence M_0, qui dépend des n échantillons précédents.

Pour le modèle N01 avec μ fixée, nous obtenons la même conclusion que précédemment, puisque l'écart entre la position μ (μ = MT par exemple) et la moyenne empirique m est pris en compte (cf. tableaux 4.5 et 4.6).

Si la moyenne μ pour le modèle N01 est considérée comme connue, mais variable d'un lot à l'autre, nous prenons l'estimation empirique soit $\mu = m_1$ pour chacun des échantillons et dans ce cas, la variance prédictive a posteriori de σ^2 $\mathbb{E}(\sigma^2) = v_1$ est une pondération entre la variance a priori $\mathbb{E}(\sigma^2) = v_0$ et la variance empirique v :

$$v_1 = \frac{k_0}{k_1} v_0 + \frac{k}{k_1} v \tag{5.42}$$

Avec le modèle N10, à l'exception de la méthode des moments 00, les autres méthodes donnent des écart-types quasiment identiques pour le lot moyen n°17 ou extrême n°18. En effet, la variance de la moyenne $v_1 = \mathbb{V}(\mu)$ est inférieure à v/k, et a moins d'influence sur la variance prédictive $V_1 = v_1 + v$ que la variance empirique v (cf. tableaux 4.3 et 4.4).

5.4.2. Moyenne prédictive

Pour une méthode quelconque utilisée, $M_1 = m_1$ est toujours située entre $M_0 = m_0$ et m, grâce à la formule de pondération des modèles N11 et N10 :

$$m_1 = \frac{k_0}{k_1} m_0 + \frac{k}{k_1} m \tag{5.43}$$

$$m_1 = \frac{(1/v_0) m_0 + (k/v) m}{1/v_0 + k/v} \tag{5.44}$$

où les poids du passé et du présent sont représentés respectivement par (m_0, k_0) et (m,k) pour le modèle N11, et par (m_0, $1/v_0$) et (m,k/v) pour le modèle N10.

Nous avons appris au paragraphe 4.6 que v_0 peut être représenté sous forme $v_0 = \sigma_0^2/k_0$. La formule de pondération pour le modèle N10 devient :

$$m_1 = \frac{(k_0/\sigma_0^2)m_0 + (k/v)m}{k_0/v_0 + k/v}$$

Pour la méthode VV, ce paramètre v_0 dépend de la dispersion des moyennes et pour d'autres méthodes, le paramètre σ_0^2 est une sorte de moyenne des variances. Comme le poids de l'information a priori est proportionnel à $1/v_0$ ou de k_0/σ_0^2, et la dispersion des moyennes est supérieure à la moyenne des variances (cf. tableaux 5.6.1 et 5.6.2), l'estimateur prédictif obtenu par la méthode VV est donc très sensible pour le lot moyen ou extrême.

La méthode séquentielle SS donne une taille cumulée $k_0 = \sum k_i$ plus élevée que les autres méthodes. La correction par l'échantillon **x** est donc faible et l'estimateur prédictif est moins sensible que les autres estimateurs pour les modèles N10 et N11.

Le modèle N01 n'étudie que la variabilité de l'écart-type et nous avons donc toujours $M_1 = m$, la moyenne empirique. Les moyennes du coefficient K sont considérées comme connues soit les moyennes empiriques soit constantes. Dans ce dernier cas, nous prendrons par exemple $M_0 = MM = \frac{1}{n}\sum m_i$ la moyenne des moyennes des échantillons antérieurs, ou $M_0 = MT$ la moyenne totale de toutes les mesures mélangées ; ceci est aussi la moyenne des moyennes pondérées par les tailles associées : $MT = \frac{1}{KT}\sum_{i=1}^{n} k_i m_i$ avec $KT = \sum k_i$ la taille totale.

D'après ces remarques, nous constatons que les estimateurs bayésiens sont une pondération entre les informations du passé et du présent, dont les influences dépendent principalement des paramètres de "taille" : k_0 et k. Ainsi, il conviendra de limiter k_0 ou d'augmenter le nombre de prélèvement k tels que le rapport k/k_0 ne soit pas trop petit, afin que l'information a priori ne pèse pas trop par rapport à l'information provenant du lot. Surtout pour la méthode séquentielle SS, nous avons $k_0 = n \cdot KM \approx 5n$. Limiter le poids du passé, donc le paramètre k_0, revient à limiter le nombre n, donc la durée de la période correspondante.

Quant à l'information provenant du lot à tester, plus la taille k de prélèvements est grande, plus le résultat s'approche à la réalité (μ,σ^2). En effet, les résultats (M_1, V_1) des modèles N11, N10 et N01 avec $\mu_1 = m_1$ et (m,v) se convergent lorsque la taille est grande, donc $k_1 = k_0 + k$ est grande. Alors, les distributions bayésiennes correspondantes $T(m_1, v_1, k_1+1)$, $N(m_1, v_1)$ et $T(m_1, c_1^2 = b_1/(2a_1), t_1 = 2a_1 = k_0 + k + 1)$ s'approchent respectivement de la distribution empirique $N(m,v)$, du fait que la loi de Student $T(m,v,t)$ devient asymptotiquement une loi normale $N(m,v)$, lorsque le degré de liberté t est assez grand. Mais pour la distribution bayésienne du modèle N01 avec μ considérée comme constante, la variance a toujours une quantité supplémentaire de l'écart $(m-\mu)$ pour la taille k soit grande soit petite (cf. tableau 4.6).

La deuxième conclusion que nous pouvons tirer de toutes ces remarques est que le modèle N11 et le modèle N01 avec μ constante seront appliqués aux problèmes pour lesquels on exige une valeur cible M_0, qui, étant en quelque sorte une moyenne des moyennes, est mobile. Comme dans le cas du contrôle des boulons HR, nous ne nous intéressons pour le moment ni à une référence de la moyenne, ni à l'écart entre la moyenne du lot et cette valeur, l'étude du modèle N11 et du modèle N01 avec μ constante n'est utile que pour la justification des modèles.

5.5. COMPARAISON DES METHODES

5.5.1. Etude de la moyenne et la variance ensemble

Rappelons que si X est une variable aléatoire de distribution normale, et si l'on étudie à la fois les paramètres de la moyenne et de l'écart-type $(\mu, \sigma,)$, alors la distribution du couple $(\mu, \eta = 1/(2\sigma^2))$ est normale-gamma, $NG(m,v,k)$. En combinant la variabilité de $\theta = (\mu,\eta)$, la distribution prédictive de X est une loi de Student T(m,v,k+1), cela veut dire que la variable $Y = (X-m)/\sqrt{v}$ suit une loi de Student centrée T(k+1).

Nous réalisons, dans les tableaux 5.9.1 et 5.9.2, le passage de la distribution a priori π_0 à la distribution a posteriori π_1 en transformant leurs paramètres $(m_0, v_0, k_0) \longrightarrow (m_1, v_1, k_1)$ par différentes méthodes.

Etudier l'influence de la distribution prédictive a priori p_0 sur la distribution prédictive a posteriori p_1 revient en effet à étudier celle de π_0 sur π_1, du fait que $M_i = m_i$ et $V_i = v_i \frac{k_i + 1}{k_i - 1}$, $i = 0, 1$. Cette étude permettra de distinguer ces méthodes afin de mieux les utiliser.

Rappelons que nous avons résumé et comparé les 5 estimateurs du modèle normal-gamma dans le tableau 4.2. Avec l'application numérique réalisée ici, nous confirmons les résultats obtenus.

	Paramètres des distributions a priori et a posteriori								
N11	A priori (m_0, v_0, k_0)			+ Lot (m, v, k) ==>			A posteriori (m_1, v_1, k_1)		
SS	163.05	139.51	60.00				162.82	132.52	65.00
EE	164.18	151.07	5.56				162.20	102.81	10.56
OO	—			160.00	40.00	5	—		
MM	164.18	24.13	5.11				162.11	36.35	10.11
VV	164.73	138.77	0.17				160.16	43.96	5.17

Tableau 5.9.1 Passage de l'a priori à l'a posteriori pour le lot n°17

	Paramètres des distributions a priori et a posteriori								
N11	A priori (m_0, v_0, k_0)			+ Lot (m, v, k) ==>			A posteriori (m_1, v_1, k_1)		
SS	160.80	89.43	60.00				159.20	114.04	65.00
EE	161.48	95.85	5.56				151.31	170.21	10.56
OO	160.80	82.10	23.40	140.00	10.00	5	157.14	132.16	28.40
MM	161.48	27.24	4.03				149.59	131.72	9.03
VV	161.43	73.10	0.43				141.68	48.16	5.43

Tableau 5.9.2 Passage de l'a priori à l'a posteriori pour le lot n°18

La méthode séquentielle SS et la méthode des moments OO utilisent de façon directe et indirecte l'ensemble des mesures mélangées. L'estimateur de la moyenne a priori est $m_0 = MT$, la moyenne totale, et celui de la variance a priori v_0 dépend de la variance totale VMT et du paramètre k_0 :

$$VMT = \frac{1}{KT}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\left(x_j^i - MT\right)^2 = \frac{1}{KT}\sum_{i=1}^{n} k_i\left(v_i + (m_i - MT)^2\right) \qquad (5.51)$$

avec $k_0 = KT = \sum k_i$ pour la méthode SS et $k_0 = 3(1+2/C4)$ pour la méthode OO.

Le paramètre k_0 est relativement grand, puisque pour la méthode OO, le coefficient d'aplatissement C4 est souvent très proche de 0. La correction par l'information provenant du lot est donc faible. Comme les statistiques MT, VMT, KT et C4 sont calculées à partir de l'ensemble des mesures mélangées, plus la taille d'un échantillon est grande, plus cet échantillon prend un poids important dans l'évaluation de la distribution a priori.

Par contre, les méthodes des échantillons équilibrés EE et des doubles moments MM donnent aussi le même estimateur $m_0 = MM$, la moyenne des moyennes. Leur différence est calculée par les estimateurs v_0 et k_0:

pour EE :

$$v_0 = VMM = \frac{1}{n}\sum\left(v_i + (m_i - MM)^2\right)$$

$$k_0 = KH = \left(\frac{1}{n}\sum k_i^{-1}\right)^{-1}$$

pour MM :

$$v_0 = VH\frac{k_0 + 1}{k_0} \qquad (5.52)$$

$$k_0 = 2\frac{HM^2}{HV} - 1$$

où HM et HV est la moyenne et la variance des précisions $\{h_i = (2v_i)^{-1}\}$ et VH est la moyenne harmonique des variances : $VH = \left(\frac{1}{n}\sum v_i^{-1}\right)^{-1} = 1/(2HM)$.

Les deux valeurs du paramètre v_0 sont éloignées, puisque la méthode EE donne comme estimateur la somme de la moyenne des variances $VM = \frac{1}{n}\sum v_i$ et de la variance des moyennes $MV = \frac{1}{n}\sum(m_i - MM)^2$ soit VMM = VM + MV, tandis que la méthode MM donne à une constante près la moyenne harmonique des variances VH. Le paramètre k_0 ne dépend que de la dispersion des précisons.

Par la méthode des doubles maxima de vraisemblance VV, l'estimateur des

paramètres est le suivant :

$$m_0 = MHM \qquad v_0 = VH\frac{k_0 + 1}{k_0}$$

$$k_0^{-1} \approx \log(HM) - HL + 2\left(MH4 - MHM^2 HM\right) = \log(HM) - HL + \frac{Vm}{VH} \tag{5.53}$$

où la statistique MHM est la moyenne des moyennes pondérées par les précisions ou par l'inverse des variances.

Dans le cas des boulons HR, le paramètre k_0 est relativement petit, parce qu'il prend en compte en même temps la dispersion de la logarithme des précisions et le rapport entre la variance des moyennes pondérées par les précisions Vm et la variance harmonique VH (cf. paragraphe 4.5.3). Ce rapport Vm/VH est de l'ordre du carré de la moyenne.

Dans cette méthode, il n'y a pas linéarité des opérations avec l'hypothèse de la distribution normale de la variable étudiée. Si l'analyse est effectuée à partir des données originales : $X \times 10^{-3}$, alors la valeur k_0 sera plus grande, mais la différence entre les deux méthodes MM et VV, dérivées toutes les deux des méthodes du quasi-échantillon, sera réduite dans notre cas. Ainsi, nous choisissons ici la méthode MM.

Remarquons également que la variance prédictive a priori $V_0 = v_0 \frac{k_0 + 1}{k_0 - 1}$ n'existe pas lorsque k_0 est inférieur à 1.

5.5.2. Etude de la moyenne

Nous étudions maintenant la variable μ de la moyenne du coefficient K, dont la distribution est normale $N(m, v)$. Cependant, l'évaluation de la loi distribution a priori de μ ne dépend pas uniquement de l'ensemble des moyennes des échantillons antérieurs, mais l'information a priori est aussi constituée par l'ensemble des triplets $\{(m_i, v_i, k_i)\ ;\ i = 1, 2, \ldots, n\}$. Nous distinguons dans ce modèle normal deux cas suivants :

(a) la variance varie d'un lot à l'autre,

(b) elle est considérée comme constante,

dans ce premier cas, nous prenons l'estimation empirique de la variance.

Après la réception d'un lot n°17 ou n°18, la distribution de la moyenne μ est modifiée par la statistique (m,v,k) de l'échantillon prélevé dans ce lot; ceci est toujours une loi de moyenne m_1 et de variance v_1 :

$$\begin{cases} m_1 = \dfrac{k\dfrac{m}{v} + \dfrac{m_0}{v_0}}{\dfrac{k}{v} + \dfrac{1}{v_0}} \\ v_1 = \dfrac{1}{\dfrac{k}{v} + \dfrac{1}{v_0}} \end{cases} \Longleftrightarrow \begin{cases} m_1 = \dfrac{khm + h_0 m_0}{kh + h_0} \\ v_1 = \dfrac{1}{kh + h_0} \end{cases} \qquad (5.54)$$

où $h_0 = 1/(2v_0)$ et $h = 1/(2v)$ sont les paramètres de précision.

Nous en déduissons que, pour la moyenne prédictive $M_1 = m_1$, les poids des informations du passé et du présent sont respectivement représentés par $1/v_0$ et k/v ou par h_0 et kh. Si la variance s'écrit sous forme $v_0 = \sigma^2/k_0$, le rapport des poids du passé et du présent est

$$\tau = \frac{h_0}{kh} = \frac{v/\sigma^2}{k/k_0} \qquad (5.55)$$

Les résultats des modèles N10 correspondant aux deux cas (a) et (b), sont donnés respectivement dans les tableaux 5.11.1 et 5.12.1 pour le lot n°17 et dans les tableaux 5.11.2 et 5.12.2 pour le lot n°18.

