

**Inférence bayésienne sous un a priori normal-gamma dans
différents contextes et pour des fonctions de la moyenne et de la
variance**

par

Jean-Philippe MORISSETTE

mémoire présenté au Département de mathématiques en vue de l'obtention
du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, mars 2019

Le 12 mars 2019

Le jury a accepté le mémoire de Monsieur Jean-Philippe Morissette dans sa version finale.

Membres du jury

Professeur Éric Marchand

Directeur de recherche

Département de mathématiques

Professeur Bernard Colin

Membre interne

Département de mathématiques

Professeur Félix Camirand-Lemyre

Président-rapporteur

Département de mathématiques

SOMMAIRE

Dans ce mémoire, on obtient différents résultats portant sur l'inférence bayésienne de divers modèles. Des résultats analytiques sont obtenus tout au long du mémoire.

Le chapitre 1 servira dans un premier temps à démontrer dans quels contextes d'application les différents modèles que l'on étudiera peuvent servir. Dans un deuxième temps, ce chapitre se verra être une introduction à quelques distributions et fonctions spéciales que l'on retrouve dans le reste du mémoire. On y présentera également quelques notions sur l'inférence prédictive. Le chapitre 2 porte sur l'inférence prédictive sous le modèle de loi multinormale à variance connue ainsi que sous le modèle gamma à paramètre d'échelle inconnu. On exploitera une décomposition faisant intervenir la loi a priori non informative. Plusieurs cas particuliers seront détaillés. On terminera ce chapitre en faisant une analyse du modèle de loi normale avec $(\theta, \eta) \sim \mathcal{NG}$, c'est-à-dire l'a priori normal-gamma. On discutera et interprétera les lois a posteriori tout en abordant l'inférence prédictive sous ce modèle. Le chapitre 3 reprend le modèle normal-gamma, mais dans un contexte d'indépendance pour la loi a priori. On étudiera les marginales des deux composantes du modèle et cette contribution nous paraît tout à fait originale. Tout d'abord, on verra que la marginale a posteriori de la moyenne fait intervenir une convolution entre une loi normale centrée réduite et une loi student. On explorera par la suite différentes propriétés de cette densité, notamment sa forme, l'impact du choix des différents hyperparamètres,

ainsi que son espérance et sa variance. Quelques cas limites seront également abordés. Puis, de façon un peu plus concise, la même approche sera utilisée pour étudier la marginale a posteriori de la variance. Enfin, le chapitre 4 présente l'inférence bayésienne lorsqu'on considère un quantile, un coefficient de variation (ainsi que son inverse) et un coefficient de discrimination. L'approche choisie ici est de considérer tout d'abord un cadre général puis de s'intéresser au modèle normal-gamma. On donnera des formes analytiques pour les fonctions de densité et de la répartition, l'espérance et la variance de ces fonctions lorsque ces quantités existeront. On y trouvera également quelques exemples afin de pouvoir comprendre le rôle des hyperparamètres du modèle choisi.

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche, M. Éric Marchand, pour sa disponibilité, pour nos échanges, son ouverture et son soutien et pour m'avoir permis d'apprendre autant de choses, que ce soit par rapport à des questions académiques ou encore par rapport à des situations personnelles. Ces deux dernières années m'ont permis de grandir sur plusieurs plans et ses conseils ont su rendre plusieurs décisions plus faciles. Je me dois également de remercier mes consoeurs et confrères du département, des plus jeunes aux plus anciens, avec qui j'ai pu discuter de plusieurs problèmes à saveur mathématique, mais aussi qui m'ont permis de profiter de ces deux belles années et de me changer les idées lorsque c'était nécessaire. J'aimerais également remercier le personnel du département. J'ai pu profiter de votre grande sagesse et j'ai pu, je crois, laisser ma trace grâce aux différentes chances que vous m'avez offertes. Évidemment, je remercie le département de mathématiques de l'Université de Sherbrooke, le Fonds de recherche du Québec Nature et Technologies (FRQNT) ainsi que le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) pour leur soutien financier. Pour finir, merci à mes amis proches et à ma copine, Marie-Frédéric, pour votre soutien inépuisable et pour avoir cru en moi du début à la fin. Cette étape de ma vie aurait été bien triste sans tous ces moments passés avec vous.

Jean-Philippe Morissette
Sherbrooke, décembre 2018

TABLE DES MATIÈRES

SOMMAIRE	iii
REMERCIEMENTS	v
TABLE DES MATIÈRES	vi
LISTE DES FIGURES	ix
INTRODUCTION	1
CHAPITRE 1 — Notions préliminaires et définitions	4
1.1 La loi multivariée normale	5
1.2 Modèle linéaire	6
1.3 Fonctions spéciales et lois non usuelles	9
1.3.1 La distribution de Student décentrée	10
1.3.2 La distribution de Student multivariée	14
1.3.3 La distribution du χ^2 décentrée	16

1.3.4	La distribution de Kummer	21
1.3.5	La convolution	22
1.4	L'analyse prédictive	23
CHAPITRE 2 — Inférence a posteriori et prédictive sur différents modèles		28
2.1	Modèle avec un seul paramètre inconnu	29
2.1.1	Modèle normal avec variance connue	29
2.1.2	Modèle gamma avec paramètre d'échelle inconnu	41
2.2	Modèle normal-gamma	44
CHAPITRE 3 — L'a priori normal-gamma indépendant		50
3.1	Loi marginale a posteriori de θ	50
3.2	Marginale a posteriori de σ^2	63
CHAPITRE 4 — Inférence bayésienne pour une fonction d'une moyenne et d'une variance		70
4.1	Estimation du quantile	71
4.1.1	Décomposition en pivot	80
4.2	Inférence bayésienne pour un coefficient de variation et un coefficient de discrimination	86
4.2.1	Le coefficient de variation et son inverse	86
4.2.2	Le coefficient de discrimination	93

CONCLUSION	98
BIBLIOGRAPHIE	100

LISTE DES FIGURES

1.1	Deux densités de Student décentrées avec 5 degrés de liberté.	12
3.1	Différents graphiques qui présentent la densité (*) pour : a) $\mu = 2, v = 3$ et σ^2 varie, b) $\mu = 2, v$ varie et $\sigma^2 = 1$, c) μ varie, $v = 3$ et $\sigma^2 = 1$. Le graphique d) compare la densité produit avec celle de Student et la normale.	54
3.2	Densité d'un produit d'une normale centrée réduite avec une densité Student de paramètre $\mu = 4, v = 2$ et $\sigma^2 = 0.05$	57
3.3	Espérance d'une densité produit normale avec une densité Student pour a) $v = 3, \sigma = 1$, b) $v = 3, \mu = 2$	63
3.4	Différents graphiques qui représentent la densité marginale de σ^2 a) a priori, mais également a posteriori lorsqu'on altère : b) la taille d'échantillon n , les hyperparamètres c) α , d) β , e) μ et f) τ	67
4.1	Différents graphiques qui représentent : a) la densité a priori, b) la densité a posteriori, c) les deux densités, d) la densité a posteriori du quantile lorsque $c = 0$	78

4.2	Différents graphiques qui représentent la densité a posteriori lorsqu'on altère : a) le paramètre c , b) la taille d'échantillon n , les hyperparamètres c) α , d) β , e) λ	79
4.3	Graphique des fonctions $f(x)$ et $h(x)$	90
4.4	Différents graphiques qui représentent : a) la densité a priori, b) la densité a posteriori,	91
4.5	Différents graphiques qui représentent l'impact sur la densité a posteriori lorsqu'on altère : a) la taille d'échantillon n , les hyperparamètres b) λ , c) α , d) β	92

INTRODUCTION

La statistique permet de répondre à plusieurs questions lorsqu'on essaie de modéliser un phénomène de la vie courante. Une approche paramétrique peut s'avérer très intéressante lorsqu'il est possible d'attribuer une interprétation à chacun des paramètres, ou encore à des *hyperparamètres*, du modèle considéré. On s'efforcera alors de donner des mesures de précision sur ces paramètres. Dans d'autres circonstances, la question d'intérêt ne sera pas nécessairement l'estimation des paramètres, mais plutôt la prédiction d'une innovation ou encore d'une nouvelle réalisation du phénomène étudié. Un modèle privilégié, bien connu, et ayant un très grand potentiel d'application est le modèle linéaire. En proposant un *résumé exhaustif*, on peut plutôt considérer un modèle bivarié (X, S) , où $X \sim \mathcal{N}_d(\theta, \frac{\sigma^2}{n})$, potentiellement multivarié, et $nS^2 \sim \mathcal{G}(\frac{n-1}{2}, 2\sigma^2)$, où $\mathcal{G}(\alpha, \beta)$ représente une densité gamma de paramètre α et β . Dans ce cadre, les interrogations concernent généralement les paramètres θ et σ^2 , dictant respectivement la moyenne et la variance du phénomène à l'étude, ou encore sur une nouvelle observation, Y , qui n'a pas nécessairement la même variance que l'observation X . Plusieurs approches sont alors à notre disposition pour inférer sur le ou les paramètres inconnus.

Une façon de considérer plusieurs de ces problèmes de façon efficace est de s'intéresser à des problèmes concernant des fonctions de la moyenne et de la variance, par exemple les quantiles, un coefficient de variation ou un coefficient de discrimination. Ces quantités

sont considérées dans [DZ12] ainsi que dans [CL07], où plusieurs approches sont présentées. L'une d'entre elles est l'approche bayésienne, qui attribue aux paramètres inconnus une *distribution a priori*. On peut alors faire de l'inférence bayésienne sur le modèle que l'on a choisi en regardant la *distribution a posteriori* qui est de moins en moins influencée par le choix de la distribution a priori au fur et à mesure que la taille d'échantillon grandit. Dans ces articles, l'analyse bayésienne ne semble pas être utilisée à son plein potentiel, ce qui est mentionné dans [NCL08]. Dans la littérature, certains auteurs prennent avantage de la puissance computationnelle des ordinateurs d'aujourd'hui pour utiliser des méthodes numériques. En effet, les méthodes de Monte-Carlo par chaîne de Markov (en anglais MCMC) sont très répandues. Les auteurs de [CH03] présentent comment faire de l'estimation en suivant cette approche. À cause de la disponibilité, de l'aisance et de l'engouement pour ce genre de technologies, certains développements analytiques ont été mis de côté.

Toujours en considérant l'approche bayésienne, un choix de distribution(s) a priori appropriée(s) pour la problématique considérée est le modèle normal-gamma, comme suggéré dans le chapitre 7 de [ER04]. Ce modèle donne la possibilité de modéliser une dépendance entre les paramètres θ et σ^2 lorsque tous les deux sont inconnus. Il s'agit d'un choix commode et bien connu, plusieurs résultats étant déjà disponibles à ce sujet. Étant donné la facilité de faire le passage entre ce modèle et les modèles linéaires, sa présence est grandement justifiée dans la littérature. Néanmoins, l'hypothèse de dépendance entre les deux paramètres n'est pas justifiable dans tous les champs d'application. On propose donc développer une analyse de ce modèle dans ce cadre, ce qui ne semble pas être fait, de façon analytique, dans la littérature.

Ainsi, le premier chapitre de ce mémoire servira à mettre en place les différents outils qui seront nécessaires pour les chapitres ultérieurs. Différentes distributions particulières seront détaillées et définies tout comme la façon d'effectuer le passage entre les modèles

linéaires et les techniques proposées.

Par la suite, on présentera l'inférence sur différents modèles lorsqu'on suppose que la variance est connue. On terminera le chapitre en détaillant le choix d'a priori normal-gamma et en illustrant l'interprétation des hyperparamètres de ce modèle.

Ensuite, on considérera l'a priori normal-gamma dans un cadre indépendant. Toujours en ne faisant que des développements analytiques, on donnera des expressions pour les densités marginales de la moyenne ainsi que de la variance et on donnera quelques propriétés de ces densités.

Finalement, le dernier chapitre sera consacré à l'étude de différentes fonctions de la moyenne et de la variance. Encore une fois, on aura des formes analytiques pour les densités de ces fonctions sous un a priori quelconque, puis on illustrera le cas de l'a priori normal-gamma.

CHAPITRE 1

Notions préliminaires et définitions

Dans ce chapitre, on présente plusieurs notions et concepts nécessaires qui seront utilisés dans le reste de ce mémoire. On présente tout d'abord quelques résultats concernant la distribution multivariée normale. Dans la deuxième section, on explique comment le reste de ce mémoire peut être mis en pratique en mettant en valeur le modèle canonique, soit lorsque $X \sim \mathcal{N}_p(\mu, \sigma^2 \mathbf{1}_p)$ est indépendant de la variable aléatoire $S^2 \sim \sigma^2 \chi^2_\nu$. On mentionne dans [Car15], par exemple, l'importance de pouvoir estimer les quantiles dans le milieu de la santé. Dans la troisième section, des distributions et des fonctions non usuelles sont détaillées. Notamment, des développements et des résultats concernant des lois décentrées (χ^2 , Student) ainsi que des lois multidimensionnelles (Student) seront explorés. Pour finir, on abordera quelques différences entre la statistique classique et la statistique bayésienne en prenant soin de parler aussi d'analyse prédictive.

1.1 La loi multinormale

On aura besoin de quelques résultats importants provenant de la théorie des lois multinormales dans la prochaine partie de ce chapitre. On réfère le lecteur au premier chapitre de [Mui09] pour les détails de ces résultats ainsi qu'une introduction formelle à la théorie de la loi multinormale.

Notation 1. Soit X une variable aléatoire.

- On notera $X \sim \mathcal{F}$ (ou encore $X \sim g$), où \mathcal{F} représente un modèle statistique paramétrique quelconque (et g est une densité), lorsque la variable aléatoire X se distribue selon ce modèle \mathcal{F} (ou encore que la densité de X est g).
- Finalement, on notera parfois également $\pi(\theta) \sim \mathcal{F}$ (ou encore $\pi(\theta|x) \sim g$) pour dire que la loi a priori (ou a posteriori) de θ suit un certain modèle paramétrique \mathcal{F} (ou encore une densité g).

Définition 2. On dit qu'une variable aléatoire X suit une loi multinormale de paramètres $\mu \in \mathbb{R}^d$ et $\Sigma_{d \times d}$, une matrice définie positive ($\Sigma > 0$) à coefficients réels, si sa densité est donnée par

$$f_X(x) = \frac{e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \mathbf{1}_{(\mathbb{R}^d)}(x).$$

On notera alors $X \sim \mathcal{N}_d(\mu, \Sigma)$.

Lemme 3. Soient $X \sim \mathcal{N}_p(\mu, \Sigma)$, $A_{q \times p}$ une matrice à coefficients réels et $b_{q \times 1}$ un vecteur colonne à coefficients réels. Alors,

$$Y = AX + b \sim \mathcal{N}_q(A\mu + b, A\Sigma A^\top).$$

Remarque 4. La Définition 2 admet une extension où Σ peut ne pas être de plein rang. Dans ce cas, au Lemme 3, il est ni nécessaire que $\Sigma > 0$ ou que A soit de plein rang.

Le prochain Lemme permet de savoir le comportement des marges d'une densité multinormale et donne également une condition pour vérifier l'indépendance entre deux composantes distinctes d'un vecteur suivant une loi multinormale.

Lemme 5. *Pour $X = (X_{(1)}, X_{(2)})^\top \sim \mathcal{N}_{p_1+p_2} \left(\mu = (\mu_{(1)}, \mu_{(2)})^\top, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$, avec $\Sigma > 0$, on a :*

a) $X_i \sim \mathcal{N}_{p_i}(\mu_{(i)}, \Sigma_{ii}), i = 1, 2.$

b) $X_{(1)}$ et $X_{(2)}$ sont indépendants si et seulement si $\Sigma_{12} = 0.$

1.2 Modèle linéaire

Les modèles linéaires englobent beaucoup de domaines d'applications en science. On y retrouve notamment la régression, l'analyse de la variance ainsi que l'analyse de la covariance. On parle de modèle linéaire, lorsque le modèle présente une structure linéaire et lorsqu'on tente de prédire une variable réponse en se basant sur différents facteurs que l'on observe, qui peuvent être qualitatifs, quantitatifs ou encore un mélange de ces deux types de variables. Selon les facteurs que l'on utilise, on pourra choisir l'outil le mieux approprié pour prédire la variable réponse et estimer les paramètres du modèle de la façon la plus juste possible. Pour plus de détail sur ces modèles, on suggère [LC06] ou [CB02].

Notation 6. *Tout au long de ce mémoire, on utilisera la notation $\mathbb{1}_p$ pour désigner la matrice identité de dimension $p \times p$.*

On parlera de modèle linéaire simple lorsque la relation pourra s'écrire sous la forme $Y = Z\beta + \epsilon$, où $Y_{n \times 1} = (Y_1, \dots, Y_n)^\top$, $Z_{n \times p} = (Z_1, \dots, Z_n)^\top$, $\beta_{p \times 1} = (\beta_1, \dots, \beta_p)^\top$, $\epsilon \sim \mathcal{N}_n(0, \mathbb{1}_n \sigma^2)$ et $n > p$. On dira que Y_i est la variable réponse de la $i^{\text{ème}}$ observation,

que $Z_i = (Z_{i1}, \dots, Z_{ip})$ est le vecteur des facteurs observés pour la $i^{\text{ème}}$ observation, que β est le vecteur des coefficients des différents facteurs du modèle et finalement que ϵ_i est l'erreur d'observation liée à la $i^{\text{ème}}$ observation, qui est supposée de loi normale.

Remarque 7. *Sous les hypothèses que l'on vient d'énoncer, on trouve que $Y \sim \mathcal{N}_n(Z\beta, \mathbb{1}_n\sigma^2)$, et donc que*

$$f_Y(y) = \frac{e^{-\frac{1}{2\sigma^2}(y-Z\beta)^\top(y-Z\beta)}}{(\sqrt{2\pi}\sigma)^n}$$

Une des méthodes pour trouver un estimateur plausible de β a été proposée au début du 19^e siècle par le mathématicien français Adrien-Marie Legendre. Il s'agissait de la technique des moindres carrés ordinaires. Carl Friedrich Gauss a participé à populariser la méthode, notamment dans un ouvrage en astronomie. [Ste12]

Lemme 8. *Lorsque Z est de plein rang, on sait que l'estimateur $\hat{\beta} = (Z^\top Z)^{-1}Z^\top Y$, que l'on trouve en utilisant les moindres carrés ordinaires, est un estimateur sans biais de β . De plus, on a que $\hat{\beta} \sim \mathcal{N}_p(\beta, (Z^\top Z)^{-1}\sigma^2)$.*

Une fois que l'on dispose d'un estimateur de β , prédire la variable réponse est très simple. En effet, les prédictions associées seront seulement des points de la droite engendrée par les observations Z et l'estimation des paramètres $\hat{\beta}$, que l'on notera $\hat{Y} = Z\hat{\beta}$. Finalement, la quantité $S^2 = (Y - \hat{Y})^\top(Y - \hat{Y})$ correspond à la somme des carrés des écarts entre la vraie valeur des réponses et leurs prédictions.

Lemme 9. *On a que $S^2 \sim \sigma^2\chi_{n-p}^2$ sous les hypothèses d'un modèle linéaire simple (voir le Chapitre 4 dans [RPD01]).*

Lemme 10. *Sous les hypothèses de la Définition ??, les statistiques S^2 et $\hat{\beta}$ sont indépendantes.*

Démonstration. On va tout d'abord montrer que $(Y - \hat{Y})$ est indépendant de $Z\hat{\beta}$, pour X une matrice de plein rang. Remarquons tout d'abord que les deux statistiques peuvent s'exprimer en fonction de Y . En effet, on a $(Y - \hat{Y}) = (\mathbf{1}_n - Z(Z^\top Z)^{-1}Z^\top)Y$ et que $Z\hat{\beta} = (Z(Z^\top Z)^{-1}Z^\top)Y$. On pose $A_1 = (\mathbf{1}_n - Z(Z^\top Z)^{-1}Z^\top)$, $A_2 = (Z(Z^\top Z)^{-1}Z^\top)$, $W = (Y, Y)^\top$ et $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$.

En utilisant le Lemme 3, on a que

$$W \sim \mathcal{N}_{2n} \left((Z\beta, Z\beta)^\top, \begin{pmatrix} \mathbf{1}_n\sigma^2 & \mathbf{1}_n\sigma^2 \\ \mathbf{1}_n\sigma^2 & \mathbf{1}_n\sigma^2 \end{pmatrix} \right)$$

et que

$$AW \sim \mathcal{N}_{2n} \left(A(Z\beta, Z\beta)^\top, \begin{pmatrix} A_1\mathbf{1}_n\sigma^2A_1^\top & A_1\mathbf{1}_n\sigma^2A_2^\top \\ A_2\mathbf{1}_n\sigma^2A_1^\top & A_2\mathbf{1}_n\sigma^2A_2^\top \end{pmatrix} \right).$$

En vertu du Lemme 5, il suffit alors de vérifier que $A_1\mathbf{1}_n\sigma^2A_2^\top = 0_{n \times n}$ pour avoir l'indépendance. On a

$$\begin{aligned} A_1\mathbf{1}_n\sigma^2A_2^\top &= \frac{(\mathbf{1}_n - Z(Z^\top Z)^{-1}Z^\top)(Z(Z^\top Z)^{-1}Z^\top)^\top}{\sigma^2} \\ &= \frac{Z(Z^\top Z)^{-1^\top}Z^\top - Z(Z^\top Z)^{-1}Z^\top Z(Z^\top Z)^{-1^\top}Z^\top}{\sigma^2} \\ &= 0_{n \times n}. \end{aligned}$$

Ainsi, on a bien que $(Y - \hat{Y})$ et $Z\hat{\beta}$ sont indépendants et, de surcroit, que S^2 et $\hat{\beta}$ le sont aussi.

□

Théorème 1.1. *Dans le cadre du modèle présenté dans la définition ??, le vecteur $(\hat{\beta}, S^2)$ est une statistique exhaustive pour (β, σ^2) .*

Démonstration. On a

$$\begin{aligned}
\|Y - Z\beta\|^2 &= \|Y - \hat{Y} + \hat{Y} - Z\beta\|^2 \\
&= \|Y - \hat{Y}\|^2 + 2(Y - \hat{Y})^\top(\hat{Y} - Z\beta) + \|\hat{Y} - Z\beta\|^2 \\
&= \|Y - \hat{Y}\|^2 + \|Z\hat{\beta} - Z\beta\|^2, \text{ en utilisant que } \sum_{i=1}^n (Y_i - \hat{Y}_i) = 0, \\
&= \|Y - \hat{Y}\|^2 + (\hat{\beta} - \beta)^\top(Z^\top Z)(\hat{\beta} - \beta),
\end{aligned}$$

d'où $e^{\frac{-1}{2\sigma^2}\|Y - Z\beta\|^2} = e^{\frac{-1}{2\sigma^2}\|Y - \hat{Y}\|^2} e^{\frac{-1}{2\sigma^2}(\hat{\beta} - \beta)^\top(Z^\top Z)(\hat{\beta} - \beta)}$.

Alors, par le critère de factorisation, le vecteur $(\hat{\beta}, S^2)$ est exhaustif.

□

Dans les sections qui suivent, on utilisera plutôt un modèle où X correspond à notre vecteur d'observations et où c'est le vecteur (X, S^2) qui est exhaustif, comme présenté dans [LC06]. Ainsi, pour faire le passage de $\hat{\beta}$ vers X , deux situations sont possibles. Dans le premier cas, la matrice Z est orthogonale. Alors, il suffit de bien choisir une matrice A , elle aussi orthogonale, pour avoir $A\hat{\beta} \sim \mathcal{N}_p(\mu = A\beta, \mathbb{1}_p\sigma^2)$, où μ est la moyenne de X . Dans le deuxième cas, la matrice Z n'est pas orthogonale. Alors, il suffit de remarquer que $(Z^\top Z)^{-\frac{1}{2}}\hat{\beta} \sim \mathcal{N}_p\left((Z^\top Z)^{\frac{1}{2}}\beta, \mathbb{1}_p\sigma^2\right)$, où on utilise que $A^{\frac{1}{2}}A^{\top\frac{1}{2}} = A$. Ceci donne alors le modèle canonique.

1.3 Fonctions spéciales et lois non usuelles

Dans les prochains chapitres, on verra quelques résultats faisant intervenir des distributions non élémentaires, telles que la loi de Student décentrée, la loi du χ^2 décentrée et la distribution de Kummer, ainsi que des fonctions spéciales. Cette section a donc pour

but de familiariser le lecteur avec ces dernières. On peut également consulter [VAG03] et [vAG00] pour des propriétés et le comportement analytique des distributions décentrées.

1.3.1 La distribution de Student décentrée

On rappelle tout d'abord comment obtenir la distribution de Student usuelle.

Définition 11. Soient $Z \sim \mathcal{N}(0,1)$ et $U \sim \chi_v^2$ deux variables aléatoires indépendantes. Alors, la variable $T = \frac{Z}{\sqrt{\frac{U}{v}}}$ suit une loi de Student univariée à $v > 0$ degrés de liberté. De plus, sa densité est donnée par

$$f_T(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}} \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}. \quad (\star)$$

Notation 12. Soit $Z \sim \mathcal{N}(0,1)$ une variable aléatoire de loi normale centrée et réduite. Alors, on notera par ϕ sa fonction de densité et par Φ sa fonction de répartition.

Dans l'éventualité où la variable Z de la Définition 11 n'aurait pas une moyenne nulle, quelle distribution obtiendrait-on ? En posant $T = \frac{Z+\delta}{\sqrt{\frac{U}{v}}}$, où $\delta \in \mathbb{R}$, on a en conditionnant par rapport à U :

$$\begin{aligned} \mathbb{P}(T \leq t) &= \int_{\mathbb{R}_+} \mathbb{P}\left(Z \leq t\sqrt{\frac{u}{v}} - \delta\right) f_U(u) du \\ &= \int_{\mathbb{R}_+} \Phi\left(t\sqrt{\frac{u}{v}} - \delta\right) f_U(u) du. \end{aligned}$$

En dérivant par rapport à t , on trouve

$$\begin{aligned}
f_T(t) &= \int_{\mathbb{R}_+} \phi\left(t\sqrt{\frac{u}{v}} - \delta\right) \sqrt{\frac{u}{v}} f_U(u) du \\
&= \frac{e^{-\frac{\delta^2}{2}}}{\sqrt{2\pi v} \Gamma\left(\frac{v}{2}\right) 2^{\frac{v}{2}}} \int_{\mathbb{R}_+} e^{-u\left(\frac{t^2+v}{2v}\right)} e^{\frac{t\delta\sqrt{u}}{\sqrt{v}}} u^{\frac{v-1}{2}} du \\
&= \frac{e^{-\frac{\delta^2}{2}}}{\sqrt{2\pi v} \Gamma\left(\frac{v}{2}\right) 2^{\frac{v}{2}}} \sum_{k=0}^{\infty} \frac{\left(\frac{t\delta}{\sqrt{v}}\right)^k}{k!} \int_{\mathbb{R}_+} u^{\frac{v+k-1}{2}} e^{-u\left(\frac{t^2+v}{2v}\right)} du \\
&= \frac{e^{-\frac{\delta^2}{2}} v^{\frac{v}{2}}}{\sqrt{\pi} \Gamma\left(\frac{v}{2}\right) (t^2+v)^{\frac{v+1}{2}}} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{v+k+1}{2}\right) \delta^k 2^{\frac{k}{2}} t^k}{\Gamma(k+1) (t^2+v)^{\frac{k}{2}}}.
\end{aligned}$$

On peut donc donner la définition suivante sans ambiguïté.

Définition 13. Soient $Z \sim \mathcal{N}(0,1)$ et $V \sim \chi_v^2$ deux variables aléatoires indépendantes. On dit que $T = \frac{Z+\delta}{\sqrt{\frac{V}{v}}}$, notée $T \sim t_{v,\delta}$, suit une loi décentrée de Student avec paramètres de décentralité $\delta \in \mathbb{R}$ et degrés de liberté $v > 0$ si densité sur \mathbb{R} est donnée par :

$$f_T(t) = \frac{v^{\frac{v}{2}} e^{-\frac{\delta^2}{2}}}{\sqrt{\pi} \Gamma\left(\frac{v}{2}\right) (v+t^2)^{\frac{v+1}{2}}} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{v+k+1}{2}\right) \delta^k 2^{\frac{k}{2}} t^k}{\Gamma(k+1) (v+t^2)^{\frac{k}{2}}}.$$

Remarque 14. Lorsque $\delta = 0$, on retrouve bien la distribution de Student usuelle et la densité en (\star) .

Remarque 15. Comme on peut le voir dans [vAG00], la distribution de Student décentrée demeure une distribution unimodale. De plus, lorsque $\delta \neq 0$, il s'agit d'une distribution asymétrique (voir Figure 1.1) Lorsque $\delta > 0$, c'est la queue droite qui est plus lourde et lorsque $\delta < 0$, c'est la queue gauche qui est plus lourde.

Lemme 16. Soit $T \sim t_{v,\delta}$. Alors, l'espérance et la variance de T sont données respectivement par :

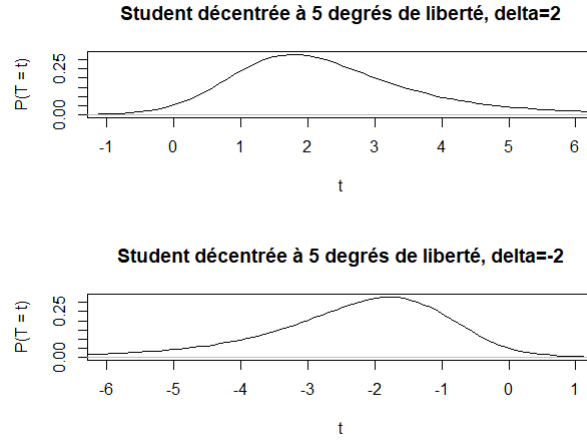


Figure 1.1 – Deux densités de Student décentrées avec 5 degrés de liberté.

$$\mathbb{E}(T) = \delta \sqrt{\frac{v}{2}} \frac{\Gamma\left(\frac{v-1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}, \text{ si } v > 1,$$

$$\mathbb{V}(T) = \frac{v(1 + \delta^2)}{v-2} - \frac{\delta^2 v}{2} \left(\frac{\Gamma\left(\frac{v-1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \right)^2, \text{ si } v > 2.$$

Démonstration. On a

$$\begin{aligned} \mathbb{E}(T) &= \mathbb{E}^U \mathbb{E} \left(\frac{Z + \delta}{\sqrt{\frac{U}{v}}} \middle| U \right) \\ &= \sqrt{v} \delta \mathbb{E} \left(U^{-\frac{1}{2}} \right) \\ &= \delta \sqrt{\frac{v}{2}} \frac{\Gamma\left(\frac{v-1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}, \end{aligned}$$

la dernière égalité étant obtenue en utilisant les résultats sur les moments d'une loi gamma. De plus, on a

$$\begin{aligned}
\mathbb{V}(T) &= \mathbb{E}^U \mathbb{V} \left(\frac{Z + \delta}{\sqrt{U}} \sqrt{v} \middle| U \right) + \mathbb{V}^U \mathbb{E} \left(\frac{Z + \delta}{\sqrt{\frac{U}{v}}} \middle| U \right) \\
&= v \mathbb{E} (U^{-1}) + v\delta^2 \mathbb{V} (U^{-\frac{1}{2}}) \\
&= \frac{v(1 + \delta^2)}{v - 2} + v\delta^2 \left(\mathbb{E} (U^{-1}) - \left(\mathbb{E} (U^{-\frac{1}{2}}) \right)^2 \right) \\
&= \frac{v(1 + \delta^2)}{v - 2} - \frac{\delta^2 v}{2} \left(\frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})} \right)^2,
\end{aligned}$$

la dernière égalité étant obtenue en utilisant les résultats sur les moments d'une loi gamma et par un réarrangement des termes.

