Correction du TD 8

Exercice 1 (Poisson et Gamma)

1. On a
$$\pi(\theta \,|\, \mathbf{X}) \propto \theta^{a-1}\,\mathrm{e}^{-b\theta}\mathbb{1}_{\theta>0}\prod_{i=1}^{n}\left(\mathrm{e}^{-\theta}\theta^{X_i}\right)$$
$$\propto \theta^{a+n\overline{X}_n-1}e^{-(b+n)\theta}\mathbb{1}_{\theta>0}.$$
Ainsi $\Pi[\cdot \,|\, \mathbf{X}]$ est une loi Gamma$(a+n\overline{X}_n, b+n)$. En particulier, on obtient que la famille des lois Gamma(a,b) est conjuguée pour le modèle des lois de Poisson $\mathcal{P}(\theta)^{\otimes n}$.
2. La moyenne et la variance a posteriori valent
$$m_{\mathbf{X}} = \frac{a+n\overline{X}_n}{b+n}, \quad \text{et} \quad v_{\mathbf{X}} = \frac{a+n\overline{X}_n}{(b+n)^2}\cdot$$
3. On applique l'inégalité de Tchebychev à la loi a posteriori $\Pi[\cdot \,|\, \mathbf{X}]$: pour $\varepsilon > 0$,
$$\mathbb{P}\left(|\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta} \,|\, \mathbf{X}]| \geq \varepsilon \,|\, \mathbf{X}\right) \leq \frac{\mathbb{E}\left[(\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta} \,|\, \mathbf{X}])^2 \,|\, \mathbf{X}\right]}{\varepsilon^2}\cdot$$
Ceci implique que
$$\Pi\left([m_{\mathbf{X}} - \varepsilon, m_{\mathbf{X}} + \varepsilon] \,|\, \mathbf{X}\right) \geq 1 - \frac{v_{\mathbf{X}}}{\varepsilon^2}\cdot$$
On en déduit que
$$I(\mathbf{X}) = \left[m_{\mathbf{X}} \pm \sqrt{\frac{v_{\mathbf{X}}}{\alpha}}\right]$$
est un intervalle de crédibilité au moins $(1-\alpha)$.
4. On applique cette fois l'inégalité de Tchebychev en conditionnant par rapport à $\boldsymbol{\theta}$: pour tout $\varepsilon > 0$, on a
$$\mathbb{P}\left(\left|\overline{X}_n - \boldsymbol{\theta}\right| \geq \varepsilon \,|\, \boldsymbol{\theta}\right) \leq \frac{\mathrm{Var}(\overline{X}_n \,|\, \boldsymbol{\theta})}{\varepsilon^2} = \frac{\boldsymbol{\theta}}{n\varepsilon^2}\cdot$$
En prenant l'espérance dans cette inégalité, on obtient
$$\mathbb{P}\left(\left|\overline{X}_n - \boldsymbol{\theta}\right| \geq \varepsilon\right) \leq \frac{a}{nb\varepsilon^2}\cdot$$
Ainsi, $\overline{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}$.
5. La longueur de $I(\mathbf{X})$ est
$$|I(\mathbf{X})| = \frac{2\sqrt{v_{\mathbf{X}}}}{\sqrt{\alpha}} = \frac{2}{\sqrt{\alpha}}\sqrt{\frac{a+n\overline{X}_n}{(b+n)^2}} = \frac{2}{\sqrt{\alpha}}\frac{\sqrt{\frac{a}{n}+\overline{X}_n}}{\frac{b}{\sqrt{n}}+\sqrt{n}}\cdot$$
Comme $\overline{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}$, on a
$$\sqrt{n}\,|I(\mathbf{X})| \xrightarrow{\mathbb{P}} \frac{2\sqrt{\boldsymbol{\theta}}}{\sqrt{\alpha}}\cdot$$
La taille de $I(\mathbf{X})$ décroît donc en $1/\sqrt{n}$.

