CORRECTION DU TD 7

EXERCICE 1 (Lois conditionnelles et vecteur gaussien)

1. (a) Par définition la densité marginale f de X s'obtient à partir de la densité jointe de (X, Y) en intégrant par rapport à y . Donc pour tout $x \in \mathbb{R}$,

$$\begin{aligned} f(x) &= \int_{\mathbb{R}} \frac{1}{2\pi} \exp(-x^2 + xy - y^2/2) dy = e^{-x^2} \int_{\mathbb{R}} \frac{1}{2\pi} e^{-(y-x)^2/2+x^2/2} dy \\ &= e^{-x^2/2} \int_{\mathbb{R}} \frac{1}{2\pi} e^{-(y-x)^2/2} dy = e^{-x^2/2} \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

Ainsi $\mathcal{L}(X) = \mathcal{N}(0, 1)$. De même, la densité g de Y est donnée par, pour tout $y \in \mathbb{R}$,

$$\begin{aligned} g(y) &= \int_{\mathbb{R}} \frac{1}{2\pi} \exp(-x^2 + xy - y^2/2) dx = \frac{e^{-y^2/2}}{2\pi} \int_{\mathbb{R}} e^{-(x-\frac{y}{2})^2+y^2/4} dx = \frac{e^{-y^2/4}}{2\pi} \int_{\mathbb{R}} e^{-(x-\frac{y}{2})^2} dx \\ &= \frac{e^{-y^2/4}}{2\sqrt{\pi}}. \end{aligned}$$

On reconnaît la densité d'une loi $\mathcal{N}(0, 2)$, qui est donc la loi de Y .

- (b) Par définition la densité conditionnelle de Y sachant $X = x$ est, si $h(x, y)$ est la densité du couple (X, Y) (ici par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$),

$$g_x(y) = \frac{h(x, y)}{\int_{\mathbb{R}} h(x, y) dy} = \frac{h(x, y)}{f(x)} = \frac{\frac{1}{2\pi} \exp(-x^2 + xy - y^2/2)}{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}},$$

d'après le calcul de $f(x)$ ci-dessus. On simplifie et on obtient

$$g_x(y) = \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2}.$$

Ainsi $\mathcal{L}(Y | X = x) = \mathcal{N}(x, 1)$. On aurait aussi pu obtenir ce résultat par la méthode du 'proportionnel à'. Utilisons cette méthode pour déterminer l'autre densité conditionnelle, $f_y(x)$. Comme on s'intéresse à une fonction de x , tout ce qui ne dépend pas de x peut aller dans la constante :

$$f_y(x) \propto_x \exp(-x^2 + xy - y^2/2) \propto_x e^{-x^2+xy} \propto_x e^{-(x-y/2)^2+y^2/4} \propto_x e^{-(x-y/2)^2}.$$

Ainsi $\mathcal{L}(X | Y = y) = \mathcal{N}(y/2, 1/2)$.

2. On rappelle que la densité d'un vecteur gaussien $\mathcal{N}(\mu, V)$ dans $\mathbb{R}^d$, avec $\mu \in \mathbb{R}^d$ et V une matrice symétrique **définie** positive de taille $d \times d$ est donnée par

$$z \mapsto \frac{1}{\sqrt{(2\pi)^d \det(V)}} \exp \left\{ -\frac{1}{2} {}^t(z - \mu) V^{-1} (z - \mu) \right\}.$$

Ici, on nous dit que $Z = {}^t[X, Y]$ (avec tu la transposée de u) a pour loi $\mathcal{N}(\mu, V)$. Il suffit d'identifier μ et V avec l'expression de $h(x, y)$ donnée par l'énoncé. Remarquons déjà que d'après la question précédente $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ ce qui suggère de poser $\mu = 0$. Pour déterminer V , on peut procéder par identification. Pour ${}^tz = {}^t[x, y]$, et $M = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ une matrice 2×2 symétrique définie positive, on a

$$\frac{1}{2} {}^tz M z = ax^2/2 + bxy + cy^2/2.$$

En identifiant avec l'expression $f(x, y)$ il vient $a = 2, b = -1, c = 1$. Ceci nous donne la matrice $V^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$. La formule d'inversion (transposée de la co-matrice) donne alors

$$V = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

On en conclut $\mathcal{L}(Z) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\right)$. Remarquons que l'on aurait aussi pu identifier V par

$$V = \begin{bmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{bmatrix}.$$

On sait que $\mathbb{E}[X] = \mathbb{E}[Y] = 0, \text{Var}(X) = 1$ et $\text{Var}(Y) = 2$. De plus, on a

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] = \mathbb{E}[X\mathbb{E}[Y | X]] = \mathbb{E}[X^2] = 1,$$

car $\mathcal{L}(Y | X) = \mathcal{N}(X, 1)$. On trouve donc directement $V = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. On peut ensuite vérifier que la densité jointe de l'énoncé est bien celle d'un vecteur gaussien centré de matrice de covariance V .