□

Remarque 17. *Notons que cette distribution est implémentée dans [R]. En effet, les fonctions `dt`, `pt`, `qt` et `rt` ont un paramètre optionnel, `ncp`, qui étend la définition de la Student à celle de la Student décentrée.*

Enfin, la distribution de Student décentrée peut être utilisée pour déterminer la puissance d'un test de Student, tel que présenté dans [JW40]. On rappelle tout d'abord ce qu'est la puissance d'un test.

Définition 18. *La puissance d'un test γ est la probabilité de rejeter l'hypothèse nulle lorsque que cette hypothèse est incorrecte.*

On considère un test de Student pour les hypothèses suivantes :

$$H_0 : \mu = \mu_0, \sigma^2 > 0$$

$$H_1 : \mu > \mu_0, \sigma^2 > 0$$

Sous H_0 , on suppose que la moyenne d'un échantillon de taille n , \bar{x} , est distribuée selon une loi normale de moyenne μ_0 et de variance $\frac{\sigma^2}{n}$. La statistique de ce test est donnée par $t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$, où s est l'écart-type de l'échantillon. Puisque $t = \frac{\sqrt{n}(\bar{x} - \mu_0)/\sigma}{\sqrt{s^2(n-1)/\sigma^2}}$ sous H_0 , la distribution de cette statistique est celle de Student avec $n - 1$ degrés de liberté. On peut également écrire :

$$\begin{aligned} t &= \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \\ &= \left(\frac{\sqrt{n}(\bar{x} - \mu_1)}{\sigma} + \frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma} \right) \div \sqrt{\frac{s^2}{\sigma^2}}, \end{aligned}$$

et donc que sous H_1 , avec $\mu = \mu_1 > \mu_0$, $t \sim t_{n-1, \delta}$, où $\delta = \frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma}$. De plus, pour un niveau de confiance α fixé, on pose t_0 tel que $\mathbb{P}(t > t_0 | H_0) = \alpha$. On peut alors calculer la puissance de ce test. On a $\mathbb{P}(\text{rejeter } H_0 | H_1) = \mathbb{P}(t_{n-1, \delta} > t_0)$.

1.3.2 La distribution de Student multivariée

En partant de la Définition 11, il est envisageable d'étendre cette définition à un contexte multidimensionnel. Ainsi, si l'on a $Z \sim \mathcal{N}_p(0, \Sigma)$, avec $\Sigma > 0$ et $U \sim \chi_v^2$ deux vecteurs aléatoires indépendantes et que l'on pose $T = \frac{Z}{\sqrt{\frac{U}{v}}} + \mu$, quelle densité obtiendrait-on ? On a, pour $t \in \mathbb{R}^p$:

$$\mathbb{P}(T_i \leq t_i, i = 1, \dots, p) = \int_{\mathbb{R}_+} \mathbb{P}\left(Z_i \leq \sqrt{\frac{u_i}{v}}(t_i - \mu_i) \mid u_i, i = 1, \dots, p\right) \frac{u^{\frac{v}{2}} e^{-\frac{u}{2}}}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} du.$$

En dérivant par rapport à t , on obtient

$$\begin{aligned}
f_T(t) &= \frac{1}{|2\pi\Sigma|^{\frac{1}{2}} 2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right) v^{\frac{p}{2}}} \int_{\mathbb{R}_+} e^{-u\left(\frac{v+(t-\mu)^\top\Sigma^{-1}(t-\mu)}{2v}\right)} u^{\frac{v+p}{2}-1} du \\
&= \frac{1}{|2\pi\Sigma|^{\frac{1}{2}} 2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right) v^{\frac{p}{2}}} \Gamma\left(\frac{v+p}{2}\right) \left(\frac{2v}{v+(t-\mu)^\top\Sigma^{-1}(t-\mu)}\right)^{\frac{v+p}{2}} \\
&= \frac{\Gamma\left(\frac{v+p}{2}\right)}{|\Sigma|^{\frac{1}{2}} \Gamma\left(\frac{v}{2}\right) \pi^{\frac{p}{2}} v^{\frac{p}{2}}} \left(1 + \frac{(t-\mu)^\top\Sigma^{-1}(t-\mu)}{v}\right)^{-\left(\frac{v+p}{2}\right)}, t \in \mathbb{R}^p.
\end{aligned}$$

La définition suivante vient donc naturellement.

Définition 19. Soient $Z \sim \mathcal{N}_p(\mu, \Sigma)$, $\mu \in \mathbb{R}^p$, $\Sigma > 0$, et $U \sim \chi_v^2$, $v > 0$, deux vecteurs indépendants. On dit que la loi de $T = \frac{Z}{\sqrt{\frac{U}{v}}} \sim t_p(v, \mu, \Sigma)$ est Student multidimensionnelle de dimension p avec paramètres de forme v et Σ , où $v > 0$ correspond au degré de liberté, et paramètre de position μ . Sa densité est alors donnée par :

$$f_T(t) = \frac{\Gamma\left(\frac{v+p}{2}\right)}{|\Sigma|^{\frac{1}{2}} \Gamma\left(\frac{v}{2}\right) \pi^{\frac{p}{2}} v^{\frac{p}{2}}} \left(1 + \frac{(t-\mu)^\top\Sigma^{-1}(t-\mu)}{v}\right)^{-\left(\frac{v+p}{2}\right)}.$$

Pour des détails sur la caractérisation de la distribution de Student multivariée, on peut consulter [Lin72].

Lemme 20. Soit $T \sim t_p(v, \mu, \Sigma)$, alors l'espérance et la matrice de covariance de T sont données respectivement par :

$$\begin{aligned}
\mathbb{E}(T) &= \mu, \text{ si } v > 1 \\
\text{Cov}(T) &= \frac{v\Sigma}{v-2}, \text{ si } v > 2.
\end{aligned}$$

1.3.3 La distribution du χ^2 décentrée

Notation 21. *Tout au long de ce mémoire, on notera la fonction indicatrice par $\mathbb{1}_{(A)}(x)$, qui correspond en fait à la fonction*

$$\mathbb{1}_{(A)}(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon.} \end{cases}$$

On rappelle tout d'abord comment obtenir la distribution du χ^2 usuelle.

Définition 22. *Soient $Z_1, \dots, Z_k \sim \mathcal{N}(0, 1)$ des variables aléatoires indépendantes. Alors, on dira que la variable $X = \sum_{i=1}^k Z_i^2 \sim \chi_k^2$, où $k \in \mathbb{N}_*$ est un paramètre de forme et correspond au degré de liberté. De plus, sa densité est donnée par*

$$f_X(x) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{([0, \infty))}(x).$$

Le prochain Lemme est connu sous le nom de la formule de dédoublement. Ce dernier sera utile afin d'obtenir une autre représentation de la densité d'une chi-deux décentrée à un degré de liberté. On définit tout d'abord la notation de Pochhammer, qu'on utilisera dans ce Lemme.

Définition 23. *On note par $(a)_k$ la notation de Pochhammer, ou encore de factoriel ascendant, qui se définit comme suit :*

$$(a)_k = \prod_{j=0}^{k-1} (a + j),$$

pour $k = 1, 2, \dots$ et on définit également $(a)_0 = 1$.

Lemme 24. *On a pour $n \in \mathbb{N}$: $(2n)! = 2^{2n} n! \left(\frac{1}{2}\right)_n$ et aussi $(2n+1)! = 2^{2n} n! \left(\frac{3}{2}\right)_n$.*

Démonstration. On a

$$\begin{aligned}
(2n)! &= 2^n n! (2n-1)(2n-3)\cdots 1 \\
&= 2^{2n} n! \frac{\Gamma\left(n + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \\
&= 2^{2n} n! \left(\frac{1}{2}\right)_n,
\end{aligned}$$

et on a également

$$\begin{aligned}
(2n+1)! &= (2n+1)(2n)! \\
&= (2n+1)2^{2n} n! \left(\frac{1}{2}\right)_n \\
&= (2n+1)2^{2n} n! \frac{\Gamma\left(n + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \\
&= \left(n + \frac{1}{2}\right) 2^{2n} n! \frac{\Gamma\left(n + \frac{1}{2}\right)}{\Gamma\left(\frac{3}{2}\right)} \\
&= 2^{2n} n! \frac{\Gamma\left(n + \frac{3}{2}\right)}{\Gamma\left(\frac{3}{2}\right)} \\
&= 2^{2n} n! \left(\frac{3}{2}\right)_n.
\end{aligned}$$

□

Une fois encore, on peut s'interroger à savoir ce qui se passerait si les différentes variables normales utilisées n'étaient pas de moyenne nulle. Dans un premier temps, que se passe-t-il si on a $Y = X^2$, où $X \sim \mathcal{N}(\mu, 1)$? On a alors

$$\mathbb{P}(Y \leq y) = \Phi(\sqrt{y} - \mu) - \Phi(-\sqrt{y} - \mu).$$

En dérivant par rapport à y , on trouve

$$\begin{aligned}
f_Y(y) &= \frac{y^{-\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{y}{2}} e^{-\frac{\mu^2}{2}} \cosh(\mu\sqrt{y}) \mathbb{1}_{((0,\infty))}(y) \\
&= y^{-\frac{1}{2}} e^{-\frac{y}{2}} e^{-\frac{\mu^2}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\mu^2 y}{2}\right)^k}{k! 2^{k+\frac{1}{2}} \Gamma\left(k + \frac{1}{2}\right)}, \tag{**}
\end{aligned}$$

l'équation (**) étant obtenue en remplaçant la fonction cosh et en utilisant la formule de dédoublement. Cette autre écriture sera réutilisée plus tard.

Remarque 25. L'équation (**) admet la représentation suivante :

$$\begin{aligned}
Y|K = k &\sim \mathcal{G}\left(k + \frac{1}{2}, 2\right) \text{ (ou } \chi_{2k+1}^2), \\
K &\sim \mathcal{P}\left(\frac{\mu^2}{2}\right),
\end{aligned}$$

où $\mathcal{P}(\lambda)$ représente la densité Poisson de paramètre λ .

On peut alors donner la définition suivante.

Définition 26. Soit $X \sim \mathcal{N}(\mu, 1)$. On dit que $Y = X^2$ suit une loi décentrée du chi-deux avec un degré de liberté si sa densité a pour expression

$$f_Y(y) = y^{-\frac{1}{2}} e^{-\frac{y}{2}} e^{-\frac{\mu^2}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\mu^2 y}{2}\right)^k}{k! 2^{k+\frac{1}{2}} \Gamma\left(k + \frac{1}{2}\right)}.$$

Dans un deuxième temps, il est naturel de vouloir étendre cette généralisation à une somme $\sum_{i=1}^r X_i^2$ où les X_i sont i.i.d. $\mathcal{N}(\mu_i, 1)$. On considère donc le cas où $Y = \sum_{i=1}^r Y_i$, où $Y_i \sim \chi_1^2(\delta_i)$, $\delta_i \in \mathbb{R}_+$, $i = 1, \dots, r$. En vertu de la Remarque 25, on sait que la représentation suivante existe ;

$$Y_i | K_i = k_i \sim \mathcal{G}\left(k_i + \frac{1}{2}, 2\right)$$

$$K_i \sim \mathcal{P}\left(\frac{\delta_i}{2}\right).$$

Ainsi, en utilisant les propriétés d'additivité des distributions Gamma et Poisson, on trouve la représentation suivante :

$$T | K = k \sim \mathcal{G}\left(\sum_{i=1}^r k_i + \frac{r}{2}, 2\right)$$

$$K = \sum_{i=1}^r K_i \sim \mathcal{P}\left(\frac{\sum_{i=1}^r \delta_i}{2}\right). \quad (\star \star \star)$$

Enfin, si on écrit la densité associée à ce mélange, il vient

$$f_Y(y) = \sum_{k=0}^{\infty} \frac{e^{-\frac{\delta}{2}} \left(\frac{\delta}{2}\right)^k y^{k+\frac{r}{2}-1} e^{-\frac{y}{2}}}{k! \Gamma\left(k + \frac{r}{2}\right) 2^{k+\frac{r}{2}}},$$

où $\delta = \sum_{i=1}^r \delta_i$. On obtient alors la définition suivante.

Définition 27. Soient $Y_i \sim \chi_1^2(\delta_i)$, $i \geq 1$, des variables aléatoires indépendantes. On dit que $T = \sum_{i=1}^r Y_i$ suit une loi décentrée du chi-deux à r degrés de liberté, si sa densité est donnée par

$$f_T(t) = \sum_{k=0}^{\infty} \frac{e^{-\frac{\sum_{i=1}^r \delta_i}{2}} \left(\frac{\sum_{i=1}^r \delta_i}{2}\right)^k}{k!} \frac{t^{k+\frac{r}{2}-1} e^{-\frac{t}{2}}}{\Gamma\left(k + \frac{r}{2}\right) 2^{k+\frac{r}{2}}}.$$

Remarque 28. On peut également écrire la densité de la Définition 27 de façon à faire ressortir la densité d'une χ^2 standard à $r + 2k$ degrés de liberté. En effet, on a

$$f_T(t) = e^{-\frac{\lambda}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^k}{k!} g_{\chi_{r+2k}^2}(t),$$

dont la fonction de répartition s'exprime sous la forme :

$$F_T(t) = e^{-\frac{\lambda}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^k}{k!} G_{\chi_{r+2k}^2}(t),$$

où $G_{\chi_{r+2k}^2}(t)$ désigne la fonction de répartition d'une loi χ^2 standard à $r + 2k$ degrés de liberté.

Lemme 29. Soit $X \sim \chi_r^2(\lambda)$, alors l'espérance et la variance de X sont données respectivement par :

$$\begin{aligned}\mathbb{E}(X) &= r + \lambda \\ \mathbb{V}(X) &= 2(r + 2\lambda).\end{aligned}$$

Démonstration. On utilise la représentation $(\star \star \star)$. En premier lieu, on a

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}^K \mathbb{E}(T|K) \\ &= \mathbb{E}(2K + r) \\ &= \lambda + r.\end{aligned}$$

En second lieu, on a

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}^K \mathbb{V}(T|K) + \mathbb{V}^K \mathbb{E}(T|K) \\ &= \mathbb{E}\left(4\left(k + \frac{r}{2}\right)\right) + \mathbb{V}(2K + 2r) \\ &= 2(r + 2\lambda).\end{aligned}$$

□

Remarque 30. La distribution du χ^2 décentrée peut également s'obtenir en tant que forme quadratique. En effet, comme on peut le voir dans [Mui09], on a que si $X \sim \mathcal{N}_p(\mu, \Sigma)$, alors

1. $(X - \mu)^\top \Sigma^{-1} (X - \mu) \sim \chi_p^2$,
2. $X^\top \Sigma^{-1} X \sim \chi_p^2(\delta = \mu^\top \Sigma^{-1} \mu)$.

1.3.4 La distribution de Kummer

Définition 31. On note par ${}_1F_1(a, b, z)$ la fonction hypergéométrique confluyente de Kummer de type I, aussi appelée la série hypergéométrique généralisée. La série hypergéométrique généralisée, qui converge pour $z \in \mathbb{R}$, est la suivante :

$${}_1F_1(a, b, z) = \sum_{k \geq 0} \frac{(a)_k z^k}{(b)_k k!},$$

où $(a)_k$ est la notation de Pochhammer, qu'on appelle également la factorielle décroissante.

Définition 32. On note $\Phi(a, b, c)$ la fonction hypergéométrique confluyente de Kummer de type II, qui correspond à la fonction suivante :

$$\Phi(a, b, c) = \frac{1}{\Gamma(a)} \int_{\mathbb{R}_+} t^{a-1} (1+t)^{b-a-1} e^{-ct} dt, a > 0, b \in \mathbb{R}, c > 0.$$

Dans [HV16], on reprend la fonction de la Définition 32 pour en déduire une densité de probabilité.

Définition 33. On dit qu'une variable aléatoire X suit une loi de Kummer de type II de paramètres $a > 0$, $b \in \mathbb{R}$ et $c > 0$, si sa densité est donnée par

$$f_X(x) = \frac{x^{a-1} (1+x)^{-a-b} e^{-cx}}{\Gamma(a) \Phi(a, 1-b, c)} \mathbf{1}_{((0, \infty))}(x).$$

Remarque 34. Si l'on pose $c = 0$ dans la Définition 33, on a

$$f_X(x) = \frac{x^{a-1} (1+x)^{-a-b} \Gamma(a+b)}{\Gamma(a) \Gamma(b)} \mathbf{1}_{((0, \infty))}(x),$$

qui correspond en fait à une densité Bêta de type II de paramètres a et b .

De plus, si l'on pose $b = -a$ dans la Définition 33, on a

$$f_X(x) = \frac{x^{a-1}e^{-cx}}{\Gamma(a)\Phi(a, 1+a, c)} \mathbb{1}_{((0, \infty))}(x) = \frac{x^{a-1}e^{-cx}}{\Gamma(a)c^a},$$

qui correspond à une densité gamma de paramètres a et c .

1.3.5 La convolution

Ici, on définit l'opérateur de convolution, puis on présente une proposition montrant l'utilisation classique de la convolution en probabilités ou en statistique.

Définition 35 ([Bre10], chapitre 4). *On définit $\mathcal{L}^1(\Omega)$ comme étant l'espace des fonctions intégrables de Ω dans \mathbb{R} . De plus, on définit son extension en p dimensions par l'ensemble suivant :*

$$\mathcal{L}^p(\Omega) = \{f : \Omega \rightarrow \mathbb{R} \mid f \text{ est mesurable et } |f|^p \in \mathcal{L}^1(\Omega)\}.$$

Définition 36 ([Bre10], p.104). *Soit $f \in \mathcal{L}^1(\mathbb{R}^n)$ et $g \in \mathcal{L}^p(\mathbb{R}^n)$ avec $1 \leq p \leq \infty$. Alors, pour $x \in \mathbb{R}^n$, la fonction $y \mapsto f(x-y)$ est intégrable sur \mathbb{R}^n presque partout par rapport à la mesure choisie et l'on définit la convolution de f et g par*

$$(f \star g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy$$

Proposition 37. *Soient X de densité f_X et Y de densité g_Y deux variables aléatoires indépendantes. Alors, la variable aléatoire $Z = X + Y$ admet la densité suivante*

$$f_Z(z) = (f_X \star g_Y)(z).$$

Démonstration. Voici une démonstration pour la cas $n = 1$. La proposition étant vraie pour $n > 1$, on réfère le lecteur Pour $z \in \mathbb{R}$, on a :

$$\mathbb{P}(X + Y \leq z) = \int_{-\infty}^{\infty} \mathbb{P}(Y \leq z - x) f_X(x) dx$$

On obtient directement le résultat en dérivant l'équation par rapport à z . □

Remarque 38. *La Proposition 37 est également vraie dans un contexte multivarié et l'idée de la démonstration reste la même.*

Exemple 39. *Soient $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ deux variables aléatoires indépendantes. Soit $Z = X + Y$. Par la Proposition 37, on a que*

$$\begin{aligned} f_Z(z) &= (f_X \star g_Y)(z) \\ &= \int_{\mathbb{R}} \frac{e^{-\frac{(x-t-\mu_X)^2}{2\sigma_X^2}}}{\sqrt{2\pi\sigma_X^2}} \frac{e^{-\frac{(t-\mu_Y)^2}{2\sigma_Y^2}}}{\sqrt{2\pi\sigma_Y^2}} dt \\ &= \frac{e^{-\frac{(x-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}}}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}}, \end{aligned}$$

et donc on a que $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

1.4 L'analyse prédictive

Avant d'aborder les concepts de l'analyse prédictive, on se doit de mentionner que deux écoles de pensée distinctes sont présentes dans le monde de la statistique : l'école fréquentiste, qui est souvent la façon classique de présenter la statistique, ainsi que l'école bayésienne. On rappellera tout d'abord que la statistique est un outil permettant d'inférence, soit d'interpréter, de décrire ou encore d'analyser un phénomène aléatoire. Elle peut également être utile lors de la prédiction pour ce type d'évènements. [Rob06]

Concept. Dans l'approche fréquentiste paramétrique, on a accès à un échantillon (x_1, \dots, x_n) et l'on désire inférer sur les paramètres θ , qui sont inconnus, permettant de modéliser la situation. On dira que la densité de ce modèle est donnée par une certaine fonction $f(x|\theta)$ et que cette dernière devrait correspondre à la fonction de vraisemblance $L(\theta|x)$. [CB02] Dans ce contexte, les tests d'hypothèses faisant intervenir des p -valeurs sont omniprésents, ainsi que les méthodes d'estimation basées sur le maximum de vraisemblance. [CL08]

Concept. Dans l'approche bayésienne, on affecte une distribution, appelée loi a priori, à certaines quantités inconnues du modèle. [CL08] Le choix de cette distribution peut être justifié par l'intuition d'experts, par des expériences antérieures ou encore il s'agira d'un choix par défaut (on parlera dans ce cas de loi a priori non informative). Encore une fois, on a accès à un échantillon (x_1, \dots, x_n) et l'on désire inférer sur les paramètres θ inconnus. Par la suite, on utilisera le théorème de Bayes (d'où l'appellation) afin d'obtenir ce qu'on appelle une loi a posteriori, laquelle nous permettra d'inférer sur les paramètres θ du modèle.

Voici un exemple permettant de comparer les deux approches.

Exemple 40. On considère un échantillon $x = (x_1, \dots, x_n)$ où $x_i \sim \text{Ber}(\theta)$. Du côté fréquentiste, une approche consiste à considérer comme estimé de θ la valeur qui maximise la vraisemblance. On a donc

$$\begin{aligned}
L(\theta|x) &= \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} \\
&= \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i},
\end{aligned}$$

$$\begin{aligned}
(\log L(\theta|x))' &= l(\theta|x) \\
&= \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1-\theta}.
\end{aligned}$$

Si on résout l'équation $l(\theta|x) = 0$, on trouve aisément que $\theta = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$.

Considérons maintenant l'approche bayésienne. Un choix en absence d'information sur θ serait de poser $\theta \sim U(0,1)$. On trouve alors que la loi a posteriori est donnée par

$$\begin{aligned}
\pi(\theta|x) &= \frac{\pi(\theta) \prod_{i=1}^n f(x_i|\theta)}{\int_0^1 \pi(\theta) \prod_{i=1}^n f(x_i|\theta) d\theta} \\
&= \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{\int_0^1 \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} d\theta} \\
&= \frac{\theta^{\sum_{i=1}^n x_i + 1} (1-\theta)^{n-\sum_{i=1}^n x_i + 1}}{B(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1)}, \text{ (où } B(\cdot, \cdot) \text{ est la fonction Bêta)}
\end{aligned}$$

ce qui correspond à une densité Bêta($\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1$).

On trouve donc du côté fréquentiste l'estimé $\frac{\sum_{i=1}^n x_i}{n}$. Cet estimé est sans biais, c'est-à-dire que $\mathbb{E}_\theta \left(\frac{\sum_{i=1}^n x_i}{n} \right) = \theta$ et sa variance est $\mathbb{V} \left(\frac{\sum_{i=1}^n x_i}{n} \right) = \frac{\theta(1-\theta)}{n}$. Du côté bayésien on trouve plutôt une densité et la variabilité de θ s'exprime à travers cette dernière. Remarquons que la moyenne de cette densité correspond à $\frac{\sum_{i=1}^n x_i + 1}{n+2}$ et donc qu'à la limite, la différence entre ces deux résultats est nulle. Ceci est dû au fait que la loi a posteriori résume l'information sur θ ; elle incorpore le choix a priori et l'observé x .

Dans ce mémoire, c'est l'approche bayésienne qui sera utilisée. On formalise le contexte bayésien en donnant la définition d'une règle de Bayes.

Définition 41. *Pour une loi a priori $\pi(\theta)$ et une fonction de perte $L(\cdot, \cdot)$, on dit que δ_π est une règle de Bayes si $\forall x \in \mathbb{X}$, $\delta_\pi(x)$ minimise $\mathbb{E}(L(\theta, d)|x)$ pour un certain $d \in D$, où D est l'espace des décisions.*

Enfin, la statistique bayésienne permet de se pencher sur la problématique de la prédiction d'une façon très intuitive. On supposera ici qu'on travaille avec les variables aléatoires $X \sim p_\theta$ et $Y \sim q_\theta$, variables indépendantes conditionnellement à θ , c'est-à-dire que $\mathbb{P}(X = x, Y = y|\theta) = \mathbb{P}(X = x|\theta)\mathbb{P}(Y = y|\theta)\forall x, y$. Le but de l'analyse prédictive est de tirer avantage des observations disponibles, ici issues de la densité p_θ afin de faire une prédiction pour la variable Y qui prend en compte la variabilité du paramètre θ . Encore mieux, on estimera la densité prédictive de Y , qui correspond à la distribution d'une innovation conditionnellement à l'observation d'un échantillon x , notée $\hat{q}_\theta(y|x)$. En effet, en supposant que la loi a priori $\pi(\theta)$ est absolument continue par rapport à une mesure σ -finie $\nu(\theta)$, on a

$$\begin{aligned}\hat{q}_\theta(y|x) &= \int_{\Theta} q(y, \theta|x) d\nu(\theta) \\ &= \int_{\Theta} q(y|\theta, x)\pi(\theta|x) d\nu(\theta) \\ &= \int_{\Theta} q(y|\theta)\pi(\theta|x) d\nu(\theta).\end{aligned}$$

Exemple 42. *Dans le contexte de l'Exemple 40, calculons la densité prédictive d'une nouvelle observation $y \sim \text{Ber}(\theta)$. On a*

$$\begin{aligned}\hat{q}(y|x) &= \int_0^1 \theta^y (1-\theta)^{1-y} \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{B(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1)} d\theta \\ &= \frac{B(\sum_{i=1}^n x_i + y + 1, n + 1 - \sum_{i=1}^n x_i - y + 1)}{B(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1)}, y = 0, 1.\end{aligned}$$

La densité prédictive de y est donc de loi Bêta-Binomiale($1, \sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1$).

CHAPITRE 2

Inférence a posteriori et prédictive sur différents modèles

Dans le monde de la statistique, plusieurs modèles sont disponibles afin de modéliser un phénomène. Certains de ces modèles, plus standards, ont été étudiés sous plusieurs angles différents, que ce soit en considérant tous les paramètres comme étant inconnus ou encore en utilisant une information disponible afin de réduire la complexité du modèle. Dans la première section de ce chapitre, on considère les modèles normal et gamma lorsqu'un seul paramètre est inconnu. Ceci permettra de faire une brève revue de la littérature et aussi de mieux comprendre les outils à notre disposition lorsqu'on fait de l'inférence bayésienne. Différents cas limites seront développés et quelques résultats concernant la loi du χ^2 décentrée seront abordés, démontrant l'importance de cette distribution. On terminera le chapitre par une section abordant le modèle normal-gamma en version multivariée, où cette fois, on aura deux paramètres inconnus. Les différentes propriétés de ce modèle seront exploitées au chapitre 4 se qui viendra alléger les calculs.

2.1 Modèle avec un seul paramètre inconnu

Nous verrons dans cette section différents résultats lorsque le paramètre de position pour une densité normale est inconnu puis lorsque le paramètre d'échelle est inconnu dans le cas d'une densité gamma. Nous présenterons également quelques exemples intéressants pour chacun de ces cas.

2.1.1 Modèle normal avec variance connue

Dans cette sous-section, on s'intéresse à l'inférence a posteriori pour θ et l'analyse prédictive pour le modèle

$$X \sim \mathcal{N}_p(\theta, \sigma_x^2 \mathbf{1}_p), Y \sim \mathcal{N}_p(\theta, \sigma_y^2 \mathbf{1}_p) \quad (\star)$$

avec σ_x^2 et σ_y^2 connues, et $p \geq 1$.

Tout d'abord, on introduit un lemme qui nous sera utile pour plusieurs calculs dans cette section.

Lemme 43. $\int_{\mathbb{R}^p} e^{\frac{-(x-\theta)^\top(x-\theta)}{2\sigma^2}} e^{\frac{-(y-\theta)^\top(y-\theta)}{2\tau^2}} d\theta = e^{\frac{-(x-y)^\top(x-y)}{2(\tau^2+\sigma^2)}} \frac{\sqrt{2\pi}}{\sqrt{\frac{\tau^2+\sigma^2}{\tau^2\sigma^2}}}.$

Démonstration. On a que

$$\begin{aligned} \frac{(x-\theta)^\top(x-\theta)}{\sigma^2} + \frac{(y-\theta)^\top(y-\theta)}{\tau^2} &= \left(\frac{\tau^2 + \sigma^2}{\tau^2\sigma^2} \right) \left(\theta - \frac{x\tau^2 + y\sigma^2}{\sigma^2 + \tau^2} \right)^\top \left(\theta - \frac{x\tau^2 + y\sigma^2}{\sigma^2 + \tau^2} \right) \\ &\quad + \frac{(x-y)^\top(x-y)}{2(\sigma^2 + \tau^2)}. \end{aligned}$$

Ainsi, on obtient

$$\begin{aligned}
\int_{\mathbb{R}^p} e^{-\frac{(x-\theta)^\top(x-\theta)}{2\sigma^2}} e^{-\frac{(y-\theta)^\top(y-\theta)}{2\tau^2}} d\theta &= e^{-\frac{(x-y)^\top(x-y)}{2(\tau^2+\sigma^2)}} \int_{\mathbb{R}^p} e^{-\frac{1}{2}\left(\frac{\tau^2+\sigma^2}{\tau^2\sigma^2}\right)\left(\theta-\frac{x\tau^2+y\sigma^2}{\sigma^2+\tau^2}\right)^\top\left(\theta-\frac{x\tau^2+y\sigma^2}{\sigma^2+\tau^2}\right)} d\theta \\
&= e^{-\frac{(x-y)^\top(x-y)}{2(\tau^2+\sigma^2)}} \frac{\sqrt{2\pi}}{\sqrt{\frac{\tau^2+\sigma^2}{\tau^2\sigma^2}}}.
\end{aligned}$$

□

Lemme 44. *Pour le modèle (\star) et pour l'a priori uniforme $\pi_U(\theta) \equiv 1$, la densité a posteriori de θ est de loi normale $\pi(\theta|x) \sim \mathcal{N}_p(x, \sigma_x^2 \mathbf{1}_p)$ et la densité prédictive a posteriori de Y est également de loi normale $\hat{q}_U(y|x) \sim \mathcal{N}_p((x, (\sigma_x^2 + \sigma_y^2) \mathbf{1}_p))$.*

Démonstration. Dans un premier temps, on a que

$$\begin{aligned}
\pi(\theta|x) &\propto f(x|\theta)\pi(\theta) \\
&\propto e^{-\frac{1}{2}(\theta-x)^\top(\sigma_x^2 \mathbf{1}_p)^{-1}(\theta-x)}.
\end{aligned}$$

Dans un deuxième temps, on a que

$$\begin{aligned}
\hat{q}_U(y|x) &= \int_{\mathbb{R}^p} p(y|\theta)\pi(\theta|x)d\theta \\
&= \frac{1}{(2\pi\sigma_x\sigma_y)^p} \int_{\mathbb{R}^p} e^{-\frac{1}{2\sigma_y^2}(y-\theta)^\top(y-\theta)} e^{-\frac{1}{2\sigma_x^2}(x-\theta)^\top(x-\theta)} d\theta \\
&= \frac{e^{-\frac{1}{2(\sigma_x^2+\sigma_y^2)}(y-x)^\top(y-x)}}{(2\pi(\sigma_x^2 + \sigma_y^2))^{\frac{p}{2}}},
\end{aligned}$$

la dernière égalité étant obtenue en appliquant le lemme 43.