Exercice 2 (Régions PHD)

1. Notons g la densité a posteriori et G la fonction de répartition associée. Par les hypothèses faites sur g, G établit une bijection (continue et strictement croissante) de $\mathbb{R}$ vers $]0,1[$. De plus, par symétrie de g autour de m, toute région de plus haute densité est nécessairement de la forme $[m-u, m+u]$ avec l'ordonnée $y = g(m-u) = g(m+u)$. La probabilité associée est :
$$G(m+u) - G(m-u) = 2G(m+u) - 1.$$
Pour tout $0 < \alpha < 1$, l'aspect bijectif de G donne
$$2G(m+u) - 1 = 1 - \alpha \iff m + u = G^{-1}(1-\alpha/2),$$
qui est le quantile d'ordre $(1-\alpha/2)$ de la loi en question. Par symétrie de G, on obtient aussi $m-u = G^{-1}(\alpha/2)$, i.e. le quantile d'ordre $\alpha/2$ de la loi en question. Ceci prouve le résultat demandé.
2. Pour que les deux types de régions ne coïncident pas, on peut penser à une loi a posteriori multimodale. Par exemple, de façon caricaturale, prenons pour la loi de $\boldsymbol{\theta}$ un mélange équiprobable de deux gaussiennes normées et de moyennes respectives $-m$ et $+m$, i.e. de densité
$$\pi(\theta) = \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{(\theta+m)^2}{2}} + \frac{1}{\sqrt{2\pi}}e^{-\frac{(\theta-m)^2}{2}}\right).$$
Supposons maintenant X indépendante de $\boldsymbol{\theta}$, alors les lois a priori et a posteriori sont les mêmes. Si m est "grand", la région PHD de niveau 95% est approximativement
$$\mathcal{H} \approx [-m-2, -m+2] \cup [m-2, m+2],$$
qui est l'union de deux intervalles et ne peut donc coïncider avec l'intervalle défini par les quantiles d'ordres $2,5\%$ et $97,5\%$ de la loi π.
On peut aussi considérer le schéma suivant : $\boldsymbol{\theta} \sim \mathcal{E}(1)$ et X indépendante de $\boldsymbol{\theta}$. Alors la région PHD de niveau $(1-\alpha)$ est $[0, -\log\alpha]$, tandis que l'intervalle de crédibilité donné par les quantiles est $[-\log(1-\alpha/2), -\log(\alpha/2)]$.
3. Comme cela a été vu plusieurs fois, la loi a posteriori est $\Pi[\cdot \,|\, \mathbf{X}] = \mathcal{N}\left(\frac{n\overline{X}_n}{n+1}, \frac{1}{n+1}\right)$. Ainsi, en notant q_α le quantile d'ordre $1-\alpha/2$ d'une $\mathcal{N}(0,1)$, on obtient que
$$\mathbb{P}\left(\left|\boldsymbol{\theta} - \frac{n\overline{X}_n}{n+1}\right| \leq \frac{q_\alpha}{\sqrt{n+1}} \,|\, \mathbf{X}\right) = 1 - \alpha\,.$$
Ainsi
$$J(\mathbf{X}) = \left[\frac{n\overline{X}_n}{n+1} \pm \frac{q_\alpha}{\sqrt{n+1}}\right]$$
est une région de crédibilité de niveau exactement $1-\alpha$. De plus, comme la densité a posteriori est continue, symétrique par rapport à $\frac{n\overline{X}_n}{n+1}$, strictement croissante sur $\left]-\infty, \frac{n\overline{X}_n}{n+1}\right]$, il s'agit d'après la question 1 d'une région PHD.

Exercice 3 (Méthodes de Monte-Carlo et de rejet)

1. La densité a posteriori s'écrit, pour tout $\theta \in \mathbb{R}$:
$$\pi(\theta|\mathbf{X}) = \frac{\pi(\theta)p_\theta(\mathbf{X})}{f(\mathbf{X})},$$
avec :

— pour la loi a priori (loi de Cauchy) : $\pi(\theta) = \frac{1}{\pi(1+\theta^2)}$;
— pour la vraisemblance : $p_\theta(\mathbf{X}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-\theta)^2}{2}}$;
— pour la loi marginale des observations :

$$f(\mathbf{X}) = \int_{\mathbb{R}} \pi(t) p_t(\mathbf{X})\,\mathrm{d}t = \int_{\mathbb{R}} \frac{1}{\pi(1+t^2)} \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-t)^2}{2}} \right) \mathrm{d}t.$$