N10	Paramètres des distributions a priori et a posteriori					
$\sigma_1^2 = v_1$	A priori (m_0, v_0)		+ Lot (m,v,k) ==>		A posteriori (m_1, v_1)	
SS	163.786	0.324	160.000	40.00 5	163.639	0.312
EE	164.728	3.631			163.252	2.498
MM	164.180	3.362			162.943	2.367
VV	164.728	115.435			160.307	7.482

Tableau 5.11.1 Passage de l'a priori à l'a posteriori avec le lot n°17

N10	Paramètres des distributions a priori et a posteriori						
σ^2=cte	A priori (m_0, v_0)		+ Lot (m,v,k) ==>			A posteriori (m_1, v_1)	
SS	163.050	0.451				162.887	0.427
EE	164.180	4.883	160.000	40.00	5	162.600	3.024
MM=VV	164.480	123.942				160.253	7.515

Tableau 5.12.1 Passage de l'a priori à l'a posteriori avec le lot n°17

N10	Paramètres des distributions a priori et a posteriori						
$\sigma_1^2 = v_1$	A priori (m_0, v_0)		+ Lot (m,v,k) ==>			A posteriori (m_1, v_1)	
SS	161.117	0.345				158.007	0.295
EE	161.432	3.929				147.230	1.325
MM	161.480	3.638	140.000	10.00	5	147.620	1.290
VV	161.432	47.575				140.865	1.919

Tableau 5.11.2 Passage de l'a priori à l'a posteriori avec le lot n°18

N10	Paramètres des distributions a priori et a posteriori						
σ^2=cte	A priori (m_0, v_0)		+ Lot (m,v,k) ==>			A posteriori (m_1, v_1)	
SS	160.800	0.484				156.743	0.390
EE	161.480	5.315	140.000	10.00	5	145.873	1.453
MM=VV	161.480	66.324				140.629	1.941

Tableau 5.12.2 Passage de l'a priori à l'a posteriori avec le lot n°18

Remarquons que la méthode des moments OO n'est pas adaptée à ce modèle. En effet, nous avons supposé que la variance σ^2 est connue ; or comme cette

méthode est basée sur l'ensemble des mesures mélangées, nous prenons donc pour σ^2 la variance totale VMT et l'estimateur de la variance de la moyenne μ est $v_0 = \mathbb{V}(\mu) = VMT - \sigma^2 = 0$.

Nous constatons que pour la méthode SS, la correction par l'information du présent (m,v,k) est moins importante, à cause du paramètre de taille k_0 = KT. Les rapports des poids du passé KT/VHT ou KT/VT et du présent k/v sont bien supérieurs aux rapports obtenus par d'autres méthodes pour le lot moyen ou extrême. Les statistiques VT et VHT sont respectivement la moyenne et la moyenne harmonique des variances $\{v_i\}$ pondérées par les tailles $\{k_i\}$.

cas	méthode	rapport τ	n°17	n°18
(a)	SS	$\frac{KT}{VHT}/\frac{k}{v}$	24.67	5.78
	EE	$\frac{KH}{VH}/\frac{k}{v}$	2.20	0.51
	MM	$\frac{KM}{VH}/\frac{k}{v}$	2.38	0.55
	VV	$\frac{1}{Vm}/\frac{k}{v}$	0.07	0.04
(b)	SS	$\frac{KT}{VT}/\frac{k}{v}$	17.75	4.13
	EE	$\frac{KH}{VM}/\frac{k}{v}$	1.64	0.38
	MM = VV	$\frac{1}{MV}/\frac{k}{v}$	0.06	0.03

Tableau 5.13 Rapport des poids du passé et du présent

Dans le cas (a), les méthodes EE et MM donnent des résultats très proches l'un de l'autre pour le lot moyen ou extrême. Les estimateurs de la moyenne $\mathbb{E}(\mu)$ et de la variance $\mathbb{V}(\mu)$ sont : (m_0, v_0) = (MHM, VH/KH) et (MM, VH/KM). Dans le cas des boulons HR, nous avons en effet des tailles peu variables et la variation des précisions (ou variances) est aussi faible. Les résultats a priori sont quasiment les mêmes :

— la moyenne des moyennes est presque égale à la moyenne des moyennes pondérées par les précisions : MHM ≈ MM ;

— la moyenne harmonique des tailles est proche de la moyenne des tailles : k_0= KH = 5.56 pour la méthode EE, et k_0= KM = 6 pour la méthode MM, donc les variances v_0= VH/k_0 sont très voisines

où VH est la moyenne harmonique des variances. Ainsi, les prédictions a posteriori sont très voisins et les poids du passé et du présent sont à peu près équilibrés (VH,k_0) et (v,k), par ces deux méthodes et pour les deux cas du modèle normal, où k = 5 est la taille de l'échantillon.

Dans les deux cas (a) et (b), les résultats obtenus par la méthode EE sont très proches. Comme l'estimateur de la moyenne $\mathbb{E}(\mu)$ et la variance $\mathbb{V}(\mu)$ est : (m_0,v_0) = (MHM, VH/KH) et (MM, VM/KH), l'écart entre les deux moyennes VM et VH n'est pas grand en général dans le cas pratique.

Nous constatons que pour les deux cas, les résultats prédictifs obtenus par la méthode VV sont aussi très voisins. Nous avons montré que les estimateurs de la moyenne $\mathbb{E}(\mu)$ et de la variance $\mathbb{V}(\mu)$ sont: (m_0,v_0) = (MHM, Vm) et (MM, MV), donc le poids du passé v_0^{-1} est l'inverse de la variance des moyennes: 1/MV dans le cas (b), et de la variance des moyennes pondérées par les précisions : 1/Vm pour le cas (a). Ces deux variances des moyennes MV et Vm se comportent de la même manière, lorsque la variation des précisions est plus faible que celle des moyennes dans le cas des boulons HR. Au contraire avec les autres méthodes, le rapport des poids du passé et du présent est très grand : τ^{-1}= v_0(k/v). Nous avons en effet v/k = 8.00 et 2.00 pour les deux lots, mais v_0= Vm = 115.435 et v_0= MV = 123.942 pour le lot n°17, Vm = 47.575 et MV = 66.324 pour le lot n°18. Ainsi, la correction par l'information du présent est beaucoup plus importante, de sorte que le résultat prédictif est très proche du résultat empirique : M_1= m_1≈ m.

D'après ces remarques, nous pouvons conclure que dans l'étude du coefficient K des boulons HR, il suffit de considérer le premier cas (a), où la variance σ^2 est considérée comme connue et variable d'un lot à l'autre.

Pour le coefficient K, la moyenne et la variance prédictives a priori avant la réception d'un lot sont : (M_0,V_0) = (m_0,v_0+ σ^2) (cf. tableau 5.7).

Dans le cas de l'estimation empirique, σ^2 peut être l'une des valeurs VHT, VT, VH et VM, ... etc. Nous obtenons alors $v_0 = \sigma^2/k_0$ et $V_0 = S_0^2 = \sigma^2(1+1/k_0)$.

La distribution prédictive (a posteriori) du coefficient K pout le lot a pour moyenne et variance $(M_1, V_1) = (m_1, v_1 + v)$ avec $V_1 = S_1^2$ et $v = s^2$. Nous en déduisons que la variance prédictive V_1 est toujours supérieure à la variance empirique v, la variance de la moyenne $v_1 = V(\mu)$ étant une quantité supplémentaire.

Autrement dit, la variance prédictive intègre non seulement la variation du coefficient K à l'intérieur du lot mais aussi la variation entre les lots.

N10	Moyenne et Ecart-type Prédictifs					
$\sigma_1^2 = v_1$	A priori (M_0, S_0)		+ Lot (m,s,k) ==>		A posteriori (M_1, S_1)	
SS	163.786	4.448	160.000	6.325 5	163.639	6.349
EE	164.728	4.879			163.252	6.519
MM	164.180	4.851			162.943	6.509
VV	164.728	11.645			160.307	6.891

Tableau 5.14.1 Prédiction avant et après la réception du lot n°17

N10	Moyenne et Ecart-type Prédictifs					
$\sigma_1^2 = v_1$	A priori (M_0, S_0)		+ Lot (m,s,k) ==>		A posteriori (M_1, S_1)	
SS	161.117	4.590	140.000	3.162 5	158.007	3.209
EE	161.432	5.075			147.230	3.365
MM	161.480	5.046			147.620	3.360
VV	161.432	8.331			140.865	3.452

Tableau 5.14.2 Prédiction avant et après la réception du lot n°18

D'après les tableaux ci-dessus, pour les distributions a priori et a posteriori de la moyenne μ, il existe une différence importante entre les résultats obtenus par les méthodes SS, EE et VV, à cause du poids du passé. Mais, les variances prédictives V_1 sont quasiment identiques, du fait que

$$V_1 = v_1 + v = v\left(1 + \frac{1}{v/v_0 + k}\right) = v\left(1 + \frac{1}{k} \times \frac{1}{1+\tau}\right) \qquad (5.56)$$

(cf. formule (5.55)). Dans le cas où les poids du passé et du présent sont équilibrés, $\tau \approx 1$ par exemple, nous avons $V_1 = v[1+1/(2k)]$. Au contraire, si $\tau < 0.1$ comme dans le cas de la méthode VV, la variance V_1 sera légèrement plus élevée, et nous prendrons $(1+\tau)$ au lieu de 2 ; si τ est bien supérieur à 1, $\tau \gg 1$, la variance V_1 sera relativement faible comme dans le cas de la méthode SS. Mais, cette différence devient négligeable par rapport à celle de la moyenne M_1.

Nous constatons que le modèle prédictif (M_1, V_1), associé à la méthode VV, est plus proche du modèle classique (m, v). Pour la moyenne prédictive $M_1 = m_1$, nous avons montré ci-dessus que la correction par la moyenne empirique soit importante. Pour la variance prédictive, on a $V_1 = v[1+1/(k+v/v_0)]$ presque égale à $v(1+1/k)$, puisque v_0 est beaucoup plus grand que v, noté par $v_0 \gg v$. Ainsi nous conclurons que

(1) ce modèle bayésien est très sensible à la moyenne, mais pas à la variance ;

(2) plus l'écart entre les lots est élevé, surtout entre les moyennes, plus le modèle bayésien est proche du modèle classique ;

(3) plus la taille de l'échantillon du lot est grande, plus le modèle se rapproche du modèle classique (ou empirique).

Pour la méthode SS, la conclusion est inverse : le modèle bayésien est beaucoup moins sensible aux variations de la moyenne.

Quant à la méthode EE, nous rappelons que (m_0, v_0) = (MHM, VH/KH) ou (MM, VM/KH). Ces deux estimateurs dépendent des précisions. Nous pouvons conclure que :

(1) ce modèle prend en compte des poids du passé et du présent équilibrés a priori parce que les moyennes KH, KM et k, ainsi que VH, VM et v, sont du même ordre de grandeur. Si la correction paraît plus grande ou plus petite par rapport à l'a priori m_0 ou à la moyenne empirique m, nous étudierons en détails l'information a priori et l'échantillon considéré, ainsi que les influences respectives du passé et du présent ;

(2) plus la précision d'un échantillon pris en compte est élevée, plus cet échantillon prend un poids important dans l'a priori, et dans l'a posteriori ;

(3) plus la taille de l'échantillon du lot est grande par rapport à la moyenne harmonique KH, plus le modèle bayésien tend vers le modèle classique.

Nous avons montré ci-dessus que la différence entre la méthode EE et la méthode MM est faible, mais la méthode MM ne prend pas en compte les précisions pour la moyenne m_0= MM dans le cas (a). Dans le cas (b), les estimateurs obtenus par les méthodes MM et VV sont égaux.

5.5.3. Etude de la variance

Nous étudions cette fois-ci la variable de précision η (ou variance σ^2) seule. Comme pour le modèle normal, nous distinguons aussi deux cas pour le modèle gamma NO1. L'information a priori pour la variable η est consituée soit par l'ensemble des précisions avec les tailles associées : $\{(h_i, k_i)\}$, lorsque les moyennes sont estimées de façon empirique : $\mu_i = m_i$, soit par l'ensemble des triplets $\{(m_i, v_i, k_i)\}$ ou $\{(m_i, h_i, k_i)\}$, lorsque la moyenne μ est considérée comme constante μ = MM ou MT par exemeple. Dans les deux cas, la moyenne prédictive du coefficient K est la moyenne empirique m.

Nous savons que la distribution de la précision η est une distribution gamma $G(a,b)$, dont la moyenne est : $\mathbb{E}(\eta) = a/b$ et la distribution prédictive du coefficient K est une loi de Student T(m, c^2= b/(2a), t= 2a) (cf. tableau 5.7). Nous en déduisons que la variance prédictive est $V = c^2 \frac{t}{t-2} = \frac{b}{2(a-1)}$ si le paramètre est supérieur à 1.

Si nous notons les paramètres de la distribution a priori par (a_0, b_0), après la réception d'un lot n°17 ou n°18, la distribution de la précision η est modifiée par la statistique (v,k) de l'échantillon prélevé dans ce lot, et reste une distribution gamma de paramètres a_1 et b_1, définis par

$$\begin{cases} a_1 = a_0 + \dfrac{k}{2} \\ b_1 = b_0 + kv \end{cases} \tag{5.57}$$

d'où la moyenne prédictive de la précision $E^{\pi_1}(\eta) = a_1/b_1$. La moyenne prédictive a posteriori est la suivante :

$$V_1 = \frac{t_1}{t_1 - 1} c_1^2 = \frac{b_1}{2(a_1 - 1)} = \frac{b_0 + kv}{2a_0 - 2 + k} \tag{5.58}$$

Les paramètres (a_0, b_0) obtenus par les méthodes SS et EE peuvent s'écrire de manière à ce que : $a_0 = k_0/2 + 1$ et $b_0 = k_0 v_0$, donc nous avons $V_1 = \frac{k_0 v_0 + kv}{k_0 + k}$; pour les méthodes MM et VV, nous avons $b_0 = 2a_0 v_0$ avec $v_0 =$ VH et $V_1 = \frac{2a_0 v_0 + kv}{2a_0 - 2 + k}$ (cf. tableaux 4.5 et 4.6).

Remarquons que la moyenne estimée par la méthode des moments OO est la moyenne totale MT, donc cette méthode n'est valable que pour le modèle N01 dans le cas où μ = MT (cf. tableau 4.6).

Pour alléger la présentation, les distributions a priori π_0 et a posteriori π_1, sont respectivement donnés dans deux tableaux à l'annexe A.3, pour les lots n°17 et n°18, ainsi que ceux des distributions prédictives a priori $p_0 = T(m_0, c_0^2, t_0 = 2a_0)$ et a posteriori $p_1(x) = T(m_1, c_1^2, t_1 = 2a_1)(x)$.

Le paramètre du degré de liberté "t" intervient par exemple à l'estimation d'un intervalle de tolérance, intervalle soit bilatéral symétrique soit unilatéral "à gauche" (inférieur) ou "à droite" (supérieur). Il s'agit de calculer comme une valeur caractéristique des fractiles de la distribution prédictive avec un risque de premier espèce α donné. Ces valeurs caractéristiques détermineront la qualité selon les conditions d'acceptation.

Pour une distribution de Student $T(m, c^2, t)$, au niveau de confiance $1-\alpha$, les marges de l'intervalle étudié dépendent des fractiles t_p de la manière

suivante :

— bilatéral symétrique : $(m - t_p c,\ m + t_p c)$ avec $p = 1 - \alpha/2$

— unilatéral à gauche : $(-\infty,\ m + t_p c)$ avec $p = 1 - \alpha$

— unilatéral à droite : $(m - t_p c,\ +\infty)$ avec $p = 1 - \alpha$

Comme $v = s^2 = \frac{t}{t-2} \times c^2$ (cf. tableau 5.7), la marge de l'intervalle est égale à $t_p c = t_p s \times \sqrt{\frac{t}{t-2}}$. Notons cette valeur du critère par $x_p(t) = t_p \sqrt{\frac{t}{t-2}}$.

Dans le cas de la distribution normale du coefficient K de moyenne m et d'écart-type s, pour le même niveau de confiance $p = (1-\alpha)$, la marge est déterminée par le fractile de la distribution normale, noté par convention u_p. La différence entre des degrés de liberté se voit en comparant les valeurs correspondantes $x_p(t)$ pour l'intervalle bilatéral, par exemple, de même niveau de confiance p = 95%, avec celle de la distribution normale u_p étant comme référence.