EXERCICE 2 (Lois conjuguées)

1. On part de $\Pi = \text{Gamma}(a, b)$. Dans le modèle $\mathcal{P} = \{\text{Gamma}(p, \lambda)^{\otimes n}, \lambda > 0\}$, la loi a posteriori $\Pi[\cdot | \mathbf{X}]$ a pour densité

$$\pi(\lambda | \mathbf{X}) \propto \pi(\lambda) \prod_{i=1}^n p_\lambda(X_i) \propto \lambda^{a-1} e^{-b\lambda} \mathbb{1}_{\lambda>0} \prod_{i=1}^n \lambda^p e^{-\lambda X_i} \mathbb{1}_{X_i>0} \propto \lambda^{np+a-1} e^{-(b+n\bar{X}_n)\lambda} \mathbb{1}_{\lambda>0}.$$

Ainsi $\Pi[\cdot | \mathbf{X}] = \text{Gamma}(a + np, b + n\bar{X}_n)$. La moyenne a posteriori vaut $\frac{a+np}{b+n\bar{X}_n}$ (la moyenne d'une $\text{Gamma}(a, b)$ est a/b).

2. On procède de même. On part de $\Pi = \text{Beta}(a, b)$. Dans le modèle $\mathcal{P} = \{\text{Geom}(\theta)^{\otimes n}, \theta \in]0, 1[\}$, la loi a posteriori $\Pi[\cdot | \mathbf{X}]$ a pour densité

$$\pi(\theta | \mathbf{X}) \propto \pi(\theta) \prod_{i=1}^n p_\theta(X_i) \propto \theta^{a-1} (1-\theta)^{b-1} \mathbb{1}_{0<\theta<1} \prod_{i=1}^n \theta (1-\theta)^{X_i-1} \propto \theta^{a+n-1} (1-\theta)^{b+n\bar{X}_n-n-1} \mathbb{1}_{0<\theta<1}.$$

Ainsi $\Pi[\cdot | \mathbf{X}] = \text{Beta}(a + n, b + n\bar{X}_n - n)$. La moyenne a posteriori vaut $\frac{a+n}{a+b+n\bar{X}_n}$ (la moyenne d'une $\text{Beta}(a, b)$ est $a/(a+b)$).

3. Soit $\Pi = \mathcal{N}(\mu, \sigma^2)$. Il suffit de montrer que, si on prend Π comme loi a priori sur θ dans le modèle $\{\mathcal{N}(\theta, v)^{\otimes n}, \theta \in \mathbb{R}\}$, la loi a posteriori $\Pi[\cdot | \mathbf{X}]$ est encore gaussienne. Ici $\mathbf{X} = (X_1, \dots, X_n)$, et $\mathbf{X} | \theta \sim \mathcal{N}(\theta, v)^{\otimes n}$. On a

$$\begin{aligned} d\Pi(\theta) &= \pi(\theta) d\theta, & \pi(\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\theta-\mu)^2}. \\ dP_\theta(x) &= p_\theta(x) dx, & p_\theta(x) &= \frac{1}{\sqrt{2\pi v}} e^{-\frac{1}{2v}(x-\theta)^2}. \end{aligned}$$

La densité a posteriori s'écrit

$$\begin{aligned} \pi(\theta | \mathbf{X}) &\propto \pi(\theta) \prod_{i=1}^n p_\theta(X_i) \propto \exp \left\{ -\frac{1}{2\sigma^2}(\theta - \mu)^2 - \frac{1}{2v} \sum_{i=1}^n (X_i - \theta)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2}(\theta^2 - 2\mu\theta) - \frac{1}{2v}(n\theta^2 - 2n\bar{X}_n\theta) \right\} \\ &\propto \exp \left\{ -\frac{\sigma^{-2} + nv^{-1}}{2} \left(\theta^2 - 2\theta \frac{\mu\sigma^{-2} + nv^{-1}\bar{X}_n}{\sigma^{-2} + nv^{-1}} \right) \right\} \\ &\propto \exp \left\{ -\frac{\sigma^{-2} + nv^{-1}}{2} \left(\theta - \frac{\mu\sigma^{-2} + nv^{-1}\bar{X}_n}{\sigma^{-2} + nv^{-1}} \right)^2 \right\}. \end{aligned}$$

Ainsi

$$\Pi[\cdot | \mathbf{X}] = \mathcal{N} \left(\frac{\mu\sigma^{-2} + nv^{-1}\bar{X}_n}{\sigma^{-2} + nv^{-1}}, \frac{1}{\sigma^{-2} + nv^{-1}} \right). \quad (1)$$

Donc la famille considérée est conjuguée. La moyenne a posteriori vaut $\mathbb{E}[\theta | \mathbf{X}] = \frac{\mu\sigma^{-2} + nv^{-1}\bar{X}_n}{\sigma^{-2} + nv^{-1}}$. On peut noter que cette moyenne a posteriori est une moyenne pondérée entre la moyenne μ de la loi a priori et la moyenne empirique $\bar{X}_n$ des observations :

$$\mathbb{E}[\theta | \mathbf{X}] = \frac{\sigma^{-2}}{\sigma^{-2} + nv^{-1}} \mu + \frac{nv^{-1}}{\sigma^{-2} + nv^{-1}} \bar{X}_n.$$

En particulier, plus le nombre n d'observations est grand et moins la loi a priori affecte la loi a posteriori.