On peut noter que la loi a posteriori n'est pas une loi classique. De plus, d'après le cours, l'estimateur de Bayes pour la perte quadratique, noté $\widehat{\theta}_n$, est la moyenne a posteriori :

$$\widehat{\theta}_n = \mathbb{E}[\boldsymbol{\theta} \mid \mathbf{X}] = \int_{\mathbb{R}} \theta \; \pi(\theta|\mathbf{X})\,\mathrm{d}\theta.$$

On ne dispose pas de formule analytique évidente pour calculer cette intégrale, d'où le recours aux méthodes Monte-Carlo dans la suite.

2. La moyenne a posteriori s'écrit de façon générale :

$$\widehat{\theta}_n = \mathbb{E}[\boldsymbol{\theta} \mid \mathbf{X}] = \int_{\mathbb{R}} \theta \pi(\theta|\mathbf{X})\,\mathrm{d}\theta = \int_{\mathbb{R}} \theta \left\{ \frac{p_\theta(\mathbf{X})\pi(\theta)}{\int_{\mathbb{R}} p_t(\mathbf{X})\pi(t)\,\mathrm{d}t} \right\} \mathrm{d}\theta = \frac{\int_{\mathbb{R}} \theta p_\theta(\mathbf{X})\pi(\theta)\,\mathrm{d}\theta}{\int_{\mathbb{R}} p_t(\mathbf{X})\pi(t)\,\mathrm{d}t},$$

où $p_\theta(\mathbf{X}) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(X_i-\theta)^2}{2}} \right)$ est la vraisemblance de θ. Numérateur et dénominateur peuvent être vus comme des espérances par rapport à la loi a priori sur $\boldsymbol{\theta}$, à savoir :

$$\widehat{\theta}_n = \varphi(\mathbf{X}), \qquad \text{où} \qquad \varphi(x) = \frac{\mathbb{E}\left[\boldsymbol{\theta} p_{\boldsymbol{\theta}}(x)\right]}{\mathbb{E}\left[p_{\boldsymbol{\theta}}(x)\right]}.$$

La notation $\varphi(x)$ a été introduite ici pour signifier que l'espérance est prise uniquement par rapport à $\boldsymbol{\theta}$ suivant la loi a priori Π (ici la loi de Cauchy) et non par rapport à $\Pi(\cdot|\mathbf{X})$.
Dès lors, ces deux quantités peuvent être estimées par une méthode Monte-Carlo basée sur la simulation de N variables $(\boldsymbol{\theta}_i)_{1\le i\le N}$ i.i.d. selon la loi a priori (et indépendamment de $X_1, \ldots, X_n$), ce qui conduit à l'estimateur :

$$\widehat{\theta}_n^N = \frac{\frac{1}{N}\sum_{j=1}^N \boldsymbol{\theta}_j \; p_{\boldsymbol{\theta}_j}(\mathbf{X})}{\frac{1}{N}\sum_{j=1}^N p_{\boldsymbol{\theta}_j}(\mathbf{X})} = \frac{\sum_{j=1}^N \boldsymbol{\theta}_j \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\theta}_j)^2\right\}}{\sum_{j=1}^N \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\theta}_j)^2\right\}}.$$

Sachant $\mathbf{X}$, la loi forte des grands nombres assure la convergence p.s. de $\widehat{\theta}_n^N$ vers $\widehat{\theta}_n$ lorsque N tend vers l'infini.

3. La méthode de rejet pour échantillonner suivant $\Pi[\cdot \mid \mathbf{X}]$ avec pour loi instrumentale Π peut être résumée de la manière suivante. Soit $m \ge 1$ telle que pour tout $\theta \in \mathbb{R}$, $\frac{\pi(\theta \mid \mathbf{X})}{\pi(\theta)} \le m$. Opérer :
   (a) Échantillonner $\boldsymbol{\theta} \sim \Pi \perp\!\!\!\perp U \sim \mathcal{U}([0,1])$ (et indépendamment des variables aléatoires précédentes).
   (b) Si $\frac{\pi(\boldsymbol{\theta} \mid \mathbf{X})}{m\pi(\boldsymbol{\theta})} \ge U$, retourner $\boldsymbol{\theta}$, sinon recommencer à l'étape (a).