$\alpha = 0.05$							$u_p = 1.960$			
t	5	10	15	20	30	40	60	80	100	200
x_p	2.874	2.349	2.206	2.140	2.077	2.047	2.017	2.003	1.994	1.977

Tableau 5.15 Valeurs du critère pour la distribution de Student

A travers cette remarque, nous voyons la convergence de la distribution de Student vers une distribution normale : $T(m, c^2, t) \longrightarrow N(m, v)$, lorsque le degré de liberté est assez grand, et nous avons dans ce cas $x_p(t) \longrightarrow u_p$.

Mais l'objectif du contrôle des boulons n'impose pas, pour le moment, un tel seuil inférieur ou supérieur. Nous ne calculons alors que les écart-types prédictifs a priori S_0 et a posteriori S_1 avec les moyennes associées M_0 et M_1 dans les tableaux 5.16.1 et 5.16.2. Dans le cas où la moyenne est constante, nous considérons ici deux cas : μ = MM et μ = MT.

Si nous estimons la moyenne de la variance σ^2 à partir de la distribution a posteriori $\pi(\eta|\mathbf{x})$, alors $\mathbb{E}^{\pi}(\sigma^2) = c^2$. C'est le paramètre "t" qui fait la différence entre cette valeur et la variance prédictive du coefficient K, leur relation étant $V = \mathbb{E}^{p}(X) = \frac{t}{t-2}\mathbb{E}^{\pi}(\sigma^2)$. Cette différence peut ne pas être négligeable lorsque la valeur de "t" est petite. Par exemple, pour le modèle gamma si nous appliquons la méthode EE, MM ou VV, la valeur de t_0 varie entre 1.21 et 7.56 et la valeur de t_1 varie entre 6.21 et 12.56 (cf. annexe 4). Ainsi, nous prenons les écart-types prédictifs du coefficient K.

L'étude des écart-types prédictifs a priori et de leur influence sur les écart-types prédictifs a posteriori est suffisante pour comparer les méthodes utilisées, dans le but d'en choisir une, qui sera le mieux adaptée au cas concret.

NO1		Moyenne et écart-type prédictifs				
		A priori (M_0, S_0)		+ Lot (m,s,k) ==>	A poster. (M_1, S_1)	
$\mu_1 = m_1$	SS	160.00	5.200	160.00 6.32 5	160.00	5.295
	EE		5.208			5.764
	MM		11.246			5.957
	VV		11.882			5.919
μ = MM	SS	161.48	11.866			11.592
	EE		12.291			10.331
	MM		—			9.014
	VV		—			8.897
μ = MT	SS	163.05	11.811			11.514
	EE		12.343			10.175
	OO		—			—
	MM		—			8.336
	VV		—			8.320

Tableau 5.16.1 Prédiction avant et après la réception du lot n°17

N01		Moyenne et écart-type prédictifs					
		A priori (M_0, S_0)		+ Lot (m,s,k) ==>		A poster. (M_1, S_1)	
$\mu_1 = m_1$	SS	140.00	5.389	140.00	3.16	5	140.00	5.252
	EE		5.434					4.503
	MM		9.126					4.461
	VV		11.197					4.482
μ = MM	SS	161.48	9.481					10.919
	EE		9.790					16.545
	MM		–					22.237
	VV		26.839					21.837
μ = MT	SS	160.80	9.457					10.798
	EE		9.814					16.136
	OO		9.457					12.405
	MM		25.284					21.156
	VV		19.717					20.985

Tableau 5.16.2 Prédiction avant et après la réception du lot n°18

Remarquons d'abord que des écart-types prédictifs a priori n'existent pas, noté dans les tableaux par le symbol "—". En effet, la variance d'une distribution de Student $T(m, c^2, t)$ est $V = S^2 = c^2 \frac{t}{t-2}$, il est clair que $t > 2$ est la condition nécessaire pour que la variance V existe. Cependant, pour l'écart-type prédictif a posteriori, un tel problème ne se pose pas pour toutes les méthodes, lorsque la taille de l'échantillon k est supérieure à 2, parce que nous avons $t_1 = 2a_1 = 2a_0 + k > 2$ (cf. formule (5.57)).

Pour un modèle quelconque, les écart-types prédictifs déterminés par la méthode séquentielle sont relativement proches de l'a priori, parce que dans la pondération, le rapport des poids des deux informations du passé et du présent sont $\tau = KT/k = 60/5 = 12$. Par contre, pour la méthode des échantillons équilibrés EE, le rapport devient alors $\tau = KH/k = 5.56/5 = 1.11$, donc

la pondération entre le passé et le présent est équilibrée.

Pour les méthodes des doubles moments MM et des doubles maxima VV, les écart-types prédictifs de chacun des modèles sont quasiment identiques pour le lot moyen ou extrême. Il suffit de considérer l'une des méthodes.

Les écart-types des modèles avec μ constante sont tous supérieurs à ceux du modèle qui considère uniquement des précisions des échantillons dans l'information a priori (dans le cas où $\mu_1 = m_1$), puisqu'ils prennent en considération l'écart entre la moyenne empirique m et la valeur μ.

Pour le coefficient K des boulons, nous choisisons le modèle N01 avec la moyenne connue et variable d'un lot à l'autre $\mu_1 = m_1$. Comme les variances des échantillons sont petites et voisines, le paramètre "$2a_0$" n'est pas très grand pour la méthode MM ou la méthode VV ; la moyenne et la moyenne harmonique des variances sont de même grandeur : VM ≈ VH. Ce sont le paramètre de "taille" : k_0 = KH et $k_0 = 2a_0$, qui fait la différence entre ces méthodes, puisque la pondération est réalisée à partir du présent (v,k) et du passé (v_0,k_0) = (VM, KH) pour la méthode EE et (VH, $2a_0$) pour la méthode MM ou VV. Comme nous avons : KH = 5.56 et $5 < 2a_0 < 6$, alors cette différence n'est pas grande. Ainsi, les parties des poids du passé et du présent représentées par les paramètres de taille k_0 et k = 5 sont presque égales.

5.6. MODELES SPECIFIQUES

5.6.1. Combinaison des modèles normal et gamma

Dans les paragraphes précédents, nous avons étudié l'estimation de la distribution du coefficient K; l'analyse est faite à l'aide des modèles : le modèle normel N10 (étude de la moyenne $\mu = \bar{K}$) et le modèle gamma N01 (étude de l'écart-type σ seul). Pour le premier modèle N10, l'écart-type prédictif est voisin de l'écart-type empirique : $S_1 \approx s\sqrt{1+1/k}$; pour le second N01, la moyenne prédictive est égale à la moyenne empirique : $M_1 = m$.

Si l'on voulais faire des prédictions simultanément sur ces deux variables μ et σ sans utiliser les estimateurs empiriques, nous proposons ici un

modèle spécifique, qui, étant la combinaison de ces deux modèles N10 et N01, permet d'analyser la moyenne et l'écart-type "indépendemment" au lieu d'utiliser le modèle normal-gamma N11, qui étudie le couple (μ,σ) en considérant que ces deux variables sont dépendantes.

Le choix d'un tel modèle est fait pour une raison principale : pour le modèle N11, la variance prédictive (bayésienne a posteriori) V_1 $(= S_1^2)$ est à peu près une pondération entre la variance a priori V_0 $(= S_0^2)$ et la variance empirique v, augmentée d'une quantité croissante avec l'écart entre la moyenne a priori M_0 et la moyenne empirique m (cf. paragraphe 5.4.1). Ce modèle sera adapté au cas où le contrôle d'un lot de production impose non seulement une évolution de la moyenne μ et de la variance σ^2, mais également d'une référence de la moyenne M_0 (valeur cible) dépendant de l'information a priori (ici n échantillons précédents). Cependant, dans le contrôle des boulons, on ne s'intéresse, pour le moment, ni à une valeur de référence de la moyenne, ni à l'écart de la moyenne du lot considéré avec cette valeur mais à une estimation de la moyenne $\mu = \bar{K}$ et de la dispersion σ d'un lot.

Dans la production de boulons, la variation de la moyenne est lente, et n'introduit pas de variation corrélative de la variance; ceci résulte de la faible valeur du coefficient d'asymétrie de la distribution des mesures, ici C3 = 0.340. En effet le coefficient de corrélation empirique entre la moyenne m et la variance v estimé à partir d'un échantillon est donné par :

$$\rho = \frac{\mathrm{Cov}(m,v)}{\sqrt{\mathrm{Var}(m)}\sqrt{\mathrm{Var}(v)}} = \frac{(n-1)C3}{n\sqrt{C4 + 2}} \qquad (5.61)$$

Ici $\rho = 0.216$. Le test statistique suivant permet d'accepter l'hypothèse de non-corrélation. D'après la théorie, si nous faisons l'hypothèse de la normalité du coefficient K, $\rho = 0$. Dans notre cas, nous considérons n = 18 échantillons. Nous avons donc 18 couples d'observations de la moyenne et de la variance $\{(m_i, v_i)\}$. En adoptant le risque de 1er espèce $\alpha = 0.05$, le seuil au-delà duquel on rejetera l'hypothèse nulle, c'est à dire $\rho = 0$, est $\rho_0 = \rho(\alpha,n) = 0.4438$. Avec $\rho = 0.216 < \rho_0$, nous pouvons accepter l'hypothèse de non-corrélation. Pour d'autre valeurs $\alpha = 0.01$, 0.02 ou 0.10, nous avons toujours $\rho < \rho_0$ (cf. annexe 5, [49] et [50]).

Nous traitons donc séparément les deux estimations par les méthodes des

échantillons équilibrés EE et des doubles moments MM :

— pour la variable μ, nous adaptons le modèle normal N10 avec une variance variable d'un lot à l'autre $\sigma_i^2 = v_i$ pour prévoir les moyennes a priori M_0 et a posteriori M_i;

— pour la variable σ, nous adaptons le modèle gamma N01 avec une moyenne variable d'un lot à l'autre $\mu_i = m_i$ pour prévoir les écart-types a priori S_0 et a posteriori S_i.

en adoptant une distribution normale comme distribution prédictive de moyenne M_i et d'écart-type S_i, $i = 0,1$. Ce modèle est noté N22. Avec les deux méthodes utilisées, nous obtenons alors deux cas différents : N22-EE et N22-MM.

5.6.2. Calcul des résultats prédictifs et Représentation graphique

Nous avons montré dans les tableaux 4.4 et 4.6 que pour les méthodes EE et MM, les estimateurs de la moyenne et de la variance a posteriori sont respectivement les suivants :

$$\begin{cases} M_i = \dfrac{KH \cdot MHM/VH + km/v}{KH/VH + k/v} \\ V_i = \dfrac{KH \cdot VM + kv}{KH + k} \end{cases} \qquad \begin{cases} M_i = \dfrac{KM \cdot MM/VH + km/v}{KM/VH + k/v} \\ V_i = \dfrac{2a_0 VH + kv}{2a_0 - 2 + k} \end{cases} \tag{5.62}$$

où MHM est la moyenne des moyennes pondérées par les précisions associées et $a_0 = HM^2/HV$ est l'inverse du coefficient de variation empirique de la précision avec HV la variance et HM la moyenne.

Nous présentons ici les résultats a priori et a posteriori pour le lot moyen n°17 et le lot extrême n°18 par rapport à leur information a priori. Ces résultats sont donnés dans le tableau 5.17.

D'après ce tableau, nous constatons que les méthodes EE et MM donnent presque le même résultat sur la variance et aussi sur la moyenne. Ainsi nous pouvons adopter l'une de ces deux méthodes. Notons que les méthodes EE et MM

donnent les poids du passé et du présent un peu près équilibrés : (VH, KH), (VH, KM) et (v,k), donc les estimateurs prédictifs sont relativement moins sensibles à la variation à court terme de la moyenne que la méthode VV et plus sensibles que la méthode SS, ceci correspond mieux aux critères attendus lors du contrôle des boulons.

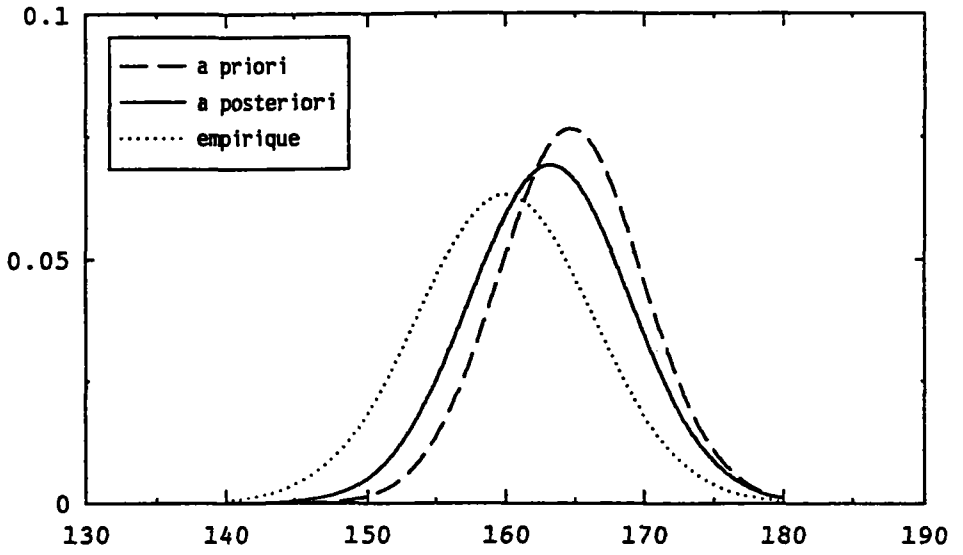


Figure 5.3 Prédiction du coefficient K pour le lot n°17 (N22-EE)

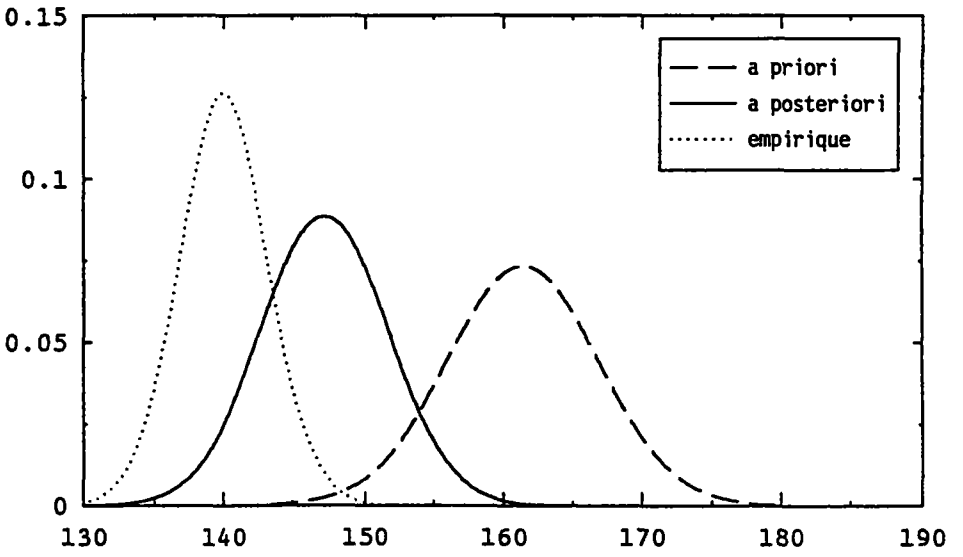


Figure 5.4 Prédiction du coefficient K pour le lot n°18 (N22-EE)

Lot	Modèle	Méthode	A priori (Mo,So)		Lot (m,s)		A poster. (M1,S1)	
n°17	N22	EE	164.728	5.208	160.000	6.325	163.252	5.764
		MM	164.180	11.246			162.943	5.957
n°18		EE	161.432	5.434	140.000	3.162	147.230	4.503
		MM	161.480	9.126			147.620	4.461

Tableau 5.17 Prédiction avant et après la réception des lots n°17 et n°18

5.6.3. Enrichissement d'un échantillon et Sensibilité du modèle

Afin d'évaluer les influences respectives de l'information du passé et de l'information du présente (échantillon considéré) et la différence entre les approches bayésienne et classique, nous allons étudié brièvement la sensibilité des modèles.