EXERCICE 3 (A priori impropre)

1. Notons que

$$\int_{\mathbb{R}} \prod_{i=1}^n p_\theta(X_i) d\theta = \int_{\mathbb{R}} \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-\theta)^2}{2}} \right) d\theta \leq \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_1-\theta)^2}{2}} d\theta = 1 < +\infty.$$

Ainsi le dénominateur dans la formule de Bayes est fini et la loi a posteriori est bien définie. Pour $\theta \in \mathbb{R}$, la densité a posteriori est donnée par

$$\pi(\theta | X_1, \dots, X_n) = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-\theta)^2}{2}}}{\int_{\mathbb{R}} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-\theta)^2}{2}} d\theta} \propto e^{-\frac{n}{2}(\theta - \bar{X}_n)^2}.$$

On en déduit $\Pi[\cdot | \mathbf{X}] = \mathcal{N}(\bar{X}_n, \frac{1}{n})$.

2. (a) On prend $\Theta = \mathbb{N}^*$ l'ensemble des valeurs possibles pour le nombre total de tramways dans la ville. Sachant qu'il y a N tramways dans la ville, le numéro du premier tramway observé est supposé uniformément distribué entre 1 et N . Plus formellement

$$N \sim \Pi$$

$$X \mid N \sim \text{Unif}(\{1, \dots, N\}),$$

où Π est une loi a priori sur $\mathbb{N}^*$ à choisir.

- (b) Si l'on ne veut vraiment rien supposer a priori, on est tenté de prendre l'a priori impropre $\Pi = \sum_{m \geq 1} \delta_m$, i.e. attribuer à chaque nombre de tramways possible le même poids 1. Le problème avec ce choix est que, pour tout $X \in \mathbb{N}^*$, on a

$$\int_{\mathbb{N}^*} p_n(X) d\Pi(n) = \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{1}_{X \in \{1, \dots, n\}} = \sum_{n=X}^{\infty} \frac{1}{n} = +\infty.$$

Cet a priori est donc « trop impropre » : la loi a posteriori n'est pas définie.

- (c) On considère maintenant, pour $k \geq 2$ donné,

$$N \sim \Pi$$

$$\mathbf{X} = (X_1, \dots, X_k) \mid N \sim \text{Unif}(\{1, \dots, N\})^{\otimes k},$$

avec à nouveau $\Pi = \sum_{m \geq 1} \delta_m$. On a alors, pour $\mathbf{X} = (X_1, \dots, X_k) \in (\mathbb{N}^*)^k$,

$$\int_{\mathbb{N}^*} p_n(\mathbf{X}) d\Pi(n) = \sum_{n=1}^{\infty} \prod_{i=1}^k \left(\frac{1}{n} \mathbb{1}_{X_i \in \{1, \dots, n\}} \right) = \sum_{n=X_{(k)}}^{\infty} \frac{1}{n^k} = \zeta(k) - \sum_{n=1}^{X_{(k)}-1} \frac{1}{n^k},$$

où $X_{(k)} = \max(X_1, \dots, X_k)$ et ζ est la fonction zêta de Riemann, définie sur $]1, \infty[$. Dès que $k \geq 2$, cette somme est finie et la loi a posteriori est bien définie. On a alors, pour tout $n \in \mathbb{N}^*$,

$$\pi(n \mid X_1, \dots, X_k) \propto \frac{1}{n^k} \mathbb{1}_{n \geq X_{(k)}}.$$

Pour essayer de répondre à la question du nombre total de tramways dans la ville, on peut par exemple considérer la moyenne a posteriori $\mathbb{E}[N \mid \mathbf{X}]$ (nous verrons pourquoi dans la suite du cours), qui existe dès que $k \geq 3$ et vaut alors :

$$\mathbb{E}[N \mid \mathbf{X}] = \frac{\sum_{n=X_{(k)}}^{\infty} \frac{1}{n^{k-1}}}{\sum_{n=X_{(k)}}^{\infty} \frac{1}{n^k}} = \frac{\zeta(k-1) - \sum_{n=1}^{X_{(k)}-1} \frac{1}{n^{k-1}}}{\zeta(k) - \sum_{n=1}^{X_{(k)}-1} \frac{1}{n^k}}.$$

Par exemple, lorsque $k = 3$ et $X_{(3)} = 100$, $\mathbb{E}[N \mid \mathbf{X}] \approx 199$.

Remarque : Si $\alpha > 1$ et $n_0 > 1$, l'encadrement série/intégrale (faire un dessin) donne

$$\frac{1}{(\alpha-1)n_0^{\alpha-1}} = \int_{n_0}^{\infty} \frac{1}{x^\alpha} dx \leq \sum_{n=n_0}^{\infty} \frac{1}{n^\alpha} \leq \int_{n_0-1}^{\infty} \frac{1}{x^\alpha} dx = \frac{1}{(\alpha-1)(n_0-1)^{\alpha-1}},$$

d'où, si $k \geq 3$ et $X_{(k)} > 1$,

$$\frac{k-1}{k-2} \left(1 - \frac{1}{X_{(k)}}\right)^{k-2} (X_{(k)} - 1) \leq \frac{\sum_{n=X_{(k)}}^{\infty} \frac{1}{n^{k-1}}}{\sum_{n=X_{(k)}}^{\infty} \frac{1}{n^k}} \leq \frac{k-1}{k-2} \left(1 + \frac{1}{X_{(k)}+1}\right)^{k-2} X_{(k)}.$$