Or, pour tout $\theta \in \mathbb{R}$,

$$\frac{\pi(\theta \mid \mathbf{X})}{\pi(\theta)} = \frac{p_\theta(\mathbf{X})}{f(\mathbf{X})} \le \frac{p_{\widehat{\theta}_n^{MV}}(\mathbf{X})}{f(\mathbf{X})}.$$

On peut donc choisir $m = \frac{p_{\widehat{\theta}_n^{MV}}(\mathbf{X})}{f(\mathbf{X})}$. Le rapport d'acceptation devient alors

$$\frac{\pi(\theta \mid \mathbf{X})}{m\pi(\theta)} = \frac{p_\theta(\mathbf{X})}{p_{\widehat{\theta}_n^{MV}}(\mathbf{X})}.$$

L'intérêt de cette méthode est de ne pas faire appel à la densité marginale $f(\mathbf{X})$, qui est souvent hors de portée.

Remarque : ceci implique néanmoins de pouvoir calculer facilement l'estimateur du maximum de vraisemblance $\widehat{\theta}_n^{MV}$ ainsi que la vraisemblance $p_\theta(\mathbf{X})$ en tout point θ.

4. Pour calculer le rapport d'acceptation dans le cas du modèle bayésien de l'exercice, il faut :
   (a) calculer l'estimateur du maximum de vraisemblance $\widehat{\theta}_n^{MV}$, qui est tout simplement $\widehat{\theta}_n^{MV} = \overline{X}_n$;
   (b) exprimer le rapport :
$$\frac{\pi(\theta|\mathbf{X})}{m\pi(\theta)} = \frac{p_\theta(\mathbf{X})}{p_{\widehat{\theta}_n^{MV}}(\mathbf{X})} = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^n(\mathbf{X}_i-\theta)^2\right\}}{\exp\left\{-\frac{1}{2}\sum_{i=1}^n(\mathbf{X}_i-\overline{\mathbf{X}}_n)^2\right\}} = \exp\left\{-\frac{n}{2}(\overline{\mathbf{X}}_n-\theta)^2\right\}.$$

Exercice 4 (Famille gaussienne à moyenne et variance inconnues)

1. Le produit de la vraisemblance et de l'a priori (impropre) s'écrit
$$\begin{aligned}
p_{\mu,\sigma^2}(\mathbf{X})\sigma^{-2} &= (2\pi)^{-n/2}(\sigma^2)^{-\frac{n}{2}-1}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\mu)^2\right\}\\
&= (2\pi)^{-n/2}(\sigma^2)^{-\frac{n}{2}-1}\exp\left\{-\frac{n}{2\sigma^2}\left(\mu^2-2\overline{X}_n\mu+\frac{1}{n}\sum_{i=1}^n X_i^2\right)\right\}\\
&= (2\pi)^{-n/2}(\sigma^2)^{-\frac{n}{2}-1}\exp\left\{-\frac{n}{2\sigma^2}(\mu-\overline{X}_n)^2-\frac{ns_{\mathbf{X}}}{2\sigma^2}\right\},
\end{aligned}$$
avec $s_{\mathbf{X}} := \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2 > 0$ presque sûrement puisque, presque sûrement, les X_i ne sont pas tous égaux. Pour que la loi a posteriori $\mathcal{L}((\mu,\sigma^2)\,|\,\mathbf{X})$ soit bien définie, il faut que l'intégrale de la quantité précédente par rapport à (μ,σ^2) soit finie. C'est bien le cas car, par le théorème de Fubini-Tonelli, cette intégrale double est proportionnelle à
$$\begin{aligned}
&\int_{\mathbb{R}_+^*}(\sigma^2)^{-\frac{n}{2}-1}e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}}\left(\int_{\mathbb{R}}e^{-\frac{n}{2\sigma^2}(\mu-\overline{X}_n)^2}\,\mathrm{d}\mu\right)\mathrm{d}\sigma^2\\
&= \sqrt{\frac{2\pi}{n}}\int_{\mathbb{R}_+^*}(\sigma^2)^{-\frac{n+1}{2}}e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}}\,\mathrm{d}\sigma^2\\
&= \sqrt{\frac{2\pi}{n}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\left(\frac{ns_{\mathbf{X}}}{2}\right)^{\frac{n-1}{2}}},
\end{aligned}$$
où l'on a reconnu la densité d'une variable IG $\left(\frac{n-1}{2},\frac{ns_{\mathbf{X}}}{2}\right)$ (noter qu'il faut $n \geq 2$ pour que cela soit bien défini).