En effet, si l'on enrichit un échantillon en augmentant le nombre de prélèvements dans le lot à tester, on diminue l'influence de l'estimation a priori sur celle a posteriori. Lorsque la taille de l'échantillon du lot étudié est grande, le modèle bayésien devient très proche du modèle classique pour toutes les méthodes utilisées ; en effet, (M_1,V_1) et (m,v) convergent vers la même limite (μ,σ^2) (les vraies valeurs). Dans le cas d'échantillons de grande taille, nous prenons plutôt des méthodes classiques moins sophistiquées. L'intérêt de l'estimation bayésienne est de résoudre éventuellement des problèmes où l'information provenant du lot ne semble pas "suffisante" pour effectuer l'analyse classique, comme dans le cas de petits échantillons.

Pour limiter le volume de ce rapport, nous présentons ici, à titre d'exemple, le cas N22-EE en supposant que le couple de moyenne et variance empiriques (m,v) reste le même, lors de l'enrichissement de l'échantillon du lot n°17 ou n°18.

La convergence du couple (M_1,V_1) des moyenne et variance bayésiennes ou prédictives vers le couple (m,v) apparait plus claire sur les graphes 5.5,

et 5.6 que dans le tableau 5.18. Les distributions prédictives sont montrées par la figure 4.2 au chapitre IV. D'après la formule (5.62), les vitesses de convergence sont :

$$\begin{cases} \dfrac{dM_1}{dk} = (m - M_0) \times \dfrac{h}{KH \cdot H1} \times \left(1 + \dfrac{k \cdot h}{KH \cdot H1}\right)^{-2} \\ \dfrac{dV_1}{dk} = \dfrac{v - V_0}{KH} \left(1 + k/KH\right)^{-2} \end{cases} \tag{5.63}$$

où M_0= MHM et V_0= VM. Donc elles sont l'ordre k^{-2} et dépendent également des écarts entre m et M_0 et entre v et V_0.

M_0, V_0		164.728	27.126	161.432	29.526
m, v		160.	40.	140.	10.
M_1, V_1		n° 17		n° 18	
k	5	163.252	33.224	147.230	20.277
	10	162.478	35.402	144.349	16.974
	15	162.002	36.521	143.109	15.277
	20	161.679	37.201	142.420	14.245
	25	161.446	37.659	141.980	13.550
	30	161.270	37.988	141.676	13.051
	35	161.132	38.236	141.453	12.675
	40	161.021	38.430	141.282	12.381
	45	160.930	38.585	141.147	12.146
	50	160.845	38.713	141.038	11.953
	60	160.733	38.909	140.872	11.655
	80	160.572	39.164	140.661	11.268
	100	160.469	39.322	140.532	11.028
	+∞	160.	40.	140.	10.

Tableau 5.18 Variation de l'estimation en fonction de la taille

Nous voyons que l'écart entre les modèles bayésien et classique décroît rapidement et que cet écart est faible même pour k relativement petit.

Lorsque le résultat d'un échantillon est contredit l'a priori, il suffit souvent de enrichir cet échantillon très peu pour obtenir un choix de deux informations.

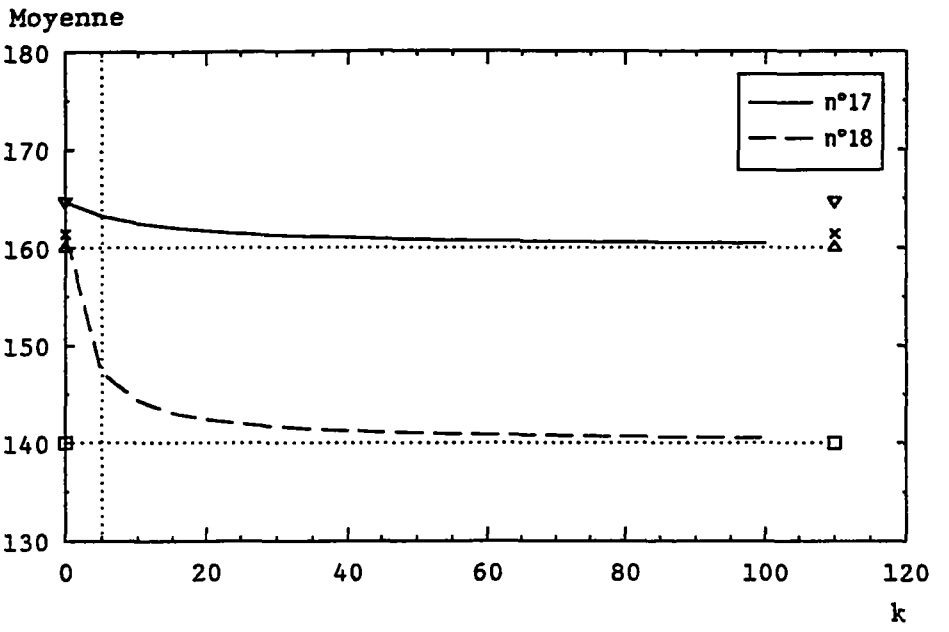


Figure 5.5 Variation de la moyenne en fonction de la taille

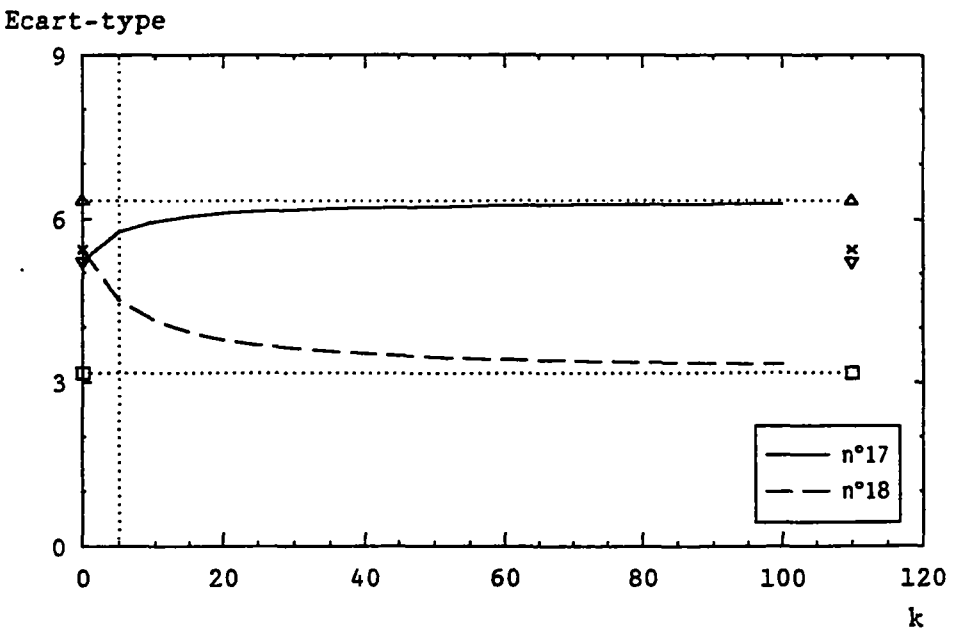


Figure 5.6 Variation de l'écart-type en fonction de la taille

5.7. CONCLUSION

Il résulte de toutes les remarques faites précédemment que la contribution fondamentale de l'approche bayésienne est en quelque sorte de pondérer les informations du passé et du présent. Cette pondération coïncide bien avec une intuition raisonnable sur les événements.

En résumé, le contrôle est effectué par échantillonnage et le but de l'analyse statistique est donc d'estimer de la manière la plus précise les vraies valeurs des statistiques du coefficient K et d'optimiser éventuellement la fréquence d'échantillonnages et le nombre des contrôles. Les techniques bayésiennes permettront d'adopter un processus optimal de contrôle statistique de la qualité, augmentant la faibilité de l'analyse avec la même fréquence d'échantillonnages, ou réduisant le volume d'essais sans perte de fiabilité. Nous pourrons compléter cette étude par l'application des autres techniques statistiques. Par exemple, une analyse des données permettrait de faire un regroupement des boulons de caractéristiques différentes en un petit nombre de familles homogènes. Le contrôle serait alors effectué par famille homogène, et un échantillon serait prélevé de manière aléatoire dans un lot quelconque de la famille considérée.

Nous supposons une continuité des productions et l'analyse est réalisée en exploitant les informations plus récentes et en tenant compte des variabilités de la moyenne et de l'écart-type du coefficient K. Lorsque ces variabilités sont faibles, les approches classique et bayésienne donneront les résultats quasiment identiques ; lorsque ces variabilités augmentent, l'approche bayésienne permet d'accroître la fiabilité de l'estimation.

NOTATION

SYMBOLES GENERAUX

X	—— variable aléatoire d'une distribution $\mathbb{P}_\theta$
θ	—— état de la nature, paramètre de la distribution $\mathbb{P}_\theta$
$(\Omega, \mathcal{B})$	—— espace des échantillons avec une σ-algèbre $\mathcal{B}$
$\mathbb{P}_\theta$	—— distribution de probabilité sur l'espace probabilisable $(\Omega, \mathcal{B})$ conditionnelle au paramètre θ
$F(x\|\theta)$ $f(x\|\theta)$	—— fonctions de répartition et de densité de la variable X
P p	—— fonctions de répartition et de densité (marginale) de X
$P(\cdot\|x)$ $p(\cdot\|x)$	—— fonctions de répartition et de densité (marginale) de X a posteriori
$(\Theta, \mathcal{A})$	—— espace des valeurs du paramètre inconnu θ avec une σ-algèbre $\mathcal{A}$
Π	—— distribution de probabilité du paramètre θ sur l'espace probabilisable $(\Theta, \mathcal{A})$ ou fonction de répartion de θ
π	—— densité de probabilité de θ
$\pi(\cdot\|x)$	—— densité de probabilité de θ a posteriori
ξ	—— paramètre de la distribution de θ (ou hyperparamètre)
μ_π, θ_π	—— moyenne de θ
σ_π	—— écart-type de θ
$I(\theta)$	—— information de Fisher

$\mathbb{R}^+ = \{x \in \mathbb{R};\ x \geq 0\}$

$\mathbb{R}^{+*} = \{x \in \mathbb{R};\ x > 0\}$

$\mathcal{D}$	—— ensemble des décisions possibles
$\mathcal{F}$, $\mathcal{K}$	—— familles de distributions
$I_A(a)$	—— fonction indicatrice définie par $I_A(a) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon} \end{cases}$
$\propto$	—— symbole proportionel

$C_k^x = \frac{k!}{x!(k-x)!}$ —— fonction combinatoire $(0 \le x \le k)$

$\mathbb{E}(\cdot)$ —— fonction de moyenne (on utilise aussi le symbole "$\overline{}$" par convention) définie par $\mathbb{E}(\mathbf{x}) = \bar{\mathbf{x}} = \frac{1}{n}\sum x_i$

$\mathbb{V}(\cdot)$ —— fonction de variance définie par $\mathbb{V}(\mathbf{x}) = \frac{1}{n}\sum(x_i - \mathbb{E}(\mathbf{x}))^2$

$\mathbb{S}(\cdot)$ —— fonction d'écart-type définie par $\mathbb{S}(\mathbf{x}) = \sqrt{\mathbb{V}(\mathbf{x})}$

$\Gamma(\cdot)$ —— fonction gamma définie par $\Gamma(\alpha) = \int_{\mathbb{R}^+} x^{\alpha-1}e^{-x}dx$

$\psi(\cdot)$ —— focntion digamma $\psi(x) = \frac{d\log\Gamma(x)}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}$

CARACTERISTIQUES D'UNE DISTRIBUTION

μ —— moyennes théorique

σ^2 —— variances théorique

η —— précisions théorique $\eta = 1/(2\sigma^2)$

$\delta = \sigma/\mu$ —— coefficients de variation théorique

Mo —— mode (valeur pour laquelle f attaint le maximum)

Me —— médiane (valeur à laqualle la probabilité d'être inférieur est 50%)

$\upsilon_i = E[X^i]$ —— moment d'ordre i avec $\upsilon_1 = \mu$ la moyenne

$\mu_i = E[(X-\upsilon_1)^i]$ —— moment centrés d'ordre i avec $\mu_1 = 0$ et $\mu_2 = \sigma^2$

$\beta_1 = \mu_3^2/\mu_2^3 = \mu_3^2/\sigma^6$ —— coefficient de Pearson

$\beta_2 = \mu_4/\mu_2^2 = \mu_4/\sigma^4$ ....

$\gamma_1 = \mu_3/\mu_2^{3/2} = \pm\sqrt{\beta_1}$ —— coefficient d'asymétrie

$\gamma_2 = \mu_4/\mu_2^2 - 3 = \beta_2 - 3$ —— coefficient d'aplatissement

STATISTIQUES D'UN ECHANTILLON $\mathbf{x} = (x_1, x_2, \ldots, x_k)$

$L(\theta;\mathbf{x}) = f(\mathbf{x}|\theta)$ —— fonction de varisemblance

$\tilde{T}(\mathbf{x})$ —— statistique exhaustive d'une distribution exponentielle

$T(\mathbf{x})$ —— statistique exhaustive (générale) ou statistique exhautive normalisée de $\tilde{T}(\mathbf{x})$

$g(T(\mathbf{x})|\theta)$ —— noyou de la fonction $f(\mathbf{x}|\theta)$ par rapport à θ

k —— taille de l'échantillon $\mathbf{x}$

$m = \frac{1}{k}\sum x_i$ —— espérance mathématique ou moyenne empirique

$v = \frac{1}{k}\sum (x_i - m)^2$ —— variance empirique (de variance minimale)

$v' = \frac{1}{k-1}\sum (x_i - m)^2$ —— variance empirique sans biais

$h = \frac{1}{2v}$, $s = \sqrt{v}$ et $d = s/m$ —— précision, écart-type et coefficient de variation empiriques

$m_G = \left(\prod x_i\right)^{1/k}$ —— moyenne géométrique

$m_{-1} = \frac{1}{k}\sum x_i^{-1}$ —— moment d'orre -1 (ou moment réciproque)

$m_H = 1/m_{-1}$ —— moyenne harmonique

$m_{\pm\alpha} = \frac{1}{k}\sum x_i^{\pm\alpha}$ —— moment d'ordre $\pm\alpha$ ($\alpha > 0$)

$l = \frac{1}{2}\left(m_H^{-1} - m^{-1}\right)^{-1}$ —— statistique

$r = \log\left(m/m_G\right)$ ...

$s = \log\left(m_G/m_H\right)$ ...

$t = \log\left(m_G/\alpha\right)$ ...