Donc si $X_{(k)} \gg 1$, on obtient

$$\mathbb{E}[N \mid \mathbf{X}] = \frac{\sum_{n=X_{(k)}}^{\infty} \frac{1}{n^{k-1}}}{\sum_{n=X_{(k)}}^{\infty} \frac{1}{n^k}} \approx \frac{k-1}{k-2} X_{(k)}.$$

En revenant à l'exemple numérique ci-dessus où $k = 3$ et $X_{(3)} = 100$, on retrouve bien $\mathbb{E}[N \mid \mathbf{X}] \approx 200$.

EXERCICE 4 (Bayésien empirique)

1. Modèle exponentiel.

(a) La loi marginale de X_1 a pour densité

$$\begin{aligned} f_{\lambda}(x) &= \int_{\Theta} p_{\theta}(x) d\Pi_{\lambda}(\theta) = \int_0^{+\infty} \theta e^{-\theta x} \mathbf{1}_{x \geq 0} \lambda e^{-\lambda \theta} d\theta = \frac{\lambda}{\lambda + x} \mathbf{1}_{x \geq 0} \int_0^{+\infty} \theta(\lambda + x) e^{-(\lambda + x)\theta} d\theta \\ &= \frac{\lambda}{\lambda + x} \mathbf{1}_{x \geq 0} \mathbb{E}[\mathcal{E}(\lambda + x)] = \frac{\lambda}{(\lambda + x)^2} \mathbf{1}_{x \geq 0}. \end{aligned}$$

(b) La densité marginale de $(X_1, \dots, X_n)$ s'écrit, calculée aux points observés,

$$\begin{aligned} \int_{\Theta} \prod_{i=1}^n p_{\theta}(X_i) d\Pi_{\lambda}(\theta) &= \int_0^{\infty} \theta^n e^{-n\bar{X}_n \theta} \lambda e^{-\lambda \theta} d\theta = \frac{\lambda}{\lambda + n\bar{X}_n} \int_0^{\infty} \theta^n (\lambda + n\bar{X}_n) e^{-(\lambda + n\bar{X}_n)\theta} d\theta \\ &= \frac{\lambda}{\lambda + n\bar{X}_n} \mathbb{E}[(\mathcal{E}(\lambda + n\bar{X}_n))^n] = \frac{n! \lambda}{(\lambda + n\bar{X}_n)^{n+1}}. \end{aligned}$$

Remarque : dans l'intégrale ci-dessus, pour laquelle on a utilisé le moment d'ordre n d'une loi exponentielle, on peut aussi passer par la densité d'une loi $\gamma(n+1, \lambda + n\bar{X}_n)$.

(c) Il suffit donc de trouver le point où le maximum de cette densité est atteint, ce qui revient à déterminer le maximum de

$$\psi(\lambda) = \log(n!) + \log \lambda - (n+1) \log(\lambda + n\bar{X}_n).$$

On annule la dérivée et on trouve que $\hat{\lambda} = \bar{X}_n$, après avoir vérifié qu'il s'agit bien d'un maximum (la dérivée est positive puis négative donc c'est bien le cas). La loi a posteriori finale suggérée par la méthode est donc $\Pi_{\hat{\lambda}}[\cdot \mid \mathbf{X}]$. On calcule la densité a posteriori pour tout λ fixé

$$\pi_{\lambda}(\theta \mid \mathbf{X}) \propto \theta^n e^{-(\lambda + n\bar{X}_n)\theta} \mathbf{1}_{\theta \geq 0}.$$

Donc $\Pi_{\lambda}[\cdot \mid \mathbf{X}]$ est une loi $\gamma(n+1, \lambda + n\bar{X}_n)$. Ainsi, la pseudo-loi a posteriori est une loi

$$\gamma(n+1, (n+1)\bar{X}_n).$$

On remarque que la pseudo-moyenne a posteriori est $1/\bar{X}_n$. Elle coïncide donc ici avec l'estimateur du maximum de vraisemblance pour θ dans ce modèle.

2. Modèle Poisson.

(a) La variable X_1 est à valeurs dans $\mathbb{N}$. Pour tout $x \in \mathbb{N}$, sa densité en x par rapport à la mesure de comptage est

$$f_{\lambda}(x) = \int_{\Theta} p_{\theta}(x) d\Pi_{\lambda}(\theta) = \int_0^{+\infty} \frac{\theta^x}{x!} e^{-\theta} \lambda e^{-\lambda \theta} d\theta = \frac{\lambda}{x!} \int_0^{+\infty} \theta^x e^{-(\lambda+1)\theta} d\theta,$$

donc

$$f_{\lambda}(x) = \frac{\lambda}{x!} \frac{\Gamma(x+1)}{(\lambda+1)^{x+1}} = \frac{\lambda}{(\lambda+1)^{x+1}} = \left(\frac{1}{\lambda+1}\right)^x \frac{\lambda}{\lambda+1}.$$

C'est bien une loi géométrique à valeurs dans $\mathbb{N}$ et de paramètre $\lambda/(\lambda+1)$.