2. Le calcul vient d'être fait. En effet, la densité $\pi(\sigma^2\,|\,\mathbf{X})$ de la loi $\mathcal{L}(\sigma^2\,|\,\mathbf{X})$ est
$$\begin{aligned}
\pi(\sigma^2\,|\,\mathbf{X}) = \int_{\mathbb{R}}p_{\mu,\sigma^2}(\mathbf{X})\sigma^{-2}\,\mathrm{d}\mu &\propto (\sigma^2)^{-\frac{n}{2}-1}e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}}\sqrt{2\pi\frac{\sigma^2}{n}}\\
&\propto (\sigma^2)^{-\frac{n+1}{2}}e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}}.
\end{aligned}$$
On reconnaît bien la densité d'une loi IG$(\frac{n-1}{2},\frac{ns_{\mathbf{X}}}{2})$.

3. Par le même raisonnement qu'en question 1, la densité $\pi(\mu\,|\,\sigma^2,\mathbf{X})$ de la loi $\mathcal{L}(\mu\,|\,\sigma^2,\mathbf{X})$ est
$$\pi(\mu\,|\,\sigma^2,\mathbf{X}) \propto p_{\mu,\sigma^2}(\mathbf{X})\sigma^{-2} \propto \exp\left\{-\frac{n}{2\sigma^2}(\mu-\overline{X}_n)^2\right\}.$$

On en déduit que $\mathcal{L}(\mu \,|\, \sigma^2, \mathbf{X}) = \mathcal{N}(\overline{X}_n, \sigma^2/n)$. Or on a vu plus haut que $\mathcal{L}(\sigma^2 \,|\, \mathbf{X}) = \mathrm{IG}(\frac{n-1}{2}, \frac{ns_{\mathbf{X}}}{2})$. Autrement dit, d'après le rappel de l'énoncé, par définition d'une loi normale inverse-gamma, $\mathcal{L}(\mu, \sigma^2 \,|\, \mathbf{X}) = \mathrm{NIG}\left(\overline{X}_n, n, \frac{n-1}{2}, \frac{ns_{\mathbf{X}}}{2}\right)$.

4. Pour déterminer $\mathcal{L}(\mu \,|\, \mathbf{X})$, on procède comme pour celle de $\mathcal{L}(\sigma^2 \,|\, \mathbf{X})$: on détermine la loi jointe de $(\mu, \mathbf{X})$ en intégrant en σ^2 la densité de $(\mu, \sigma^2, \mathbf{X})$, puis on en déduit la loi de μ sachant $\mathbf{X}$. La densité de $(\mu, \mathbf{X})$ est proportionnelle à

$$\begin{aligned}
\int_{\mathbb{R}_+^*} p_{\mu,\sigma^2}(\mathbf{X})\sigma^{-2}\,\mathrm{d}\sigma^2 &\propto \int_{\mathbb{R}_+^*} (\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\frac{n}{2\sigma^2}(\mu - \overline{X}_n)^2 - \frac{ns_{\mathbf{X}}}{2\sigma^2}\right\} \mathrm{d}\sigma^2 \\
&\propto \int_{\mathbb{R}_+^*} (\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\sigma^{-2}\left(\frac{n}{2}(\mu - \overline{X}_n)^2 + \frac{ns_{\mathbf{X}}}{2}\right)\right\} \mathrm{d}\sigma^2 \\
&\propto \left(\frac{n}{2}(\mu - \overline{X}_n)^2 + \frac{ns_{\mathbf{X}}}{2}\right)^{-\frac{n}{2}} \\
&\propto \left(\frac{(\mu - \overline{X}_n)^2}{s_{\mathbf{X}}} + 1\right)^{-\frac{n}{2}}
\end{aligned}$$