$m_{ln} = \frac{1}{k}\sum \log x_i$ —— espérance mathématique de lnX

$v_{ln} = \frac{1}{k}\sum \left(\log x_i - m_{ln}\right)^2$ —— variance empirique de lnX

STATISTIQUES DE n ECHANTILLONS $\mathbf{x}_1, x_2, \ldots, \mathbf{x}_n$

$\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_n$ —— échantillon composé ou "total"

$\Longleftrightarrow \left(\bigcup_i \bigcup_j x_j^i\right)$ avec $k = \sum k_i$

$m_i, v_i, h_i, \ldots$ —— moyenne, variance, précision, ... de $\mathbf{x}_i$

n —— nombre d'échantillons

1/ Statistiques au sens total

$KT = k = \sum k_i$ —— effectif total

$MT = \frac{1}{KT}\sum_i k_i m_i = \frac{1}{KT}\sum_i \sum_j x_j^i$ —— moyenne totale

$VMT = \frac{1}{KT}\sum k_i \left(v_i + (m_i - MT)^2\right) = \frac{1}{KT}\sum_i \sum_j \left(x_j^i - MT\right)^2$ —— variance totale

$D2 = VMT/MT^2$ — coefficient total de variation

$V3 = \frac{1}{KT}\sum_i \sum_j (x_j^i - MT)^3$ — moment centré total d'ordre 3

$V4 = \frac{1}{KT}\sum_i \sum_j (x_j^i - MT)^4$ — moment centré total d'ordre 4

$C3 = V3/VMT^{3/2}$ — coefficient total d'asymétrie

$C4 = V4/VMT^2 - 3$ — coefficient total d'aplatissement

$VT = \frac{1}{KT}\sum k_i v_i$ — moyenne de $\{v_i\}$ pondérées par $\{k_i\}$

$HT = \frac{1}{KT}\sum k_i h_i$ — moyenne de $\{h_i\}$ pondérées par $\{k_i\}$

$VHT = 1/(2HT)$

$MHT = \sum k_i h_i m_i / \sum k_i h_i$ — moyenne de $\{m_i\}$ pondérées par $\{k_i h_i\}$

$MIT = \frac{1}{KT}\sum_i k_i (m_{-1})_i^{-1} = \frac{1}{KT}\sum_i \sum_j (x_j^i)^{-1}$ — moment récipoque total

$MGT = (\prod_i m_{Gi}^{k_i})^{1/\Sigma k_i} = (\prod_i \prod_j x_j^i)^{1/KT}$ — moyenne géométrique totale

<u>2/ Moyennes des statistiques</u> m_i, v_i, h_i, ..., pour $i = 1,2,...n$

$KH = (\frac{1}{n}\sum k_i^{-1})^{-1}$ — moyenne harmonique de $\{k_i\}$

$KM = (\frac{1}{n}\sum k_i)$ — moyenne de $\{k_i\}$

$MM = \frac{1}{n}\sum m_i$ — moyenne de $\{m_i\}$

$VMM = \frac{1}{n}\sum (v_i + (m_i - MM)^2)$

$VM = \frac{1}{n}\sum v_i$ — moyenne de $\{v_i\}$

$HM = \frac{1}{n}\sum h_i$ — moyenne de $\{h_i\}$

$VH = 1/(2HM) = (\frac{1}{n}\sum v_i^{-1})^{-1}$ — moyenne harmonique de $\{v_i\}$

$MHM = \sum h_i m_i / \sum h_i$ — moyenne de $\{m_i\}$ pondérées par $\{h_i\}$

$MH = MIM^{-1} = (\frac{1}{n}\sum m_i^{-1})^{-1}$ — moyenne harmonique de $\{m_i\}$

$MIH = \frac{1}{n}\sum h_i/m_i$

<u>3/ Notations associées</u>

$MHO = \frac{1}{n}\sum h_i m_i^2 \log(h_i m_i)$ $\qquad$ $MRO = \frac{1}{n}\sum r_i^{-1} \log(r_i m_i)^{-1}$

$$\begin{cases} MM = \frac{1}{n}\sum m_i \\ MV = \frac{1}{n}\sum (m_i - MM)^2 \\ ML = \frac{1}{n}\sum \log m_i \end{cases}$$

$$\begin{cases} LM = \frac{1}{n}\sum l_i \\ LV = \frac{1}{n}\sum (l_i - LM)^2 \\ LL = \frac{1}{n}\sum \log l_i \end{cases}$$

$$\begin{cases} MIM = \frac{1}{n}\sum m_i^{-1} \\ MIV = \frac{1}{n}\sum \left(m_i^{-1} - MIM\right)^2 \end{cases}$$

$$\begin{cases} MLM = \frac{1}{n}\sum l_i/m_i \\ MLV = \frac{1}{n}\sum \left(l_i/m_i - MLM\right)^2 \end{cases}$$

$$\begin{cases} HM = \frac{1}{n}\sum h_i \\ HV = \frac{1}{n}\sum \left(h_i - HM\right)^2 \\ HL = \frac{1}{n}\sum \log h_i \end{cases}$$

$$\begin{cases} RI1 = \frac{1}{n}\sum r_i^{-1} \\ RI2 = \frac{1}{n}\sum \left(r_i^{-1} - RI1\right)^2 \\ RI3 = \frac{1}{n}\sum \log r_i^{-1} \end{cases}$$

$$\begin{cases} MH1 = \frac{1}{n}\sum h_i m_i \\ MH2 = \frac{1}{n}\sum \left(h_i m_i - MH1\right)^2 \\ MH3 = \frac{1}{n}\sum \log(h_i m_i) = HL + ML \end{cases}$$

$$\begin{cases} MR1 = \frac{1}{n}\sum \left(r_i m_i\right)^{-1} \\ MR2 = \frac{1}{n}\sum \left(\left(r_i m_i\right)^{-1} - MR1\right)^2 \\ MR3 = \frac{1}{n}\sum \log\left(r_i m_i\right)^{-1} = RI3 - ML \end{cases}$$

$$\begin{cases} MH4 = \frac{1}{n}\sum h_i m_i^2 \\ MH5 = \frac{1}{n}\sum \left(h_i m_i^2 - MH4\right)^2 \\ MH6 = \frac{1}{n}\sum \log\left(h_i m_i^2\right) = HL + 2ML \end{cases}$$

$$\begin{cases} MR4 = \frac{1}{n}\sum \left(r_i m_i^2\right)^{-1} \\ MR5 = \frac{1}{n}\sum \left(\left(r_i m_i^2\right)^{-1} - MR4\right)^2 \\ MR6 = \frac{1}{n}\sum \log\left(r_i m_i^2\right)^{-1} = RI3 - 2ML \end{cases}$$

$$\begin{cases} MH7 = \frac{1}{n}\sum h_i m_i^3 \\ MH8 = \frac{1}{n}\sum \left(h_i m_i^3 - MH7\right)^2 \\ MH9 = \frac{1}{n}\sum \log\left(h_i m_i^3\right) = HL + 3ML \end{cases}$$

$$\begin{cases} TM = \frac{1}{n}\sum \left(1+\sqrt{1+2h_i m_i^2}\right) \\ TV = \frac{1}{n}\sum \left(1+\sqrt{1+2h_i m_i^2} - TM\right)^2 \\ TL = \frac{1}{n}\sum \log\left(1+\sqrt{1+2h_i m_i^2}\right) \end{cases}$$

$$\begin{cases} MGM = \frac{1}{n}\sum \left(m_i^{-1}\Gamma\left(1 \pm \frac{1}{\alpha_i}\right)\right)^{\mp\alpha_i} \\ MGV = \frac{1}{n}\sum \left[\left(m_i^{-1}\Gamma\left(1 \pm \frac{1}{\alpha_i}\right)\right)^{\mp\alpha_i} - MGM\right]^2 \\ MGL = \frac{1}{n}\sum \log\left(m_i^{-1}\Gamma\left(1 \pm \frac{1}{\alpha_i}\right)\right)^{\mp\alpha_i} \end{cases}$$

$$\begin{cases} TIM = \frac{1}{n}\sum t_i^{-1} \\ TIV = \frac{1}{n}\sum \left(t_i^{-1} - TM\right)^2 \\ TIL = \frac{1}{n}\sum \log t_i^{-1} \end{cases}$$

$$\begin{cases} MAM = \frac{1}{n}\sum m_{\pm\alpha i}^{-1} \\ MAV = \frac{1}{n}\sum \left(m_{\pm\alpha i}^{-1} - MAM\right)^2 \\ MAL = \frac{1}{n}\sum \log\left(m_{\pm\alpha i}\right)^{-1} \end{cases}$$

ANNEXES

A.1. DETERMINATION DE LA LOI A PRIORI $Be(\alpha,\beta)$ DU TAUX θ

Notations : (α,β) —— paramètres de la distribution bêta

$\bar{\theta}_\pi = \bar{\theta}_{\pi 0}$ —— moyenne

$\hat{\theta}_\pi = \hat{\theta}_{\pi 0}$ —— maximum de vraisemblance ou mode

$\theta_{.90}$ et $\theta_{.95}$ —— α-fractiles correspondant aux probabilités $\alpha = 0.90$ et 0.95.

n —— nombre des expériences

k —— nombre de tirages de la variable $X(k,\theta)$

x —— nombre d'unités défectueuses parmi k objets

$\hat{\theta}$ —— taux empirique total

$\bar{\theta}$ —— moyenne des taux empiriques

Cas 1. Nombre de tirages constant k

Information a priori : $n = 4$ expériences du même nombre de tirages k mais d'unités défectueuses différentes $x = 1,3,5,7$

Statistiques de $\{k_i\}$, $\{x_i\}$ et $\{\theta_i = x_i/k_i\}$:

$$KT = \sum k_i = 4k \qquad KM = \frac{1}{n}\sum k_i = k \qquad KH = \left(\frac{1}{n}\sum k_i^{-1}\right)^{-1} = k$$

$$m = \frac{1}{n}\sum x_i = 4 \qquad v = \frac{1}{n}\sum (x_i - m)^2 = 5$$

$$\hat{\theta} = m/KM = 4/k \qquad \bar{\theta} = \frac{1}{n}\sum \theta_i = 4/k$$

$$d_\pi = \frac{\sqrt{V(\theta)}}{E(\theta)} = \frac{\sqrt{V(x)}}{E(x)} = \sqrt{5}/4$$

Estimateurs de (α,β) :

$$SS : \begin{cases} \alpha = n\cdot m + 1 = 17 \\ \beta = n(KM-m) + 1 = 4n - 15 \end{cases}$$

$$EE : \begin{cases} \alpha = m + 1 = 5 \\ \beta = KH - m + 1 = k - 3 \end{cases}$$

$$\text{OO} : \quad \begin{cases} \alpha = \dfrac{m^2(1-\hat{\theta})-v\hat{\theta}}{v-m(1-\hat{\theta})} = 4\dfrac{4k-21}{k+16} \\ \beta = \alpha(1/\hat{\theta}-1) \quad = (k-4)\dfrac{4k-21}{k+16} \end{cases}$$

$$\text{MM} : \quad \begin{cases} \alpha = d_\pi^{-2}-(d_\pi^{-2}+1)\bar{\theta} = 4\dfrac{4k-21}{5k} \\ \beta = \alpha(1/\bar{\theta}-1) \quad = (k-4)\dfrac{4k-21}{5k} \end{cases}$$

Cas 2. Nombre d'unités défectueuses constant x

<u>Information a priori</u> : 4 expériences des mêmes unités défectueuses x mais des nombres de tirages k = 30, 40, 60, 70

<u>Statistiques</u> de $\{k_i\}$, $\{x_i\}$ et $\{\theta_i\}$:

KT = 200 KM = 50 KH = 44.80000

m = x v = 0

$\hat{\theta}$ = x/KM = 0.02x $\bar{\theta}$ = x/KH = 0.02232x

$$d_\pi = \frac{\sqrt{V(1/k)}}{E(1/k)} = 1/8.85834$$

<u>Estimateurs de (α,β)</u> :

$$\text{SS} : \quad \begin{cases} \alpha = 4x + 1 \\ \beta = 201 - 4x \end{cases}$$

$$\text{EE} : \quad \begin{cases} \alpha = x + 1 \\ \beta = 45.8 - x \end{cases}$$

$$\text{OO} : \quad \begin{cases} \alpha = -x < 0 \\ \beta = x - KM < 0 \end{cases} \quad \text{non valable } (v = 0)$$

$$\text{MM} : \quad \begin{cases} \alpha = d_\pi^{-2} - (d_\pi^{-2}+1)\bar{\theta} = 8.85834 - 0.22005x \\ \beta = \alpha(1/\bar{\theta} - 1) \quad = 396.85354/x + 0.22005x - 18.71667 \end{cases}$$

Cas 1.	k	10	20	30	40	50	60	80	100	200
	$\hat{\theta} = \bar{\theta}$	.400	.200	.133	.100	.080	.067	.050	.040	.020
SS	α	17								
	β	25	65	105	145	185	225	305	385	785
	$\bar{\theta}_\pi$	.405	.207	.139	.105	.084	.070	.053	.042	.021
	$\hat{\theta}_\pi = \hat{\theta}$									
	$\theta_{.90}$	.502	.266	.172	.137	.110	.092	.069	.056	.028
	$\theta_{.95}$	.531	.284	.183	.147	.118	.099	.075	.060	.030
EE	α	5								
	β	7	17	27	37	47	57	77	97	197
	$\bar{\theta}_\pi$	.417	.227	.156	.119	.096	.081	.061	.049	.025
	$\hat{\theta}_\pi = \bar{\theta}$									
	$\theta_{.90}$	.560	.345	.242	.186	.151	.127	.096	.078	.039
	$\theta_{.95}$	.650	.384	.271	.210	.171	.144	.109	.088	.045
OO	α	2.923	6.556	8.609	9.929	10.85	11.53	12.46	13.07	14.43
	β	4.385	26.22	55.96	89.36	124.76	161.37	236.71	313.67	706.87
	$\bar{\theta}_\pi = \hat{\theta}$									
	$\hat{\theta}_\pi$	.362	.181	.126	.092	.074	.062	.046	.037	.019
	$\theta_{.90}$	.633	.292	.189	.140	.113	.092	.068	.054	.025
	$\theta_{.95}$	.696	.323	.208	.154	.124	.100	.075	.059	.028
MM	α	1.52	2.36	2.64	2.78	2.864	2.92	2.99	3.032	3.116
	β	2.28	9.44	17.02	25.02	32.936	40.38	56.81	72.768	152.68
	$\bar{\theta}_\pi = \bar{\theta}$									
	$\hat{\theta}_\pi$	.289	.139	.092	.069	.055	.046	.034	.028	.014
	$\theta_{.90}$	.718	.355	.235	.176	.141	.117	.088	.070	.035
	$\theta_{.95}$	.795	.402	.274	.206	.165	.137	.103	.082	.041

Cas 2.	x	1	2	3	4	5	6	8	10	20
	$\hat{\theta}$	.020	.040	.060	.080	.100	.120	.160	.200	.400
SS	α	5	9	13	17	21	25	33	41	81
	β	197	193	189	185	181	177	169	161	121
	$\bar{\theta}_{\pi}$	.025	.045	.064	.084	.104	.124	.163	.203	.401
	$\bar{\theta}$	.022	.045	.067	.089	.112	.134	.179	.223	.446
EE	α	2	3	4	5	6	7	9	11	21
	β	44.8	43.8	42.8	41.8	40.8	39.8	37.8	35.8	25.8
	$\bar{\theta}_{\pi}$	.043	.064	.085	.107	.128	.150	.192	.235	.449
MM	α	8.64	8.42	8.20	7.98	7.76	7.75	7.10	6.66	6.46
	β	378.36	180.15	114.23	81.38	61.75	48.75	32.65	23.17	5.53
	$\bar{\theta}_{\pi} = \hat{\theta}$									