(b) De même, la densité marginale de $(X_1, \dots, X_n)$ s'écrit, calculée aux points observés,

$$\begin{aligned} \int \prod_{i=1}^n p_{\theta}(X_i) d\Pi_{\lambda}(\theta) &= \int \frac{\theta^{n\bar{X}_n}}{\prod_{i=1}^n X_i!} e^{-n\theta} \lambda e^{-\lambda\theta} d\theta = \frac{\lambda}{\prod_{i=1}^n X_i!} \int \theta^{n\bar{X}_n} e^{-(n+\lambda)\theta} d\theta \\ &= \frac{\lambda}{\prod_{i=1}^n X_i!} \frac{\Gamma(n\bar{X}_n + 1)}{(\lambda + n)^{1+n\bar{X}_n}}. \end{aligned}$$

(c) Il suffit donc de trouver le point où le maximum de cette densité est atteint, ce qui revient à déterminer le maximum de

$$\psi(\lambda) = \log \lambda - (1 + n\bar{X}_n) \log(\lambda + n).$$

On annule la dérivée et on trouve que, si $\bar{X}_n \neq 0$, alors $\hat{\lambda} = 1/\bar{X}_n$, après avoir vérifié qu'il s'agit bien d'un maximum (la dérivée est positive puis négative donc c'est bien le cas). La loi a posteriori finale suggérée par la méthode est donc $\Pi_{\hat{\lambda}}[\cdot | \mathbf{X}]$. On calcule la densité a posteriori pour tout λ fixé

$$\pi_{\lambda}(\theta | \mathbf{X}) \propto \theta^{n\bar{X}_n} e^{-(n+\lambda)\theta} \mathbf{1}_{\theta \geq 0}.$$

Donc $\Pi_{\lambda}[\cdot | \mathbf{X}]$ est une loi Gamma($n\bar{X}_n + 1, n + \lambda$). On en conclut

$$\Pi_{\hat{\lambda}}[\cdot | \mathbf{X}] = \gamma \left(n\bar{X}_n + 1, \frac{n\bar{X}_n + 1}{\bar{X}_n} \right).$$

On remarque que la pseudo-moyenne a posteriori est $\bar{X}_n$. Elle coïncide donc à nouveau avec l'estimateur du maximum de vraisemblance pour θ dans ce modèle.

3. Modèle gaussien.

(a) **Première méthode :**

Pour calculer l'estimateur du maximum de vraisemblance marginale, on n'a pas besoin de déterminer exactement la densité $f_{\mu}(\mathbf{X}) = f_{\mu}(X_1, \dots, X_n)$, mais seulement à une constante indépendante de μ près.

$$\begin{aligned} f_{\mu}(\mathbf{X}) &= \int \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i - \theta)^2} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta - \mu)^2} d\theta \propto_{\mu, \theta} e^{-\frac{\mu^2}{2}} \int e^{-\frac{n}{2}\theta^2 + n\bar{X}_n\theta - \frac{1}{2}\theta^2 + \theta\mu} d\theta \\ &\propto_{\mu, \theta} e^{-\frac{\mu^2}{2}} \int e^{-\frac{n+1}{2}\left(\theta - \frac{n\bar{X}_n + \mu}{n+1}\right)^2} e^{\frac{(n\bar{X}_n + \mu)^2}{2(n+1)}} d\theta \propto_{\mu} e^{-\frac{\mu^2}{2} + \frac{(n\bar{X}_n + \mu)^2}{2(n+1)}} \propto_{\mu} e^{-\frac{n}{2(n+1)}(\mu - \bar{X}_n)^2}. \end{aligned}$$

Le maximum en μ de la densité marginale de $\mathbf{X}$ est donc atteint pour $\mu = \bar{X}_n$.

Deuxième méthode :

Soient $\boldsymbol{\theta} \sim \mathcal{N}(\mu, 1)$ et $\mathbf{X} | \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}\mathbf{1}, I_n)$. Il apparaît que $\mathbf{X} - \boldsymbol{\theta}\mathbf{1}$ et $\boldsymbol{\theta}$ sont indépendants puisque par définition $\mathbf{X} - \boldsymbol{\theta}\mathbf{1} | \boldsymbol{\theta} \sim \mathcal{N}(0, I_n)$, loi qui ne dépend pas de $\boldsymbol{\theta}$. Ainsi $\mathbf{X} = \boldsymbol{\theta}\mathbf{1} + (\mathbf{X} - \boldsymbol{\theta}\mathbf{1}) \stackrel{\mathcal{L}}{=} \boldsymbol{\theta}\mathbf{1} + \mathbf{Z}$, où $\mathbf{Z} \sim \mathcal{N}(0, I_n)$, $\boldsymbol{\theta}\mathbf{1} \sim \mathcal{N}(\mu\mathbf{1}, J_n)$, où $J_n = \mathbf{1}^t\mathbf{1}$ est la matrice uniquement composée de 1, et $\mathbf{Z} \perp \boldsymbol{\theta}$. Le vecteur $\mathbf{X}$ est donc gaussien de loi $\mathcal{N}(\mu\mathbf{1}, \Sigma)$ avec $\Sigma = I_n + J_n$. Puisque $J_n = nP$, où P est la matrice de projection sur $\text{Vect}(\mathbf{1})$, J_n admet n comme valeur propre simple (associée au vecteur propre $\mathbf{1}$) et 0 comme valeur propre de multiplicité $(n-1)$, donc Σ admet $(n+1)$ comme valeur propre simple (associée au vecteur propre $\mathbf{1}$) et 1 comme valeur propre de multiplicité $(n-1)$. En particulier Σ est inversible et $\Sigma^{-1}\mathbf{1} = \frac{1}{n+1}\mathbf{1}$.