□

Il est également intéressant de considérer une borne inférieure comme contrainte pour θ dans le cas univarié. La présence de contraintes paramétriques survient dans plusieurs contextes (voir par exemple [M⁺02]) et a fait l'objet de nombreux travaux (comme par exemple [DVE09] et [MS04]).

Proposition 45. Pour le modèle $\bar{X}|\theta \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$ et l'a priori $\theta \sim \mathcal{N}(\mu, \tau^2)\mathbb{1}_{\mathbb{R}_+}(\theta)$, la densité a posteriori de θ est de loi normale tronquée sur \mathbb{R}_+ $\mathcal{N}(\mu', \tau'^2)$ et la densité prédictive a posteriori de $Y \sim \mathcal{N}(\theta, \sigma^2)$ est égale à :

$$\hat{q}_\pi(y|\bar{x}) = \frac{\frac{e^{-\frac{(y-\mu')^2}{2(\tau'^2+\sigma^2)}}}{\sqrt{2\pi(\tau'^2+\sigma^2)}} \left(1 - \Phi\left(\frac{-M}{S}\right)\right),$$

$$\text{où } \mu' = \frac{\bar{x}\tau^2 + \mu\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}, \quad \tau'^2 = \frac{\tau^2\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}, \quad M = \frac{y\tau'^2 + \mu'\sigma^2}{\tau'^2 + \sigma^2} \text{ et } S^2 = \frac{\tau'^2\sigma^2}{\tau'^2 + \sigma^2}.$$

Démonstration. Comme la fonction indicatrice ne dépend pas des données, on déduit le résultat pour la densité a posteriori de θ avec le Lemme 44 et en prenant $p = 1$.

Pour la densité prédictive a posteriori de Y , il suffit d'appliquer dans un premier temps le Lemme 43, puis d'utiliser la relation entre les fonctions de densité et de répartition de la loi normale centrée et réduite. On a que

$$\begin{aligned} \hat{q}_\pi(y|\bar{x}) &= \int_{\mathbb{R}_+} \frac{e^{-\frac{(y-\theta)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \frac{e^{-\frac{(\theta-\mu')^2}{2\tau'^2}}}{\sqrt{2\pi}\tau'} d\theta \\ &= \frac{e^{-\frac{(y-\mu')^2}{2(\tau'^2+\sigma^2)}}}{2\pi\sigma\tau'} \int_{\mathbb{R}_+} e^{-\frac{1}{2}\left(\frac{\tau'^2+\sigma^2}{\tau'^2\sigma^2}\right)\left(\theta - \frac{y\tau'^2 + \mu'\sigma^2}{\tau'^2 + \sigma^2}\right)^2} d\theta \\ &= \frac{e^{-\frac{(y-\mu')^2}{2(\tau'^2+\sigma^2)}}}{\sqrt{2\pi}\sigma\tau'} S \int_{\mathbb{R}_+} \frac{\phi\left(\frac{\theta-M}{S}\right)}{S} d\theta \\ &= \frac{e^{-\frac{(y-\mu')^2}{2(\tau'^2+\sigma^2)}}}{\sqrt{2\pi(\tau'^2+\sigma^2)}} \left(1 - \Phi\left(\frac{-M}{S}\right)\right). \end{aligned}$$

□

Théorème 2.1. [GLX⁺06] Pour $X|\theta \sim \mathcal{N}_p(\theta, \sigma_x^2\mathbb{1}_p)$, $Y|\theta \sim \mathcal{N}_p(\theta, \sigma_y^2\mathbb{1}_p)$ prises comme étant conditionnellement indépendantes et n'importe quelle loi a priori $\pi(\theta)$ absolument

continue par rapport à une mesure σ -finie $\nu(\theta)$, à condition que la loi a posteriori existe, on a que la densité prédictive a posteriori de Y est donnée par

$$\hat{q}_\pi(y|x) = \frac{m_{Z,\pi}(z)}{m_{X,\pi}(x)} \hat{q}_U(y|x),$$

où $\hat{q}_U(y|x)$ est donnée dans le Lemme 44, $Z = \frac{X\sigma_y^2 + Y\sigma_x^2}{\sigma_x^2 + \sigma_y^2}$ et $m_{W,\pi}(w)$ est la marginale de W donnée par $\int f(w|\theta)\pi(\theta)d\nu(\theta)$.

Démonstration. Remarquons tout d'abord que $\hat{q}_\pi(y|x) = \frac{f(x,y)}{m_{X,\pi}(x)}$. Il suffit alors de calculer la densité conjointe de X et Y . On a que

$$\begin{aligned} f(x,y) &= \int_{\Theta} f(x|\theta)f(y|\theta)\pi(\theta)d\nu(\theta) \\ &= \frac{e^{\frac{-1}{2(\sigma_x^2 + \sigma_y^2)}(y-x)^\top(y-x)}}{(2\pi\sigma_x\sigma_y)^p} \int_{\Theta} e^{\frac{-1}{2}\left(\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2\sigma_y^2}\right)\left(\theta - \frac{x\sigma_y^2 + y\sigma_x^2}{\sigma_x^2 + \sigma_y^2}\right)^2} \pi(\theta)d\nu(\theta) \\ &= \frac{e^{\frac{-1}{2(\sigma_x^2 + \sigma_y^2)}(y-x)^\top(y-x)}}{(2\pi(\sigma_x^2 + \sigma_y^2))^{\frac{p}{2}}} \int_{\Theta} f(z|\theta)\pi(\theta)d\nu(\theta) \\ &= \hat{q}_U(y|x)m_{Z,\pi}(z). \end{aligned}$$

On obtient l'avant dernière équation en remarquant que $\mathbb{V}(Z) = \frac{\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2}$ et en ajoutant les termes manquants pour obtenir la densité de $Z|\theta$.

□

Remarque 46. Dans le Théorème précédent, rien n'exclut qu'on utilise une loi a priori discrète, ou de façon plus générale une densité a priori qui n'est pas absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^p .

Exemple 47. Trouvons la densité prédictive a posteriori de Y dans le cadre du Théorème 2.1 avec l'a priori $\pi(\theta) \sim \mathcal{N}_p(\mu, \tau^2 \mathbf{1}_p)$, $\|\theta\| \leq m$ où $m > 0$, c'est-à-dire que la norme

de θ est plus petite ou égale à une constante positive m . Commençons par calculer la marginale de X . En posant C_θ comme étant la constante de normalisation de la normale tronquée, on a que

$$\begin{aligned} \frac{m_{X,\pi}(x)}{C_\theta} &= \int_{\mathbb{R}^p} \frac{e^{\frac{-1}{2\sigma_x^2}(x-\theta)^\top(x-\theta)}}{(2\pi\sigma_x^2)^{\frac{p}{2}}} \frac{e^{\frac{-1}{2\tau^2}(\theta-\mu)^\top(\theta-\mu)}}{(2\pi\tau^2)^{\frac{p}{2}}} \mathbf{1}_{(B_m)}(\theta) d\theta, \text{ où } B_m = \{\theta \in \mathbb{R}^p : \|\theta\| \leq m\}, \\ &= \frac{e^{\frac{-1}{2(\tau^2+\sigma_x^2)}(x-\mu)^\top(x-\mu)}}{(2\pi\sigma_x\tau)^p} \int_{\mathbb{R}^p} e^{\frac{-1}{2}\left(\frac{\tau^2+\sigma_x^2}{\tau^2\sigma_x^2}\right)\left(\theta-\frac{x\tau^2+\mu\sigma_x^2}{\sigma_x^2+\tau^2}\right)^\top\left(\theta-\frac{x\tau^2+\mu\sigma_x^2}{\sigma_x^2+\tau^2}\right)} \mathbf{1}_{(B_m)}(\theta) d\theta. \end{aligned}$$

L'intégrale présente à la dernière ligne peut être réécrite comme étant une probabilité, soit $\mathbb{P}(\|\theta\| \leq m|X)$, où $\theta \sim \mathcal{N}_p\left(\frac{x\tau^2+\mu\sigma_x^2}{\sigma_x^2+\tau^2}, \frac{\tau^2\sigma_x^2}{\sigma_x^2+\tau^2}\mathbf{1}_p\right)$. Or, on a que

$$\frac{\theta}{\sqrt{\tau'}} \sim \mathcal{N}_p\left(\frac{\mu'}{\sqrt{\tau'}}, \mathbf{1}_p\right) \Rightarrow \frac{\|\theta\|^2}{\tau'} \sim \chi_p^2\left(\frac{\|\mu'\|^2}{\tau'}\right),$$

avec $\mu' = \frac{x\tau^2+\mu\sigma_x^2}{\sigma_x^2+\tau^2}$ et $\tau' = \frac{\tau^2\sigma_x^2}{\sigma_x^2+\tau^2}$. Ainsi, si l'on note par $F_{k,\Delta}(\cdot)$ la fonction de répartition d'une loi $\chi_k^2(\Delta)$, on trouve que

$$\mathbb{P}(\|\theta\| \leq m|X) = \mathbb{P}\left(\frac{\|\theta\|^2}{\tau'} \leq \frac{m^2}{\tau'} \middle| X\right) = F_{p, \frac{\|\mu'\|^2}{\tau'}}\left(\frac{m^2}{\tau'}\right),$$

et l'on obtient finalement que $\frac{m_{X,\pi}(x)}{C_\theta} = \frac{e^{\frac{-1}{2(\tau^2+\sigma_x^2)}(x-\mu)^\top(x-\mu)}}{(2\pi\sigma_x\tau)^p} F_{p, \frac{\|\mu'\|^2}{\tau'}}\left(\frac{m^2}{\tau'}\right)$.

De même, on trouve que la loi marginale de Z est

$$\frac{m_{Z,\pi}(z)}{C_\theta} = \frac{e^{\frac{-1}{2(\tau^2+\sigma_z^2)}(z-\mu)^\top(z-\mu)}}{(2\pi\sigma_z\tau)^p} F_{p, \frac{\|\mu'_z\|^2}{\tau'_z}}\left(\frac{m^2}{\tau'_z}\right),$$

où $\mu'_z = \frac{z\tau^2+\mu\sigma_z^2}{\sigma_z^2+\tau^2}$ et $\tau'_z = \frac{\tau^2\sigma_z^2}{\sigma_z^2+\tau^2}$.

On trouve alors que

$$\hat{q}_\pi(y|x) = e^{-\frac{1}{2} \frac{(\tau^2 + \sigma_x^2) \left\| y - \frac{\mu \sigma_x^2 + \tau^2 x}{\tau^2 + \sigma_x^2} \right\|^2}{(\tau^2 + \sigma_y^2) \sigma_x^2 + \tau^2 \sigma_y^2}} \left(\frac{\tau^2 + \sigma_x^2}{2\pi((\tau^2 + \sigma_y^2) \sigma_x^2 + \tau^2 \sigma_y^2)} \right)^{\frac{p}{2}} \frac{F_{p, \frac{\|\mu'_z\|^2}{\tau'^2}} \left(\frac{m^2}{\tau'^2} \right)}{F_{p, \frac{\|\mu'\|^2}{\tau'^2}} \left(\frac{m^2}{\tau'^2} \right)},$$

qui est tout simplement une densité normale $\mathcal{N}_p \left(\mu_y = \frac{\mu \sigma_x^2 + \tau^2 x}{\tau^2 + \sigma_x^2}, \tau_y = \left(\frac{\tau^2 + \sigma_x^2}{(\tau^2 + \sigma_y^2) \sigma_x^2 + \tau^2 \sigma_y^2} \right)^{-1} \mathbf{1}_p \right)$ pondérée. De plus, on observe que

$$\begin{aligned} \lim_{m \rightarrow \infty} \hat{q}_\pi(y|x) &\sim \mathcal{N}_p(\mu_y, \tau_y \mathbf{1}_p), \\ \lim_{\tau \rightarrow \infty} \hat{q}_\pi(y|x) &= e^{-\frac{\|y-x\|^2}{2(\sigma_x^2 + \sigma_y^2)}} \left(\frac{1}{2\pi(\sigma_x^2 + \sigma_y^2)} \right)^{\frac{p}{2}} \frac{F_{p, \frac{\|z\|^2}{\sigma_z^2}} \left(\frac{m^2}{\sigma_z^2} \right)}{F_{p, \frac{\|x\|^2}{\sigma_x^2}} \left(\frac{m^2}{\sigma_x^2} \right)}, \\ \lim_{\tau, m \rightarrow \infty} \hat{q}_\pi(y|x) &= \hat{q}_U(y|x). \end{aligned}$$

Dans le second cas, lorsque $\tau \rightarrow \infty$, on a que la densité a priori π se comporte comme une densité uniforme sur la boule $B_m = \{\theta \in \mathbb{R}^p : \|\theta\| \leq m\}$. Ainsi, dans le cas où $p = 1$, on peut alors réécrire la densité obtenue ci-dessus de manière équivalente :

$$\lim_{\tau \rightarrow \infty} \hat{q}_\pi(y|x) = \frac{e^{-\frac{(y-x)^2}{2(\sigma_x^2 + \sigma_y^2)}}}{2\pi(\sigma_x^2 + \sigma_y^2)} \frac{\Phi\left(\frac{m}{\sigma_z} - \sqrt{\frac{\|z\|}{\sigma_z^2}}\right) - \Phi\left(-\frac{m}{\sigma_z} - \sqrt{\frac{\|z\|}{\sigma_z^2}}\right)}{\Phi\left(\frac{m}{\sigma_x} - \sqrt{\frac{\|x\|}{\sigma_x^2}}\right) - \Phi\left(-\frac{m}{\sigma_x} - \sqrt{\frac{\|x\|}{\sigma_x^2}}\right)}.$$

Le résultat qui suit, présenté dans [Str03], utilise une approche quelque peu différente. Tout d'abord, introduisons le concept de règle de Bayes ainsi que l'identité de Stein [Ste81].

Lemme 48 (Identité de Stein). *Soient $X \sim \mathcal{N}_p(\theta, \sigma^2 \mathbf{1}_p)$ et $g(x) = (g_1(x), \dots, g_p(x))^\top$ une fonction de \mathbb{R}^p dans \mathbb{R}^p telle que $\mathbb{E}(\|g(X)\|^2) < \infty$. Alors, $\mathbb{E}((X - \theta)^\top g(X)) = \sigma^2 \mathbb{E}(\operatorname{div} g(X))$.*

Démonstration. On commence par considérer le cas univarié $X \sim \mathcal{N}(\theta, \sigma^2)$. On a

$$\begin{aligned}\mathbb{E}((X - \theta)g(X)) &= \int_{\mathbb{R}} (x - \theta) \frac{e^{-\frac{(x-\theta)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} g(x) dx \\ &= -\sigma^2 \frac{e^{-\frac{(x-\theta)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} g(x) \Bigg|_{-\infty}^{\infty} + \sigma^2 \int_{\mathbb{R}} \frac{e^{-\frac{(x-\theta)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} g'(x) dx \\ &= 0 + \sigma^2 \mathbb{E}(g'(X)).\end{aligned}$$

Pour le cas $X \sim \mathcal{N}_p(\theta, \sigma^2 \mathbf{1}_p)$, on a

$$\begin{aligned}\mathbb{E}((X - \theta)^\top g(X)) &= \mathbb{E}\left(\sum_{i=1}^p (X_i - \theta_i) g_i(X_i)\right) \\ &= \sum_{i=1}^p \mathbb{E}((X_i - \theta_i) g_i(X_i)) \\ &= \sigma^2 \sum_{i=1}^p \mathbb{E}\left(\frac{\partial}{\partial X_i} g_i(X_i)\right), \text{ qui correspond au cas univarié} \\ &= \sigma^2 \mathbb{E}\left(\sum_{i=1}^p \frac{\partial}{\partial X_i} g_i(X_i)\right) \\ &= \sigma^2 \mathbb{E}(\operatorname{div} g(X)).\end{aligned}$$

□

La représentation qui suit pour la règle de Bayes peut être retrouvée dans [GMS87].

Théorème 2.2. *Soit $X|\theta \sim \mathcal{N}_p(\theta, \sigma^2 \mathbf{1}_p)$ et supposons également que l'on dispose de la fonction de perte quadratique $L(\theta, d) = \|d - \theta\|^2$. Alors, pour n'importe quelle loi a priori $\pi(\theta)$, la règle de Bayes, qui correspond dans ce cas à l'espérance a posteriori de θ lorsque $\mathbb{V}(\theta|x)$ existe, est donnée par*

$$\delta_\pi(x) = \mathbb{E}(\theta|x) = x + \sigma^2 \frac{\nabla m(x)}{m(x)},$$

où $m(x) = \int_{\Theta} \frac{1}{\sigma^p} \phi\left(\frac{x-\theta}{\sigma}\right) \pi(\theta) d\nu(\theta)$ et ∇ désigne le gradient d'un vecteur par rapport à x .

Démonstration. Remarquons tout d'abord qu'on peut écrire $\mathbb{E}(\theta|x) = \mathbb{E}(x + \theta - x|x) = x + \mathbb{E}(\theta - x|x)$. De plus, on a pour $i = 1, \dots, p$, en utilisant le fait que $\phi'(t) = -t\phi(t)$:

$$\begin{aligned} \sigma^2 \frac{\frac{\partial}{\partial x_i} m(x)}{m(x)} &= \sigma^2 \frac{\int_{\Theta} \frac{\partial}{\partial x_i} \left(\frac{\phi\left(\frac{x-\theta}{\sigma}\right) \pi(\theta)}{\sigma^p} \right) d\nu(\theta)}{m(x)} \\ &= \sigma^2 \int_{\Theta} \frac{(\theta_i - x_i) \phi\left(\frac{x-\theta}{\sigma}\right) \pi(\theta)}{\sigma^{p+2} m(x)} d\nu(\theta) \\ &= \int_{\Theta} (\theta_i - x_i) \pi(\theta|x) d\nu(\theta) \\ &= \mathbb{E}[(\theta_i - x_i)|x]. \end{aligned}$$

□

Les résultats techniques qui suivent seront primordiaux pour les prochains exemples.

Lemme 49. Soit $F_{k,\lambda}$ la fonction de répartition d'une $\chi_k^2(\lambda)$ avec $k > 0$ et $\lambda > 0$. Alors, la dérivée partielle de cette fonction par rapport à λ est donnée par

$$\frac{\partial F_{k,\lambda}(x)}{\partial \lambda} = \frac{F_{k+2,\lambda}(x) - F_{k,\lambda}(x)}{2}, \lambda > 0, x > 0.$$

Démonstration. On a, par la représentation donnée à la Remarque 28, que $F_{k,\lambda}(x) = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{(\frac{\lambda}{2})^j}{j!} G_{k+2j}(x)$, où G_k est la répartition d'une χ_k^2 . Alors,

$$\begin{aligned} \frac{\partial F_{k,\lambda}(x)}{\partial \lambda} &= \frac{-e^{-\frac{\lambda}{2}}}{2} \sum_{j=0}^{\infty} \frac{(\frac{\lambda}{2})^j}{j!} G_{k+2j}(x) + e^{-\frac{\lambda}{2}} \sum_{j=1}^{\infty} \frac{(\frac{\lambda}{2})^{j-1}}{2(j-1)!} G_{k+2j}(x) \\ &= \frac{-F_{k,\lambda}(x)}{2} + e^{-\frac{\lambda}{2}} \sum_{l=0}^{\infty} \frac{(\frac{\lambda}{2})^l}{2(l)!} G_{k+2+2l}(x) \\ &= \frac{F_{k+2,\lambda}(x) - F_{k,\lambda}(x)}{2}. \end{aligned}$$

□

Les deux prochains corollaires peuvent se déduire directement du Lemme 49. Ils s'avèrent utiles si l'on désire obtenir des formes analytiques pour des densités du χ^2 décentrées où le degré de liberté est un nombre impair (voir Exemple 51).

Corollaire 50. [Coh88] Soient $f_{k,\lambda}(x)$ la fonction de densité d'une $\chi_k^2(\lambda)$ et $g_{k,\lambda}(x) = e^{\frac{x+\lambda}{2}} f_{k,\lambda}(x)$, où $k > 0$. Alors, la dérivée partielle de f par rapport à λ est donnée par

$$\frac{\partial}{\partial \lambda} f_{k,\lambda}(x) = \frac{f_{k+2,\lambda}(x) - f_{k,\lambda}(x)}{2}.$$

De plus, la dérivée partielle de g par rapport à λ est donnée par

$$\frac{\partial}{\partial \lambda} g_{k,\lambda}(x) = \frac{g_{k+2,\lambda}(x)}{2}.$$

Démonstration. La première équation s'obtient directement en utilisant la relation présentée au Lemme 49 et en prenant la dérivée partielle par rapport à x de chaque côté. Pour ce qui est de la deuxième expression, on a

$$\begin{aligned} \frac{\partial}{\partial \lambda} g_{k,\lambda}(x) &= \frac{\partial}{\partial \lambda} \left(e^{\frac{x+\lambda}{2}} f_{k,\lambda}(x) \right) \\ &= \frac{e^{\frac{x+\lambda}{2}}}{2} f_{k+2,\lambda}(x) + e^{\frac{x+\lambda}{2}} \left(\frac{f_{k+2,\lambda}(x) - f_{k,\lambda}(x)}{2} \right) \\ &= \frac{e^{\frac{x+\lambda}{2}} f_{k+2,\lambda}(x)}{2} \\ &= \frac{g_{k+2,\lambda}(x)}{2}. \end{aligned}$$

□

Voici un court exemple qui montre comment on peut utiliser quelques-unes des relations que l'on vient d'établir.

Exemple 51. Comme on peut écrire,

$$f_{1,\lambda}(x) = \frac{x^{-\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{x}{2}} e^{-\frac{\lambda}{2}} \cosh(\sqrt{x\lambda}), \text{ (voir la section 1.3.3)}$$

et que

$$\begin{aligned} \frac{\partial}{\partial \lambda} g_{1,\lambda}(x) &= \frac{\partial}{\partial \lambda} \frac{\cosh(\sqrt{x\lambda})}{\sqrt{2\pi x}} \\ &= \frac{\sinh(\sqrt{x\lambda})}{2\sqrt{2\pi\lambda}}, \end{aligned}$$

on trouve en utilisant le Corollaire 50 que $f_{3,\lambda}(x) = \frac{e^{-\frac{x}{2}} e^{-\frac{\lambda}{2}} \sinh(\sqrt{x\lambda})}{\sqrt{2\pi\lambda}}$.

Lemme 52. Soient $F_{k,\lambda}(x)$ et $f_{k,\lambda}(x)$ les fonctions de répartition et de densité d'une $\chi_k^2(\lambda)$. Alors, on a la relation suivante :

$$\frac{F_{k+2,\lambda}(x) - F_{k,\lambda}(x)}{2} = -f_{k+2,\lambda}(x). \quad (2.1)$$

Démonstration. On a tout d'abord que

$$\begin{aligned} F_{k+2,\lambda}(x) - F_{k,\lambda}(x) &= e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^j}{j!} [G_{k+2+2j}(x) - G_{k+2j}(x)] \\ -2f_{k+2,\lambda}(x) &= e^{-\frac{\lambda}{2}} \sum_{i=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^i}{i!} \left[-2 \frac{x^{i+\frac{k}{2}} e^{-\frac{x}{2}}}{\Gamma\left(\frac{k}{2} + i + 1\right) 2^{\frac{k}{2}+i+1}} \right]. \end{aligned}$$

Ainsi, il suffit de montrer que

$$G_{k+2+2j,\lambda}(x) - G_{k+2j,\lambda}(x) = -2 \frac{x^{j+\frac{k}{2}} e^{-\frac{x}{2}}}{\Gamma\left(\frac{k}{2} + j + 1\right) 2^{\frac{k}{2}+j+1}}. \quad (\bullet)$$

Or, on a que

$$\begin{aligned}
G_p(x) &= \int_0^{\frac{x}{2}} \frac{t^{\frac{p}{2}-1} e^{-t}}{\Gamma\left(\frac{p}{2}\right)} dt \\
&= \frac{-e^{-\frac{x}{2}} \left(\frac{x}{2}\right)^{\frac{p}{2}-1}}{\Gamma\left(\frac{p}{2}\right)} + G_{p-2}(x),
\end{aligned}$$

en faisant l'intégrale par partie. Ainsi, on trouve que $G_{k+2+2j,\lambda}(x) - G_{k+2j}(x) = \frac{-e^{-\frac{x}{2}} x^{\frac{k}{2}+j}}{\Gamma\left(\frac{k}{2}+j+1\right) 2^{\frac{k}{2}+j}}$, ce qui montre le résultat. □

Exemple 53. On considère la règle de Bayes pour $X|\theta \sim \mathcal{N}_p(\theta, \sigma^2 \mathbf{1}_p)$ avec la loi a priori $\pi(\theta) = \mathbf{1}_{(B_m)}(\theta)$, où $B_m = \{\theta \in \mathbb{R}^p \mid \|\theta\| \leq m\}$. Alors, il suffit de calculer la loi marginale de X ainsi que son gradient pour pouvoir appliquer le Théorème 2.2. On a

$$\begin{aligned}
m_\pi(x) &= \int_{B_m} (2\pi\sigma^2)^{-\frac{p}{2}} e^{-\frac{\|x-\theta\|^2}{2\sigma^2}} d\theta \\
&= \mathbb{P}(\|Y\|^2 \leq m^2), \text{ où } Y \sim \mathcal{N}_p(x, \sigma^2 \mathbf{1}_p) \\
&= F_{p, \frac{\|x\|^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right),
\end{aligned}$$

où $F_{k,\Delta}$ est la fonction de répartition d'une loi $\chi_k^2(\Delta)$. Alors, en utilisant le Lemme 49 et le Théorème 2.2, on trouve que

$$\begin{aligned}
\delta_\pi(x) &= x + \sigma^2 \frac{2x \left[F_{p+2, \frac{\|x\|^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right) - F_{p, \frac{\|x\|^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right) \right]}{2\sigma^2 F_{p, \frac{\|x\|^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right)}, \tag{**} \\
&= x \frac{F_{p+2, \frac{\|x\|^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right)}{F_{p, \frac{\|x\|^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right)}.
\end{aligned}$$

Si on se restreint au cas $p = 1$ et lorsque $\theta \in (a, b)$, il y a donc une troncature à gauche et à droite, on trouve de façon analogue que $m(x) = \frac{1}{2(b-a)} \left(\Phi\left(\frac{x-a}{\sigma}\right) - \Phi\left(\frac{x-b}{\sigma}\right) \right)$, $m'(x) = \frac{1}{2(b-a)\sigma} \left(\phi\left(\frac{x-a}{\sigma}\right) - \phi\left(\frac{x-b}{\sigma}\right) \right)$, d'où

$$\begin{aligned} \delta_\pi(x) &= x + \sigma^2 \frac{\phi\left(\frac{x-a}{\sigma}\right) - \phi\left(\frac{x-b}{\sigma}\right)}{\sigma \left(\Phi\left(\frac{x-a}{\sigma}\right) - \Phi\left(\frac{x-b}{\sigma}\right) \right)} \\ &= x + \sigma \frac{\phi\left(\frac{x-a}{\sigma}\right) - \phi\left(\frac{x-b}{\sigma}\right)}{\Phi\left(\frac{x-a}{\sigma}\right) - \Phi\left(\frac{x-b}{\sigma}\right)}. \end{aligned} \quad (\bullet\bullet)$$

On a alors une expression permettant de traiter différents cas où $\theta \in C \subset \mathbb{R}$. Un des cas intéressants est celui où $C = B_m$. Dans ce cas, on trouve, à partir de $(\bullet\bullet)$, que

$$\delta_\pi(x) = x + \sigma \frac{\phi\left(\frac{x+m}{\sigma}\right) - \phi\left(\frac{x-m}{\sigma}\right)}{\Phi\left(\frac{x+m}{\sigma}\right) - \Phi\left(\frac{x-m}{\sigma}\right)}. \quad (\star\star\star)$$

Enfin, on peut se demander comment les représentations $(\star\star)$ et $(\star\star\star)$ sont équivalentes pour $p = 1$ et $\theta \in B_m$. Dans un premier temps, on peut réécrire $(\star\star)$ de la façon suivante :

$$\begin{aligned} \delta_\pi(x) &= x + 2x \frac{-f_{3, \frac{x^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right)}{F_{1, \frac{x^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right)} \\ &= x + 2x \frac{-f_{3, \frac{x^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right)}{\Phi\left(\frac{x+m}{\sigma}\right) - \Phi\left(\frac{x-m}{\sigma}\right)}. \end{aligned}$$

Pour ce qui est du numérateur, on a, selon l'Exemple 51, que

$$\begin{aligned}
f_{3, \frac{x^2}{\sigma^2}}\left(\frac{m^2}{\sigma^2}\right) &= \frac{e^{-\frac{m^2}{2\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi} \sqrt{\frac{x^2}{\sigma^2}}} \left(\frac{e^{\frac{xm}{\sigma^2}} - e^{-\frac{xm}{\sigma^2}}}{2} \right) \\
&= \frac{-\sigma}{2x} \left(\frac{e^{-\frac{(x+m)^2}{2\sigma^2}}}{\sqrt{2\pi}} - \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi}} \right) \\
&= \frac{-\sigma}{2x} \left(\phi\left(\frac{x+m}{\sigma}\right) - \phi\left(\frac{x-m}{\sigma}\right) \right),
\end{aligned}$$

d'où

$$\delta_\pi(x) = x + \sigma \frac{\phi\left(\frac{x+m}{\sigma}\right) - \phi\left(\frac{x-m}{\sigma}\right)}{\Phi\left(\frac{x+m}{\sigma}\right) - \Phi\left(\frac{x-m}{\sigma}\right)}.$$

2.1.2 Modèle gamma avec paramètre d'échelle inconnu

Dans cette sous-section, on va continuer d'explorer le même type de scénario. Cette fois, la base du modèle sera la loi gamma plutôt que la loi normale, c'est-à-dire qu'on a $X|\beta \sim \mathcal{G}(\alpha, \beta)$. On cherchera ici aussi à identifier les lois a posteriori pour β ainsi que la loi prédictive d'une nouvelle observation $Y|\beta \sim \mathcal{G}(\gamma, \beta)$ lorsque les paramètres α et γ sont connus.

Un premier cas intéressant survient avec le choix de la loi a priori non informative usuelle pour β . En fait, le résultat qui suit est un peu l'équivalent du Lemme 44 mais dans un contexte de loi gamma.