où l'on a utilisé l'expression de la densité d'une loi inverse gamma $z \mapsto \frac{b^a}{\Gamma(a)} z^{-a-1} e^{-b/z}$ avec

$$a = \frac{n}{2} \qquad \text{et} \qquad b = \frac{n}{2}(\mu - \overline{X}_n)^2 + \frac{ns_{\mathbf{X}}}{2}\,.$$

En utilisant l'expression de la densité d'une loi de Student $f_p(u) \propto \left(1 + \frac{u^2}{p}\right)^{-\frac{p+1}{2}}$, et en utilisant le fait que si Z est une variable de densité g et si $a \in \mathbb{R}$ et $b > 0$ sont deux constantes, alors $\frac{Z-a}{b}$ a pour densité $u \mapsto bg(bu + a)$, on obtient

$$\mathcal{L}\left(\frac{\mu - \overline{X}_n}{\sqrt{\frac{s_{\mathbf{X}}}{n-1}}} \,\Big|\, \mathbf{X}\right) = \mathcal{T}(n-1)\,,$$

où $\mathcal{T}(n-1)$ est la loi de Student à $n-1$ degrés de liberté. En notant z_α le quantile d'ordre $(1-\alpha/2)$ d'une loi de Student à $(n-1)$ degrés de liberté (qui est l'opposé du quantile d'ordre $\alpha/2$ puisque la loi de Student est symétrique), on en conclut que

$$\mathbb{P}\left(\left|\frac{\mu - \overline{X}_n}{\sqrt{\frac{s_{\mathbf{X}}}{n-1}}}\right| \le z_\alpha \,\Big|\, \mathbf{X}\right) = 1 - \alpha\,,$$

et donc que

$$I(\mathbf{X}) = \left[\overline{X}_n \pm \sqrt{\frac{s_{\mathbf{X}}}{n-1}} z_\alpha\right]$$

est un intervalle de crédibilité pour μ au niveau $1 - \alpha$.

Exercice 5 (Famille gaussienne conjuguée dans $\mathbb{R}^d$)

1. Soit $\mathbf{X} \,|\, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)^{\otimes n}$ et $\boldsymbol{\theta} \sim \Pi = \mathcal{N}(0, \Lambda)$. La formule de Bayes donne

$$\pi(\theta \,|\, \mathbf{X}) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} {}^t(X_i - \theta)\Sigma^{-1}(X_i - \theta) - \frac{1}{2}\,{}^t\theta\Lambda^{-1}\theta\right\}.$$

Il suffit de regrouper les termes en θ pour former une nouvelle forme quadratique dans l'exponentielle, de la forme $-\frac{1}{2}\,{}^t(\theta - m_{\mathbf{X}})\Sigma_{\mathbf{X}}^{-1}(\theta - m_{\mathbf{X}})$. On cherche donc $m_{\mathbf{X}}, \Sigma_{\mathbf{X}}$ tels que, en développant chaque forme quadratique

$$ {}^t\theta(n\Sigma^{-1} + \Lambda^{-1})\theta - 2\,{}^t(n\overline{X}_n)\Sigma^{-1}\theta = {}^t\theta\Sigma_{\mathbf{X}}^{-1}\theta - 2\,{}^t m_{\mathbf{X}}\Sigma_{\mathbf{X}}^{-1}\theta. $$

En identifiant, il suffit donc de poser

$$ \Sigma_{\mathbf{X}} = (n\Sigma^{-1} + \Lambda^{-1})^{-1} $$

et de choisir $m_{\mathbf{X}}$ tel que

$$ 2\,{}^t m_{\mathbf{X}}\Sigma_X^{-1} = 2n\,{}^t\overline{X}_n\Sigma^{-1}, $$

ce qui donne $m_{\mathbf{X}} = n\Sigma_{\mathbf{X}}\Sigma^{-1}\overline{X}_n$.