Remarque : On aurait aussi pu simplement affirmer qu'en prenant $\boldsymbol{\theta} \sim \mathcal{N}(\mu, 1)$ indépendant des variables $(Y_1, \dots, Y_n)$ elles-mêmes i.i.d. selon une loi $\mathcal{N}(0, 1)$, alors la construction bayésienne de l'énoncé revient à dire qu'en loi, on a $X_i = \boldsymbol{\theta} + Y_i$ pour tout i . Les $(n+1)$ variables $\boldsymbol{\theta}$ et

$(Y_1, \dots, Y_n)$ étant gaussiennes indépendantes, le vecteur $\mathbf{X}$ est gaussien. Par construction les X_i sont de même loi et de moyenne μ . Par ailleurs, un calcul immédiat donne $\text{Var}(X_i) = 2$ pour tout i , et $\text{Cov}(X_i, X_j) = 1$ pour tout $i \neq j$, ce qui correspond bien à la matrice Σ .

Cherchons maintenant $\hat{\mu}$, l'estimateur du maximum de vraisemblance marginal. Puisque Σ est inversible, $\mathbf{X}$ a une densité f_μ par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$, définie pour tout $x \in \mathbb{R}^n$ par

$$f_\mu(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2} {}^t(x-\mu\mathbf{1})\Sigma^{-1}(x-\mu\mathbf{1})}.$$

Maximiser en μ la vraisemblance $f_\mu(\mathbf{X})$ revient à minimiser

$${}^t(\mathbf{X} - \mu\mathbf{1})\Sigma^{-1}(\mathbf{X} - \mu\mathbf{1}),$$

ce qui équivaut à minimiser, en tenant compte du fait que $\Sigma^{-1}\mathbf{1} = \frac{1}{n+1}\mathbf{1}$,

$$({}^t\mathbf{1}\Sigma^{-1}\mathbf{1})\mu^2 - 2({}^t\mathbf{X}\Sigma^{-1}\mathbf{1})\mu = \frac{n}{n+1}\mu^2 - 2\frac{n}{n+1}\bar{X}_n\mu.$$

On obtient l'estimateur $\hat{\mu} = \bar{X}_n$.

- (b) La loi a posteriori correspondant à la loi a priori $\Pi_\mu = \mathcal{N}(\mu, 1)$ est, par le calcul habituel (voir l'équation (1)), une loi $\mathcal{N}\left(\frac{n\bar{X}_n + \mu}{n+1}, \frac{1}{n+1}\right)$, on obtient comme pseudo-loi a posteriori la loi $\mathcal{N}\left(\bar{X}_n, \frac{1}{n+1}\right)$. On remarque que cette loi est centrée en le maximum de vraisemblance $\bar{X}_n$ pour ce modèle.

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

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

$$\pi(\boldsymbol{\theta} \mid \mathbf{X}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (X_i - \boldsymbol{\theta})' \Sigma^{-1} (X_i - \boldsymbol{\theta}) - \frac{1}{2} (\boldsymbol{\theta} - \mu)' \Lambda^{-1} (\boldsymbol{\theta} - \mu) \right\}.$$

Il suffit de regrouper les termes en $\boldsymbol{\theta}$ pour former une nouvelle forme quadratique dans l'exponentielle, de la forme $-\frac{1}{2}(\boldsymbol{\theta} - m_{\mathbf{X}})' \Sigma_{\mathbf{X}}^{-1} (\boldsymbol{\theta} - m_{\mathbf{X}})$. On cherche donc $m_{\mathbf{X}}, \Sigma_{\mathbf{X}}$ tels que, en développant chaque forme quadratique,

$$\boldsymbol{\theta}'(n\Sigma^{-1} + \Lambda^{-1})\boldsymbol{\theta} - 2(\mu'\Lambda^{-1} + (n\bar{X}_n)'\Sigma^{-1})\boldsymbol{\theta} = \boldsymbol{\theta}'\Sigma_{\mathbf{X}}^{-1}\boldsymbol{\theta} - 2m_{\mathbf{X}}'\Sigma_{\mathbf{X}}^{-1}\boldsymbol{\theta}.$$

En identifiant, il suffit donc de poser

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

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

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

ce qui donne $m_{\mathbf{X}} = \Sigma_{\mathbf{X}}(\Lambda^{-1}\mu + n\Sigma^{-1}\bar{X}_n)$.