Lemme 54. *Pour $X|\beta \sim \mathcal{G}(\alpha, \beta)$, $Y|\beta \sim \mathcal{G}(\gamma, \beta)$ prises comme étant conditionnellement indépendantes et l'a priori $\pi_0(\beta) = \frac{\mathbb{1}_{(\mathbb{R}_+)}}{\beta}(\beta)$, la densité a posteriori de β est de loi inverse gamma $\mathcal{IG}(\alpha, x)$ et la densité prédictive a posteriori de Y est de la forme*

$$\hat{q}_{\pi_0}(y|x) = \frac{\Gamma(\alpha + \gamma)}{\Gamma(\alpha)\Gamma(\gamma)} \left(1 + \frac{x}{y}\right)^{1-\gamma} \left(1 + \frac{y}{x}\right)^{-\alpha} \frac{1}{x+y} \mathbb{1}_{((0,\infty))}(y).$$

Démonstration. Dans un premier temps, on a pour $\beta > 0$:

$$\begin{aligned} \pi_0(\beta|x) &\propto f(x|\beta)\pi_0(\beta) \\ &\propto \frac{e^{-\frac{x}{\beta}}}{\beta^{\alpha+1}}. \end{aligned}$$

Dans un deuxième temps, on a que

$$\begin{aligned} \hat{q}_{\pi_0}(y|x) &= \int_{\mathbb{R}_+} f(y|\beta)\pi(\beta|x)d\beta \\ &= \int_{\mathbb{R}_+} \frac{y^{\gamma-1}e^{-\frac{y}{\beta}}}{\Gamma(\gamma)\beta^\gamma} \frac{x^\alpha e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha+1}} d\beta \\ &= \frac{y^{\gamma-1}x^\alpha}{\Gamma(\gamma)\Gamma(\alpha)} \int_{\mathbb{R}_+} \frac{e^{-\frac{(x+y)}{\beta}}}{\beta^{\alpha+\gamma+1}} d\beta \\ &= \frac{\Gamma(\alpha + \gamma)}{\Gamma(\alpha)\Gamma(\gamma)} \left(1 + \frac{x}{y}\right)^{1-\gamma} \left(1 + \frac{y}{x}\right)^{-\alpha} \frac{1}{x+y}. \end{aligned}$$

□

Remarque 55. *La densité prédictive a posteriori sous π_0 peut se réécrire de la façon suivante*

$$\hat{q}_{\pi_0}(y|x) = \frac{\Gamma(\alpha + \gamma)}{\Gamma(\alpha)\Gamma(\gamma)} \frac{1}{x} \left(\frac{y}{x}\right)^{\gamma-1} \left(1 + \frac{y}{x}\right)^{-(\alpha+\gamma)},$$

ce qui correspond à une densité Bêta de type II de paramètres (γ, α) et de paramètre d'échelle x .

Pour d'autres a priori, la représentation suivante fut obtenue par L'Moudden et coll. dans [LMKS17].

Théorème 2.3. *Pour $X|\beta \sim \mathcal{G}(\alpha, \beta)$, $Y|\beta \sim \mathcal{G}(\gamma, \beta)$ prises comme étant conditionnellement indépendantes par rapport à β et n'importe quelle loi a priori $\pi(\beta)$ absolument continue par rapport à une mesure σ -finie $\nu(\beta)$, la densité prédictive a posteriori de Y est donnée par*

$$\hat{q}_\pi(y|x) = \hat{q}_{\pi_0}(y|x) \frac{m_{Z,\pi}(z)}{m_{X,\pi}(x)} \frac{z}{x},$$

où $\hat{q}_{\pi_0}(y|x)$ est donnée au Lemme 54, $Z = X + Y$ et $m_{W,\pi}(w) = \int_{\Theta} f(w|\beta)\pi(\beta)d\nu(\beta)$.

Démonstration. On utilisera encore une fois la décomposition $\hat{q}_\pi(y|x) = \frac{f(x,y)}{m_{X,\pi}(x)}$. On a que

$$\begin{aligned} f(x,y) &= \int_{\mathbb{R}_+} f(x|\beta)f(y|\beta)\pi(\beta)d\nu(\beta) \text{ (par indépendance)} \\ &= \frac{x^{\alpha-1}y^{\gamma-1}}{\Gamma(\alpha)\Gamma(\gamma)} \int_{\mathbb{R}_+} \frac{e^{-\frac{x+y}{\beta}}}{\beta^{\alpha+\gamma}} \pi(\beta)d\nu(\beta) \\ &= \frac{x^{\alpha-1}y^{\gamma-1}}{\Gamma(\alpha)\Gamma(\gamma)} \frac{\Gamma(\alpha+\gamma)}{(x+y)^{\alpha+\gamma-1}} \int_{\mathbb{R}_+} f(z|\beta)\pi(\beta)d\nu(\beta) \\ &= \hat{q}_{\pi_0}(y|x) m_{Z,\pi}(z) \frac{z}{x}, \end{aligned}$$

d'où le résultat. □

Exemple 56. *Trouvons la densité prédictive a posteriori de Y dans le cadre du Théorème 2.3 avec l'a priori $\pi(\beta) \sim \mathcal{IG}(\alpha_0, \beta_0)$. Calculons dans un premier temps la loi marginale de X . On a que*

$$\begin{aligned} m_{X,\pi}(x) &= \int_{\mathbb{R}_+} \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} \frac{\beta_0^{\alpha_0}e^{-\frac{\beta_0}{\beta}}}{\Gamma(\alpha_0)\beta^{\alpha_0+1}} d\beta \\ &= \frac{\Gamma(\alpha+\alpha_0)}{\Gamma(\alpha)\Gamma(\alpha_0)} \frac{1}{x} \left(\frac{x}{x+\beta_0} \right)^\alpha \left(\frac{\beta_0}{x+\beta_0} \right)^{\alpha_0}, \text{ pour } x > 0. \end{aligned}$$

De la même façon, on trouve que

$$m_{Z,\pi}(x+y) = \frac{\Gamma(\alpha + \alpha_0 + \gamma)}{\Gamma(\alpha + \gamma)\Gamma(\alpha_0)} \frac{1}{x+y} \left(\frac{x+y}{x+y+\beta_0} \right)^{\alpha+\gamma} \left(\frac{x+\beta_0}{x+y+\beta_0} \right)^{\alpha_0}, \text{ pour } x+y > 0.$$

Enfin, on peut trouver la densité prédictive de Y a posteriori, qui est de la forme

$$\hat{q}_\pi(y|x) = \frac{\Gamma(\alpha + \alpha_0 + \gamma)}{\Gamma(\alpha + \alpha_0)\Gamma(\gamma)} \frac{1}{x+\beta_0+y} \left(1 + \frac{x+\beta_0}{y} \right)^{1-\gamma} \left(1 + \frac{y}{x+\beta_0} \right)^{-(\alpha+\alpha_0)}.$$

On remarque qu'il s'agit d'une densité Bêta de type II de paramètres $(\gamma, \alpha + \alpha_0)$ et de paramètre d'échelle $x + \beta_0$, le cas $\alpha_0 = \beta_0 = 0$ nous ramenant au Lemme 54.

2.2 Modèle normal-gamma

Cette sous-section porte sur un modèle connu, le modèle normal-gamma. On présentera quelques propriétés importantes de ce modèle et on tirera profit de ces propriétés au cours des prochains chapitres.

Définition 57. On dit que le couple (u, v) est de loi normale-gamma avec paramètre $\mu \in \mathbb{R}^p$, $\lambda > 0$, $\alpha > 0$ et $\beta > 0$, que l'on note $(u, v) \sim \mathcal{N}_p \mathcal{G}(\mu, \lambda, \alpha, \beta)$ si

$$\begin{aligned} u|v &\sim \mathcal{N}_p \left(\mu, \frac{1}{\lambda v} \mathbf{1}_p \right), \\ v &\sim \mathcal{G}(\alpha, \beta). \end{aligned}$$

Il en découle que la densité conjointe de (u, v) est alors donnée par

$$f(u, v) = \frac{\beta^\alpha \lambda^{\frac{p}{2}}}{\Gamma(\alpha)(2\pi)^{\frac{p}{2}}} v^{\alpha-1+\frac{p}{2}} e^{-\beta v} e^{-\frac{1}{2}(u-\mu)^\top (\lambda v \mathbf{1}_p)(u-\mu)}, u \in \mathbb{R}^p, v > 0.$$

Le prochain résultat donne la marginale de u .

Lemme 58. *Soit $(u, v) \sim \mathcal{N}_p \mathcal{G}(\mu, \lambda, \alpha, \beta)$. Alors, la loi marginale de u est :*

$$u \sim t_p \left(2\alpha, \mu, \frac{\beta}{\alpha\lambda} \mathbb{1}_p \right).$$

Démonstration. On a

$$\begin{aligned} f(u) &= \int_{\mathbb{R}_+} \frac{\beta^\alpha \lambda^{\frac{p}{2}}}{\Gamma(\alpha)(2\pi)^{\frac{p}{2}}} v^{\alpha+\frac{p}{2}-1} e^{-v(\beta+\frac{1}{2}(u-\mu)^\top(\lambda\mathbb{1}_p)(u-\mu))} dv \\ &= \frac{\beta^\alpha \lambda^{\frac{p}{2}}}{\Gamma(\alpha)(2\pi)^{\frac{p}{2}}} \Gamma\left(\frac{2\alpha+p}{2}\right) \left(\beta + \frac{1}{2}(u-\mu)^\top(\lambda\mathbb{1}_p)(u-\mu)\right)^{-\left(\frac{2\alpha+p}{2}\right)} \\ &= \frac{\Gamma\left(\frac{2\alpha+p}{2}\right)}{\Gamma\left(\frac{2\alpha}{2}\right)(2\alpha\pi)^{\frac{p}{2}}} \left|\frac{\beta}{\alpha\lambda} \mathbb{1}_p\right|^{-\frac{1}{2}} \left(1 + \frac{1}{2\alpha}(u-\mu)^\top \left(\frac{\beta}{\alpha\lambda} \mathbb{1}_p\right)^{-1} (u-\mu)\right)^{-\left(\frac{2\alpha+p}{2}\right)}. \end{aligned}$$

On reconnaît alors la densité de Student de dimension p , telle que vue au chapitre 1, avec les paramètres donnés. \square

Lemme 59. *Soient X_1, \dots, X_n iid $\mathcal{N}_p(\theta, \eta^{-1})$. Soit $(\theta, \eta) \sim \mathcal{N}_p \mathcal{G}(\mu, \lambda, \alpha, \beta)$. Soient également $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^\top (X_i - \bar{X})$. Alors, la loi a posteriori de (θ, η) est également normale-gamma de dimension p avec les paramètres : $\mu_0 = \frac{\bar{x}n + \mu\lambda}{\lambda + n}$, $\lambda_0 = \lambda + n$, $\alpha_0 = \frac{n-1+p}{2} + \alpha$ et $\beta_0 = \beta + \frac{ns^2 + n\lambda(\bar{x}-\mu)^\top(\bar{x}-\mu)}{2}$.*

Démonstration. On a

$$\begin{aligned} f_{\theta, \eta | \bar{X}, S}(\theta, \eta) &\propto \frac{1}{\left|\frac{\eta^{-1}}{n} \mathbb{1}_p\right|^{\frac{1}{2}}} e^{\frac{-1}{2}(\bar{x}-\theta)^\top \left(\frac{\eta^{-1}}{n} \mathbb{1}_p\right)^{-1} (\bar{x}-\theta)} \frac{e^{-\frac{ns^2\eta}{2}}}{(\eta^{-1})^{\frac{n-1}{2}}} \eta^{\alpha-1+\frac{p}{2}} e^{-\beta\eta} e^{\frac{-1}{2}(\theta-\mu)^\top(\lambda\eta\mathbb{1}_p)(\theta-\mu)} \\ &\propto \eta^{\frac{n-1}{2} + \alpha - 1 + \frac{p}{2}} e^{-\eta\left(\beta + \frac{ns^2}{2}\right)} e^{\frac{-\eta}{2}(n(\bar{x}-\theta)^\top(\bar{x}-\theta) + \lambda(\theta-\mu)^\top(\theta-\mu))} \\ &\propto \eta^{\frac{n-1}{2} + \alpha - 1 + \frac{p}{2}} e^{-\eta\left(\beta + \frac{ns^2}{2}\right)} e^{\frac{-1}{2}\left(\frac{(\bar{x}-\mu)^\top(\bar{x}-\mu)}{\lambda+n}\right)(n\lambda\eta) + \eta(n+\lambda)\left(\theta - \frac{\bar{x}n + \mu\lambda}{\lambda+n}\right)^\top \left(\theta - \frac{\bar{x}n + \mu\lambda}{\lambda+n}\right)} \\ &\propto \eta^{\frac{n}{2} + \alpha - 1 + \frac{p}{2} + \frac{p}{2} - \frac{1}{2}} e^{-\eta\left(\beta + \frac{ns^2 + (\bar{x}-\mu)^\top(\bar{x}-\mu)n\lambda}{2}\right)} e^{\frac{-\eta}{2}(n+\lambda)\left(\theta - \frac{\bar{x}n + \mu\lambda}{\lambda+n}\right)^\top \left(\theta - \frac{\bar{x}n + \mu\lambda}{\lambda+n}\right)}, \theta \in \mathbb{R}^p, \eta > 0. \end{aligned}$$

On reconnaît alors la densité d'une normale-gamma de dimension p avec les paramètres donnés. □

Remarque 60. *Voici quelques remarques à propos du Lemme 59.*

- Selon la Définition 57, on a que $\eta|\bar{X}, S \sim \mathcal{G}\left(\alpha + \frac{n-1+p}{2}, \beta + \frac{ns^2+n\lambda(x-\mu)^\top(x-\mu)}{2}\right)$.
- Le paramètre μ_0 s'exprime comme une combinaison convexe de la moyenne expérimentale et la moyenne a priori. En effet, on a $\mu_0 = \bar{x}\left(\frac{n}{\lambda+n}\right) + \mu\left(1 - \frac{n}{\lambda+n}\right)$. On voit bien que lorsque $n \rightarrow \infty$, $\mu_0 \rightarrow \bar{x}$ et donc que l'influence de la loi a priori finit par disparaître.
- Il y a une mise à jour additive avec l'inverse des variances : $\frac{1}{\mathbb{V}(\theta_i|\eta)} + \frac{1}{\mathbb{V}(\bar{X}_i|\eta, \theta)} = \frac{1}{\mathbb{V}(\theta_i|\eta, \bar{X}, S)}$, où \bar{X}_i représente la i -ième composante de \bar{X} , pour $i = 1, \dots, d$. En effet, on a $\lambda\eta + n\eta = \eta(\lambda + n)$.

Corollaire 61. *Sous les conditions du Lemme 59, la loi a posteriori de θ est la suivante :*

$$\pi(\theta|\bar{X}, S) \sim t_p\left(2\alpha_0, \mu_0, \frac{\beta_0}{\lambda_0\alpha_0}\mathbb{1}_p\right).$$

Démonstration. Immédiat en appliquant le Lemme 58 et le Lemme 59. □

Théorème 2.4. *Pour $Y|\theta, \eta \sim \mathcal{N}_p(\theta, \eta^{-1})$ et $(\theta, \eta) \sim \mathcal{N}_p\mathcal{G}(\mu, \lambda, \alpha, \beta)$, la densité prédictive de Y est $t_p\left(2\alpha_0, \mu_0, \frac{\beta_0(1+\lambda_0)}{\lambda_0\alpha_0}\mathbb{1}_p\right)$.*

Démonstration. Afin de faciliter les calculs, on pose $A = \frac{y+\lambda_0\mu_0}{\lambda_0+1}$ et $B = \beta_0 + \frac{\lambda_0}{2(\lambda_0+1)}(y-\mu_0)^2$. On a, par la définition de la densité prédictive et à l'aide du Lemme 59, que

$$\begin{aligned}
\hat{q}(y|\bar{x}, s) &= \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} f_y(y|\theta, \eta) \pi(\theta, \eta|\bar{x}, s) d\theta d\eta \\
&= \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} \frac{\eta^{\frac{p}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(y-\theta)^\top (\eta^{-1} \mathbb{1}_p)^{-1} (y-\theta)} \frac{\beta_0^{\alpha_0} \lambda_0^{\frac{p}{2}}}{\Gamma(\alpha_0) (2\pi)^{\frac{p}{2}}} \eta^{\alpha_0-1+\frac{p}{2}} e^{-\beta_0 \eta} e^{-\frac{1}{2}(\theta-\mu_0)^\top (\lambda_0 \eta \mathbb{1}_p) (\theta-\mu_0)} d\eta d\theta \\
&= \frac{\beta_0^{\alpha_0} \lambda_0^{\frac{p}{2}}}{\Gamma(\alpha_0) (2\pi)^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}_+} \eta^{\alpha_0+p-1} e^{-\eta \left(\beta_0 + \frac{(y-\theta)^\top (y-\theta) + \lambda_0 (\theta-\mu_0)^\top (\theta-\mu_0)}{2} \right)} d\eta d\theta \\
&= \frac{\beta_0^{\alpha_0} \lambda_0^{\frac{p}{2}}}{\Gamma(\alpha_0) (2\pi)^p} \Gamma(\alpha_0 + p) \int_{\mathbb{R}^p} \left(\beta_0 + \frac{(y-\theta)^\top (y-\theta) + \lambda_0 (\theta-\mu_0)^\top (\theta-\mu_0)}{2} \right)^{-(\alpha_0+p)} d\theta \\
&= \frac{\beta_0^{\alpha_0} \lambda_0^{\frac{p}{2}} \Gamma(\alpha_0 + p)}{\Gamma(\alpha_0) (2\pi)^p} \int_{\mathbb{R}^p} \left(B + \frac{(1+\lambda_0) (\theta-A)^\top (\theta-A)}{2} \right)^{-(\alpha_0+p)} d\theta \\
&= \frac{\beta_0^{\alpha_0} \lambda_0^{\frac{p}{2}} \Gamma(\alpha_0 + p)}{\Gamma(\alpha_0) (2\pi)^p B^{\alpha_0+p}} \int_{\mathbb{R}^p} \left(1 + \frac{(1+\lambda_0) (\theta-A)^\top (\theta-A) (2\alpha_0+p)}{2B(2\alpha_0+p)} \right)^{-\left(\frac{2\alpha_0+p+p}{2}\right)} d\theta \\
&= \frac{\beta_0^{\alpha_0} \lambda_0^{\frac{p}{2}} \Gamma(\alpha_0 + p)}{\Gamma(\alpha_0) (2\pi)^p B^{\alpha_0+p}} \frac{\Gamma\left(\frac{2\alpha_0+p}{2}\right)}{\Gamma(\alpha_0 + p)} (2\alpha_0+p)^{\frac{p}{2}} \pi^{\frac{p}{2}} \left(\frac{2B}{(1+\lambda_0)(2\alpha_0+p)} \right)^{\frac{p}{2}} \quad (\diamond) \\
&= \frac{\Gamma\left(\frac{2\alpha_0+p}{2}\right)}{\Gamma\left(\frac{2\alpha_0}{2}\right)} \left(\frac{\alpha_0 \lambda_0}{1+\lambda_0} \right)^{\frac{p}{2}} \frac{1}{(2\pi \alpha_0 \beta_0)^{\frac{p}{2}}} \left(1 + \frac{(y-\mu_0)^\top (y-\mu_0) \alpha_0 \lambda_0}{2\alpha_0 \beta_0 (1+\lambda_0)} \right)^{-\left(\frac{2\alpha_0+p}{2}\right)}.
\end{aligned}$$

On obtient (\diamond) en utilisant le fait que l'intégrale de l'étape précédente n'est autre que le noyau d'une loi de Student. On reconnaît alors la densité d'une Student de dimension p avec les paramètres attendus. \square

Le prochain résultat, semblable à celui présenté dans [Ait75], considère le cas où la loi a priori, dite non-informative, est $\pi(\theta, \eta) = \frac{1}{\eta} \mathbb{1}_{(\mathbb{R} \times \mathbb{R}_+)}(\theta, \eta)$.

Exemple 62. *Trouvons la densité prédictive a posteriori de Y dans le cadre du Théorème 2.4 avec l'a priori $\pi_0(\theta, \eta) \sim \mathcal{N}_d \mathcal{G}(\mu, \lambda, \alpha, \beta)$ où $\alpha \rightarrow 0$ et $\beta \rightarrow 0$. Dans ce cas, la densité a posteriori est*

$$\pi(\theta, \eta|\bar{x}, s) \sim \mathcal{N}_d \mathcal{G} \left(\frac{\bar{x}n + \mu\lambda}{\lambda + n}, \lambda + n, \frac{n-1+d}{2}, \frac{ns^2 + n\lambda(\bar{x}-\mu)^\top (\bar{x}-\mu)}{2} \right).$$

On trouve donc que la densité prédictive a posteriori de Y est

$$\hat{q}_{\pi_0}(y|\bar{x}, s) \sim t_d \left(n - 1 + d, \frac{\bar{x}n + \mu\lambda}{\lambda + n}, \frac{ns^2 + n\lambda(\bar{x} - \mu)^\top(\bar{x} - \mu)}{(\lambda + n)(n - 1 + d)} \right).$$

Remarque 63. Le modèle que l'on a présenté dans cette section est plutôt complexe. En effet, celui-ci est basé sur 4 hyperparamètres, soit μ , λ , α et β . Il est naturel de se demander l'impact qu'ont ces paramètres, mais aussi de quelle façon et quelle interprétation on peut attribuer à chacun de ceux-ci. On suppose ici que l'on a un couple $(\theta, \eta) \sim \mathcal{N}_p\mathcal{G}(\mu, \lambda, \alpha, \beta)$. On explique dans [Lee97] comment interpréter et choisir les paramètres d'un modèle normale- χ^2 , ce qui est analogue à notre modèle.

- Selon la Définition 57, l'interprétation de μ est plutôt simple : la densité a priori de $\theta|\eta$ est normale de moyenne μ . Ainsi, si l'on dispose d'une étude ultérieure ou encore que l'on a fait une préétude exploratoire, on peut trouver une valeur potentielle de μ aisément.
- Selon la Définition 57, les paramètres α et β devraient tout simplement permettre de bien modéliser η . On peut alors choisir ces paramètres en remarquant que $\alpha = \left(\frac{\bar{\eta}}{s_\eta}\right)^2$ et $\beta = \frac{s_\eta^2}{\bar{\eta}}$, où $\bar{\eta}$ est la moyenne envisagée de η et s_η est l'écart-type envisagé de η .
- Selon la Définition 57, le paramètre λ aurait un impact sur la densité de $\theta|\eta$. En effet, celui-ci joue sur la variance de cette densité. Ce paramètre peut alors être vu comme étant la taille de l'échantillon de la préétude. Cette affirmation est également appuyée par le Lemme 59, où l'on trouve que le paramètre a posteriori $\lambda_0 = \lambda + n$. Ceci semble vouloir dire que la taille d'échantillon a été mise à jour.

Exemple 64. [Lee97] Une station expérimentale fait pousser du blé et s'intéresse à l'effet d'une hormone de croissance sur le rendement par parcelle de blé. La distribution a priori pour la variance sur les parcelles pourrait avoir une moyenne de 300 avec un écart-type de 160. Pour ce qui est de la moyenne, on s'attend à ce qu'elle tourne autour de 110, selon un petit échantillon de taille 15. Alors, on trouve dans un premier temps que

$\alpha = \frac{\bar{\eta}^2}{s_{\eta}^2} = \frac{300^2}{160^2} \approx 3,5156$ et que $\beta = \frac{s_{\eta}^2}{e\bar{a}} \approx 85,33$. Ensuite, puisque la moyenne tourne autour de 110 selon un échantillon de taille 15, on trouve que $\mu = 110$ et que $\lambda = 15$. La densité a priori correspondante est alors $(\theta, \eta) \sim \mathcal{NG}(110, 15, 3.5156, 85.33)$.

Supposons qu'on observe les rendements de douze parcelles où l'on avait utilisé l'hormone de croissance :

141, 102, 73, 171, 137, 91, 81, 157, 146, 69, 121, 134.

On trouve alors que $\bar{x} \approx 118.58$ et que $s^2 \approx 1087.08$. Ainsi, la densité a posteriori est donnée par $(\theta, \eta | \bar{x}, s) \sim \mathcal{NG}(113.81, 27, 9.5156, 13238.41)$. De plus, on a que $(\theta | \bar{x}, s) \sim t(19.0312, 113.81, 51.52712193)$ et que $(\eta | \bar{x}, s) \sim \mathcal{G}(9.5156, 13238.41)$.

CHAPITRE 3

L'a priori normal-gamma indépendant

Le modèle normal-gamma présenté à la fin du dernier chapitre est facile à utiliser en pratique étant donné qu'il s'agit d'un modèle conjugué. Par contre, on risque de rencontrer des scénarios où la modélisation de la dépendance a priori entre les paramètres θ et σ^2 n'a pas de sens. Il faut alors adapter les outils et développer des résultats, ce qui sera fait dans ce chapitre. Dans la première section, on abordera la loi marginale a posteriori de θ . Cette densité sera normalisée par une constante faisant intervenir une convolution. On donnera également quelques propriétés ainsi que des cas limites de cette densité. Dans la seconde section, on parlera de la loi marginale a posteriori de σ^2 et on survolera certaines de ses propriétés.

3.1 Loi marginale a posteriori de θ

Le résultat qui suit traite le cas d'un a priori à composantes indépendantes pour (θ, σ^2) . Bien que la représentation standard soit avec l'a priori \mathcal{NG} , le cas à composantes indépendantes est tout autant intéressant. Ce nouveau résultat fait intervenir une convolution

(voir Chapitre 1) pour exprimer la constante de normalisation de la loi a posteriori de θ .

Proposition 65. Soient \bar{X} et S des variables conditionnellement indépendantes par rapport aux paramètres θ et σ^2 telles que $\bar{X}|\theta, \sigma^2 \sim \mathcal{N}_p\left(\theta, \frac{\sigma^2}{n} \mathbb{1}_p\right)$ et $nS^2|\theta, \sigma^2 \sim \mathcal{G}\left(\frac{n-1}{2}, 2\sigma^2\right)$. On suppose ici que $\theta|\sigma^2 \sim \mathcal{N}_p(\mu, \tau^2 \mathbb{1}_p)$ et $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$. Alors, la densité a posteriori de $\frac{\theta-\mu}{\tau}$ admet la représentation

$$\pi(t|\bar{x}, s) = \frac{\phi_p(t) t_{n-1+2\alpha, 0, \tau^2 n \frac{n-1+2\alpha}{ns^2+2\beta}}(w-t)}{C_1}, \quad (\star)$$

où

$$w = \frac{\bar{x} - \mu}{\tau} \quad \text{et} \quad C_1 = \left(\phi_p \star t_{n-1+2\alpha, 0, \tau^2 n \frac{n-1+2\alpha}{ns^2+2\beta}} \right)(w).$$

Démonstration. On a que

$$\begin{aligned} \pi(\theta|\bar{x}, s) &\propto \int_0^\infty \frac{e^{-\frac{1}{2}(\bar{x}-\theta)^\top \left(\frac{\sigma^2}{n} \mathbb{1}_p\right)^{-1} (\bar{x}-\theta)} (ns^2)^{\frac{n-1}{2}-1} e^{-\frac{ns^2}{2\sigma^2}} e^{-\frac{1}{2}(\theta-\mu)^\top (\tau^2 \mathbb{1}_p)^{-1} (\theta-\mu)} \beta^\alpha e^{-\frac{\beta}{\sigma^2}}}{(2\pi)^{\frac{p}{2}} \left|\frac{\sigma^2}{n} \mathbb{1}_p\right|^{\frac{1}{2}} \Gamma\left(\frac{n-1}{2}\right) (2\sigma^2)^{\frac{n-1}{2}} (2\pi)^{\frac{p}{2}} |\tau^2 \mathbb{1}_p|^{\frac{1}{2}} \Gamma(\alpha) (\sigma^2)^{\alpha+1}} d\sigma^2 \\ &\propto \int_0^\infty \frac{e^{-\frac{1}{2} \left[(\bar{x}-\theta)^\top \left(\frac{\sigma^2}{n} \mathbb{1}_p\right)^{-1} (\bar{x}-\theta) + \frac{ns^2}{\sigma^2} + (\theta-\mu)^\top (\tau^2 \mathbb{1}_p)^{-1} (\theta-\mu) + \frac{2\beta}{\sigma^2} \right]}}{(\sigma^2)^{\frac{p+n-1+2\alpha}{2}+1}} d\sigma^2 \end{aligned}$$

Or, comme on a que $(\bar{x} - \theta)^\top \mathbb{1}_p (\bar{x} - \theta) = \|\bar{x} - \theta\|^2$, on obtient

$$\begin{aligned} \pi(\theta|\bar{x}, s) &\propto e^{-\frac{\|\theta-\mu\|^2}{2\tau^2}} \int_0^\infty \frac{e^{-\frac{1}{2\sigma^2} [n\|\bar{x}-\theta\|^2 + ns^2 + 2\beta]}}{(\sigma^2)^{\frac{p+n-1+2\alpha}{2}+1}} d\sigma^2 \\ &\propto e^{-\frac{\|\theta-\mu\|^2}{2\tau^2}} \frac{1}{\left(\frac{n\|\bar{x}-\theta\|^2 + ns^2 + 2\beta}{2}\right)^{\frac{p+n-1+2\alpha}{2}}} \quad (\diamond) \\ &\propto e^{-\frac{\|\theta-\mu\|^2}{2\tau^2}} \left(1 + \frac{1}{n-1+2\alpha} (\bar{x} - \theta)^\top \left(\frac{n(n-1+2\alpha)}{ns^2+2\beta} \mathbb{1}_p\right) (\bar{x} - \theta)\right)^{-\left(\frac{p+n-1+2\alpha}{2}\right)}. \end{aligned}$$

L'égalité (\diamond) s'obtenant en remarquant de l'intégrale de l'équation précédente n'est en fait que le noyau d'une loi gamma-inverse. Un peu de la même façon, on peut voir de la dernière équation qu'on retrouve le noyau d'une loi multinormale de moyenne μ et de matrice de variance-covariance $\tau^2 \mathbb{1}_p$ ainsi que le noyau d'une loi Student multivarié de dimension p ayant $n-1+2\alpha$ degrés de liberté, de paramètre de position \bar{x} et de paramètre d'échelle $n \frac{n-1+2\alpha}{ns^2+2\beta}$. En centrant la densité de la loi multinormale, on obtient, en posant $t = \frac{\theta-\mu}{\tau}$:

$$\pi(t|\bar{x}, s) \propto e^{-\frac{\|t\|^2}{2}} \left(1 + \frac{1}{n-1+2\alpha} \left(t - \left(\frac{\bar{x}-\mu}{\tau} \right) \right)^\top \left(\tau^2 n \frac{n-1+2\alpha}{ns^2+2\beta} \mathbb{1}_p \right) \left(t - \left(\frac{\bar{x}-\mu}{\tau} \right) \right) \right)^{-\frac{p+n-1+2\alpha}{2}},$$

ce qui est (\star). Finalement, la constante de normalisation est bien donnée par :

$$\begin{aligned} C_1 &= \int_{\mathbb{R}^p} \phi_p(t) t_{p(n-1+2\alpha, 0, \tau^2 n \frac{n-1+2\alpha}{ns^2+2\beta})}(w-t) dt \\ &= \left(\phi_p * t_{p(n-1+2\alpha, 0, \tau^2 n \frac{n-1+2\alpha}{ns^2+2\beta})} \right)(w). \end{aligned}$$

□

La forme de la densité dans le cas univarié, exprimée comme un produit de densités de loi normale et Student, ainsi que la constante de normalisation exprimée en termes d'une convolution en un point, s'apparente à la densité d'une somme (ou d'une convolution) de ces lois étudiée dans [PN13].