A.2. DETERMINATION DES ESTIMATEURS DES LOIS A PRIORI

Tableau 1. Estimation séquentielle

Tableau 2. Estimation des échantillons équilibrés

Tableau 3. Estimation des moments

Tableau 4. Estimation des double momnets

Tableau 5. Estimation des maxima-momnets

Tableau 6. Estimation des momnets-maxima

Tableau 7. Estimation des double maxima de vraisemblance

TABLEAU 1. ESTIMATION SÉQUENTIELLE

$f(x\|\theta)$	$\pi(\theta)$	$KT = \sum k_i$ effectif total (commun)	
$N(\mu,\sigma^2)$	$NG(m,v,k)$ μ , η	$m = MT$ $v = VMT$ $k = KT$	
	$N(m,v)$ μ	$m = MHT$ $v = VHT/KT$ $(\sigma_i^2 = v_i)$	$m = MT$ $v = VT/KT$ $(\sigma^2 = cte)$
	$G(a,b)$ $\eta = \frac{1}{2}\sigma^{-2}$	$a = KT/2 + 1$ $b = KT \cdot VT$ $(\mu_i = m_i)$	$a = KT/2 + 1$ $b = KT \cdot VMT$ $(\mu = MT)$
$X \approx\approx\approx LN(\alpha,\beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha,\beta^2)$		$(m^y, v^y, k) \longrightarrow (m_{ln}, v_{ln}, k)$	
$NI(\mu,\beta^2)$	$NG(a,b,k)$ α , λ	$a = MT^{-1}$ $b = MHHT^{-1} - MT^{-1}$ $k = KT$	$MHHT = \left(\frac{1}{KT}\sum_i k_i m_{Hi}^{-1}\right)^{-1}$
	$N(m,v)$ $\alpha = 1/\mu$	$m = MT^{-1}$ $v = \frac{1}{2KT \cdot MT}$	
	$G(a,b)$ $\lambda = \frac{1}{2}\beta^{-2}$	$a = KT/2 + 1$ $b = KT\left(MT(\mu^{-1} - MT^{-1})^2 + (MHHT^{-1} - MT^{-1})\right)$	
$G(\alpha,\beta)$	$GG(m,r,k)$ α , β	$m = MT$ $r = \log MT - \log MGT$ $k = KT$	$MGT = \left(\prod_i m_{Gi}^{k_i}\right)^{1/\Sigma k_i}$
	$G(a,b)$ β	$a = KT \cdot AMT + 1$ $b = KT \cdot MT$	$AMT = \frac{1}{KT}\sum_i k_i \alpha_i$
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow Y = X^{-1} \approx\approx\approx G(\alpha,\beta)$		$(m^y, m_G^y, k) \longrightarrow (m_H^{-1}, m_G^{-1}, k)$	
$W(\alpha,\beta)$	$G(\alpha,\beta)$ $\lambda = \beta^{\mp\alpha}$	$a = KT + 1$ $b = KT \cdot MAT$	$MAT = \frac{1}{KT}\sum_i k_i (m_\alpha)_i$
$Fr(\alpha,\beta)$		$a = KT + 1$ $b = KT \cdot MAIT$	$MAIT = \frac{1}{KT}\sum_i k_i (m_{-\alpha})_i$
$P(\alpha,\beta)$	$G(a,b)$ β	$a = KT+1$ $b = KT(\log MGT - \log AGT)$	$AGT = \left(\prod_i \alpha_i^{k_i}\right)^{1/\Sigma k_i}$

TABLEAU 2. ESTIMATION DES ECHANTILLONS EQUILIBRES

$f(x\|\theta)$	$\pi(\theta)$	$KH = \left(\frac{1}{n}\sum k_i\right)^{-1}$ moyenne harmonique
$N(\mu,\sigma^2)$	$NG(m,v,k)$ $\mu\ ,\ \eta$	$m = MM$ $v = VMM$ $k = KH$
	$N(m,v)$ μ	$m = MHM$ $v = VH/KH$ $(\sigma_i^2 = v_i)$ $m = MM$ $v = VM/KH$ $(\sigma^2 = VM)$
	$G(a,b)$ $\eta = \frac{1}{2}\sigma^{-2}$	$a = KH/2 + 1$ $b = KH\cdot VM$ $(\mu_i = m_i)$ $a = KH/2 + 1$ $b = KH\cdot VMM$ $(\mu = MM)$
$X \approx\approx\approx LN(\alpha,\beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\mu,\sigma^2)$		$(m^y, v^y, k) \longrightarrow (m_{ln}, v_{ln}, k)$
$NI(\mu,\beta^2)$	$NG(a,b,k)$ $\alpha\ ,\ \lambda$	$a = MM^{-1}$ $b = MHH^{-1} - MM^{-1}$ $k = KH$ $MHH = \left(\frac{1}{n}\sum_i m_{Hi}^{-1}\right)^{-1}$
	$N(m,v)$ $\alpha = 1/\mu$	$m = MM^{-1}$ $v = \frac{1}{2KH\cdot MM}$
	$G(a,b)$ $\lambda = \frac{1}{2}\beta^{-2}$	$a = KM/2 + 1$ $b = KH\left(MM(\mu^{-1} - MM^{-1})^2 + (MHH^{-1} - MM^{-1})\right)$
$G(\alpha,\beta)$	$GG(m,r,k)$ $\alpha\ ,\ \beta$	$m = MM$ $r = \log MM - \log MG$ $k = KH$ $MG = \left(\prod_i m_{Gi}\right)^{1/n}$
	$G(a,b)$ β	$a = KH\cdot AM + 1$ $b = KH\cdot MM$ $AM = \frac{1}{n}\sum_i \alpha_i$
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow Y = X^{-1} \approx\approx\approx G(\alpha,\beta)$		$(m^y, m_G^y, k) \longrightarrow (m_H^{-1}, m_G^{-1}, k)$
$W(\alpha,\beta)$	$G(\alpha,\beta)$ $\lambda = \beta^{\mp\alpha}$	$a = KH + 1$ $b = KH\cdot MA$ $MA = \frac{1}{n}\sum_i (m_\alpha)_i$
$Fr(\alpha,\beta)$		$a = KH + 1$ $b = KH\cdot MAI$ $MAI = \frac{1}{n}\sum_i (m_{-\alpha})_i$
$P(\alpha,\beta)$	$G(a,b)$ β	$a = KH + 1$ $b = KH(\log MG - \log AG)$ $AG = \left(\prod_i \alpha_i\right)^{1/n}$

TABLEAU 3. ESTIMATION DES MOMENTS

$f(x\|\theta)$	$\pi(\theta)$	$KT = \sum k_i$ effectif total (commun)
$N(\mu,\sigma^2)$	$NG(m,v,k)$ $\mu,\ \eta$	$m = MT$ $v = VMT\frac{k-1}{k+1}$ $k = 3 + 6/C4$
	$N(m,v)$ μ	$m = MT$ $v = VMT - \sigma^2$
	$G(a,b)$ $\eta = \frac{1}{2}\sigma^{-2}$	$a = 2 + 3/C4$ $b = 2VMT\cdot(a-1)$
$X \approx\approx\approx LN(\alpha,\beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha,\beta^2)$		$(m^y, v^y, k) \longrightarrow (m_{ln}, v_{ln}, k)$
$NI(\mu,\beta^2)$	$NG(a,b,k)$ $\alpha,\ \lambda$	Non valable!
	$N(m,v)$ $\alpha = 1/\mu$	Non valable!
	$G(a,b)$ $\lambda = \frac{1}{2}\beta^{-2}$	$a = 2 + \frac{3\sqrt{VMT}}{B1-3\sqrt{VMT}}$ $b = 2VMT(a-1)/MT^3$
$G(\alpha,\beta)$	$GG(m,r,k)$ $\alpha,\ \beta$	Non valable!
$G(\alpha,\beta)$	$G(a,b)$ β	$a = \frac{2\cdot D2+1-1/\alpha}{D2 - 1/\alpha}$ $b = MT\cdot(a-1)/\alpha$
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow Y = X^{-1} \approx\approx\approx G(\alpha,\beta)$		$(m^y, m_G^y, k) \longrightarrow (m_H^{-1}, m_G^{-1}, k)$
$W(\alpha,\beta)$ $Fr(\alpha,\beta)$	$G(\alpha,\beta)$ $\lambda = \beta^{\mp\alpha}$	$\frac{\Gamma(a\mp 2/\alpha)\Gamma(a)}{\Gamma(a\mp 1/\alpha)^2} = (D2+1)\frac{\Gamma(1\pm 1/\alpha)^2}{\Gamma(1\pm 2/\alpha)}$ $b = \left[\frac{MT\cdot\Gamma(a)}{\Gamma(1\pm 1/\alpha)\Gamma(a\pm 1/\alpha)}\right]^{\alpha}$
$P(\alpha,\beta)$	$G(a,b)$ β	$a = \frac{2\cdot DA2}{DA2 - 1}$ $b = MAT\cdot(a-1)$ MAT et $DA = \sqrt{DA2}$: moyenne et coefficient de variation de $\{\ln(x_j^i/\alpha)\}$

TABLEAU 4. ESTIMATION DES DOUBLE MOMENTS

$f(x\|\theta)$	$\pi(\theta)$	$E(\beta) = \frac{1}{n}\sum\beta_i$	$V(\beta) = \frac{1}{n}{}_i - E(\beta))^2$
$N(\mu,\sigma^2)$	$NG(m,v,k)$ $\mu\ ,\ \eta$	$m = MM$ $v = VH\frac{k+1}{k}$ $k = 2HM^2/HV - 1$	$m = MM \qquad c = MV/VH)$ $v = MV\cdot(k-1$ $k = \frac{1}{2c}(1+c)\sqrt{1+\frac{4c}{(1+c)^2}}$
	$N(m,v)$ μ	$m = MM$ $v = VH/KM \qquad (\sigma_i^2 = v_i)$	$m = MM$ $v = MV \qquad (\sigma^2 = cte)$
	$G(a,b)$ $\eta = \frac{1}{2}\sigma^{-2}$	$a = HM^2/HV$ $b = a/HM \qquad (\mu_i = m_i)$	$a = HMb^2/H$ $b = a/HMb \qquad (\mu = MM)$
$X \approx\approx\approx LN(\alpha,\beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha,\beta^2)$			$(m^y, v^y, k) - (m_{ln}, v_{ln}, k)$
$NI(\mu,\beta^2)$	$NG(a,b,k)$ $\alpha\ ,\ \lambda$	$a = MIM$ $b = \frac{k+1}{k}(2MH1)^{-1}$ $k = 2MH1^2/MH2 - 1$	$a = MIM \qquad = 2MH1\frac{MIV}{MIM})$ $b = (k-1)MIIM$ $k = \frac{1}{2c}(1+c)\sqrt{1+\frac{4c}{(1+c)^2}}$
	$N(m,v)$ $\alpha = 1/\mu$	$m = MIM$ $v = MIV$	
	$G(a,b)$ $\lambda = \frac{1}{2}\beta^{-2}$	$a = MH7^2/MH8$ $b = a/MH7$	
$G(\alpha,\beta)$	$GG(m,r,k)$ $\alpha\ ,\ \beta$	$m = \frac{1}{MH1}\left(MH4+\frac{1}{2k}\right)$ $r = \frac{1}{4k}MH4/MH5$ $k = 2\frac{MH4^2}{MH5} - 3$	$m = \frac{1}{2k}\times\frac{MH1}{MH2}$ $r = \frac{1}{4MH4}(1+)$ $k = \frac{1}{2MH4}\left(\frac{M}{M} - 1\right)$
	$G(a,b)$ β	$a = MH1^2/MH2$ $b = \frac{1}{2}MH1/MH2$	
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow Y = X^{-1} \approx\approx\approx G(\alpha,\beta)$			$(m^y, m_G^y, k) - (m_H^{-1}, m_G^{-1}, k)$
$W(\alpha,\beta)$ $Fr(\alpha,\beta)$	$G(a,b)$ $\lambda = \beta^{\mp\alpha}$	$a = E(\beta)^2/V(\beta)$ $b = a/E(\beta)$	$\beta_i = m_i^{-1}\Gamma\left(\frac{1}{\alpha_i}\right)$
$P(\alpha,\beta)$	$G(a,b)$ β	$a = E(\beta)^2/V(\beta)$ $b = a/E(\beta)$	$\beta_i = 1 + \sqrt{m_i^2}$

TABLEAU 5. ESTIMATION DES MAXIMA-MOMENTS

$f(x\|\theta)$	$\pi(\theta)$	$E(\beta) = \frac{1}{n}\sum\beta_i$	$V(\beta) = \frac{1}{n}\sum(\beta_i - E(\beta))^2$
$N(\mu,\sigma^2)$	$NG(m,v,k)$ $\mu\ ,\ \eta$	$m = MM$ $v = VH\frac{k+1}{k}$ $k = 2HM^2/HV - 1$	$m = MM$ $(c = MV/VH)$ $v = MV\cdot(k-1)$ $k = \frac{1}{2c}(1+c)\left[1+\sqrt{1+\frac{4c}{(1+c)^2}}\right]$
	$N(m,v)$ μ	$m = MHM$ $v = VH/KM$ $(\sigma_i^2 = v_i)$	$m = MM$ $v = MV$ $(\sigma^2 = cte)$
	$G(a,b)$ $\eta = \frac{1}{2}\sigma^{-2}$	$a = HM^2/HV$ $b = a/HM$ $(\mu_i = m_i)$	$a = HMb^2/HVb$ $b = a/HMb$ $(\mu = MM)$
$X \approx\approx\approx LN(\alpha,\beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha,\beta^2)$			$(m^y, v^y, k) \longrightarrow (m_{ln}, v_{ln}, k)$
$NI(\mu,\beta^2)$	$NG(a,b,k)$ $\alpha\ ,\ \lambda$	$a = MIM$ $b = \frac{k+1}{k}(2LM)^{-1}$ $k = 2LM^2/LV - 1$	$a = MIM$ $(c = 2LM\frac{MIV}{MIM})$ $b = (k-1)MIV/MIM$ $k = \frac{1}{2c}(1+c)\left[1+\sqrt{1+\frac{4c}{(1+c)^2}}\right]$
	$N(m,v)$ $\alpha = 1/\mu$	$m = MIM$ $v = MIV$	
	$G(a,b)$ $\lambda = \frac{1}{2}\beta^{-2}$	$a = LM^2/LV$ $b = a/LM$	
$G(\alpha,\beta)$	$GG(m,r,k)$ $\alpha\ ,\ \beta$	$m = \frac{1}{MR1}\left(RI1+\frac{2}{k}\right)$ $r = \frac{2}{k}RI1/RI2$ $k = 2\frac{RI1^2}{RI2} - 3$	$m = \frac{2}{k}\times\frac{MR1}{MR2}$ $r = \frac{1}{RI2}(1+3/k)$ $k = \frac{2}{MR1}\left(\frac{RI1^2}{MR2} - 1\right)$
	$G(a,b)$ β	$a = MR1^2/MR2$ $b = \frac{1}{2}MR1/MR2$	
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow Y = X^{-1} \approx\approx\approx G(\alpha,\beta)$			$(m^y, m_G^y, k) \longrightarrow (m_H^{-1}, m_G^{-1}, k)$
$W(\alpha,\beta)$ $Fr(\alpha,\beta)$	$G(a,b)$ $\lambda = \beta^{\mp\alpha}$	$a = E(\beta)^2/V(\beta)$ $b = a/E(\beta)$	$\beta_i = (m_{\pm\alpha i})^{-1}$
$P(\alpha,\beta)$	$G(a,b)$ β	$a = E(\beta)^2/V(\beta)$ $b = a/E(\beta)$	$\beta_i = (\log(m_{Gi}/\alpha_i))^{-1}$