2. La famille $\{\mathcal{N}(\mu, \Lambda), \mu \in \mathbb{R}^d, \Lambda \in S_d^{++}(\mathbb{R})\}$ est donc conjuguée pour le modèle $\{\mathcal{N}(\boldsymbol{\theta}, \Sigma)_{\boldsymbol{\theta} \in \mathbb{R}^d}\}$ pour toute matrice $\Sigma \in S_d^{++}(\mathbb{R})$ fixée.

EXERCICE 6 (A priori de Jeffreys)

1. Notons $\psi = \varphi^{-1}$ et $q_\eta = p_{\psi(\eta)}$, soit $p_\theta = q_{\varphi(\theta)}$. L'information de Fisher $\mathbf{J}(\eta)$ dans le modèle paramétré par η s'exprime en fonction de celle $\mathbf{I}(\theta)$ dans le modèle paramétré par θ . En effet, $q'_\eta = \psi'(\eta)p'_{\psi(\eta)}$, donc

$$\mathbf{J}(\eta) = \int_E \frac{q'_\eta(x)^2}{q_\eta(x)} d\mu(x) = \psi'(\eta)^2 \int_E \frac{p'_{\psi(\eta)}(x)^2}{p_{\psi(\eta)}(x)} d\mu(x) = \psi'(\eta)^2 \mathbf{I}(\psi(\eta)).$$

Prenons Π l'a priori de Jeffreys dans le modèle paramétré par θ , de densité $\pi(\theta) = \frac{1}{\Lambda} \sqrt{\mathbf{I}(\theta)}$ avec Λ donné dans l'énoncé. Cherchons la densité de $\Pi \circ \varphi^{-1}$, la mesure image de Π par φ . Soit f une fonction mesurable bornée. Par le théorème de transfert et la définition de Π , on a

$$\int_{\varphi(\Theta)} f(\eta) d\Pi \circ \varphi^{-1}(\eta) = \int_{\Theta} f(\varphi(\theta)) d\Pi(\theta) = \frac{1}{\Lambda} \int_{\Theta} f(\varphi(\theta)) \sqrt{\mathbf{I}(\theta)} d\theta.$$

En effectuant le changement de variable $\eta = \varphi(\theta)$, on obtient

$$\begin{aligned} \frac{1}{\Lambda} \int_{\Theta} f(\varphi(\theta)) \sqrt{\mathbf{I}(\theta)} d\theta &= \frac{1}{\Lambda} \int_{\varphi(\Theta)} f(\eta) \sqrt{\mathbf{I}(\varphi^{-1}(\eta))} |(\varphi^{-1})'(\eta)| d\eta \\ &= \frac{1}{\Lambda} \int_{\varphi(\Theta)} f(\eta) \sqrt{\mathbf{J}(\eta)} d\eta, \end{aligned}$$

où l'on a utilisé l'expression de $\mathbf{J}(\eta)$ donnée plus haut. Ainsi la mesure image $\Pi \circ \varphi^{-1}$ a pour densité $\eta \mapsto \frac{1}{\Lambda} \sqrt{\mathbf{J}(\eta)}$. C'est donc bien l'a priori de Jeffreys dans le modèle paramétré par η .

2. (a) L'information de Fisher est donnée par

$$\forall \theta \in]0, 1[, \mathbf{I}(\theta) = \frac{1}{\theta(1-\theta)}.$$

Ainsi la loi a priori de Jeffreys est celle dont la densité $\pi(\theta)$ est proportionnelle à

$$\pi(\theta) \propto \sqrt{\mathbf{I}(\theta)} = \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}.$$

On reconnaît la densité d'une loi Beta($\frac{1}{2}, \frac{1}{2}$).

- (b) Ici $\mathbf{I}(\theta) = 1$ pour tout $\theta \in \mathbb{R}$. L'a priori de Jeffreys est donc la mesure de Lebesgue sur $\mathbb{R}$. Il s'agit d'un a priori impropre.
(c) Pour tout $\theta \in \mathbb{R}_+^*$, la log-vraisemblance est donnée par

$$\ell_\theta(X) = -\theta + X \ln(\theta) - \ln(X!),$$

et en dérivant par rapport à θ , on obtient le score

$$\ell'_\theta(X) = -1 + \frac{X}{\theta}.$$

Cette variable possède bien un moment d'ordre 2 et l'on a

$$\mathbf{I}(\theta) = \mathbb{E}_\theta [\ell'_\theta(X)^2] = \frac{1}{\theta^2} \text{Var}_\theta(X) = \frac{1}{\theta}.$$

L'a priori de Jeffreys est l'a priori impropre de densité $\theta \mapsto \theta^{-1/2}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}_+^*$.

EXERCICE 7 (★) (Famille de Dirichlet et modèle multinomial)

1. (a) Soit $Z = (Z_1, \dots, Z_K) \sim \text{Dir}(a)$, où $a = (a_1, \dots, a_K)$.