Remarque 66. *Bien qu'il soit intéressant d'exprimer la constante de normalisation C_1 sous forme de convolution, on peut trouver une autre forme pour cette dernière. En effet, si l'on considère la fonction $f_{\mu, v, \sigma^2}(x) = e^{-\frac{\|x\|^2}{2}} \left(1 + \frac{(x-\mu)^\top (x-\mu)}{v\sigma^2} \right)^{-\left(\frac{v+p}{2}\right)}$, on a alors*

$$\begin{aligned}
C_1 &= \int_{\mathbb{R}^p} f_{\mu, v, \sigma^2}(x) dx \\
&= \int_{\mathbb{R}^p} e^{-\frac{\|x-\mu\|^2}{2}} \left(1 + \frac{x^\top x}{v\sigma^2}\right)^{-\left(\frac{v+p}{2}\right)} dx \\
&= (2\pi)^{\frac{p}{2}} \mathbb{E} \left[\left(1 + \frac{X^\top X}{v\sigma^2}\right)^{-\left(\frac{v+p}{2}\right)} \right], \text{ où } X \sim \mathcal{N}_p(\mu, \mathbf{1}_p).
\end{aligned}$$

Alors, si l'on pose $Y = X^\top X$, on a que $Y \sim \chi_p^2(\mu^\top \mu)$. En utilisant la représentation sous forme de mélange de cette variable aléatoire, telle que présentée au Chapitre 1 dans (***) à la page 19, on a

$$\begin{aligned}
Y|L &\sim \chi_{p+2L}^2 \sim \mathcal{G}\left(\frac{p}{2} + L, 2\right) \\
L &\sim \mathcal{P}\left(\frac{\mu^\top \mu}{2}\right).
\end{aligned}$$

On remarque également que $Z|L \sim \mathcal{G}\left(\frac{p}{2} + L, \frac{2}{v\sigma^2}\right)$, où $Z = \frac{Y}{v\sigma^2}$. Ainsi, on a

$$\begin{aligned}
\mathbb{E} \left[(1 + Z)^{-\left(\frac{v+p}{2}\right)} \right] &= \mathbb{E}^L \mathbb{E} \left[(1 + Z)^{-\left(\frac{v+p}{2}\right)} \middle| L \right] \\
&= \mathbb{E}^L \left[\frac{1}{\Gamma\left(\frac{p}{2} + L\right) \left(\frac{2}{v\sigma^2}\right)^{\frac{p}{2} + L}} \int_{\mathbb{R}_+} (1 + z)^{-\left(\frac{v+p}{2}\right)} z^{\frac{p}{2} + L - 1} e^{-z\left(\frac{v\sigma^2}{2}\right)} dz \right] \\
&= \mathbb{E}^L \left[\frac{\Phi\left(\frac{p}{2} + L, L + 1 - \frac{v}{2}, \frac{v\sigma^2}{2}\right) (v\sigma^2)^{\frac{p}{2} + L}}{2^{\frac{p}{2} + L}} \right],
\end{aligned}$$

la dernière égalité étant obtenue en reconnaissant une densité de Kummer de type II (voir la Définition 33). On trouve donc une représentation différente pour la constante de normalisation de $f_{\mu, v, \sigma^2}(x)$, soit

$$C_1 = \left(\frac{v\sigma^2}{2}\right)^{\frac{p}{2}} \sum_{L \geq 0} \Phi\left(\frac{p}{2} + L, L + 1 - \frac{v}{2}, \frac{v\sigma^2}{2}\right) \frac{e^{-\frac{\mu^\top \mu}{2}} \left(\frac{v\sigma^2 \mu^\top \mu}{4}\right)^L}{L!}.$$

Exemple 67. On a voulu voir l'impact des différents paramètres de la densité donnée en (*) dans le cas univarié $p = 1$. Pour les trois graphiques qui suivent, on a utilisé comme paramètres de base $\mu = 2$, $v = 3$ et $\sigma^2 = 1$. On a altéré un seul paramètre à la fois, de façon à bien saisir son impact.

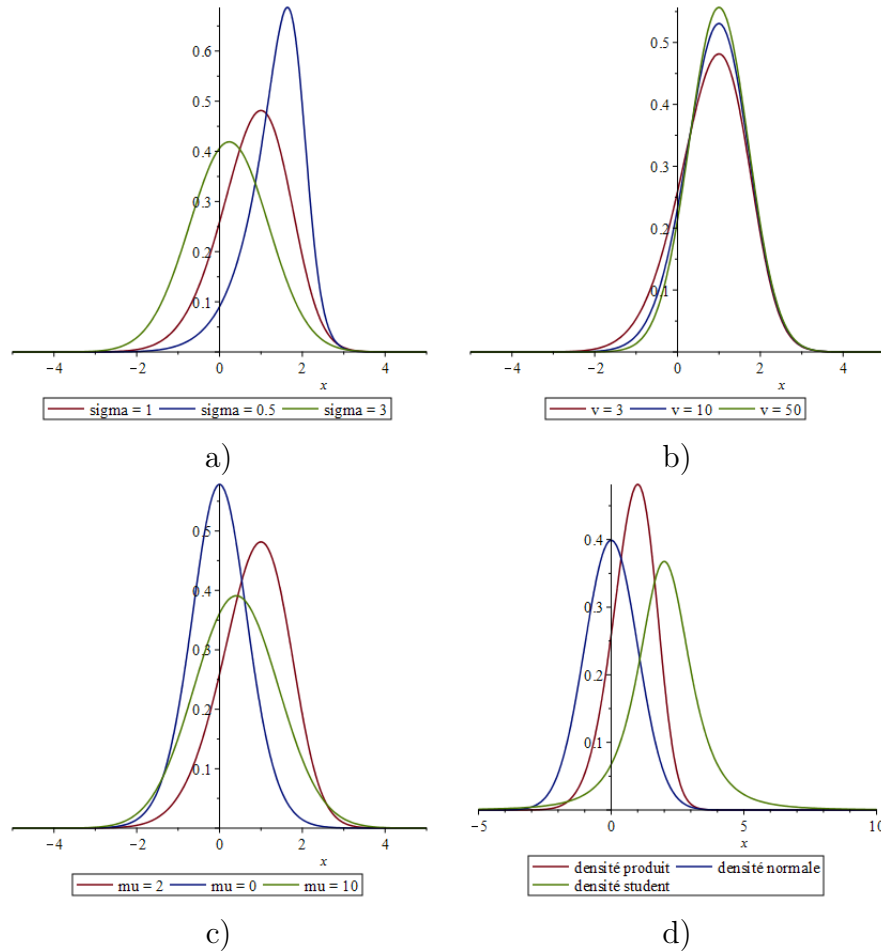


Figure 3.1 – Différents graphiques qui présentent la densité (*) pour : a) $\mu = 2$, $v = 3$ et σ^2 varie, b) $\mu = 2$, v varie et $\sigma^2 = 1$, c) μ varie, $v = 3$ et $\sigma^2 = 1$. Le graphique d) compare la densité produit avec celle de Student et la normale.

En regardant le graphique a) de la Figure 3.1, on voit que pour des petites valeurs de σ , la densité résultante semble plus proche de la densité de Student, alors que pour des grandes

valeurs de σ , la densité résultante se rapproche de la densité normale centrée réduite. Ceci est dû au fait que le paramètre σ contrôle la dispersion de la densité de Student, et que cette dispersion joue un rôle important lorsqu'on multiplie les deux densités ensemble. En effet, si la densité de Student devient très dispersée, alors elle deviendra de plus en plus aplatie, ce qui, éventuellement, revient à multiplier la densité normale par une densité presque uniforme.

Évidemment, la position de la densité Student est également importante, comme on peut le voir avec le graphique c) de la Figure 3.1. On voit que lorsque $|\mu|$ est petite, la distribution aura tendance à être asymétrique et que la masse de la fonction s'éloigne du point de centralité. Par contre, si jamais μ est trop grand, alors la densité résultante ressemblera plus ou moins à une densité de loi normale centrée et réduite. Ceci est dû au fait qu'il n'y a pas vraiment d'intervalle où les deux densités ont une masse importante en même temps, comme on peut le voir dans le graphique d). Ainsi, la densité normale est multipliée par l'une des queues de la densité Student, qui, à ce moment, est presque horizontale.

Finalement, on peut voir grâce au graphique b) de la Figure 3.1 qu'augmenter le degré de liberté a un impact sur la forme de la densité, diminuant son asymétrie.

Suite à la Proposition 65 et des applications potentielles, il est pertinent d'étudier les propriétés d'une densité. Une des premières questions est celle de sa forme et plus particulièrement le nombre de modes. En effet, pour $p = 1$, l'unimodalité d'une loi a posteriori simplifie la détermination d'un intervalle de crédibilité le plus court (en anglais *HPD*).

Lemme 68. Soient $f \equiv e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)}$ le noyau de la densité produit et $x_0 = \mu - \sqrt{3v}\sigma$. Alors,

- Si $x_0 < 0$, alors f est unimodale.

- Si $x_0 > 0$ et $8v\sigma^2 > v + 1$, alors f est unimodale.
- Si $x_0 > 0$ et $8v\sigma^2 < v + 1$, alors f peut être soit unimodale ou bimodale.

Démonstration. On pose

$$\begin{aligned} H(x) &= -(\ln f(x))' \\ &= x + \frac{(v+1)(x-\mu)}{v\sigma^2 + (x-\mu)^2}. \end{aligned}$$

Alors, il suffit de s'intéresser au nombre de racines réelles que H possède pour déterminer si f admet un seul mode ou plus d'un. Le cas où $\mu = 0$ garantit qu'il n'y aura qu'un seul mode puisque dans ce cas les deux distributions que l'on multiplie ont leurs modes en 0. On se permet ainsi, sans perte de généralité, de considérer seulement les cas où $\mu > 0$. Les dérivées premières et secondes de H sont alors respectivement données par

$$\begin{aligned} H'(x) &= 1 + (v+1) \frac{(v\sigma^2 - (x-\mu)^2)}{(v\sigma^2 + (x-\mu)^2)^2} \\ H''(x) &= -\frac{2(v+1)(x-\mu)(3v\sigma^2 - (x-\mu)^2)}{(v\sigma^2 + (x-\mu)^2)^3} \end{aligned}$$

Une simple analyse de la fonction H permet de constater que $H(0) < 0$ et que $H(\mu) > 0$. Ainsi, la partie du support qui est d'intérêt est $(0, \mu)$. Comme les racines de H'' sont $\mu, \mu \pm \sqrt{3v}\sigma$, il suffit de considérer les cas où $x_0 = \mu - \sqrt{3v}\sigma$ est positif. En effet, si $x_0 < 0$, alors la fonction H sera convexe sur $(0, \mu)$ et n'admettra qu'une racine réelle. De plus, on a que $H'(x_0) = \frac{8v\sigma^2 - (v+1)}{8v\sigma^2}$ et donc on trouve que $H'(x_0) < 0 \iff 8v\sigma^2 < v + 1$.

Dans le cas où $x_0 > 0$ et $8v\sigma^2 > v + 1$, on a que $H'(x)$ décroît sur $(0, x_0)$ et que $H'(x)$ croît sur (x_0, μ) . Ainsi, H n'admet qu'une seule racine réelle dans $(0, \mu)$.

Enfin, pour le cas où $x_0 > 0$ et $8v\sigma^2 < v + 1$, on trouve des cas de figure où H admet trois racines réelles.

□

Exemple 69. Les conditions du Lemme 68 pour l'unimodalité de f_{μ,v,σ^2} sont satisfaites pour beaucoup de triplets (μ, v, σ) . Pour faire suite au dernier commentaire de la preuve du Lemme 68, on donne un cas de figure où la distribution est bimodale.

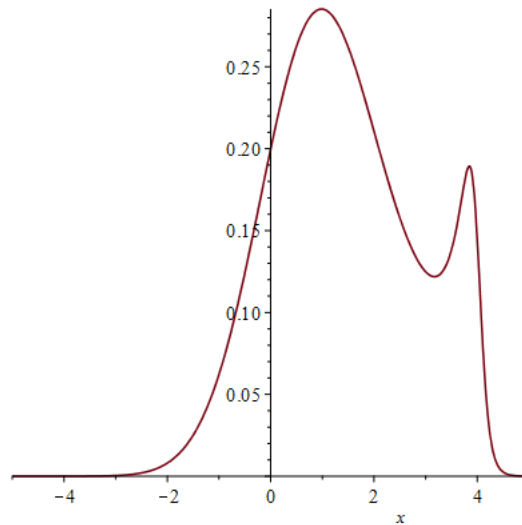


Figure 3.2 – Densité d'un produit d'une normale centrée réduite avec une densité Student de paramètre $\mu = 4$, $v = 2$ et $\sigma^2 = 0.05$.

Remarque 70. Lorsqu'on parle de densité de Student, il peut être intéressant de considérer le cas où le degré de liberté devient particulièrement grand. Ainsi, il est naturel de vouloir considérer

$$\begin{aligned} \lim_{v \rightarrow \infty} \frac{\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx}{\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx} &= \frac{\int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{\int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx} \\ &= \frac{\sqrt{1 + \sigma^2} e^{-\frac{x^2}{2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma^2 \sqrt{2\pi} e^{-\frac{\mu^2}{2(\sigma^2+1)}}}. \end{aligned}$$

La permutation de l'intégrale et de la limite dans les derniers calculs étant justifié par le théorème de convergence dominée de Lebesgue. Ainsi, on peut calculer l'espérance par rapport à cette densité, soit

$$\begin{aligned} e^{\frac{\mu^2}{2(\sigma^2+1)}} \int_{\mathbb{R}} \sqrt{\frac{1+\sigma^2}{\sigma^2}} \frac{x e^{-\frac{x^2}{2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}} dx &= \int_{\mathbb{R}} \sqrt{\frac{1+\sigma^2}{\sigma^2}} \frac{x e^{-\frac{1}{2}\left(\frac{1+\sigma^2}{\sigma^2}\right)\left(x-\frac{\mu\sigma^2}{\sigma^2+1}\right)^2}}{\sqrt{2\pi}} dx \\ &= \frac{\mu\sigma^2}{\sigma^2+1}. \end{aligned}$$

On trouve également que la variance de cette densité correspond à $\frac{\sigma^2}{\sigma^2+1}$. De même, on trouve les mêmes expressions lorsqu'on considère le cas multivarié, la seule différence étant que μ sera alors un vecteur.

Remarque 71. Un autre cas limite intéressant est de regarder ce qu'il se passe lorsque $\sigma^2 \rightarrow 0$. On s'intéresse donc à la quantité

$$\begin{aligned} R &= \lim_{\sigma^2 \rightarrow 0} \frac{\int_{\mathbb{R}} x e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx}{\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx} \\ &= \lim_{\sigma^2 \rightarrow 0} \frac{\mathbb{E}\left(X e^{-\frac{X^2}{2}}\right)}{\mathbb{E}\left(e^{-\frac{X^2}{2}}\right)}, \text{ où } X \sim t_{v,\mu,\sigma}. \end{aligned}$$

Lorsque $\sigma^2 \rightarrow 0$, X devient une variable aléatoire dégénérée en μ , c'est-à-dire que $X \xrightarrow{\text{Prob.}} \mu$. De plus, comme les fonctions $x e^{-\frac{x^2}{2}}$ et $e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$, sont toutes les deux bornées par 1 en valeur absolue, on a que $\mathbb{E}\left(X e^{-\frac{X^2}{2}}\right) \rightarrow \mu e^{-\frac{\mu^2}{2}}$ et que $\mathbb{E}\left(e^{-\frac{X^2}{2}}\right) \rightarrow e^{-\frac{\mu^2}{2}}$, et ce toujours grâce au théorème de convergence dominée de Lebesgue. On trouve finalement que $R \rightarrow \mu$ lorsque $\sigma^2 \rightarrow 0$. Si l'on considère plutôt le cas multivarié, on trouvera que la limite de cette espérance est le vecteur μ . En effet, les mêmes conditions seront vérifiées.

On termine cette section en donnant l'espérance et la variance pour les densités ayant la forme donnée en (\star) lorsque $p = 1$.

Lemme 72. Soit $X \sim \frac{e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)}}{C_1}$ où C_1 est la constante qui fait de cette fonction une densité. Alors,

$$\mathbb{E}(X) = \mu \left(1 - \frac{1}{C_1} e^{-\frac{\mu^2}{2}} (v\sigma^2)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right) \sum_{k \geq 0} \frac{(\mu^2 v \sigma^2)^k}{2^{2k} k!} \Phi\left(k + \frac{3}{2}, k + \frac{5}{2} - \frac{(v+1)}{2}, \frac{v\sigma^2}{2}\right) \right),$$

où Φ est la fonction hypergéométrique confluyente de type II de la Définition 32 qui est reliée à la densité de Kummer.

Démonstration. On a

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{C_1} \int_{\mathbb{R}} x e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx \\ &= \frac{1}{C_1} \int_{\mathbb{R}} (x+\mu) e^{-\frac{(x+\mu)^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx \\ &= \frac{1}{C_1} \int_{\mathbb{R}} x e^{-\frac{(x+\mu)^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx + \mu. \end{aligned}$$

On note ici que l'espérance existe puisque $\int_{\mathbb{R}} |x| e^{-\frac{(x+\mu)^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} \leq \int_{\mathbb{R}} |x| e^{-\frac{(x+\mu)^2}{2}} < \infty$.

Ainsi, on a

$$\begin{aligned} \int_{\mathbb{R}} x e^{-\frac{(x+\mu)^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx &= e^{-\frac{\mu^2}{2}} \int_{\mathbb{R}} x e^{-\frac{x^2}{2}} e^{-x\mu} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx \\ &= e^{-\frac{\mu^2}{2}} \sum_{j \geq 0} \frac{(-\mu)^j}{j!} \int_{\mathbb{R}} x^{j+1} e^{-\frac{x^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx. \end{aligned}$$

En posant $I_j = \int_{\mathbb{R}} x^{j+1} e^{-\frac{x^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx$, on a que

$$I_j = \int_{\mathbb{R}_+} (-1)^{j+1} x^{j+1} e^{-\frac{x^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx + \int_{\mathbb{R}_+} x^{j+1} e^{-\frac{x^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx.$$

On voit que $I_j = 0$ lorsque j est pair. Sinon, c'est-à-dire lorsqu'on peut écrire $j = 2k + 1$, on a

$$\begin{aligned} I_{2j+1} &= 2 \int_{\mathbb{R}_+} x^{j+1} e^{-\frac{x^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx \\ &= \int_{\mathbb{R}_+} (yv\sigma^2)^{\frac{2k+1}{2}} e^{-y\left(\frac{v\sigma^2}{2}\right)} (1+y)^{-\left(\frac{v+1}{2}\right)} v\sigma^2 dy \\ &= (v\sigma^2)^{k+\frac{3}{2}} \Gamma\left(k + \frac{3}{2}\right) \Phi\left(k + \frac{3}{2}, k + \frac{5}{2} - \frac{(v+1)}{2}, \frac{v\sigma^2}{2}\right). \end{aligned}$$

On trouve finalement que

$$\begin{aligned} \int_{\mathbb{R}} x e^{-\frac{(x+\mu)^2}{2}} \left(1 + \frac{x^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)} dx &= e^{-\frac{\mu^2}{2}} \sum_{k \geq 0} \frac{(-\mu)^{2k+1}}{(2k+1)!} I_{2k+1} \\ &= -\mu e^{-\frac{\mu^2}{2}} (v\sigma^2)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right) \sum_{k \geq 0} \frac{(\mu^2 v\sigma^2)^k}{2^{2k} k!} \\ &\quad \times \Phi\left(k + \frac{3}{2}, k + \frac{5}{2} - \frac{(v+1)}{2}, \frac{v\sigma^2}{2}\right), \end{aligned}$$

d'où le résultat. □

Lemme 73. Soit $X \sim \frac{e^{-\frac{\|x\|^2}{2}} \left(1 + \frac{\|x-\mu\|^2}{v\sigma^2}\right)^{-\left(\frac{v+p}{2}\right)}}{C_1}$. Alors,

$$\begin{aligned} \mathbb{E}(X^\top X) &= \frac{\sqrt{\pi}}{C_1} e^{-\frac{\|\mu\|^2}{2}} (v\sigma^2)^{\frac{3}{2}} \sum_{L \geq 0} \frac{\left(\frac{\|\mu\|^2 v\sigma^2}{4}\right)^L (L + \frac{1}{2})}{L!} \Phi\left(L + \frac{3}{2}, L + \frac{5}{2} - \frac{(v+1)}{2}, \frac{v\sigma^2}{2}\right) \\ &\quad + 2\mu \mathbb{E}(X) - \|\mu\|^2. \end{aligned}$$

Démonstration. On a

$$\begin{aligned}\mathbb{E}(X^\top X) &= \frac{1}{C_1} \int_{\mathbb{R}^p} x^\top x e^{-\frac{x^\top x}{2}} \left(1 + \frac{(x - \mu)^\top (x - \mu)}{v\sigma^2}\right)^{-\left(\frac{v+p}{2}\right)} dx \\ &= \frac{1}{C_1} \int_{\mathbb{R}^p} x^\top x e^{-\frac{(x+\mu)^\top (x+\mu)}{2}} \left(1 + \frac{x^\top x}{v\sigma^2}\right)^{-\left(\frac{v+p}{2}\right)} dx + 2\mu \mathbb{E}(X) + \mu^\top \mu.\end{aligned}$$

Il suffit donc de calculer l'intégrale présente dans le premier terme. On a

$$\int_{\mathbb{R}^p} x^\top x e^{-\frac{(x+\mu)^\top (x+\mu)}{2}} \left(1 + \frac{x^\top x}{v\sigma^2}\right)^{-\left(\frac{v+p}{2}\right)} dx = (2\pi)^{\frac{p}{2}} \mathbb{E}\left(x^\top x \left(1 + \frac{x^\top x}{v\sigma^2}\right)^{-\left(\frac{v+p}{2}\right)}\right),$$

où $X \sim \mathcal{N}(-\mu, \mathbb{1}_p)$. En posant $Y = X^\top X \sim \chi_p^2(\mu^\top \mu)$ et en utilisant la représentation de la loi du χ^2 décentrée sous forme de mélange, c'est-à-dire

$$\begin{aligned}Y|L &\sim \chi_{p+2L}^2 \sim \mathcal{G}\left(\frac{p}{2} + L, 2\right) \\ L &\sim \mathcal{P}\left(\frac{\mu^\top \mu}{2}\right),\end{aligned}$$

on a que

$$\begin{aligned}
\mathbb{E}^L \mathbb{E} \left(Y \left(1 + \frac{Y}{v\sigma^2} \right)^{-\left(\frac{v+p}{2}\right)} \middle| L \right) &= \mathbb{E}^L \left(\int_{\mathbb{R}_+} y \left(1 + \frac{y}{v\sigma^2} \right)^{-\left(\frac{v+p}{2}\right)} \frac{y^{\frac{p}{2}+L-1} e^{-\frac{y}{v\sigma^2}}}{\Gamma\left(\frac{p}{2}+L\right) 2^{\frac{p}{2}+L}} dy \right) \\
&= \mathbb{E}^L \left(\frac{(v\sigma^2)^{\frac{p+2}{2}+L}}{\Gamma\left(\frac{p}{2}+L\right) 2^{\frac{p}{2}+L}} \int_{\mathbb{R}_+} z^{\frac{3}{2}+L-1} (1+z)^{-\left(\frac{v+p}{2}\right)} e^{-z\left(\frac{v\sigma^2}{2}\right)} dz \right) \\
&= \mathbb{E}^L \left[\frac{(v\sigma^2)^{\frac{p+2}{2}+L} \Gamma\left(L + \frac{p+2}{2}\right)}{2^{\frac{p}{2}+L} \Gamma\left(L + \frac{p}{2}\right)} \right. \\
&\quad \left. \times \Phi \left(L + \frac{p+2}{2}, L + \frac{p+4}{2} - \frac{(v+p)}{2}, \frac{v\sigma^2}{2} \right) \right] \\
&= e^{-\frac{\mu^\top \mu}{2}} \frac{(v\sigma^2)^{\frac{p+1}{2}}}{2^{\frac{p}{2}}} \sum_{L \geq 0} \frac{\left(\frac{\mu^\top \mu v \sigma^2}{4}\right)^L (L + \frac{p}{2})}{L!} \\
&\quad \times \Phi \left(L + \frac{p+2}{2}, L + \frac{p+4}{2} - \frac{(v+p)}{2}, \frac{v\sigma^2}{2} \right),
\end{aligned}$$

d'où le résultat. □

Corollaire 74. Soit $X \sim \frac{e^{-\frac{x^2}{2}} \left(1 + \frac{(x-\mu)^2}{v\sigma^2}\right)^{-\left(\frac{v+1}{2}\right)}}{C_1}$. Alors,

$$\begin{aligned}
\mathbb{V}(X) &= \frac{\sqrt{\pi}}{C_1} e^{-\frac{\mu^2}{2}} (v\sigma^2)^{\frac{3}{2}} \sum_{L \geq 0} \frac{\left(\frac{\mu^2 v \sigma^2}{4}\right)^L (L + \frac{1}{2})}{L!} \Phi \left(L + \frac{3}{2}, L + \frac{5}{2} - \frac{(v+1)}{2}, \frac{v\sigma^2}{2} \right) \\
&\quad - \mu^2 + \mathbb{E}(X) (2\mu - \mathbb{E}(X))
\end{aligned}$$

Démonstration. L'expression de la variance découle directement des Lemmes 72 et 73. □

Exemple 75. On s'intéresse ici au comportement de $\mathbb{E}(X)$ lorsqu'on fait varier les paramètres μ et σ . On a utilisé comme paramètres de base $\mu = 2$, $v = 3$ et $\sigma = 1$.

Dans le graphique a) de la Figure 3.3, on voit que $\mathbb{E}(X)$ est une fonction impaire de μ . Pour ce qui est du graphique b), on voit que $\mathbb{E}(X)$ décroît lorsque σ augmente. De

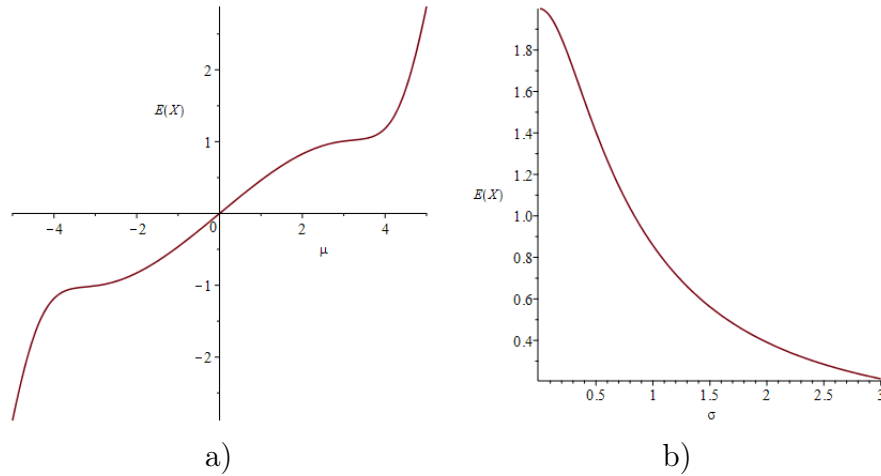


Figure 3.3 – Espérance d’une densité produit normale avec une densité Student pour a) $v = 3, \sigma = 1$, b) $v = 3, \mu = 2$.

plus, le graphique illustre la propriété $\mathbb{E}(X) \rightarrow \mu$ lorsque $\sigma \rightarrow 0$, tel que mentionné dans la Remarque 71. Enfin, bien que ce ne soit pas illustré ici, on a que lorsque $\sigma \rightarrow \infty$, $\mathbb{E}(X) \rightarrow 0$. Ceci est dû au fait que la composante Student de la fonction devient alors uniforme et on récupère alors l’espérance de la loi normale centrée en 0.

Ainsi, on a donné quelques propriétés de base concernant la densité produit telle que présentée à la Proposition 65. On a préféré présenter sa forme, son espérance, sa variance ainsi qu’une brève analyse de ces paramètres pour une forme standard de la densité plutôt que celle de la proposition qui fait intervenir plus de paramètres.

3.2 Marginale a posteriori de σ^2

Pour compléter le travail de la section précédente, on présente la densité marginale a posteriori de σ^2 . On doit tout d’abord introduire le Lemme qui suit afin de s’assurer de pouvoir intervertir sommation et intégrale au cours de cette sous-section.

Notation 76. On a déjà introduit au Chapitre 1 la fonction ${}_1F_1(a, b, z) = \sum_{k \geq 0} \frac{(a)_k z^k}{(b)_k k!}$.