TABLEAU 6. ESTIMATION DES MOMENTS-MAXIMA

$f(x\|\theta)$	$\pi(\theta)$	$E(\beta) = \frac{1}{n}\sum\beta_i$ $\qquad$ $E(\log\beta) = \frac{1}{n}\sum\log\beta_i$
$N(\mu,\sigma^2)$	$NG(m,v,k)$ $\mu\ ,\ \eta$	$m = MHM$ $v = VH\frac{k+1}{k}$ $k \approx \left(\log HM - HL + 2(MH4 - MHM^2HM)\right)^{-1}$
	$N(m,v)$ μ	$m = MHM$ $\qquad$ $m = MM$ $v = MH4/HM - MHM^2 = Vm$ $\qquad$ $v = MV$ $\qquad$ $(\sigma^2=cte)$
	$G(a,b)$ $\eta = \frac{1}{2}\sigma^{-2}$	$a \approx \frac{1}{2}\left(\log HM - HL\right)^{-1}$ $\qquad$ $a \approx \frac{1}{2}\left(\log HMb - HLb\right)^{-1}$ $b = a/HM$ $\qquad$ $b = a/HMb$
$X \approx\approx\approx LN(\alpha,\beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha,\beta^2)$ $\qquad$ $(m^y, v^y, k) \longrightarrow (m_{ln}, v_{ln}, k)$		
$NI(\mu,\beta^2)$	$NG(a,b,k)$ $\alpha\ ,\ \lambda$	$a = \frac{1}{2k\cdot MH7}\left((1+16k^2MH7\cdot MH4)^{-1} - 1\right)$ $b = \frac{k+1}{k}(2MH7)^{-1}$ $\frac{1}{k+1} - 4MH7\times a \approx \log MH7 - MH9 - MH4$
	$N(m,v)$ $\alpha = 1/\mu$	$m = MHM^{-1}$ $v = MIH/MH1 - m^2$
	$G(a,b)$ $\lambda = \frac{1}{2}\beta^{-2}$	$a \approx \frac{1}{2}\left(\log MH7 - MH8\right)^{-1}$ $b = a/MH7$
$G(\alpha,\beta)$	$GG(m,r,k)$ $\alpha\ ,\ \beta$	$m = \frac{1}{MH1}\left(MH4 + \frac{1}{2k}\right)$ $r = \frac{1}{4MH4}(1+3/k)$ $k = 2\left(\log MH4 - MH6 + 2(MH4\cdot MH3 - MH0)\right)^{-1}$
	$G(a,b)$ β	$a = \frac{1}{2}\left(\log MH1 - MH3\right)^{-1}$ $b = a/(2MH1)$
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow Y = X^{-1} \approx\approx\approx G(\alpha,\beta)$ $\qquad$ $(m^y, m_G^y, k) \longrightarrow (m_H^{-1}, m_G^{-1}, k)$		
$W(\alpha,\beta)$ $Fr(\alpha,\beta)$	$G(a,b)$ $\lambda = \beta^{\mp\alpha}$	$a = \frac{1}{2}\left(\log E(\beta) - E(\log\beta)\right)^{-1}$ $\qquad$ $\beta_i = m_i^{-1}\Gamma\left(1 \pm \frac{1}{\alpha_i}\right)$ $b = a/E(\beta)$
$P(\alpha,\beta)$	$G(a,b)$ β	$a = \frac{1}{2}\left(\log E(\beta) - E(\log\beta)\right)^{-1}$ $\qquad$ $\beta_i = 1 + \sqrt{2h_i m_i^2}$ $b = a/E(\beta)$

TABLEAU 7. ESTIMATION DES DOUBLE MAXIMA DE VRAISEMBLANCE

$f(x\|\theta)$	$\pi(\theta)$	$E(\beta) = \frac{1}{n}\sum\beta_i$ $\qquad$ $E(\log\beta) = \frac{1}{n}\sum\log\beta_i$
$N(\mu,\sigma^2)$	$NG(m,v,k)$ $\mu\ ,\ \eta$	$m = MHM$ $v = VH\frac{k+1}{k}$ $k \approx \left(\log HM - HL + 2(MH4 - MHM^2HM)\right)^{-1}$
	$N(m,v)$ μ	$m = MHM$ $\qquad$ $m = MM$ $v = MH4/HM - MHM^2 = Vm$ $\qquad$ $v = MV$ $\qquad$ $(\sigma^2=cte)$
	$G(a,b)$ $\eta = \frac{1}{2}\sigma^{-2}$	$a \approx \frac{1}{2}\left(\log HM - HL\right)^{-1}$ $\qquad$ $a \approx \frac{1}{2}\left(\log HMb - HLb\right)^{-1}$ $b = a/HM$ $\qquad$ $b = a/HMb$
$X \approx\approx\approx LN(\alpha,\beta^2) \Longrightarrow Y = \ln X \approx\approx\approx N(\alpha,\beta^2)$ $\qquad$ $(m^y, v^y, k) \longrightarrow (m_{ln}, v_{ln}, k)$		
$NI(\mu,\beta^2)$	$NG(a,b,k)$ $\alpha\ ,\ \lambda$	$a = \frac{1}{2k\cdot LM}\left((1+16k^2LM\cdot MLM)^{1/2} - 1\right)$ $b = \frac{1}{2LM}\times\frac{k+1}{k}$ $\frac{1}{k+1} - 4LM\times a \approx \log L1M - LL - MLM$
	$N(m,v)$ $\alpha = 1/\mu$	$m = MLM/LM$ $v = MLV/LM - m^2$
	$G(a,b)$ $\lambda = \frac{1}{2}\beta^{-2}$	$a \approx \frac{1}{2}\left(\log LM - LL\right)^{-1}$ $b = a/LM$
$G(\alpha,\beta)$	$GG(m,r,k)$ $\alpha\ ,\ \beta$	$m = \frac{1}{MR1}\left(RI1 + \frac{2}{k}\right)$ $r = \frac{1}{RI1}(1+3/k)$ $k = 2\left(\log RI1 - RI3 + 2(RI1\cdot MR1 - MR0)\right)^{-1}$
	$G(a,b)$ β	$a = \frac{1}{2}\left(\log MR1 - MR3\right)^{-1}$ $b = 2a/MR1$
$X \approx\approx\approx GI(\alpha,\beta) \Longrightarrow Y = X^{-1} \approx\approx\approx G(\alpha,\beta)$ $\qquad$ $(m^y, m^y_G, k) \longrightarrow (m_H^{-1}, m_G^{-1}, k)$		
$W(\alpha,\beta)$ $Fr(\alpha,\beta)$	$G(a,b)$ $\lambda = \beta^{\mp\alpha}$	$a = \frac{1}{2}\left(\log E(\beta) - E(\log\beta)\right)^{-1}$ $\qquad$ $\beta_i = \left(m_{\pm\alpha i}\right)^{-1}$ $b = a/E(\beta)$
$P(\alpha,\beta)$	$G(a,b)$ β	$a = \frac{1}{2}\left(\log E(\beta) - E(\log\beta)\right)^{-1}$ $\qquad$ $\beta_i = \left(\log(m_{Gi}/\alpha_i)\right)^{-1}$ $b = a/E(\beta)$

ANNEXE 3. HISTROGRAMMES DES BOULONS 1988

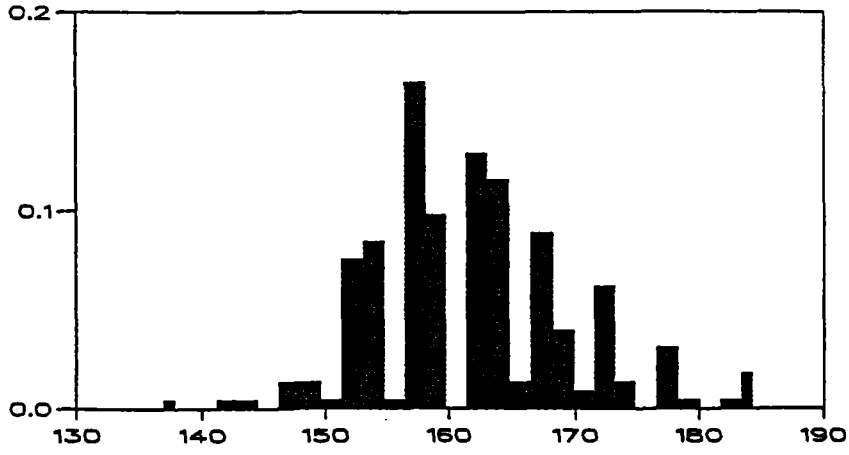


Figure 1. Valeurs K de la qualité HR1 (ϕ = 20mm)

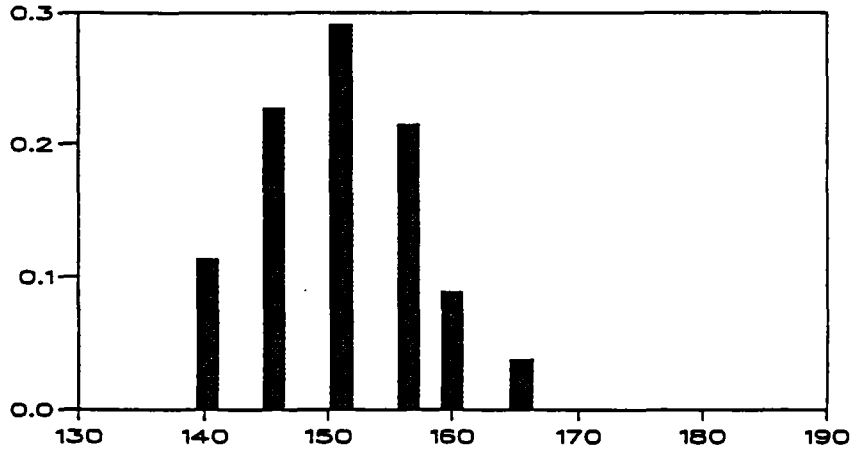


Figure 2. Valeurs K de la qualité HR1Z (ϕ = 18mm)

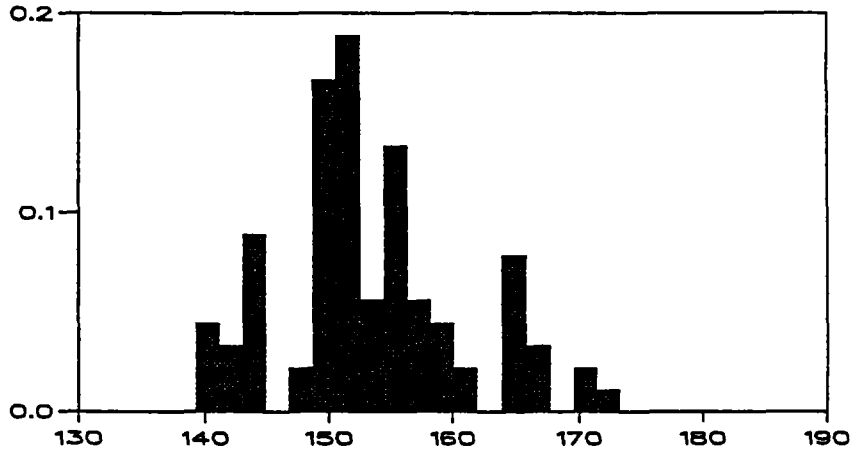


Figure 3. Valeurs K de la qualité HR2Z (ϕ = 22mm)

A.4. RESULTATS DU COEFFICIENT K DES BOULONS

Tableau 1. Prédictions avant et après la réception du lot n°17 (par le modèle normal-gamma N11)

Tableau 2. Passage de l'a priori à l'a posteriori pour le lot n°17 (réalisé par le modèle gamma N01)

Tableau 3. Prédictions avant et après la réception du lot n°17 (par le modèle normal-gamma N01)

Tableau 4. Prédictions avant et après la réception du lot n°18 (par le modèle normal-gamma N11)

Tableau 5. Passage de l'a priori à l'a posteriori pour le lot n°18 (réalisé par le modèle gamma N01)

Tableau 6. Prédictions avant et après la réception du lot n°18 (par le modèle normal-gamma N01)

N11	Moyenne et Ecart-type Prédictifs, Degré de liberté "t"								
	A priori (M_0, S_0, t_0)			+ Lot (m,s,k) ==>			A posteriori (M_1, S_1, t_1)		
SS	163.05	12.01	61.00				162.82	11.69	66.00
EE	164.18	14.74	6.56				162.20	11.15	11.56
OO	—			160.00	6.32	5	—		
MM	164.18	5.99	6.11				162.11	6.66	11.11
VV	164.73	—	1.17				160.16	8.07	6.17

Tableau 1. Prédictions avant et après la réception du lot n°17

N01		Paramètres des distributions a priori et a posteriori						
		A priori (a_0, b_0)		+ Lot (m,v,k) ==>			A posteriori (a_1, b_1)	
μ = MT 163.05	SS	31.000	8370.850				33.500	8617.363
	EE	3.778	846.358				6.278	1092.871
	OO	—					—	
	MM	0.727	63.017				3.227	309.530
	VV	0.749	64.887				3.249	311.399
μ = MM 164.18	SS	31.000	8447.464	160.00	40.00	5	33.500	8734.826
	EE	3.778	839.264				6.278	1126.626
	MM	0.607	55.004				3.107	342.366
	VV	0.737	66.817				3.237	354.179
$\mu_i = m_i$ emp.	SS	31.000	1622.600				33.500	1822.600
	EE	3.778	150.700				6.278	350.700
	MM	2.516	123.197				5.553	323.197
	VV	2.948	128.801				5.692	328.801

Tableau 2. Passage de l'a priori à l'a posteriori avec le lot n°17

NO1		Paramètres des distributions de Student						
		A priori (m_0, v_0, t_0)			+ Lot (m, v, k) ==>		A poster. (M_1, v_1, t_1)	
μ = MT	SS	163.05	135.01	62.00	160.00	40.0 5	160.00	128.62 67.00
	EE		112.02	7.56				87.04 12.56
	OO		—					—
	MM		43.32	1.45				47.95 6.45
	VV		43.32	1.50				47.92 6.50
μ = MM	SS	161.48	136.25	62.00				130.37 67.00
	EE		111.08	7.56				89.73 12.56
	MM		43.32	1.21				55.10 6.21
	VV		43.32	1.47				54.70 6.47
$\mu_1 = m_1$	SS	160.00	26.17	62.00				27.20 67.00
	EE		19.95	7.56				27.93 12.56
	MM		20.17	6.11				29.10 11.11
	VV		20.17	6.38				28.88 11.38

Tableau 3. Prédictions avant et après la réception du lot n°17

N11	Moyenne et Ecart-type Prédictifs, Degré de liberté "t"								
	A priori (M_0, S_0, t_0)			+ Lot (m,s,k) ==>			A posteriori (M_1, S_1, t_1)		
SS	160.80	9.62	61.00				159.20	10.84	66.00
EE	161.48	11.74	6.56				151.31	14.35	11.56
OO	160.80	9.46	24.40	140.00	3.16	5	157.14	11.91	29.40
MM	161.48	6.72	5.03				149.59	12.83	10.03
VV	161.48	——	1.43				141.68	8.36	6.43

Tableau 4. Prédictions avant et après la réception du lot n°18

NO1		Paramètres des distributions a priori et a posteriori						
		A priori (a_0, b_0)		+ Lot (m,v,k) ==>			A posteriori (a_1, b_1)	
μ = MT 160.80	SS	31.000	5365.600				33.500	7578.800
	EE	3.778	535.067				6.278	2748.267
	OO	12.199	2003.025				14.699	4216.225
	MM	1.064	82.209				3.564	2295.409
	VV	1.110	85.763	140.00	10.00	5	3.610	2298.963
μ = MM 161.48	SS	31.000	5393.344				33.500	7750.296
	EE	3.778	532.498				6.278	2889.450
	MM	0.957	73.132				3.457	2430.084
	VV	1.056	80.678				3.556	2437.630
μ = m_1 emp.	SS	31.000	1742.600				33.500	1792.600
	EE	3.778	164.033				6.278	214.033
	MM	2.516	109.840				5.016	159.840
	VV	2.948	128.695				5.448	178.695

Tableau 5. Passage de l'a priori à l'a posteriori avec le lot n°18

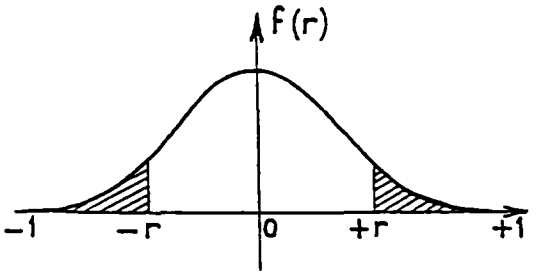
N01		Paramètres des distributions de Student						
		A priori (m_0, v_0, t_0)			+ Lot (m, v, k) ==>		A poster. (m_1, v_1, t_1)	
μ = MT	SS	160.80	86.54	62.00	140.00	10.0 5	140.00	113.12 67.00
	EE		70.82	7.56				218.89 12.56
	OO		82.10	24.40				143.42 29.30
	MM		38.62	2.13				322.00 7.13
	VV		38.62	2.22				318.39 7.22
μ = MM	SS	161.48	86.99	62.00				115.68 67.00
	EE		70.48	7.56				230.13 12.56
	MM		38.20	1.91				351.45 6.91
	VV		38.20	2.11				342.75 7.11
$\mu_1 = m_1$	SS	140.00	28.11	62.00				26.76 67.00
	EE		21.71	7.56				17.05 12.56
	MM		21.83	5.03				15.93 10.03
	VV		21.83	5.90				16.40 10.90

Tableau 6. Prédictions avant et après la réception du lot n°18

A.5. TABLE DE DISTRIBUTION DU COEFFICIENT DE CORRELATION

Cette table concerne le coefficient de corrélation r estimé à partir d'un échantillon prélevé dans une population normale à deux variables dans laquelle le coefficient de corrélation est nul ($\rho = 0$).