2. La famille n'est pas conjuguée car la loi a posteriori n'est en général pas d'espérance nulle.
Remarque. Si l'on considère la loi a priori $\mathcal{N}(\mu, \Lambda)$, avec $\mu \in \mathbb{R}^d$ et $\Lambda \in S_d^{++}(\mathbb{R})$, alors le même raisonnement que ci-dessus permet de montrer que la loi a posteriori est $\mathcal{N}(m_{\mathbf{X}}, \Sigma_{\mathbf{X}})$ avec

$$ \Sigma_{\mathbf{X}} = (n\Sigma^{-1} + \Lambda^{-1})^{-1} $$

et

$$ m_{\mathbf{X}} = \Sigma_{\mathbf{X}}(\Lambda^{-1}\mu + n\Sigma^{-1}\overline{X}_n), $$

ce qui prouve que la famille $\{\mathcal{N}(\mu, \Lambda), \mu \in \mathbb{R}^d, \Lambda \in S_d^{++}(\mathbb{R})\}$ est conjuguée.

Exercice 6 ($\star$) (Urne de Polya)

1. Soit $n \geq 1$ et $0 \leq k \leq n$. En utilisant la formule

$$ \mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \,|\, A_1)\ldots\mathbb{P}(A_n \,|\, A_1 \cap \cdots \cap A_{n-1})\,, $$

on obtient

$$ \begin{aligned} &\mathbb{P}\left(I_1 = \cdots = I_k = 1, I_{k+1} = \cdots = I_n = 0\right) \\ &\quad = \frac{a}{a+b} \times \cdots \times \frac{a+k-1}{a+b+k-1} \times \frac{b}{a+b+k} \times \cdots \times \frac{b+n-k-1}{a+b+n-1}\,. \end{aligned} $$

2. Soit $(i_1, \ldots, i_n) \in \{0,1\}^n$ et notons $k = \sum_j i_j$ le nombre de 1 dans la suite. Lorsque l'on écrit la probabilité d'obtenir $(i_1, \ldots, i_n)$ comme produit des probabilités de chaque tirage sachant les tirages précédents, le dénominateur est égal à

$$ (a+b)(a+b+1)\ldots(a+b+n-1) $$

le $\ell^{\text{ème}}$ terme de ce produit correspondant au nombre total de boules dans l'urne à l'instant du $\ell^{\text{ème}}$ tirage. Quant au numérateur, quitte à réordonner les termes, il est égal à

$$ a(a+1)\ldots(a+k-1).b(b+1)\ldots(b+n-k-1)\,. $$

En effet, au moment où l'on tire la $\ell^{\text{ème}}$ boule noire, il y a $a+\ell-1$ boules noires dans l'urne, et au moment où l'on tire la $\ell^{\text{ème}}$ boule blanche, il y a $b+\ell-1$ boules blanches dans l'urne. La loi de $(I_1, \ldots, I_n)$ ne dépend donc pas de l'ordre dans lequel on tire les boules mais uniquement du nombre de boules noires tirées. On dit que la loi de $(I_1, \ldots, I_n)$ est *échangeable*.

3. Par la question précédente, pour tout $k \in \{0,\ldots,n\}$ et pour tout $(i_1,\ldots,i_n) \in \{0,1\}^n$ tel que $\sum_{j=1}^n i_j = k$, on a, en choisissant la permutation des indices telles que toutes les boules noires arrivent d'abord, et toutes les boules blanches ensuite,

$$\begin{aligned}&\mathbb{P}\left(I_1 = i_1,\ldots,I_n = i_n\right)\\ &\qquad = \mathbb{P}\left(I_1 = \cdots = I_k = 1, I_{k+1} = \cdots = I_n = 0\right) .\end{aligned}$$