On peut également définir une fonction analogue à cette dernière, soit

$${}_0F_1(a, -, z) = \sum_{k \geq 0} \frac{(a)_k z^k}{k!}.$$

Lemme 77. Soit $f_n(x) = \frac{\theta^n}{n!} \frac{e^{-\frac{x}{\beta}} x^{\alpha+n-1}}{(x+1)^n} \mathbf{1}_{([0, \infty[)}(x)$, où $\theta \geq 0$. Alors,

$$\sum_{n \geq 0} \int_{\mathbb{R}_+} f_n(x) dx < \infty.$$

Démonstration. On a

$$\begin{aligned} \sum_{n \geq 0} \int_{\mathbb{R}_+} f_n(x) dx &= \sum_{n \geq 0} \frac{\theta^n}{n!} \int_{\mathbb{R}_+} \frac{e^{-\frac{x}{\beta}} x^{\alpha+n-1}}{(x+1)^n} dx \\ &\leq \sum_{n \geq 0} \frac{\theta^n}{n!} \Gamma(\alpha+n) \beta^{\alpha+n} \\ &= \Gamma(\alpha) \beta^\alpha \sum_{n \geq 0} \frac{(\theta\beta)^n}{n!} \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \\ &= \Gamma(\alpha) \beta^\alpha {}_0F_1(\alpha; -; \theta\beta), \end{aligned}$$

d'où le résultat. □

Proposition 78. Soient \bar{X} et S des variables indépendantes telles que $\bar{X} | \theta, \sigma^2 \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$ et $nS^2 | \theta, \sigma^2 \sim \mathcal{G}\left(\frac{n-1}{2}, 2\sigma^2\right)$. On suppose ici que $\theta | \sigma^2 \sim \mathcal{N}(\mu, \tau^2)$ et $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$. Alors, la densité a posteriori de σ^2 est donnée par :

$$\pi(\sigma^2 | \bar{x}, s) = \frac{1}{C_2} \frac{e^{-\frac{(ns^2+2\beta)}{2\sigma^2}} e^{-\frac{n(\bar{x}-\mu)^2}{2(n\tau^2+\sigma^2)}}}{(\sigma^2)^{\frac{n-1}{2}+\alpha+1} \sqrt{n\tau^2+\sigma^2}}$$

où

$$C_2 = \sum_{j=0}^{\infty} \frac{\left(\frac{-n(\bar{x}-\mu)^2}{2}\right)^j}{j!} \left(\frac{1}{n\tau^2}\right)^{\frac{n-1}{2}+\alpha+j+\frac{1}{2}} \Gamma\left(\frac{n-1}{2} + \alpha + j + \frac{1}{2}\right) \\ \times \Phi\left(\frac{n-1}{2} + \alpha + j + \frac{1}{2}; 1 + \frac{n-1}{2} + \alpha; \frac{nS^2 + 2\beta}{2n\tau^2}\right).$$

Démonstration. On a

$$\pi(\theta, \sigma^2 | \bar{x}, s) \propto \frac{e^{-\frac{(n(\bar{x}-\theta)^2 + ns^2 + 2\beta)}{2\sigma^2}} e^{-\frac{(\theta-\mu)^2}{2\tau^2}}}{(\sigma^2)^{\frac{n}{2}+\alpha+1}}, \theta \in \mathbb{R}, \sigma^2 > 0,$$

d'où

$$\pi(\sigma^2 | \bar{x}, s) \propto \int_{\mathbb{R}} \frac{e^{-\frac{(n(\bar{x}-\theta)^2 + ns^2 + 2\beta)}{2\sigma^2}} e^{-\frac{(\theta-\mu)^2}{2\tau^2}}}{(\sigma^2)^{\frac{n}{2}+\alpha+1}} d\theta \\ \propto \frac{e^{-\frac{(ns^2 + 2\beta)}{2\sigma^2}} e^{-\frac{n(\bar{x}-\mu)^2}{2(n\tau^2 + \sigma^2)}}}{(\sigma^2)^{\frac{n-1}{2}+\alpha+1} \sqrt{n\tau^2 + \sigma^2}}.$$

En exprimant la seconde exponentielle sous la forme d'une série, et en posant $a = \frac{n-1}{2} + \alpha$, $b = \frac{ns^2 + 2\beta}{2}$ et $c = n\tau^2$, on obtient

$$\pi(\sigma^2 | \bar{x}, s) = \frac{1}{C_2} \sum_{j=0}^{\infty} \frac{\left(\frac{-n(\bar{x}-\mu)^2}{2}\right)^j}{j!} \frac{e^{-\frac{b}{\sigma^2}}}{(\sigma^2)^{a+1} (c + \sigma^2)^{j+\frac{1}{2}}},$$

d'où

$$\begin{aligned}
C_2 &= \int_{\mathbb{R}_+} \sum_{j=0}^{\infty} \frac{\left(\frac{-n(\bar{x}-\mu)^2}{2}\right)^j}{j!} \frac{e^{-\frac{b}{\sigma^2}}}{(\sigma^2)^{a-1} (c + \sigma^2)^{j+\frac{1}{2}}} d\sigma^2 \\
&= \sum_{j=0}^{\infty} \frac{\left(\frac{-n(\bar{x}-\mu)^2}{2}\right)^j}{j!} \int_{\mathbb{R}_+} \frac{e^{-\frac{b}{\sigma^2}}}{(\sigma^2)^{a-1} (c + \sigma^2)^{j+\frac{1}{2}}} d\sigma^2,
\end{aligned}$$

la dernière égalité étant valide en vertu du Lemme 77 et du théorème de convergence dominée de Lebesgue. En remarquant que l'intégrale correspond au noyau d'une distribution de Kummer de type II, on trouve

$$\begin{aligned}
C_2 &= \sum_{j=0}^{\infty} \frac{\left(\frac{-n(\bar{x}-\mu)^2}{2}\right)^j}{j!} \left(\frac{1}{c}\right)^{a+j+\frac{1}{2}} \Gamma\left(a+j+\frac{1}{2}\right) \Phi\left(a+j+\frac{1}{2}, 1+a, \frac{b}{c}\right) \\
&= \sum_{j=0}^{\infty} \frac{\left(\frac{-n(\bar{x}-\mu)^2}{2}\right)^j}{j!} \left(\frac{1}{n\tau^2}\right)^{\frac{n-1}{2}+\alpha+j+\frac{1}{2}} \Gamma\left(\frac{n-1}{2}+\alpha+j+\frac{1}{2}\right) \\
&\quad \times \Phi\left(\frac{n-1}{2}+\alpha+j+\frac{1}{2}, 1+\frac{n-1}{2}+\alpha, \frac{ns^2+2\beta}{2n\tau^2}\right).
\end{aligned}$$

□

Exemple 79. À la manière de l'Exemple 67, on a exploré différentes combinaisons de paramètres afin de bien voir leur impact sur la densité a posteriori de σ^2 . Cette fois, les paramètres de base sont $n = 10$, $\alpha = 2$, $\beta = 2$, $\tau = 1$, $\mu = 5$, $\bar{x} = 0.81093$ et $s = 1.0220142$. Dans chacun des graphiques qui suivent, seulement l'un des paramètres a été modifié.

En jetant un coup d'oeil au graphique a) de la Figure 3.4, on voit que la moyenne a posteriori de σ^2 a augmenté. De plus, si l'on regarde également le graphique b), on remarque que plus la taille d'échantillon augmente, moins la distribution a posteriori est dispersée. La distribution semble également approchée la valeur de s^2 .

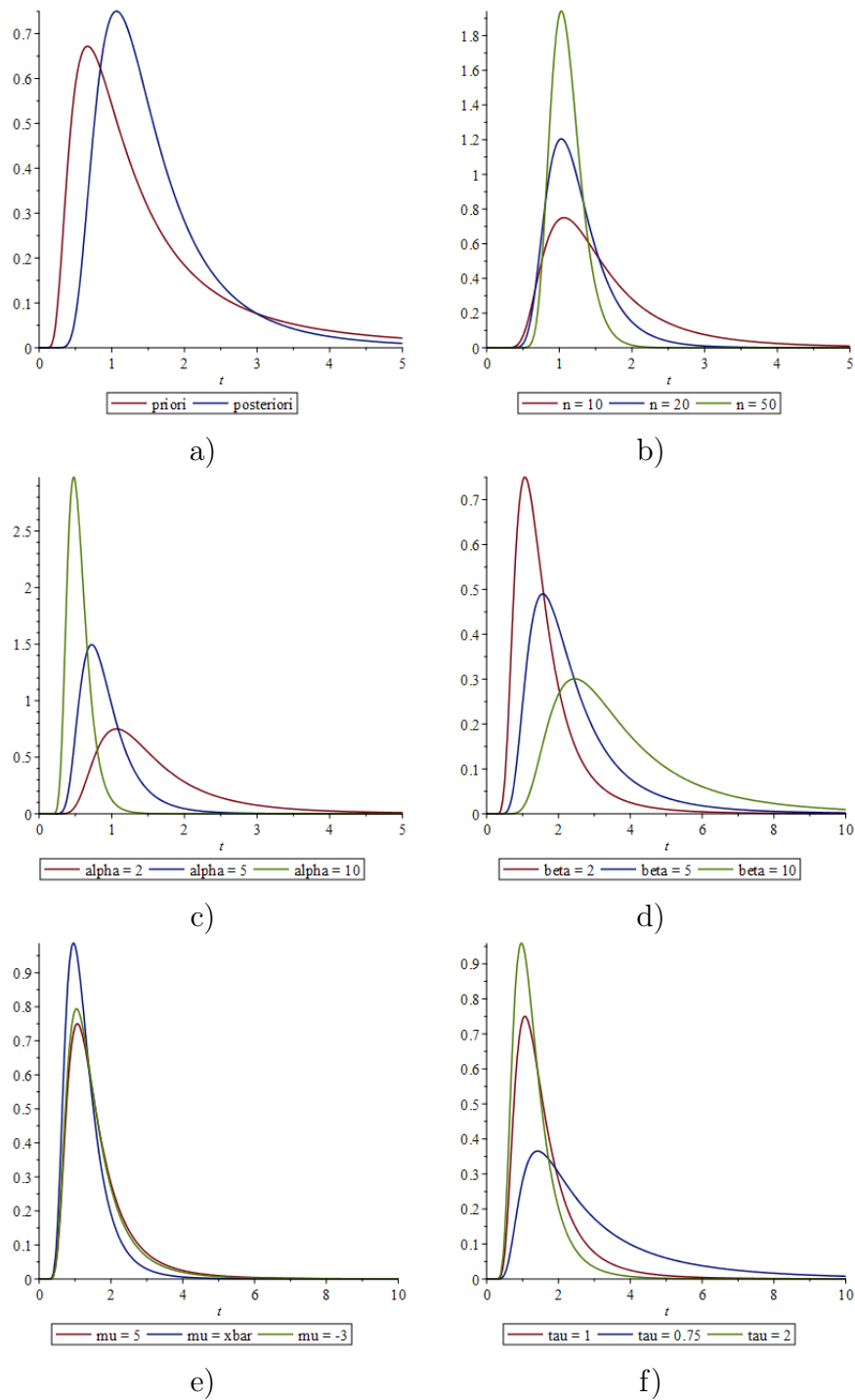


Figure 3.4 – Différents graphiques qui représentent la densité marginale de σ^2 a) a priori, mais également a posteriori lorsqu'on altère : b) la taille d'échantillon n , les hyperparamètres c) α , d) β , e) μ et f) τ .

Les graphiques c) et d) de la Figure 3.4 semblent montrer que les paramètres α et β ont des impacts opposés. En effet, lorsqu'on augmente la valeur de α , la distribution a posteriori est de moins en moins dispersée alors que l'on constate l'effet inverse pour le paramètre β . On peut aussi voir grâce au graphique f), que le paramètre τ a le même genre d'impact que le paramètre α .

Finalement, le graphique e) de la Figure 3.4 montre l'impact du paramètre μ . La courbe en bleue est la densité a posteriori qui minimise la dispersion ou encore la variance. En fait, la densité marginale a posteriori de σ^2 est alors une densité de Kummer de type II et les calculs sont alors beaucoup plus simples. En effet, si l'on remplace μ par \bar{x} dans la densité de la Proposition 78, alors une des deux exponentielles disparaît et un simple changement de variable permettra de récupérer la densité de Kummer de type II. Pour les autres valeurs de μ , la densité gagne en dispersion en fonction de la distance entre μ et \bar{x} . Remarquons que le cas $\mu = \bar{x}$ ne se réalisera pas dans le contexte étudié et la densité n'est donnée qu'à titre indicatif.

Le prochain résultat donne la moyenne et la variance a posteriori de σ^2 . Afin de faciliter les calculs qui suivent, on pose

$$\begin{aligned} C_2(a, b, c, x) &= \int \sum_{j=0}^{\infty} \frac{x^j}{j!} \frac{e^{-\frac{b}{\sigma^2}}}{(\sigma^2)^{a+1} (c + \sigma^2)^{j+\frac{1}{2}}} d\sigma^2 \\ &= \sum_{j=0}^{\infty} \frac{x^j}{j!} \left(\frac{1}{c}\right)^{a+j+\frac{1}{2}} \Gamma\left(a+j+\frac{1}{2}\right) \Phi\left(a+j+\frac{1}{2}, 1+a, \frac{b}{c}\right). \end{aligned}$$

Corollaire 80. Dans le cadre de la proposition précédente, on a que l'espérance et la variance a posteriori de σ^2 sont données respectivement par :

$$\mathbb{E}(\sigma^2 | \bar{X}, S) = \frac{C_2\left(\frac{n-1}{2} + \alpha - 1; \frac{ns^2+2\beta}{2}; n\tau^2, \frac{-n(\bar{x}-\mu)^2}{2}\right)}{C_2\left(\frac{n-1}{2} + \alpha; \frac{ns^2+2\beta}{2}; n\tau^2, \frac{-n(\bar{x}-\mu)^2}{2}\right)}$$

et

$$\mathbb{V}(\sigma^2 | \bar{X}, S) = \frac{C_2\left(\frac{n-1}{2} + \alpha - 2; \frac{ns^2+2\beta}{2}; n\tau^2, \frac{-n(\bar{x}-\mu)^2}{2}\right)}{C_2\left(\frac{n-1}{2} + \alpha; \frac{ns^2+2\beta}{2}; n\tau^2, \frac{-n(\bar{x}-\mu)^2}{2}\right)} - \left(\frac{C_2\left(\frac{n-1}{2} + \alpha - 1; \frac{ns^2+2\beta}{2}; n\tau^2, \frac{-n(\bar{x}-\mu)^2}{2}\right)}{C_2\left(\frac{n-1}{2} + \alpha; \frac{ns^2+2\beta}{2}; n\tau^2, \frac{-n(\bar{x}-\mu)^2}{2}\right)} \right)^2$$

Démonstration. Immédiat en utilisant l'écriture proposée dans la remarque précédente. □

CHAPITRE 4

Inférence bayésienne pour une fonction d'une moyenne et d'une variance

Dans ce chapitre, nous allons étudier le modèle $X_i|\theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2), i = 1, \dots, n, \pi(\theta|\sigma^2) \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right)$ et $\frac{1}{\sigma^2} \sim g$, où g est une densité quelconque, et considérer différentes fonctions d'une moyenne et d'une variance. Les approches dans la littérature pour ces problèmes semblent être computationnelles ou basées sur des méthodes de vraisemblance. Par exemple, [ER04] présentent l'échantillonneur de Gibbs ainsi que les lois *conditionnelles complètes* pour estimer la densité a posteriori d'un coefficient de variation. Les auteurs de [CL07] utilisent quand à eux des simulations pour trouver les bornes d'un intervalle de crédibilité pour un quantile. L'objectif et la contribution de ce chapitre sont de développer différentes expressions de façon analytique. Dans la première section de ce chapitre, on considère l'estimation de la densité d'un quantile a posteriori. On parlera également de l'approche du pivot pour certains cas intéressants. Dans la deuxième section de ce chapitre, c'est l'estimation de la densité a posteriori du coefficient de variation ainsi que de son inverse qui sera abordé. On terminera le chapitre en donnant quelques

résultats sur une extension multivariée, soit le coefficient de discrimination. Tout au long du chapitre, on développera des expressions pour l'espérance et la variance de ces diverses quantités et quelques représentations graphiques seront également fournies.

4.1 Estimation du quantile

Dans cette section, nous allons considérer l'inférence bayésienne pour les quantiles, d'une loi normale, de la forme $\tau = \theta + c\sigma$, $c \neq 0$ avec lois a priori : $\pi(\theta|\sigma^2) \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right)$ et $\frac{1}{\sigma^2} \sim g$. Pour le modèle $X_i \sim \mathcal{N}(\theta, \sigma^2)$, $i = 1, \dots, n$, nous déduirons la densité a posteriori, et d'autres propriétés générales, du quantile τ . Ensuite, le cas où $g \sim \mathcal{G}(\alpha, \beta)$ sera détaillé et une simulation sera présentée afin de bien illustrer ces résultats.

Remarque 81. *Dans cette section, on utilisera la reparamétrisation $\eta = \frac{1}{\sigma^2}$ afin de faciliter certaines écritures.*

Théorème 4.1. *Pour $X|\theta, \sigma^2 \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$ et $S^2|\theta, \sigma^2 \sim \mathcal{G}\left(\frac{k}{2}, 2\sigma^2\right)$ indépendantes, l'a priori $\theta|\eta \sim \mathcal{N}\left(\mu, \frac{1}{\lambda\eta}\right)$ et $\eta \sim g$, la densité et la répartition du quantile $\tau = \theta + c\sigma$ a posteriori sont respectivement données par :*

$$f_\tau(t) = e^{-c^2 \frac{\lambda'}{2}} \frac{\lambda'}{\sqrt{2\pi}} \int_{\mathbb{R}_+} \sqrt{\eta} g(\eta|x, s^2) e^{-A(t)\eta + B(t)\sqrt{\eta}} d\eta$$

$$\text{et } F_\tau(t) = \int_{\mathbb{R}_+} \Phi\left(\sqrt{\lambda'\eta}\left(t - \mu' - \frac{c}{\sqrt{\eta}}\right)\right) g(\eta|x, s^2) d\eta,$$

où $\mu' = \frac{nx + \lambda\mu}{\lambda + n}$, $\lambda' = \lambda + n$, $A(t) = \frac{\lambda'(t - \mu')^2}{2}$ et $B(t) = c\lambda'(t - \mu')$.

Démonstration. Remarquons tout d'abord que la loi de $\theta|x, s^2, \eta$ est normale $\mathcal{N}\left(\mu', \frac{1}{\lambda'\eta}\right)$.

En effet, on a que

$$\begin{aligned}
\pi(\theta|x, s^2, \eta) &\propto f(x, s^2|\theta, \eta)\pi(\theta|\eta) \\
&\propto e^{-\frac{(x-\theta)^2}{2\frac{\sigma^2}{n}}} e^{-\frac{(\theta-\mu)^2\lambda\eta}{2}} \\
&\propto e^{-\frac{1}{2}\left(\frac{n+\sigma^2\lambda\eta}{\sigma^2}\right)\left(\theta-\frac{\frac{n\bar{x}+\mu\sigma^2}{\lambda\eta}+\frac{n}{\lambda\eta}}{\sigma^2+\frac{n}{\lambda\eta}}\right)^2} \\
&\propto e^{-\frac{1}{2}\eta(\lambda+n)\left(\theta-\frac{n\bar{x}+\lambda\mu}{\lambda+n}\right)^2}.
\end{aligned}$$

On a alors

$$\begin{aligned}
\mathbb{P}(\tau \leq t|X, S^2) &= \int \mathbb{P}(\theta \leq t - c\eta^{\frac{1}{2}}|X, S^2, \eta)g(\eta|x, s^2)d\eta \\
&= \int_{\mathbb{R}_+} \Phi\left(\sqrt{\lambda'\eta}\left(t - \mu' - \frac{c}{\sqrt{\eta}}\right)\right)g(\eta|x, s^2)d\eta,
\end{aligned}$$

ce qui nous donne la fonction de répartition du quantile a posteriori. En dérivant par rapport à t , on obtient

$$\begin{aligned}
f_\tau(t) &= \int_{\mathbb{R}_+} \phi\left(\sqrt{\lambda'\eta}\left(t - \mu' - \frac{c}{\sqrt{\eta}}\right)\right)\sqrt{\lambda'\eta} g(\eta|x, s^2)d\eta \\
&= \int_{\mathbb{R}_+} \frac{e^{-\frac{(t-\mu'-\frac{c}{\sqrt{\eta}})^2\lambda'\eta}{2}}}{\sqrt{2\pi}}\sqrt{\lambda'\eta} g(\eta|x, s^2)d\eta \\
&= \sqrt{\frac{\lambda'}{2\pi}}e^{-\frac{c^2\lambda'}{2}} \int_{\mathbb{R}_+} e^{-A(t)\eta+B(t)\sqrt{\eta}}\sqrt{\eta} g(\eta|x, s^2)d\eta.
\end{aligned}$$

□

Corollaire 82. *Toujours sous le modèle du théorème 4.1, l'espérance ainsi que la variance a posteriori sont données par les expressions suivantes :*

$$\mathbb{E}(\theta + c\sigma|x, s^2) = \mu' + c \int_{\mathbb{R}_+} \eta^{-\frac{1}{2}} g(\eta|x, s^2)d\eta \quad (4.1)$$

$$\mathbb{V}(\theta + c\sigma|x, s^2) = \mathbb{V}(\theta|x, s^2) + c^2 \mathbb{V}(\sigma|x, s^2). \quad (4.2)$$

Démonstration. Dans un premier temps, on a que

$$\begin{aligned}\mathbb{E}(\theta + c\sigma|x, s^2) &= \mathbb{E}\mathbb{E}(\theta + c\sigma|x, s^2, \sigma^2) \\ &= \mathbb{E}(\mu' + c\sigma|x, s^2) \\ &= \mu' + c \int_{\mathbb{R}_+} \eta^{-\frac{1}{2}} g(\eta|x, s^2) d\eta, \text{ ce qui est 4.1.}\end{aligned}$$

Remarquons maintenant que

$$\begin{aligned}\mathbb{E}(\theta\sigma|x, s^2) &= \mathbb{E}^{\sigma|x, s^2}(\sigma \mathbb{E}(\theta|\sigma, x, s^2)) \\ &= \mu' \mathbb{E}(\sigma|x, s^2) \\ &= \mathbb{E}(\theta|x, s^2) \mathbb{E}(\sigma|x, s^2).\end{aligned}$$

Les variables θ et σ étant de covariance a posteriori nulle, d'où l'identité 4.2 qui en découle.

□

Avant de parler du cas particulier où $g \sim \mathcal{G}(\alpha, \beta)$, nous devons introduire un lemme qui nous permettra de réécrire une intégrale comme étant une somme de fonctions hypergéométriques. On pourra alors obtenir une forme explicite de la densité du quantile, ce qui s'avérera très utile, voire indispensable, lors de simulations.

Lemme 83. *Soit l'intégrale $I = \int_{\mathbb{R}_+} e^{B\sqrt{y}} y^{a-1} e^{-by} dy$, pour $a > 0$, $b > 0$ et $B \in \mathbb{R}$. Alors, on a que*

$$I = \frac{1}{b^a} \left(\Gamma(a) {}_1F_1\left(a; \frac{1}{2}; \frac{B^2}{4b}\right) + \Gamma\left(a + \frac{1}{2}\right) \frac{B}{\sqrt{b}} {}_1F_1\left(a + \frac{1}{2}; \frac{3}{2}; \frac{B^2}{4b}\right) \right)$$

Démonstration. On a

$$\begin{aligned}
I &= \int_{\mathbb{R}_+} e^{B\sqrt{y}} y^{a-1} e^{-by} dy \\
&= \int_{\mathbb{R}_+} \sum_{k=0}^{\infty} \frac{(B\sqrt{y})^k}{k!} y^{a-1} e^{-by} dy \\
&= \sum_{k=0}^{\infty} \frac{B^k}{k!} \int_{\mathbb{R}_+} y^{a+\frac{k}{2}-1} e^{-by} dy \\
&= \frac{1}{b^a} \sum_{k=0}^{\infty} \frac{B^k \Gamma\left(a + \frac{k}{2}\right)}{k! b^{\frac{k}{2}}} \\
&= \frac{1}{b^a} \left(\sum_{m=0}^{\infty} \frac{\left(\frac{B^2}{b}\right)^m \Gamma(a+m)}{(2m)!} + \frac{B}{\sqrt{b}} \sum_{m=0}^{\infty} \frac{\left(\frac{B^2}{b}\right)^m \Gamma\left(a+m+\frac{1}{2}\right)}{(2m+1)!} \right),
\end{aligned}$$

la dernière égalité s'obtenant en prenant respectivement les termes pairs puis impairs de la ligne précédente et la permutation entre le signe d'intégrale et la sommation étant valable puisque le dernier terme du développement converge. En effet, il suffit maintenant d'utiliser la formule de dédoublement telle que présentée au Lemme 24 sur chacune des sommes pour obtenir le résultat. On a

$$\begin{aligned}
\sum_{m=0}^{\infty} \frac{\left(\frac{B^2}{b}\right)^m \Gamma(a+m)}{(2m)!} &= \sum_{m=0}^{\infty} \frac{\left(\frac{B^2}{b}\right)^m \Gamma(a+m)}{2^{2m} \left(\frac{1}{2}\right)_m m!} \\
&= \Gamma(a) \sum_{m=0}^{\infty} \frac{\left(\frac{B^2}{4b}\right)^m (a)_m}{\left(\frac{1}{2}\right)_m m!} \\
&= \Gamma(a) {}_1F_1\left(a; \frac{1}{2}; \frac{B^2}{4b}\right),
\end{aligned}$$

et également que

$$\begin{aligned}
\sum_{m=0}^{\infty} \frac{\left(\frac{B^2}{b}\right)^m \Gamma\left(a+m+\frac{1}{2}\right)}{(2m+1)!} &= \Gamma\left(a+\frac{1}{2}\right) \sum_{m=0}^{\infty} \frac{\left(\frac{B^2}{b}\right)^m \left(a+\frac{1}{2}\right)_m}{2^{2m} m! \left(\frac{3}{2}\right)_m} \\
&= \Gamma\left(a+\frac{1}{2}\right) {}_1F_1\left(a+\frac{1}{2}; \frac{3}{2}; \frac{B^2}{4b}\right).
\end{aligned}$$

□

Théorème 4.2. *Pour $X|\theta, \sigma^2 \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$, $S^2|\theta, \sigma^2 \sim \mathcal{G}\left(\frac{k}{2}, 2\sigma^2\right)$ indépendants et l'a priori $(\theta, \eta) \sim \mathcal{NG}(\mu, \lambda, \alpha, \beta)$, la densité a posteriori du quantile $\tau = \theta + c\sigma$, $c \neq 0$, est donnée par*

$$\begin{aligned}
f_{\tau}(t) &= \frac{e^{-\frac{c^2\lambda_0}{2}} \sqrt{\lambda_0}}{\Gamma(\alpha_0)\beta_0^{\alpha_0}\sqrt{2\pi}} \frac{1}{\left(A(t) + \frac{1}{\beta_0}\right)^{\alpha_0+\frac{1}{2}}} \\
&\quad \times \left(\Gamma\left(\alpha_0 + \frac{1}{2}\right) {}_1F_1\left(\alpha_0 + \frac{1}{2}; \frac{1}{2}; \frac{B(t)^2}{4A(t) + \frac{4}{\beta_0}}\right) \right. \\
&\quad \left. + \frac{\Gamma(\alpha_0 + 1) B(t)}{\sqrt{A(t) + \frac{1}{\beta_0}}} {}_1F_1\left(\alpha_0 + 1; \frac{3}{2}; \frac{B(t)^2}{4A(t) + \frac{4}{\beta_0}}\right) \right),
\end{aligned}$$

où $\alpha_0 = \frac{n}{2} + \alpha$, $\beta_0 = \beta + \frac{ns^2 + n\lambda(x-\mu)^2}{2}$, $\lambda_0 = \lambda + n$, $\mu_0 = \frac{xn + \mu\lambda}{\lambda + n}$, $A(t) = \frac{\lambda_0(t-\mu_0)^2}{2}$ et $B(t) = c\lambda_0(t - \mu_0)$.

On a également que l'espérance a posteriori et la variance a posteriori sont données par :

$$\begin{aligned}
\mathbb{E}(\theta + c\sigma|x, s^2) &= \mu_0 + c \frac{\Gamma\left(\alpha_0 - \frac{1}{2}\right)}{\Gamma(\alpha_0)\sqrt{\beta_0}}, \\
\mathbb{V}(\theta + c\sigma|x, s^2) &= \frac{1}{\lambda_0\beta_0} \frac{\Gamma(\alpha_0 - 1)}{\Gamma(\alpha_0)} + \frac{c^2}{\beta_0} \left(\frac{\Gamma(\alpha_0 - 1)}{\Gamma(\alpha_0)} - \left(\frac{\Gamma\left(\alpha_0 - \frac{1}{2}\right)}{\Gamma(\alpha_0)} \right)^2 \right).
\end{aligned}$$

Démonstration. On a du Théorème 4.1

$$f_{\tau}(t) = \frac{e^{-\frac{c^2\lambda_0}{2}} \sqrt{\lambda_0}}{\Gamma(\alpha_0)\beta_0^{\alpha_0}\sqrt{2\pi}} \int_{\mathbb{R}_+} \eta^{\alpha_0-\frac{1}{2}} e^{\eta\left(-A(t)-\frac{1}{\beta_0}\right)+\sqrt{\eta}B(t)} d\eta,$$

et le résultat s'ensuit à l'aide du Lemme 83. Pour trouver l'espérance ainsi que la variance a posteriori, on utilise le Corollaire 82, la variance de la loi marginale a posteriori de θ ainsi que la propriété suivante :

$$\mathbb{E}(W^r) = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \beta^r,$$

où $W \sim \mathcal{G}(\alpha, \beta)$ et $r > -\alpha$. En effet, on a que

$$\begin{aligned} \mathbb{E}(\theta + c\sigma|x, s^2) &= \mu_0 + c \mathbb{E}\left(\eta^{-\frac{1}{2}} \middle| x, s^2\right) \\ &= \mu_0 + c \frac{\Gamma(\alpha_0 - \frac{1}{2})}{\Gamma(\alpha_0)\sqrt{\beta_0}}. \end{aligned}$$

De plus, on a que

$$\begin{aligned} \mathbb{V}(\theta|x, s^2) &= \mathbb{E}^\eta[\mathbb{V}(\theta|x, s^2, \eta)] + \mathbb{V}^\eta[\mathbb{E}(\theta|x, s^2, \eta)] \\ &= \mathbb{E}\left(\frac{1}{\lambda_0\eta} \middle| x, s^2\right) \\ &= \frac{1}{\lambda_0} \frac{\Gamma(\alpha_0 - 1)}{\beta_0\Gamma(\alpha_0)}. \end{aligned}$$

Finalement, on a que

$$\begin{aligned} \mathbb{V}(\sigma|x, s^2) &= \mathbb{V}\left(\frac{1}{\eta^{\frac{1}{2}}} \middle| x, s^2\right) \\ &= \mathbb{E}(\eta^{-1}|x, s^2) - \mathbb{E}^2\left(\eta^{-\frac{1}{2}} \middle| x, s^2\right) \\ &= \frac{\Gamma(\alpha_0 - 1)}{\beta_0\Gamma(\alpha_0)} - \left(\frac{\Gamma(\alpha_0 - \frac{1}{2})}{\sqrt{\beta_0}\Gamma(\alpha_0)}\right)^2. \end{aligned}$$

□

Exemple 84. *On s'est intéressé à comprendre l'impact des divers choix d'hyperparamètres pour le modèle choisi. Pour tout ce qui suit, on a utilisé les valeurs suivantes*

$X = 1.3684407$ et $S = 1.1220697$. De plus, on a utilisé les paramètres $\alpha = \beta = 2$, $\mu = 0$, $\lambda = 0.2$, $c = 1.5$ et $n = 10$ pour comparer les densités a priori et a posteriori. Pour les graphiques de la Figure 4.2, on a plutôt utilisé $\alpha = \beta = 1$ comme paramètres de base. On a fait varier un seul paramètre à la fois afin de pouvoir bien visualiser l'impact de sa variation.

On a voulu voir la transition entre la densité a priori et celle a posteriori. On voit bien en regardant le graphique c) de la Figure 4.1 que la loi a priori est beaucoup plus diffuse que celle a posteriori. On voit également que la loi a posteriori a une masse centrée très près de sa moyenne. En effet, les espérances de la densité a priori et de la densité a posteriori sont respectivement 0.940 et 1.530 et les variances sont respectivement 2.741 et 0.003. Le même phénomène se répète dans le graphique b) de la Figure 4.2, où plus la taille d'échantillon augmente, moins la densité est dispersée. De plus, on a que les espérances a posteriori se rapprochent de plus en plus de X , ces dernières prenant les valeurs approximatives de 1.558, 1.448, et 1.409 pour les tailles d'échantillons 10, 25 et 50 respectivement.

De plus, lorsqu'on regarde le graphique d) de la Figure 4.1, on reconnaît une densité en cloche, à la manière d'une densité de Student. Comme ce graphique correspond en fait à la densité a posteriori de θ puisque $c = 0$, ceci concorde avec le Lemme 2.20.

En regardant le graphique a) de la Figure 4.2, on remarque que la valeur de c va avoir un impact sur l'asymétrie de la densité. Une valeur négative de c donne une densité asymétrique à gauche, alors qu'une valeur positive de c donne une densité avec une asymétrie à droite.