Elle donne, en fonction du paramètre ν, les valeurs de r ayant une probabilité P = 0,10 - 0,05 - 0,02 - 0,01 d'être dépassées en valeur absolue, n étant le nombre de couples d'observations, et $\nu = n - 2$.



La table s'applique aussi au coefficient de correlation partielle dans le cas d'une loi normale à un nombre de variables k > 2. On a alors $\nu = n - k$.

Exemples : a) 15 couples d'observations (x, y), $\nu = 13$; il y a une probabilité P = 0,05 pour que le coefficient r_{xy} ait une valeur absolue supérieure à 0,51, l'échantillon étant pris dans une population où $\rho_{xy} = 0$. Si l'on a trouvé $|r| > 0,51$, l'hypothèse $\rho = 0$ doit être rejetée au niveau de signification 5 %.

b) 15 groupes d'observations (x, y, z, t) ; $\nu = 11$. Il y a une probabilité P = 0,05 pour que le coefficient de corrélation partielle $r_{xy,zt} > 0,55$ si $\rho_{xy,zt} = 0$.

ν \ P	0,10	0,05	0,02	0,01
1	0,9877	0,9969	0,9995	0,9999
2	0,9000	0,9500	0,9800	0,9900
3	0,8054	0,8783	0,9343	0,9587
4	0,7293	0,8114	0,8822	0,9172
5	0,6694	0,7545	0,8329	0,8745
6	0,6215	0,7067	0,7887	0,8343
7	0,5822	0,6664	0,7498	0,7977
8	0,5494	0,6319	0,7155	0,7646
9	0,5214	0,6021	0,6851	0,7348
10	0,4973	0,5760	0,6581	0,7079
11	0,4762	0,5529	0,6339	0,6835
12	0,4575	0,5324	0,6120	0,6614
13	0,4409	0,5139	0,5923	0,6411
14	0,4259	0,4973	0,5742	0,6226
15	0,4124	0,4821	0,5577	0,6055
16	0,4000	0,4683	0,5425	0,5897
17	0,3887	0,4555	0,5285	0,5751
18	0,3783	0,4438	0,5155	0,5614
19	0,3687	0,4329	0,5034	0,5487
20	0,3598	0,4227	0,4921	0,5368
25	0,3233	0,3809	0,4451	0,4869
30	0,2960	0,3494	0,4093	0,4487
35	0,2746	0,3246	0,3810	0,4182
40	0,2573	0,3044	0,3578	0,3932
45	0,2428	0,2875	0,3384	0,3721
50	0,2306	0,2732	0,3218	0,3541
60	0,2108	0,2500	0,2948	0,3248
70	0,1954	0,2319	0,2737	0,3017
80	0,1829	0,2172	0,2565	0,2830
90	0,1726	0,2050	0,2422	0,2673
100	0,1638	0,1946	0,2301	0,2540

— Cette table est extraite du livre [50].

BIBLIOGRAPHIE

[1] Abramowitz M. & Stegun I.A. (1972). *Handbook of Mathematical Functions with formulas graphs and math. tables.* Dover

[2] Basawa I.V. and Rao B.L.S.P. (1980). *Statistical Inference for Stochastic Process.* Academic Press, London

[3] Bayes, T. (1763). "*An essay towards solving a problem in the doctrine of chances*". Plilos. Trans. Roy. Sco. London, **53**, 370-418. Reprinted in Biometrica, 1958, **45**, 293-315. As Appendix to S.J. Press (1989)

[4] Benjamin J.R. and Cornell C.A. (1970). *Probability, Statistics and Decision for Civil Engineers.* McGraw-Hill

[5] Berger J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag. 2nd edition. (1er 1980)

[6] - Berger J.O. (1985). Discussion of 'quantifying priori opinion' by Diacinis and Ylvisaker

- Diaconis P. and Ylvisaker D. (1985). "*Quantifying priori opinion*".

In *Baysian Statistics II*, J. Bernardo, M. De Groot, D. lindley and A. Smith (Eds.) 133-156

[7] Berger J.O. (1990). "*Robust bayesian analysis : sensitivity to priori*". J. Statist. Plann. Inf., **25**, 303-328

[8] Berger J.O. and Wolpert *R.* (1988). *The Likelihood Pinciple.* Institute of Math. Statist., Monograph series, Hayward, California. 2nd Edition

[9] Bonitzer J. (1984). "*Réflexions sur les modèles statistiques de décision I°, II° parties*". Annales des PC, 2° trimestre

[10] Box D.E.P. and Tiao G.C. (1973). *Bayesian Inference in Statistical Analysis.* Addison Wesley

[11] Brown L.D. (1986). *Fondamentals of Statistical Exponential Families with applications in statistical decision theory.* Institute of Math. Statist., Monograph series, Hayward, California

[12] Case K.E. and Keats J.B. (1982). "*On the selection of a priori distribution in bayesian acceptance sampling*". J. of Quality Tech., **14**(1)

[13] Chhikara R.S. & Folks J.L. (1989). *Theory, Methodology and Applications* Marcel Dekker

[14] Crow E.L. & K. Shimizu (1988) (ed.). *Lognormal Distribution, Theory and Applications*". Marcel Dekker

[15] David H.A. (1981). *Order Statistics*. Wiley. 2nd edition (1er 1970)

[16] De Groot M. (1970). *Optimal Statistical Decisions*. McGraw-Hill

[17] Ditlevsen O. (1971). "*Bayes-philosophy and structural design*". CIB Basic Structural Engineering Requirements, Committee W23, May 17-21, Copenhagen Demark

[18] Feller W. (1975). *Introduction to Pobability Theory and its Applications*. Vol.1 and Vol.2, Wiley, N.Y.

[19] Genest C. and Zidek J.V. (1986). "*Combining probability distributions : a critique and an annotated bibliography*". Statist. Science, **1**(1), 114-148.

[20] Good I.J. (1983). *Good Thinking : The Foundations of Probability and its Applications*. Univ. of Minnesota Press, MN

[21] Gumbel E.J. (1958). *Statistics of Extremes*. Colombia Univ. Press, N.Y.

[22] Hwang J.T. (1985). "*Universal domination and stochastic domination : decision theory simultaneously under a broad class of loss function*". Ann. Statist., **13**, 295-314

[23] Jacob B. (1985). "*An application of the bayesian probability theory to a predictive model of steel strength*". 2nd International Workshop of Pavia, August 24-27

[24] Jacquet J. (1978). "*Assemblage par boulons à serrage contrôlé, méthodes de serrage et de contrôle des boulons*". Const. Métallique, **2**, 37-47

[25] Jeffrey H. (1961). *Theory of Probability*. Oxford Univ. Press, London 3rd Edition

[26] Johnson N.L. and Kotz S. (1970). *Distributions in Statistics Continuous Univariate Distributions — I*. John Wiley & Sons, N.Y.

[27] Lehmann E.L. (1983). *Theory of Point Estimation*. Wiley

[28] Lehmann E.L. (1986). *Testing Statistical Hypotheses*. Wiley

[29] Lind N.C. and Nowak A.S. (1988). "*Pooling expert opinions on probability distributions*". J. of Engin. Mecanics, **114**(2), 328-341

[30] Lindley D.V. (1965). *Introduction to Probability and Statistics from a bayesian Viewpoint (Part 1 et 2)*. Cambridge Unversity Press, Cambridge

[31] Maritz J.S. and Lwin T. (1989). *Empirical Bayes Methods*. 2nd Edition, Chapman and Hall, NY

[32] Parzen E. (1962). "*On estimation of a probability density function and mode*". Ann. Math. Statist., 33, 1065-1076

[33] Pollard A. et Rivoire (1971). *Fiabilité et Statistiques Prévisionnelles Méthode de Weibull.* Edition Eyralles, Paris

[34] Press S.J. (1989). *Bayesian Satistics : Principles, Models, and Application.* Wiley

[35] Rackwitz R.(1983). "*Predictive distribution of strength under controls*" Mat. et Const. **16**(94), 259-267

[36] Raiffa H. and Schlaifer R. (1961). *Applied Statistical Decision Theory.* The MIT Press, Cambridge

[37] Ringler J. (1979). "*Réduction des coûts d'essais de fiabilité par la pratique des techniques bayésiennes*". R.S.A., **27**(2), 55-68

[38] Robbins H. (1964). "*The empirical bayes approach to statistical problems*". Ann. Math. Statist., **35**, 1-20

[39] Robert C. (1990). "*Hidden mixtures and bayesian sampling*". Rapport tech. #115, LSTA, Université Paris VII

[40] Rutherford J.R. and Krutchkoff R.G. (1969). "*Some empirical bayes techniques in point Estimation*". Biometrika, **56**, 133-137

[41] Saporta G. (1978). *Théorie et Méthodes de la Statistique.* Technip

[42] Tiago de Oliveira J. (1983). (ed.) *Statistical Extremes and Applications.* D. Reidel, Dordrechet, Holland, NATO ASI series, Serie C, **131**, "Proceeding ..., Vimeiro Portugal, 31 August-14Semtembre, 1983"

[43] Teicher H. (1963). "*Identifitiability of finite mixtures*". Ann. Math. Statist., **34**, 1265-1269

[44] Torrent R.J. (1978). "*The log-normal distribution : a better flitness for the results of mechanical testing of materials*". Mat. et Const., **11**(64), 235-245

[45] Vessereau A. (1980). "*Modèles bayésiens et méthodes classiques en contrôle de réception*". R.S.A., **28**(4), 5-35

[46] Viertl R. (1987). (ed.) *Probability Bayesian Statistics.* Plenum Press, New York and London, "Based on the Proc. of the Inter. Sym. on Proba. and Bayesian Statist. held Sept. 23-26, 1986, in Innsbuck, Austria"

[47] Villegas C. (1977). "*On the representation of ignorance*". J. Amer. Statist. Assoc. **72**, 651-654

[48] Weibull W. (1954). "*A statistical representation of fatigue failure in solids*". Roy. Inst. Technology (Srockholm)

[49] "*Aide-mémoire pratique des techniques statistiques*". R.S.A. **34**, 1986

[50] "*Tables statistiques*". R.S.A. n° Spécial, 1971 et 1981

DOCUMENTS SUR BOULONS :

Normes de l'AFNOR

[51] NF E 27-701 (Norme Française Homologuée, Jan. 1977). Boulonnerie à Serrage Contrôlé, Spécifications Techniques

[52] NF E 27-702 (Norme Française Homologuée, Jan. 1977). Boulonnerie à Serrage Contrôlé, Essai d'Aptitude à l'Emploi

[53] NF E 22-466 (Norme Française Enregistrée, Juin 1979). Assemblages par Boulons à Serrage Contrôlé, Méthodes de Serrage et de Contrôle des Boulons

[54] Brunschwig G. (1986). "*note sur la normalisation des boulons à serrage contrôlé*".

[55] Ferveur A. et Fushs J. "*Procès verbal des essais de conformité*". AFNOR Marque NF Boulons à Serrage Contrôlé, Forges et Boulonneries d'Ars-sur-Moselle, 1er Semestre, 1989 et 2ème Semestre, 1990

[56] "*Recommandations pour la définition de la qualité et la réception des boulons à haute résistance et à serrage contrôlé*". Const. Métallique, 1, 80-90, 1973

DOCUMENTS COMMUNIQUES :

[57] Bonitzer J. (1967). "*Note sur le contrôle de qualité des fils de précontrainte*".

[58] Cohen V. "*Elément d'analyse bayésienne*".

[59] Ditlevsen O. (1990). "*State of the art, bayesian decision analysis, as a tool for structural engineering decisions*".

[60] Polleux A. (1986). "*Application de la théorie bayésienne à l'Estimation de paramètres caractéristiques de la résistance d'aciers*".

[61] Robert C. (1990). "*Analyse statistique bayésienne*".

[62] CEB-CECM-CIB-IABBE-RILEM (Nov. 1986). "*Basic notes sur strength*". R-01, Basic Notes Sur Resistance, R-02 et R-03, Joint Committee On Structural Safety

[63] Review of ISO Documents in statistical Quality Control, Prague, 1989-03

[64] La Qualité par la Maîtrise Statistique des Procédés Philosophie TAGUCHI (M.S.P. ou S.P.C. Statist. Process Control). Jounées d'Etude Organisées par la 11e Section, Technique de la S.I.A., 16-17 Fév. 1988 77-100

Documents Supplémentaires

Bazant Z.P., Fellow, ASCE and Kim J.K. (1989). "*Segmental box girder : detection probability and bayesian updating*". J. of Structural Engin., **115**(10)

Benjamin J.R. (1983). "*Risk and decision analysis applied to dams and levees*". Structural Safety, Vol.1, No.4, pp239-255

Bouleau N. (1986). *Probabilités de l'Ingénieur Variables Aléatoires et Simulation.* Hermann

Box G.E.P., Leonard Tom & Wu C.F. (ed.). (1981). *Scientific Inference, Data Analysis and Robustness.* Academic Press, NY

Csorgo M., Dawson D.A., Rao J.N. and Saleh A.K.Ad.E. (ed.). (1981). Statistics and Related Topics North-Holland

Good I.J. (1965). *The Estimation of Probabilities: an Essay on Modern Bayesian Methods.* The MIT Press, Cambrige

Hansel G. et Grouchko D. (1965). "*Prévision séquentielle par la méthode de bayes*". R.S.A., **3**.

Hasofer A.M. & Esteva L. (1985). "*Empirical bayes estimation of seismicity parameters*". Structural Safety, Vol.1, No.2, pp199-205

Ioannou G. (1988). "*Dynamic probabilistic decision processes*". J. of Const. Engin. and Manag., **115**(2)

Morlat G. (1978). "*Sur la comparaison de confiance classique et bayésienne*". R.S.A., 26(2), 33-35

Silverman B. (1986). *Density Estimation for Statistics and Data Analisys.* Chapman and Hall, NY

Vessereau A. (1978). "*Sur l'intervalle de confiance d'une proportion logique 'classique' et logique 'bayésienne'*". R.S.A. **26**(2), 1-32

Wertz W. and Scheider B. (1979). "*Statistical density estimation : a bibliography*". Inter. Statist. Review, n°47, pp155-175

Wald A. (1950). *Statistical decision theory.* Wiley, N.Y.