En réécrivant la probabilité obtenue à la question 1 à l'aide de la fonction Gamma, on obtient

$$\begin{aligned}&\mathbb{P}\left(I_1 = \cdots = I_k = 1, I_{k+1} = \cdots = I_n = 0\right)\\ &= \frac{a}{a+b} \times \cdots \times \frac{a+k-1}{a+b+k-1} \times \frac{b}{a+b+k} \times \cdots \times \frac{b+n-k-1}{a+b+n-1}\\ &= \frac{\Gamma(a+k)}{\Gamma(a)} \times \frac{\Gamma(a+b)}{\Gamma(a+b+k)} \times \frac{\Gamma(b+n-k)}{\Gamma(b)} \times \frac{\Gamma(a+b+k)}{\Gamma(a+b+n)}\\ &= \frac{B(a+k,b+n-k)}{B(a,b)} .\end{aligned}$$

4. (a) En utilisant encore une fois le fait que la probabilité de tirer $(i_1,\ldots,i_n)$ ne dépend que du nombre de 1 dans la suite mais pas de leur position dans la suite, on a, pour tout $0 \le k \le n$,

$$\begin{aligned}\mathbb{P}(X_n = k) &= \binom{n}{k}\mathbb{P}\left(I_1 = \cdots = I_k = 1, I_{k+1} = \cdots = I_n = 0\right)\\ &= \binom{n}{k}B(k+1,n-k+1)\\ &= \binom{n}{k}\frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1} .\end{aligned}$$

La loi de X_n est donc uniforme sur $\{0,\ldots,n\}$.

(b) La variable X_n/n est à valeurs dans $[0,1]$. Soit $x \in [0,1]$, alors

$$\mathbb{P}(X_n/n \le x) = \sum_{k=0}^{\lfloor nx \rfloor} \mathbb{P}(X_n = k) = \frac{\lfloor nx \rfloor + 1}{n+1} \underset{n\to\infty}{\longrightarrow} x .$$

La loi limite de X_n/n est donc la loi uniforme sur $[0,1]$. Par le Lemme de Slutsky, la loi limite de M_n est aussi la loi uniforme sur $[0,1]$.

5. (a) Pour tout $0 \le k \le n$,

$$\mathbb{P}(X_n = k) = \binom{n}{k}\frac{B(a+k,b+n-k)}{B(a,b)} .$$

Dans le scénario bayésien $\boldsymbol{\theta} \sim \mathrm{Beta}(a,b)$ et $Z_n \,\big|\, \boldsymbol{\theta} \sim \mathrm{Bin}(n,\boldsymbol{\theta})$, la loi marginale de Z_n est donnée par : pour tout $0 \le k \le n$,

$$\begin{aligned}\mathbb{P}(Z_n = k) &= \frac{1}{B(a,b)}\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\mathbb{P}_\theta(Z_n = k)d\theta\\ &= \frac{1}{B(a,b)}\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\binom{n}{k}\theta^k(1-\theta)^{n-k}d\theta\\ &= \frac{\binom{n}{k}}{B(a,b)}\int_0^1 \theta^{a+k-1}(1-\theta)^{b+n-k-1}d\theta\\ &= \binom{n}{k}\frac{B(a+k,b+n-k)}{B(a,b)} .\end{aligned}$$

Les variables X_n et Z_n suivent bien la même loi.

(b) Soit $\varepsilon > 0$. Par l'inégalité de Hoeffding,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{Z_n}{n} - \boldsymbol{\theta}\right| > \varepsilon\right) &= \mathbb{E}\left[\mathbb{P}\left(\left|\frac{Z_n}{n} - \boldsymbol{\theta}\right| > \varepsilon \,|\, \boldsymbol{\theta}\right)\right] \\ &\leq 2\exp\left(-2n\varepsilon^2\right) , \end{aligned}$$

qui est le terme général d'une série convergente. Par le lemme de Borel-Cantelli, $Z_n/n \xrightarrow{\text{p.s.}} \boldsymbol{\theta}$.

(c) On a

$$\frac{X_n}{n} \overset{\text{loi}}{=} \frac{Z_n}{n} .$$

Donc X_n/n converge en loi vers une loi $\mathrm{Beta}(a, b)$, et par le lemme de Slutsky, il en est de même de M_n. En remarquant que la suite (M_n) est une martingale, on peut montrer qu'il s'agit d'une convergence presque sûre.