En jetant un coup d'oeil aux graphiques c) et d) de la Figure 4.2, on remarque que les

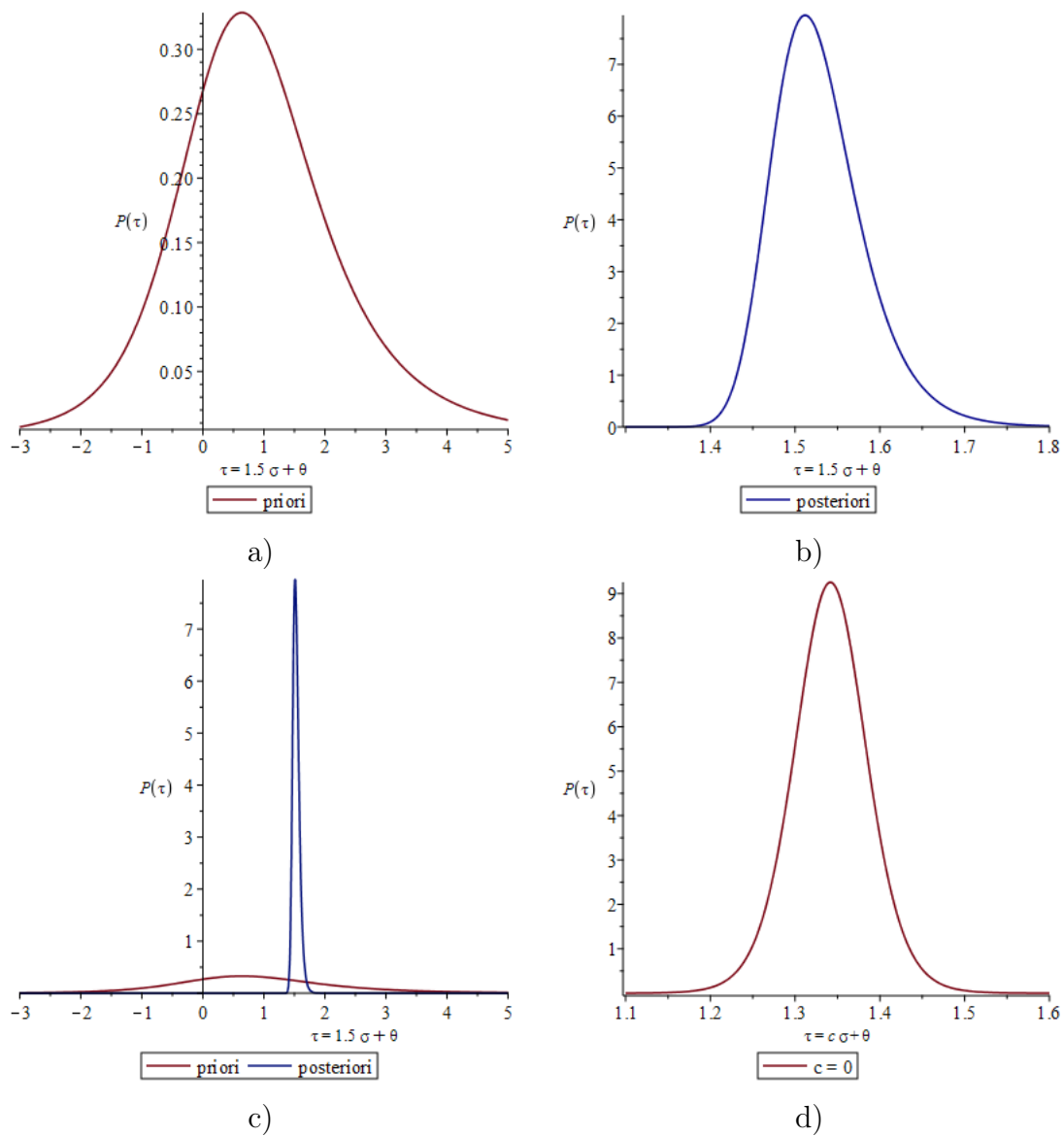


Figure 4.1 – Différents graphiques qui représentent : a) la densité a priori, b) la densité a posteriori, c) les deux densités, d) la densité a posteriori du quantile lorsque $c = 0$.

paramètres α et β semblent tous les deux avoir un impact similaire, rapprochant une bonne partie de la masse de la densité vers X , mais également en ayant un impact sur la dispersion de la densité. En fait, si on regarde bien les paramètres de la densité a posteriori, le fait de faire varier α ou β a seulement un impact sur la marginale de η a

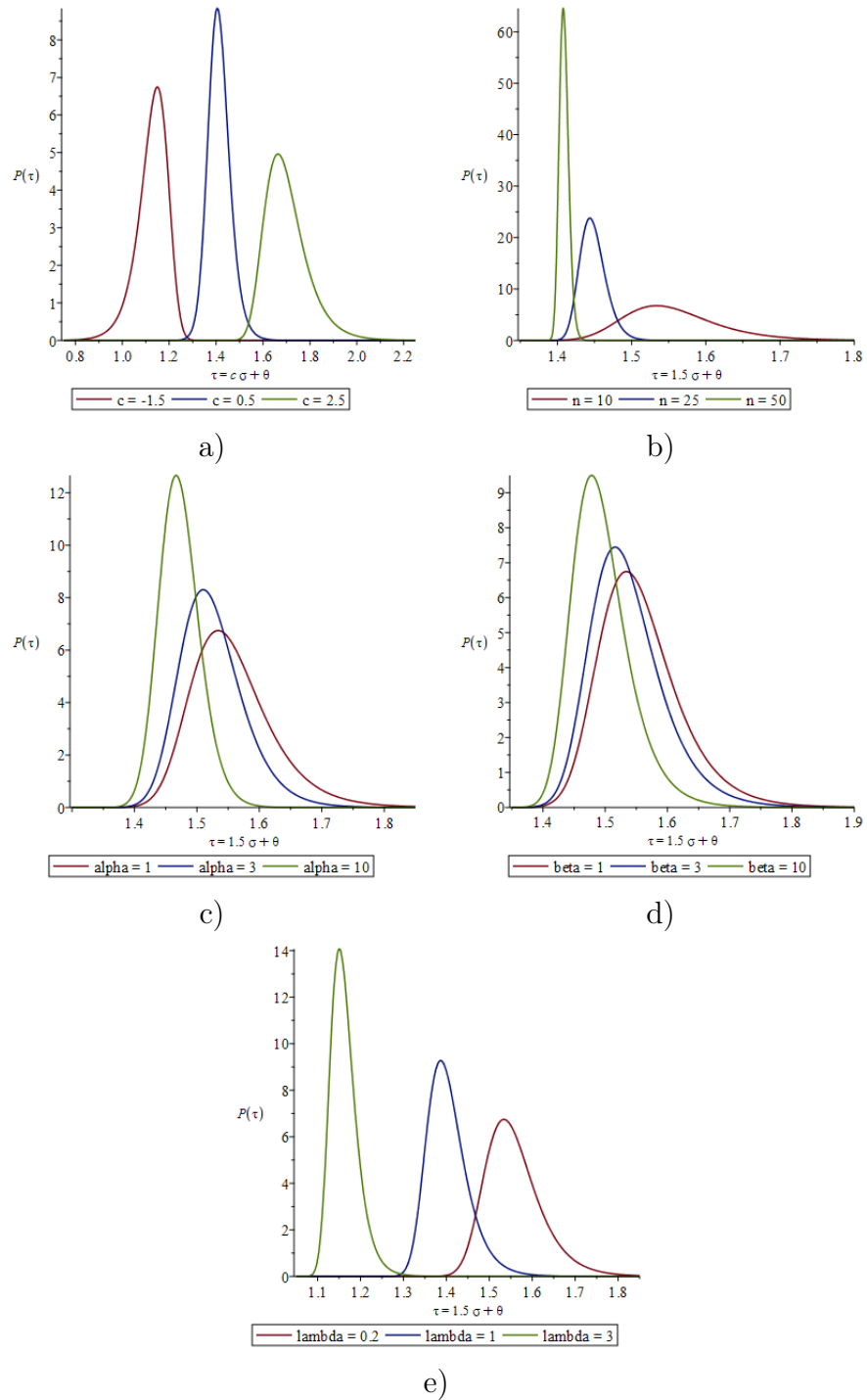


Figure 4.2 – Différents graphiques qui représentent la densité a posteriori lorsqu'on altère : a) le paramètre c , b) la taille d'échantillon n , les hyperparamètres c) α , d) β , e) λ .

posteriori. Comme $\eta \sim \mathcal{G}\left(\alpha + \frac{n}{2}, \beta + \frac{ns^2 + n\lambda(\bar{x} - \mu)^2}{2}\right)$, le fait d'augmenter α ou β met plus de masse sur de grandes valeurs de η , ce qui a un impact direct sur la variance du modèle complet en réduisant celle-ci.

Finalement, avec le graphique e) de la Figure 4.2, on voit que l'hyperparamètre λ semble lui aussi faire varier la position et la dispersion de la densité. On voit que plus λ grandit, plus la dispersion diminue. Ceci n'est pas étonnant, puisque lorsqu'on regarde la densité conditionnelle de $\theta|\eta$, la variance de cette densité a exactement le même comportement. Ce qui est plus étonnant, c'est de voir la position des densités. Les espérances des densités a posteriori du graphique e) de la Figure 2 sont 1.558, 1.404 et 1.133 pour les paramètres a priori λ prenant les valeurs 0.2, 1 et 3 respectivement. Ainsi, le fait d'augmenter λ ne veut pas dire que l'espérance a posteriori s'approche de X . Comme $\theta|\eta, \bar{x}, s \sim \mathcal{N}\left(\frac{\bar{x}n + \mu\lambda}{\lambda + n}, \frac{1}{(\lambda + n)\eta}\right)$, on voit que le fait d'augmenter λ a un impact sur la combinaison linéaire présente dans la moyenne de $\theta|\eta, \bar{x}, s$. Ceci donne plus de poids à la moyenne a priori, μ , ce qui explique ce que l'on observe.

4.1.1 Décomposition en pivot

Dans cette sous-section, on commencera par présenter la méthode du pivot, telle qu'expliquée dans [GMS16] (voir également [CB02]) pour déduire des intervalles de confiance dans un premier temps d'un point de vue fréquentiste, puis dans un deuxième temps d'un point de vue bayésien. Par la suite, on considérera un scénario particulier où les deux points de vue coïncident et, plus spécifiquement, où la crédibilité bayésienne correspond à la probabilité de recouvrement fréquentiste.

Définition 85. [CB02] Un pivot $T(X, \theta)$ est une fonction des données X et du paramètre $\theta \in \Theta \subseteq \mathbb{R}^p$ dont la loi ne dépend pas de θ .

On suppose ici que $X|\theta \sim f_\theta$, que $\tau : \mathbb{R}^p \rightarrow \mathbb{R}$ est la quantité d'intérêt et qu'il existe un pivot $T(X, \theta) = \frac{a_1(X) - \tau(\theta)}{a_2(X)}$, où $a_1(\cdot) \in \mathbb{R}$ $a_2(\cdot) > 0$. Beaucoup de pivots sont de cette forme. Par exemple, pour $X = (X_1, \dots, X_n)$ avec $X_i \sim \mathcal{N}(\mu, \sigma^2)$ indépendants et $\theta = (\mu, \sigma^2)$, la quantité $\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ est un pivot de la loi de Student avec $n - 1$ degrés de liberté. Ici, on peut poser $a_1(X) = \bar{X}$, $a_2(X) = \frac{S}{\sqrt{n}}$ et $\tau(\theta) = \mu$. On retrouve alors la forme $T(X, \theta) = \frac{a_1(X) - \tau(\theta)}{a_2(X)}$.

1. Cas fréquentiste.

Supposons que $T(X, \theta|\theta) \sim g$, g étant une densité intégrable au sens de Lebesgue, et qu'il existe des constantes c et d telles que $\int_c^d g(t)dt = 1 - \alpha$. On peut alors trouver un intervalle de confiance de niveau $1 - \alpha$. En effet, on a que

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(c \leq \frac{a_1(X) - \tau(\theta)}{a_2(X)} \leq d \middle| \theta\right) \\ &= \mathbb{P}\left(a_1(X) - da_2(X) \leq \tau(\theta) \leq a_1(X) - ca_2(X) \middle| \theta\right) \\ &= \mathbb{P}(\tau(\theta) \in I(X)), \forall \theta, \end{aligned}$$

où $I(X) = [a_1(X) - da_2(X), a_1(X) - ca_2(X)]$. On obtient alors un intervalle de confiance de probabilité de recouvrement fréquentiste $1 - \alpha$.

2. Cas bayésien.

Supposons que la densité a posteriori de $T(X, \theta)$ est donnée par h_x , où h_x est une densité intégrable au sens de Lebesgue, c'est-à-dire que $T(X, \theta)|X = x \sim h_x$. On suppose aussi qu'il existe des constantes s et t telles que $\int_s^t h_x(y)dy = 1 - \alpha$. Alors, on a que

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(s \leq \frac{a_1(X) - \tau(\theta)}{a_2(X)} \leq t \middle| X = x\right) \\ &= \mathbb{P}(a_1(X) - sa_2(X) \leq \tau(\theta) \leq a_1(X) - ta_2(X) | X = x) \\ &= \mathbb{P}(\tau(\theta) \in I(X) | X = x), \end{aligned}$$

où $I(X) = [a_1(X) - ta_2(X); a_1(X) - sa_2(X)]$. On obtient alors un intervalle de confiance bayésien avec crédibilité $1 - \alpha$.

Notons qu'en général, les fonctions g et h_x seront différentes. Néanmoins, certains contextes permettent de travailler dans un cadre où les deux fonctions sont égales et d'assurer que la crédibilité et la probabilité de recouvrement seront identiques.

Exemple 86. Voici un exemple familier avec $X = (X_1, \dots, X_n)$, où les $X_i \sim \mathcal{N}(\mu, \sigma^2)$ sont indépendants, $\theta = \mu$ et σ^2 est connue. Ainsi, on a $\bar{X} \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$ et le pivot est $T(X, \theta) = \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}}$.

Dans le cadre fréquentiste, $\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \Big| \theta \sim \mathcal{N}(0, 1)$. Ainsi, pour avoir un intervalle de confiance fréquentiste de probabilité de recouvrement $1 - \alpha$, il suffit de trouver des constantes c et d telles que

$$\begin{aligned} \mathbb{P}(c \leq Z \leq d) &= 1 - \alpha, \text{ où } Z \sim \mathcal{N}(0, 1) \\ \Leftrightarrow \mathbb{P}\left(c \leq \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \leq d \Big| \theta\right) &= 1 - \alpha, \forall \theta \end{aligned}$$

Un intervalle, qui se trouve à être celui de plus petite longueur, est celui qui sera symétrique autour de 0, c'est-à-dire, $c = \Phi^{-1}\left(\frac{\alpha}{2}\right) = z_{\frac{\alpha}{2}}$ et $d = -c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. On retrouve alors l'intervalle habituel, soit $\left[\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$.

De plus, pour l'a priori $\pi(\theta) = 1$, qui est une loi a priori impropre, on a que $\theta | X_1, \dots, X_n \sim \mathcal{N}\left(\bar{x}, \frac{\sigma^2}{n}\right)$ et donc, sous la densité $\mathcal{N}(0, 1)$, on a que $T(X, \theta | x) = \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \Big| x \sim \mathcal{N}(0, 1)$. Ainsi, pour trouver un intervalle bayésien $I(X)$, on peut prendre $s = \Phi^{-1}\left(\frac{\alpha}{2}\right)$ et $t = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = -s$, qui mène encore une fois à l'intervalle usuel. Il est important de remarquer les aspects suivants :

1. Dans cette situation, $\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \Big| \theta \sim \mathcal{N}(0, 1)$ correspond à la densité g et $T(X, \theta | x) =$

$\frac{\bar{x}-\theta}{\frac{\sigma}{\sqrt{n}}} \Big| \bar{x} \sim \mathcal{N}(0, 1)$ correspond à la densité h . On a donc que les deux densités sont identiques.

2. La crédibilité et la probabilité de recouvrement sont identiques

$$\mathbb{P}\left(\theta \in \left[\bar{x} \pm s \frac{\sigma}{\sqrt{n}}\right] \Big| x_1, \dots, x_n\right) = 1 - \alpha = \mathbb{P}\left(\theta \in \left[\bar{X} \pm c \frac{\sigma}{\sqrt{n}}\right] \Big| \theta\right),$$

pour tout $x = (x_1, \dots, x_n)$ et pour tout $\theta \in \mathbb{R}$.

Remarque 87. Dans l'exemple 86, on a mentionné que l'a priori $\pi(\theta) = 1$ est impropre, c'est-à-dire qu'elle n'est pas intégrable. En effet, on a que $\int_{\mathbb{R}} \pi(\theta) d\theta = +\infty$. Or, c'est vraiment la distribution a posteriori qui est d'intérêt lorsqu'on fait de l'inférence bayésienne. Ainsi, les choix d'a priori impropre sont valables tant et aussi longtemps que la loi a posteriori qui en découle existe. Par contre, il est primordial de se souvenir que l'a priori ne peut alors pas s'interpréter comme étant une densité de probabilité! Ce type d'a priori est également relié aux lois dites non-informatives. On suggère [Rob06] pour plus de détails.

Avec l'objectif d'estimer à nouveau un quantile $\tau = \theta + c\sigma$, $c \neq 0$, on considérera pour le reste de cette section le pivot $T = \sqrt{n} \left(\frac{\bar{X} - \theta - c\sigma}{S} \right)$. On trouvera la densité de ce pivot et on s'intéressera aussi à la densité a posteriori de T associée à la loi a priori non-informative $\pi(\theta, \sigma^2) = \frac{1}{\sigma^2} \mathbb{1}_{(\mathbb{R} \times \mathbb{R}_+)}(\theta, \sigma^2)$.

Corollaire 88. Soient X_1, \dots, X_n indépendantes de loi $\mathcal{N}(\theta, \sigma^2)$. Soient également $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ et $T = \sqrt{n} \left(\frac{\bar{X} - \theta - c\sigma}{S} \right)$. Alors, $T | \theta, \sigma \sim t_{n-1}(-\sqrt{nc})$.

Démonstration. La preuve découle directement de la Définition 13. En effet, il suffit de remarquer que T peut également s'écrire de la façon suivante,

$$T = \frac{\frac{\sqrt{n}(\bar{X} - \theta)}{\sigma} + \delta}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}}, \text{ où } \delta = -\sqrt{nc}.$$

Comme $\sqrt{n} \frac{\bar{X} - \theta}{\sigma} \sim \mathcal{N}(0, 1)$ et que $\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{G}\left(\frac{n-1}{2}, 2\right)$ sont indépendantes, on trouve que $T | \theta, \sigma^2 \sim t_{n-1}(\delta)$. □

Voici deux résultats où les lois fréquentiste et bayésienne coïncident.

Théorème 4.3. *Pour $\bar{X}|\theta, \sigma^2 \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$, $(n-1)S^2|\theta, \sigma^2 \sim \mathcal{G}\left(\frac{n-1}{2}, 2\sigma^2\right)$ avec l'a priori $\pi(\theta, \sigma^2) = \frac{1}{\sigma^2}$ et la quantité $V = \frac{\sigma}{\sqrt{(n-1)S^2}}$, les distributions de $V|\bar{x}, s^2$ et de $V|\theta, \sigma^2$ sont identiques pour tout \bar{x}, s^2, θ et σ^2 .*

Démonstration. Calculons tout d'abord la distribution de $V|\theta, \sigma^2$. On a que

$$\begin{aligned}\mathbb{P}(V \leq t|\theta, \sigma^2) &= \mathbb{P}\left(\frac{\sigma}{\sqrt{(n-1)S^2}} \leq t \mid \theta, \sigma^2\right) \\ &= \mathbb{P}\left(\frac{\sigma^2}{(n-1)S^2} \leq t^2 \mid \theta, \sigma^2\right)\end{aligned}$$

qui correspond en fait à la répartition d'une loi $\mathcal{IG}\left(\frac{n-1}{2}, \frac{1}{2\sigma^2}\right)$ évaluée au point t^2 . On trouve donc que la densité est donnée par

$$f_{\frac{1}{(n-1)S^2}}(t^2)(2t) = \frac{e^{-\frac{1}{2t^2}}}{2^{\frac{n-1}{2}-1} t^n \Gamma\left(\frac{n-1}{2}\right)}, t > 0.$$

Calculons maintenant la distribution de $V|\bar{x}, s^2$. On a que

$$\begin{aligned}\mathbb{P}(V \leq t|\bar{x}, s^2) &= \mathbb{P}\left(\frac{\sigma}{\sqrt{(n-1)S^2}} \leq t \mid \bar{x}, s^2\right) \\ &= \mathbb{P}\left(\sigma^2 \leq t^2(n-1)S^2 \mid \bar{x}, s^2\right).\end{aligned}$$

On trouve donc que la densité a posteriori de V est donnée par

$$\begin{aligned}g_{\sigma^2}(t^2(n-1)s^2)2t(n-1)s^2 &= \frac{\left(\frac{(n-1)s^2}{2}\right)^{\frac{n-1}{2}} e^{-\frac{(n-1)s^2}{2t^2(n-1)s^2}}}{\Gamma\left(\frac{n-1}{2}\right)(t^2(n-1)s^2)^{\frac{n-1}{2}+1}} 2t(n-1)s^2 \\ &= \frac{e^{-\frac{1}{2t^2}}}{2^{\frac{n-1}{2}-1} t^n \Gamma\left(\frac{n-1}{2}\right)},\end{aligned}$$

ce qui montre le résultat. □

Théorème 4.4. Pour la loi a priori $\pi(\sigma^2) = \frac{1}{\sigma^2}$, on a que $\forall \bar{x}, s^2, \theta, \sigma^2$,

$$\frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \Big| V, \bar{x}, s^2 \equiv \frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \Big| V, \theta, \sigma^2.$$

De plus, on a que

$$\sqrt{n} \left(\frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \right) \Big| \bar{x}, s^2 \equiv \sqrt{n} \left(\frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \right) \Big| \theta, \sigma^2 \sim t_{n-1}(-\sqrt{nc}).$$

Démonstration. On a que

$$\begin{aligned} \theta + c\sigma | \sigma^2, \bar{x}, s^2 &\sim \mathcal{N}\left(\bar{x} + c\sigma, \frac{\sigma^2}{n}\right) \\ \Rightarrow \frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \Big| \sigma^2, \bar{x}, s^2 &\sim \mathcal{N}\left(\frac{-c\sigma}{\sqrt{(n-1)s^2}}, \frac{\sigma^2}{n(n-1)s^2}\right) \\ \Rightarrow \frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \Big| V, \bar{x}, s^2 &\sim \mathcal{N}\left(-c\sigma, \frac{\sigma^2}{n}\right). \end{aligned}$$

On a aussi que

$$\begin{aligned} \frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \Big| \theta, \sigma, s^2 &\sim \mathcal{N}\left(\frac{-c\sigma}{\sqrt{(n-1)s^2}}, \frac{\sigma^2}{n(n-1)s^2}\right) \\ \Rightarrow \frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \Big| V, \theta, \sigma^2 &\sim \mathcal{N}\left(-c\sigma, \frac{\sigma^2}{n}\right). \end{aligned}$$

On peut alors trouver que les densités de $\sqrt{n} \left(\frac{\bar{X} - \theta - c\sigma}{\sqrt{(n-1)S^2}} \right)$ en conditionnant sur (\bar{x}, s^2) et en conditionnant sur (θ, σ^2) sont identiques, puisque $V|\bar{x}, s^2$ et $V|\theta, \sigma^2$ sont de même loi.

On trouve alors la distribution en appliquant le Corollaire 88. \square

4.2 Inférence bayésienne pour un coefficient de variation et un coefficient de discrimination

4.2.1 Le coefficient de variation et son inverse

Dans cette section, nous allons considérer les quotients de la forme $CV = \frac{\sigma}{\theta}$ et $ICV = \frac{\theta}{\sigma}$ où $\pi(\theta|\sigma^2) \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{\lambda}\right)$, $\frac{1}{\sigma^2} \sim g$ et X_1, \dots, X_n sont i.i.d. selon une loi $\mathcal{N}(\theta, \sigma^2)$. On s'intéresse aux lois a posteriori de ces quotients. Ensuite, le cas où $g \sim \mathcal{G}(\alpha, \beta)$ sera détaillé. Cette question a déjà été abordée précédemment, notamment dans [HR36].

Théorème 4.5. *Pour $X|\theta, \sigma^2 \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$ et $S^2|\theta, \sigma^2 \sim \mathcal{G}\left(\frac{k}{2}, 2\sigma^2\right)$ et pour l'a priori $\theta|\eta \sim \mathcal{N}\left(\mu, \frac{1}{\lambda\eta}\right)$ et $\eta = \frac{1}{\sigma^2} \sim g$, où g est une densité intégrable au sens de Lebesgue, la densité et la répartition a posteriori du coefficient de variation $CV = \frac{\sigma}{\theta}$, ainsi que de son inverse $ICV = \frac{\theta}{\sigma}$ sont respectivement données par :*

$$F_{ICV}(t) = \int_{\mathbb{R}_+} \Phi\left(\sqrt{\lambda'\eta}\left(\frac{t}{\sqrt{\eta}} - \mu'\right)\right) g(\eta|x, s) d\eta ; \quad (4.3)$$

$$F_{CV}(t) = 1 - F_{ICV}\left(\frac{1}{t}\right) ; \quad (4.4)$$

$$f_{ICV}(t) = \sqrt{\frac{\lambda'}{2\pi}} e^{-\frac{t^2\lambda'}{2}} \int_{\mathbb{R}_+} e^{-A(t)\eta+B(t)\sqrt{\eta}} g(\eta|x, s) d\eta ; \quad (4.5)$$

$$f_{CV}(t) = \sqrt{\frac{\lambda'}{2\pi}} \frac{e^{-\frac{\lambda'}{2t^2}}}{t^2} \int_{\mathbb{R}_+} e^{-A(t)\eta+C(t)\sqrt{\eta}} g(\eta|x, s) d\eta ; \quad (4.6)$$

où $t \in \mathbb{R}_+$, $\mu' = \frac{nx+\lambda\mu}{\lambda+n}$, $\lambda' = \lambda + n$, $A(t) = \lambda'\mu'^2$, $B(t) = \lambda'\mu't$ et $C(t) = \frac{\lambda'\mu'}{t}$.

Démonstration. On a

$$\begin{aligned}
\mathbb{P}(ICV \leq t|X, S^2) &= \int_{\mathbb{R}_+} \mathbb{P}(\theta \leq t\eta^{\frac{1}{2}}|x, s)g(\eta|x, s)d\eta \\
&= \int_{\mathbb{R}_+} \Phi\left(\sqrt{\lambda'\eta}\left(\frac{t}{\sqrt{\eta}} - \mu'\right)\right)g(\eta|x, s)d\eta,
\end{aligned}$$

ce qui nous donne (4.3). En dérivant par rapport à t , on obtient

$$\begin{aligned}
f_{ICV}(t) &= \int_{\mathbb{R}_+} \phi\left(\sqrt{\lambda'\eta}\left(\frac{t}{\sqrt{\eta}} - \mu'\right)\right)\sqrt{\lambda'}g(\eta|x, s)d\eta \\
&= \sqrt{\frac{\lambda'}{2\pi}}e^{-\frac{t^2\lambda'}{2}} \int_{\mathbb{R}_+} e^{\lambda'\mu't\sqrt{\eta}-\lambda'\mu'^2\eta}g(\eta|x, s)d\eta \\
&= \sqrt{\frac{\lambda'}{2\pi}}e^{-\frac{t^2\lambda'}{2}} \int_{\mathbb{R}_+} e^{-A(t)\eta+B(t)\sqrt{\eta}}g(\eta|x, s)d\eta.
\end{aligned}$$

Pour ce qui est du coefficient de variation, on déduit les expressions pour la densité et la répartition a posteriori du fait que $\mathbb{P}\left(\frac{\sigma}{\theta} \leq t|x, s\right) = 1 - \mathbb{P}\left(\frac{\theta}{\sigma} \leq \frac{1}{t}|x, s\right) = 1 - F_{ICV}\left(\frac{1}{t}\right)$.

□

Corollaire 89. *Sous les conditions du Théorème 4.5, l'espérance et la variance a posteriori de l'inverse du coefficient de variation sont données par les expressions suivantes :*

$$\begin{aligned}
\mathbb{E}(ICV|x, s) &= \mu' \int_{\mathbb{R}_+} \eta^{\frac{1}{2}}g(\eta|x, s)d\eta, \\
\mathbb{V}(ICV|x, s) &= \mu'^2 \mathbb{E}(\eta|x, s) + \frac{1}{\lambda'} - \mu'^2 \left(\int_{\mathbb{R}_+} \eta^{\frac{1}{2}}g(\eta|x, s)d\eta \right)^2
\end{aligned}$$

où $\mu' = \frac{nx+\lambda\mu}{\lambda+n}$ et $\lambda' = \lambda + n$, et ce lorsque le côté droit de chacune des expressions existe. De plus, l'espérance et la variance a posteriori du coefficient de variation $\frac{\sigma}{\theta}$ n'existent pas.

Démonstration. Dans un premier temps, on a que

$$\begin{aligned}
\mathbb{E}(ICV|x, s) &= \mathbb{E} \mathbb{E}(\theta \sqrt{\eta} | \eta, x, s), \\
&= \mathbb{E}(\mu' \sqrt{\eta} | x, s) \\
&= \mu' \int_{\mathbb{R}_+} \eta^{\frac{1}{2}} g(\eta | x, s) d\eta.
\end{aligned}$$

De plus, on a que

$$\begin{aligned}
\mathbb{E}(\theta^2 \eta | x, s) &= \mathbb{E} \left(\eta \mathbb{E}(\theta^2 | \eta, x, s) | x, s \right) \\
&= \mathbb{E} \left(\eta \left(\mu'^2 + \frac{1}{\lambda' \eta} \right) | x, s \right) \\
&= \mu'^2 \mathbb{E}(\eta | x, s) + \frac{1}{\lambda'},
\end{aligned}$$

ce qui nous permet de déduire la variance a posteriori de ICV . Finalement, on a que

$$\mathbb{E}(CV|x, s) = \mathbb{E} \left(\frac{1}{\sqrt{\eta}} \mathbb{E} \left(\frac{1}{\theta} \middle| \eta, x, s \right) \middle| x, s \right).$$

Or, comme le premier moment inverse d'une loi normale n'existe pas, la quantité ci-haut n'existe pas non plus et par conséquent l'espérance et la variance de CV a posteriori n'existent pas. \square

Après avoir appliqué une démarche analogue à celle présentée à la Section 4.1, nous avons cru bon de donner la densité a posteriori de l'inverse du coefficient de variation de la même façon que la densité a posteriori du quantile au Théorème 4.2, qui est donnée dans le prochain théorème.

Théorème 4.6. *Pour $X|\theta, \sigma^2 \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$ et $nS^2|\theta, \sigma^2 \sim \mathcal{G}\left(\frac{n-1}{2}, 2\sigma^2\right)$ et pour l'a priori $(\theta, \eta) \sim \mathcal{NG}(\mu, \lambda, \alpha, \beta)$, la densité a posteriori de l'inverse du coefficient de variation $ICV = \frac{\theta}{\sigma}$ est donnée par*

$$\begin{aligned}
f_{ICV}(t) &= \sqrt{\frac{\lambda_0}{2\pi}} \frac{e^{-\frac{t^2\lambda_0}{2}}}{\Gamma(\alpha_0)\beta_0^{\alpha_0}} \frac{1}{\left(A(t) + \frac{1}{\beta_0}\right)^{\alpha_0}} \\
&\times \left(\Gamma(\alpha_0) {}_1F_1\left(\alpha_0; \frac{1}{2}; \frac{B(t)^2}{4\left(A(t) + \frac{1}{\beta_0}\right)}\right) \right. \\
&\left. + \frac{\Gamma\left(\alpha_0 + \frac{1}{2}\right) B(t)}{\sqrt{A(t) + \frac{1}{\beta_0}}} {}_1F_1\left(\alpha_0 + \frac{1}{2}; \frac{3}{2}; \frac{B(t)^2}{4\left(A(t) + \frac{1}{\beta_0}\right)}\right) \right),
\end{aligned}$$

où $\mu_0 = \frac{nx + \lambda\mu}{\lambda + n}$, $\lambda_0 = \lambda + n$, $\alpha_0 = \alpha + \frac{n}{2}$, $\beta_0 = \beta + \frac{ns^2 + n\lambda(x - \mu)^2}{2}$, $A(t) = \lambda_0\mu_0^2$ et $B(t) = \lambda_0\mu_0 t$.

On a également que l'espérance a posteriori et la variance a posteriori sont données par :

$$\begin{aligned}
\mathbb{E}(ICV|X, S) &= \mu_0 \frac{\Gamma\left(\alpha_0 + \frac{1}{2}\right)}{\Gamma(\alpha_0)} \beta_0^{\frac{1}{2}} \\
\mathbb{V}(ICV|X, S) &= \mu_0^2 \beta_0 \left(\alpha_0 - \left(\frac{\Gamma\left(\alpha_0 + \frac{1}{2}\right)}{\Gamma(\alpha_0)} \right)^2 \right) + \frac{1}{\lambda_0}.
\end{aligned}$$

Démonstration. On obtient la densité de ICV en appliquant directement le Lemme 83 ainsi que le Théorème 4.5. L'espérance et la variance a posteriori sont, quant à eux, obtenues en utilisant le Corollaire 89 avec la propriété des moments d'une loi gamma. \square

Avant de passer au prochain exemple, on va regarder les fonctions $f(x) = \frac{\Gamma(x + \frac{1}{2})}{\Gamma(x)}$ et $h(x) = x - f^2(x)$, que l'on retrouve dans l'expression de l'espérance et la variance de l'ICV.

Afin de pouvoir bien interpréter la fonction $f(x)$, on introduit la *formule de Stirling*.

Définition 90. *La formule de Stirling est une approximation de la fonction factorielle ou encore de la fonction gamma. Elle dit que $\Gamma(n + 1) \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$, où la notation $a(y) \sim b(y)$ signifie ici que $\lim_{y \rightarrow \infty} \frac{a(y)}{b(y)} = 1$.*

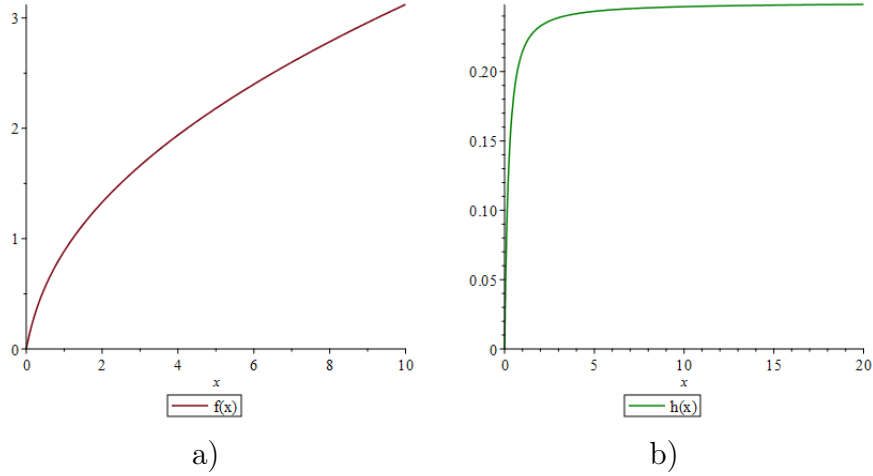


Figure 4.3 – Graphique des fonctions $f(x)$ et $h(x)$.

Dans la Figure 4.3 on voit que $f(x)$ croît de façon similaire à la fonction \sqrt{x} . En effet, par la formule de Stirling, on obtient

$$\lim_{x \rightarrow \infty} \left(1 + \frac{1}{2x}\right)^{x+1} \sqrt{\frac{x + \frac{1}{2}}{e}} = \lim_{x \rightarrow \infty} \sqrt{x + \frac{1}{2}}.$$

Pour ce qui est de la fonction $h(x)$, elle semble croître très lentement à partir de 5.

Exemple 91. *On s'est intéressé à comprendre l'impact des divers choix d'hyperparamètres pour le modèle choisi. Pour tout ce qui suit, on a utilisé les valeurs suivantes : $\bar{X} = 0.446926$ et $S = 0.631189$. De plus, on a utilisé les paramètres $\alpha = \beta = 1$, $\mu = 0$, $\lambda = 0.2$ et $n = 10$ pour l'ensemble des graphiques qui suivent. On a fait varier un seul paramètre à la fois afin de pouvoir bien visualiser l'impact de sa variation.*

En regardant la Figure 4.4, on voit que la densité a priori est encore une fois beaucoup plus dispersée que celle a posteriori. Ici, on a que les espérances de la densité a priori et de la densité a posteriori sont respectivement données par 0 et 1.878 alors que les variances sont respectivement 5 et 0.248.

Dans le graphique a) de la Figure 4.5, on remarque que plus la taille d'échantillon est grande, plus la dispersion de la densité a posteriori augmente. En effet, lorsque la taille

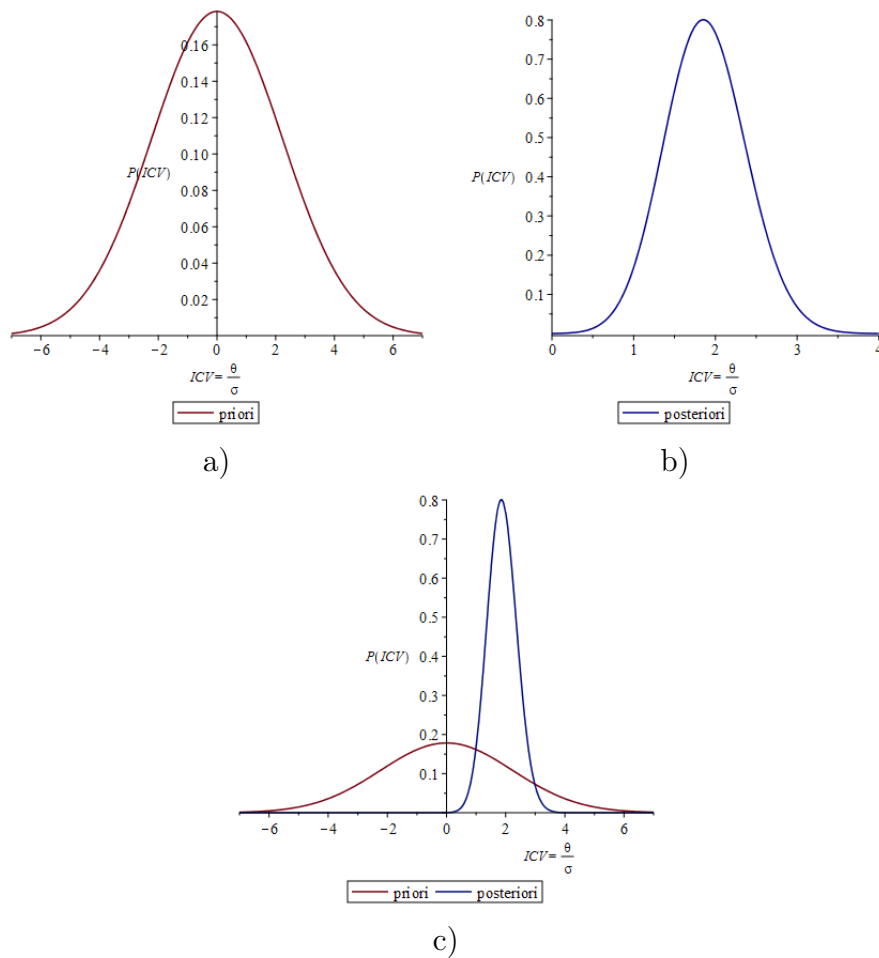


Figure 4.4 – Différents graphiques qui représentent : a) la densité a priori, b) la densité a posteriori,

d'échantillon augmente, les paramètres a posteriori $\alpha_0 = \alpha + \frac{n}{2}$, $\beta_0 = \beta + \frac{ns^2 + n\lambda(\bar{x} - \mu)^2}{2}$ et $\lambda_0 = \lambda + n$ augmentent eux aussi. Or, comme on l'a vu précédemment, lorsque α_0 et β_0 augmentent, la probabilité d'observer de grandes valeurs de η , et par conséquent de petite valeur de σ^2 , augmente-elle aussi. Comme on devra diviser par une petite valeur de σ , ceci explique pourquoi la dispersion semble augmenter avec la taille d'échantillon.

Dans le graphique a) de la Figure 4.5, on remarque que l'hyperparamètre λ ne semble

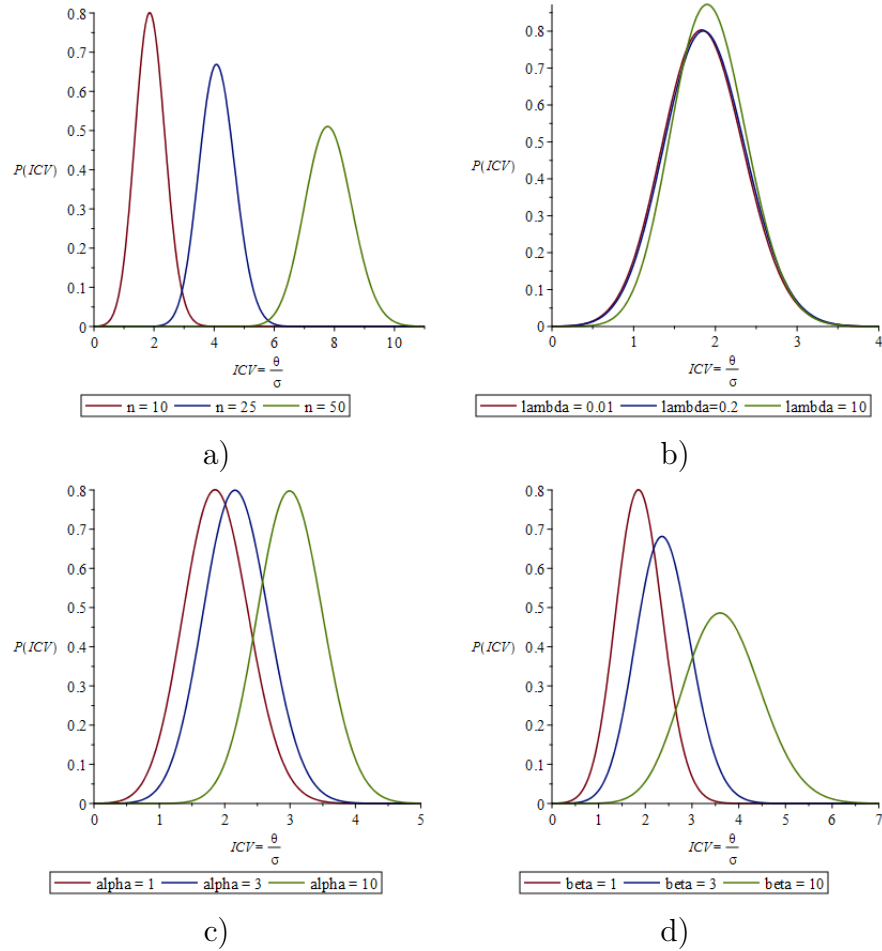


Figure 4.5 – Différents graphiques qui représentent l’impact sur la densité a posteriori lorsqu’on altère : a) la taille d’échantillon n , les hyperparamètres b) λ , c) α , d) β .

pas avoir beaucoup d’impact sur la densité a posteriori, à moins qu’on lui attribue de grandes valeurs. Dans ce cas, il vient altérer la moyenne posteriori de $\theta|\eta$. Sinon, l’impact de l’augmentation du paramètre λ se fait sentir légèrement au niveau de la variance a posteriori, comme on le voit dans l’expression de la variance a posteriori de l’ICV.

Dans le graphique c) de la Figure 4.5, on remarque que lorsque α augmente, la densité semble changer de position. On trouve que les espérances a posteriori sont 1.878, 2.180,

et 3.007, que les variances a posteriori sont 0.248, 0.249 et 0.250 pour des valeurs de α de 1, 3 et 10 respectivement. Ces valeurs semblent bien concorder avec la petite analyse que l'on a fournie avant cet exemple sur les fonctions d'un rapport de gamma et les expressions pour l'espérance et la variance a posteriori de l'ICV.

Dans le graphique d) de la Figure 4.5, on voit bien que l'augmentation du paramètre β altère la position et la dispersion de la densité a posteriori. Les expressions de l'espérance et de la variance viennent appuyer cette observation.

4.2.2 Le coefficient de discrimination

Dans cette section, nous allons considérer une extension de l'inverse du coefficient de variation, soit le coefficient de discrimination $CD = \frac{\|\theta\|^2}{\sigma^2}$, où $\pi(\theta, \sigma^2) \sim \mathcal{N}_d\left(\mu, \frac{\sigma^2}{\lambda}\right)$ et $\frac{1}{\sigma^2} \sim g$. On utilise la version canonique du modèle linéaire telle que présentée au chapitre 1. L'auteur de [L.91] parle du coefficient de discrimination et de l'admissibilité de certains estimateurs. Remarquons également que le coefficient de discrimination peut apparaître lorsqu'on considère deux populations multivariées de même variance mais ayant potentiellement des moyennes différentes. Dans ce cas, on a une première population $\mathcal{N}_p(\mu_1, \sigma^2 \mathbf{1}_p)$ et une deuxième population $\mathcal{N}_p(\mu_2, \sigma^2 \mathbf{1}_p)$. Un problème d'intérêt sera alors de regarder la différence des moyennes entre ces deux populations, soit $\theta = \mu_1 - \mu_2$. Enfin, il suffit d'utiliser la distance de Mahalanobis entre ces deux populations pour faire ressortir le coefficient de discrimination qui correspond alors à une distance euclidienne normalisée par σ^2 .

Définition 92. Soient un vecteur d'observation X , son vecteur de moyenne μ et la matrice de covariance qui lui est associée Σ . Alors, on définit la distance de Mahalanobis

comme suit :

$$\text{Dist}_M(X) = \sqrt{(X - \mu)^\top \Sigma^{-1} (X - \mu)}.$$

Si l'on considère plutôt un autre vecteur d'observation Y de même dimension que X , on peut définir la distance de Mahalanobis entre ces deux vecteurs comme suit :

$$\text{Dist}_M(X, Y) = \sqrt{(X - Y)^\top \Sigma^{-1} (X - Y)}.$$

Théorème 4.7. Pour $X|\theta, \sigma^2 \sim \mathcal{N}_d\left(\theta, \frac{\sigma^2}{n}\right)$ et $S^2|\theta, \sigma^2 \sim \mathcal{G}\left(\frac{k}{2}, 2\sigma^2\right)$ conditionnellement indépendantes et pour l'a priori $\theta|\eta \sim \mathcal{N}_d\left(\mu, \frac{1}{\lambda\eta}\mathbb{1}_d\right)$ et $\eta \sim g$, les fonctions de densité et de répartition de $CD = \frac{\|\theta\|^2}{\sigma^2}$ a posteriori sont respectivement données par :

$$F_{CD}(t) = \int_{\mathbb{R}_+} F_{Y|\eta}(\lambda't) g(\eta|x, s) d\eta, \quad (4.7)$$

$$f_{CD}(t) = \lambda' \sum_{k=0}^{\infty} \frac{\left(\frac{d}{2}\right)^k (\lambda't)^{k+\frac{d}{2}-1} e^{-\frac{\lambda't}{2}}}{k! \Gamma\left(k + \frac{d}{2}\right) 2^{k+\frac{d}{2}}} \int_{\mathbb{R}_+} e^{-\frac{\delta\eta}{2}} \eta^k g(\eta|x, s) d\eta, \quad (4.8)$$

où $\lambda' = \lambda + n$, $\mu' = \frac{nx + \lambda\mu}{\lambda + n}$, $\delta = \lambda'd\mu'$ et $Y|\eta \sim \chi_d^2(\delta\eta)$.

Démonstration. Comme on l'a vu au chapitre 2 à la Définition 57, on a que

$$\begin{aligned} \theta_i|\eta, x, s \sim \mathcal{N}\left(\mu', \frac{1}{\lambda'\eta}\right) &\Rightarrow \sqrt{\lambda'\eta} \theta_i|\eta, x, s \sim \mathcal{N}\left(\sqrt{\lambda'\eta}\mu', 1\right) \\ &\Rightarrow Y = \lambda'\eta \|\theta\|^2|\eta, x, s \sim \chi_d^2(\lambda'd\mu'\eta). \end{aligned} \quad (\star)$$

Ainsi,

$$\begin{aligned}
F_{CD}(t) &= \mathbb{P}(\|\theta\|^2 \eta \leq t | x, s) \\
&= \int_{\mathbb{R}_+} \mathbb{P}(\lambda' \eta \|\theta\|^2 \leq \lambda' t | \eta, x, s) g(\eta | x, s) d\eta \\
&= \int_{\mathbb{R}_+} F_{Y|\eta}(\lambda' t) g(\eta | x, s) d\eta.
\end{aligned}$$

On trouve alors la densité en dérivant par rapport à t , soit

$$f_{CD}(t) = \int_{\mathbb{R}_+} \lambda' f_{Y|\eta}(\lambda' t) g(\eta | x, s) d\eta = \lambda' \sum_{k=0}^{\infty} \frac{\left(\frac{d}{2}\right)^k (\lambda' t)^{k+\frac{d}{2}-1} e^{-\frac{\lambda' t}{2}}}{k! \Gamma\left(k + \frac{d}{2}\right) 2^{k+\frac{d}{2}}} \int_{\mathbb{R}_+} e^{-\frac{\delta \eta}{2}} \eta^k g(\eta | x, s) d\eta.$$

□

Corollaire 93. *Sous le modèle du Théorème 4.7, l'espérance ainsi que la variance a posteriori de $CD = \frac{\|\theta\|^2}{\sigma^2}$ sont données par les expressions suivantes :*

$$\begin{aligned}
\mathbb{E}(\|\theta\|^2 \eta | x, s) &= \frac{d + \delta \mathbb{E}(\eta | x, s)}{\lambda'}, \\
\mathbb{V}(\|\theta\|^2 \eta | x, s) &= \frac{2d + 4\delta \mathbb{E}(\eta | x, s) + \delta^2 \mathbb{V}(\eta | x, s)}{\lambda'^2},
\end{aligned}$$

où $\lambda' = \lambda + n$, $\mu' = \frac{nx + \lambda \mu}{\lambda + n}$ et $\delta = \lambda' d \mu'$.

Démonstration. Posons $Y = \lambda' \eta \|\theta\|^2$ et rappelons que $Y | \eta, x, s \sim \chi_d^2(\lambda' d \mu' \eta)$. Alors, on a que

$$\begin{aligned}
\mathbb{E}(\|\theta\|^2 \eta | x, s) &= \frac{1}{\lambda'} \mathbb{E} \mathbb{E}(Y | \eta, x, s) \\
&= \frac{d + \delta \mathbb{E}(\eta | x, s)}{\lambda'}.
\end{aligned}$$

De plus,

$$\begin{aligned}
\mathbb{V}(\|\theta\|^2 \eta | x, s) &= \frac{1}{\lambda'^2} \mathbb{V}(\lambda' \eta \|\theta\|^2 | x, s) \\
&= \frac{1}{\lambda'^2} (\mathbb{E} \mathbb{V}(Y | \eta) + \mathbb{V} \mathbb{E}(Y | \eta)) \\
&= \frac{1}{\lambda'^2} (\mathbb{E}(2d + 4\delta \eta | x, s) + \mathbb{V}(d + \delta \eta | x, s)) \\
&= \frac{1}{\lambda'^2} (2d + 4\delta \mathbb{E}(\eta | x, s) + \delta^2 \mathbb{V}(\eta | x, s)).
\end{aligned}$$

□

On termine avec une jolie représentation de la loi a posteriori du coefficient de discrimination $\frac{\|\theta\|^2}{\sigma^2}$. Avant, on se doit d'introduire une distribution qui n'est pas encore apparue dans ce mémoire.

Définition 94. *On dit qu'une variable aléatoire X suit une loi Binomiale Négative de paramètres $r > 0$ et $p \in (0, 1)$, si sa densité par rapport à la mesure de comptage est donnée par*

$$f_X(x) = \frac{\Gamma(r+x)}{x! \Gamma(r)} p^r (1-p)^x \mathbb{1}_{(\mathbb{N})}(x).$$

On notera alors $X \sim \mathcal{BN}(r, p)$.

Théorème 4.8. *Pour $X | \theta, \sigma^2 \sim \mathcal{N}_d\left(\theta, \frac{\sigma^2}{n}\right)$, $nS^2 | \theta, \sigma^2 \sim \mathcal{G}\left(\frac{k}{2}, 2\sigma^2\right)$ indépendantes et en utilisant l'a priori $(\theta, \eta) \sim \mathcal{N}_d \mathcal{G}(\mu, \lambda, \alpha, \beta)$, la densité a posteriori de $CD = \|\theta\|^2 \eta$ admet la représentation suivante*

$$\begin{aligned}
CD | K = k &\sim \mathcal{G}\left(k + \frac{d}{2}, \frac{2}{\lambda_0}\right) \\
K &\sim \mathcal{BN}\left(\alpha_0, \frac{2}{\delta \beta_0 + 2}\right),
\end{aligned}$$

où $\mu_0 = \frac{xn+\mu\lambda}{\lambda+n}$, $\lambda_0 = \lambda + n$, $\alpha_0 = \alpha + \frac{n-1+d}{2}$, $\beta_0 = \beta + \frac{ns^2+n\lambda(x-\mu)^\top(x-\mu)}{2}$ et $\delta = \lambda_0 d \mu_0$. On a également que l'espérance a posteriori et la variance a posteriori sont données par :

$$\begin{aligned}\mathbb{E}(\|\theta\|^2 \eta | x, s) &= \frac{d + \delta \alpha_0 \beta_0}{\lambda_0}, \\ \mathbb{V}(\|\theta\|^2 \eta | x, s) &= \frac{2d + 4\delta \alpha_0 \beta_0 + \delta^2 \alpha_0 \beta_0^2}{\lambda_0^2}.\end{aligned}$$

Démonstration. On obtient la densité de CD en substituant la densité $(\theta, \eta | x, s) \sim \mathcal{N}_d \mathcal{G}(\mu_0, \lambda_0, \alpha_0, \beta_0)$ dans l'équation (4.8). En effet, on a que

$$\int_{\mathbb{R}_+} e^{-\frac{\delta n}{2} \eta^k} \frac{\eta^{\alpha_0-1} e^{-\frac{\eta}{\beta_0}}}{\Gamma(\alpha_0) \beta_0^{\alpha_0}} d\eta = \frac{\Gamma(\alpha_0 + k) \left(\frac{2\beta_0}{\delta\beta_0+2}\right)^{\alpha_0+k}}{\Gamma(\alpha_0) \beta_0^{\alpha_0}}.$$

On trouve alors la densité, soit

$$\begin{aligned}f_{CD}(t) &= \frac{\lambda_0}{\beta_0^{\alpha_0} \Gamma(\alpha_0)} \left(\frac{2\beta_0}{\delta\beta_0+2}\right)^{\alpha_0} \sum_{k=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^k \left(\frac{2\beta_0}{\delta\beta_0+2}\right)^k (\lambda_0 t)^{k+\frac{d}{2}-1} e^{-\frac{\lambda_0 t}{2}} \Gamma(k + \alpha_0)}{k! 2^{k+\frac{d}{2}} \Gamma\left(k + \frac{d}{2}\right)} \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k + \alpha_0)}{\Gamma(\alpha_0) k!} \left(\frac{2}{\delta\beta_0+2}\right)^{\alpha_0} \left(\frac{\delta\beta_0}{\delta\beta_0+2}\right)^k \times \frac{t^{k+\frac{d}{2}} e^{-\frac{t}{\lambda_0}}}{\left(\frac{2}{\lambda_0}\right)^{k+\frac{d}{2}} \Gamma\left(k + \frac{d}{2}\right)}.\end{aligned}$$

Pour ce qui est de l'espérance et de la variance, on les obtient en substituant $\mathbb{E}(\eta | x, s)$ et $\mathbb{V}(\eta | x, s)$ au Corollaire 93. \square

CONCLUSION

En résumé, on a donné des formes analytiques des densités a posteriori et des densités prédictives a posteriori sous divers modèles et sous diverses lois a priori, en mettant l'emphase sur l'a priori normal-gamma et sur les différentes fonctions de la moyenne et de la variance. Quelques densités prédictives ont été données, rappelant des travaux faits au cours des dernières années. L'étude des marges du modèle lorsqu'on choisit un a priori normal-gamma indépendant a mené à quelques idées novatrices, notamment en ce qui concerne l'élicitation de ses marges a posteriori. On a vu tout au long de ce mémoire l'importance des lois décentrées, comme celle de la loi du chi-deux et de la loi de student, et des lois spéciales comme la densité de Kummer de type II. Divers exemples ont été détaillés afin de bien illustrer les résultats présentés et aussi afin de mieux comprendre l'impact des hyperparamètres dans les modèles considérés. On a également donné de nouvelles formes analytiques en ce qui concerne les densités a posteriori de fonction de la moyenne et de la variance sous un a priori normal-gamma. Toujours pour le quantile, le coefficient de variance et son inverse, ainsi que pour le coefficient de discrimination, on a décrit l'espérance, la variance, la densité ainsi que la répartition a posteriori pour un a priori conditionnellement normal avec une densité quelconque pour le paramètre qui contrôle la variance.

Des avenues possibles de recherches pourraient être de développer l'analyse prédictive

bayésienne pour les fonctions de la moyenne et de la variance présentées au dernier chapitre. Dans le même ordre d'idées, l'analyse a posteriori pour des cas de contraintes paramétriques pourrait être intéressante. On pourrait considérer des contraintes du type $\frac{\|\theta\|}{\sigma} \leq m$ où m est connue, des contraintes d'ordre du type $\theta_1 \geq \theta_2 \geq \dots \geq \theta_p$. De plus, il serait certainement intéressant de développer l'analyse a posteriori pour ces fonctions lorsqu'on considère l'a priori normal-gamma indépendant tel que présenté dans le Chapitre 3. Il n'est pas exclu de rendre disponibles certains outils de calcul et de simulation sur Maple ou encore sur R.

Finalement, on pourrait considérer des familles différentes ou encore d'autres modèles complexes. Par exemple, les mélanges de lois normales sur la variance et la moyenne, comme présenté dans [BNKS82], pourraient être explorées avec la même approche qu'on a utilisée dans ce mémoire. On considérerait alors les distributions hyperboliques généralisées et les distributions inverses gaussiennes généralisées, explorées dans [EH04], qui englobent les lois de student et les lois gamma, entre autres.

Bibliographie

- [Ait75] J. Aitchison. Goodness of prediction fit. *Biometrika*, 62(3) :547–554, 1975.
- [BNKS82] O. Barndorff-Nielsen, J. Kent, and M. Sorensen. Normal variance-mean mixtures and z distributions. *International Statistical Review*, 50 :145–159, 1982.
- [Bre10] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.
- [Car15] Andrew Carkeet. Exact parametric confidence intervals for bland-altman limits of agreement. *Optometry and Vision Science*, 92(3) :e71–e80, 2015.
- [CB02] George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.
- [CH03] Victor Chernozhukov and Han Hong. An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2) :293–346, 2003.
- [CL07] S Chakraborti and J Li. Confidence interval estimation of a normal percentile. *The American Statistician*, 61(4) :331–336, 2007.
- [CL08] Bradley P Carlin and Thomas A Louis. *Bayesian methods for data analysis*. CRC Press, 2008.
- [Coh88] Jonathan D Cohen. Noncentral chi-square : Some observations on recurrence. *The American Statistician*, 42(2) :120–122, 1988.

- [DVE09] Yiping Dou and Constance Van Eeden. Comparisons of the performances of estimators of a bounded normal mean under squared-error loss. *RevStat*, 7(3) :203–226, 2009.
- [DZ12] Allan Donner and GY Zou. Closed-form confidence intervals for functions of the normal mean and standard deviation. *Statistical Methods in Medical Research*, 21(4) :347–359, 2012. PMID : 20826501.
- [EH04] Ernst Eberlein and Ernst August v Hammerstein. Generalized hyperbolic and inverse gaussian distributions : limiting cases and approximation of processes. In *Seminar on stochastic analysis, random fields and applications IV*, pages 221–264. Springer, 2004.
- [ER04] Michael J Evans and Jeffrey S Rosenthal. *Probability and statistics : The Science of Uncertainty*. Macmillan, 2004.
- [GLX+06] Edward I George, Feng Liang, Xinyi Xu, et al. Improved minimax predictive densities under kullback–leibler loss. *The Annals of Statistics*, 34(1) :78–91, 2006.
- [GMS87] Constantine Gatsonis, Brenda MacGibbon, and William Strawderman. On the estimation of a restricted normal mean. *Statistics & Probability letters*, 6(1) :21–30, 1987.
- [GMS16] Ehssan Ghashim, Éric Marchand, and William E Strawderman. On a better lower bound for the frequentist probability of coverage of bayesian credible intervals in restricted parameter spaces. *Statistical Methodology*, 31 :43–57, 2016.
- [HR36] Walter A. Hendricks and Kate W. Robey. The sampling distribution of the coefficient of variation. *The Annals of Mathematical Statistics*, 7(3) :129–132, 09 1936.

- [HV16] Marwa Hamza and Pierre Vallois. On kummer's distribution of type two and a generalized beta distribution. *Statistics & Probability Letters*, 118 :60–69, 2016.
- [JW40] N. L. JOHNSON and B. L. WELCH. Applications of the non-central t-distribution. *Biometrika*, 31(3-4) :362–389, 1940.
- [L.91] Rukhin Andrew L. Admissible Estimators of Discriminant Coefficients. *Statistics & Risk Modeling*, 9(3) :285–296, 1991.
- [LC06] Erich L Lehmann and George Casella. *Theory of Point Estimation*. Springer Science & Business Media, 2006.
- [Lee97] Peter M Lee. *Bayesian Statistics*. Arnold Publication, 1997.
- [Lin72] Pi-Erh Lin. Some characterizations of the multivariate t distribution. *Journal of Multivariate Analysis*, 2(3) :339 – 344, 1972.
- [LMKS17] Aziz L'Moudden, Éric Marchand, Othmane Kortbi, and William E Strawderman. On predictive density estimation for gamma models with parametric constraints. *Journal of Statistical Planning and Inference*, 185 :56–68, 2017.
- [M+02] Mark Mandelkern et al. Setting confidence intervals for bounded parameters. *Statistical Science*, 17(2) :149–172, 2002.
- [MS04] Eric Marchand and William E Strawderman. Estimation in restricted parameter spaces : A review. *Lecture notes-monograph series*, pages 21–44, 2004.
- [Mui09] Robb J Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009.
- [NCL08] Saralees Nadarajah, S. Chakraborti, and J. Li. "confidence interval estimation of a normal percentile," "the american statistician," 61, 331-336 : Comment by nadarajah and response. *The American Statistician*, 62(2) :186–187, 2008.
- [PN13] Tibor K. Pogány and Saralees Nadarajah. On the convolution of normal and t random variables. *Statistics*, 47(6) :1363–1369, 2013.

- [Rob06] Christian Robert. *Le choix bayésien : Principes et pratique*. Springer Science & Business Media, 2006.
- [RPD01] John O Rawlings, Sastry G Pantula, and David A Dickey. *Applied regression analysis : a research tool*. Springer Science & Business Media, 2001.
- [Ste81] Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, pages 1135–1151, 1981.
- [Ste12] Ian Stewart. *Seventeen equations that changed the world*. Profile Books, 2012.
- [Str03] William E Strawderman. On minimax estimation of a normal mean vector for general quadratic loss. *Lecture Notes-Monograph Series*, pages 3–14, 2003.
- [vAG00] Andrea van Aubel and Wolfgang Gawronski. On the unimodality and the bell-shape of noncentral distributions. *Methods and Applications of Analysis*, 7(1) :233–250, 2000.
- [VAG03] Andrea Van Aubel and Wolfgang Gawronski. Analytic properties of noncentral distributions. *Applied Mathematics and Computation*, 141(1) :3–12, 2003.