Première méthode. Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée. On commence par remarquer qu'intégrer sur le simplexe $\mathcal{S}_K$, c'est en fait intégrer sur $\{z \in [0, 1]^{K-1}, \sum_{i=1}^{K-1} z_i \leq 1\}$ (et z_K est alors donné par $1 - z_1 - \dots - z_{K-1}$) :

$$\mathbb{E}[\varphi(Z_1)] = \frac{1}{B(a)} \int_{[0,1]^{K-1}} \varphi(z_1) \prod_{i=1}^{K-1} z_i^{a_i-1} \left(1 - \sum_{j=1}^{K-1} z_j\right)^{a_K-1} \mathbb{1}_{\{\sum_{i=1}^{K-1} z_i \leq 1\}} dz_1 \dots dz_{K-1}.$$

En intégrant d'abord sur z_1 puis sur le reste, il vient

$$\begin{aligned} & \int_{[0,1]^{K-1}} \varphi(z_1) \prod_{i=1}^{K-1} z_i^{a_i-1} \left(1 - \sum_{j=1}^{K-1} z_j\right)^{a_K-1} \mathbb{1}_{\{\sum_{i=1}^{K-1} z_i \leq 1\}} dz_1 \dots dz_{K-1} \\ &= \int_0^1 \varphi(z_1) z_1^{a_1-1} \left(\int_{[0,1-z_1]^{K-2}} \prod_{i=2}^{K-1} z_i^{a_i-1} \left(1 - z_1 - \sum_{j=2}^{K-1} z_j\right)^{a_K-1} \mathbb{1}_{\{\sum_{i=2}^{K-1} z_i \leq 1-z_1\}} dz_2 \dots dz_{K-1} \right) dz_1 \end{aligned}$$

et en effectuant, pour z_1 fixé, le changement de variable $u_j = \frac{z_j}{1-z_1}$ pour $j = 2, \dots, K-1$ (de jacobien $(1-z_1)^{K-2}$), on a

$$\begin{aligned} & \int_{[0,1-z_1]^{K-2}} \prod_{i=2}^{K-1} z_i^{a_i-1} \left(1 - z_1 - \sum_{j=2}^{K-1} z_j\right)^{a_K-1} \mathbb{1}_{\{\sum_{i=2}^{K-1} z_i \leq 1-z_1\}} dz_2 \dots dz_{K-1} \\ &= (1-z_1)^{\sum_{j=2}^K a_j-1} \int_{[0,1]^{K-2}} \prod_{i=2}^{K-1} u_i^{a_i-1} \left(1 - \sum_{j=2}^{K-1} u_j\right)^{a_K-1} \mathbb{1}_{\{\sum_{i=2}^{K-1} u_i \leq 1\}} du_2 \dots du_{K-1}. \end{aligned}$$

Enfin, en remarquant que

$$\begin{aligned} & \int_{[0,1]^{K-2}} \prod_{i=2}^{K-1} u_i^{a_i-1} \left(1 - \sum_{j=2}^{K-1} u_j\right)^{a_K-1} \mathbb{1}_{\{\sum_{i=2}^{K-1} u_i \leq 1\}} du_2 \dots du_{K-1} \\ &= B(a_2, \dots, a_K) = \frac{\prod_{j=2}^K \Gamma(a_j)}{\Gamma(\sum_{j=2}^K a_j)}, \end{aligned}$$

on obtient

$$\mathbb{E}[\varphi(Z_1)] = \frac{\Gamma(\sum_{j=1}^K a_j)}{\Gamma(a_1)\Gamma(\sum_{j=2}^K a_j)} \int_0^1 \varphi(z_1) z_1^{a_1-1} (1-z_1)^{\sum_{j=2}^K a_j-1} dz_1.$$

Autrement dit, $Z_1 \sim \text{Beta}(a_1, \sum_{j=2}^K a_j)$.

Seconde méthode. On note $f(z_1)$ la densité de Z_1 et on utilise le symbole $\propto$, sous-entendu $\propto_{z_1}$. Puisque la densité jointe de $z = (z_1, \dots, z_{K-1})$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^{K-1}$ vaut

$$f(z_1, \dots, z_{K-1}) = \frac{1}{B(a)} \prod_{i=1}^{K-1} z_i^{a_i-1} \left(1 - \sum_{j=1}^{K-1} z_j\right)^{a_K-1} \mathbb{1}_{\{z \in [0,1]^{K-1}, \sum_{i=1}^{K-1} z_i \leq 1\}},$$

ou encore

$$f(z_1, \dots, z_{K-1}) \propto z_1^{a_1-1} \prod_{i=2}^{K-1} z_i^{a_i-1} \left(1 - z_1 - \sum_{j=2}^{K-1} z_j\right)^{a_K-1} \mathbb{1}_{\{z \in [0,1]^{K-1}, \sum_{i=2}^{K-1} z_i \leq 1-z_1\}},$$

la marginalisation donne, pour tout $z_1 \in [0, 1]$,

$$f(z_1) \propto z_1^{a_1-1} \int_{[0,1-z_1]^{K-2}} \prod_{i=2}^{K-1} z_i^{a_i-1} \left(1 - z_1 - \sum_{j=2}^{K-1} z_j\right)^{a_{K-1}} \mathbb{1}_{\{\sum_{i=2}^{K-1} z_i \leq 1-z_1\}} dz_2 \dots dz_{K-1}$$

En effectuant, pour z_1 fixé, le changement de variable $u_j = \frac{z_j}{1-z_1}$ pour $j = 2, \dots, K-1$ (de jacobien $(1-z_1)^{K-2}$), on en déduit

$$f(z_1) \propto z_1^{a_1-1} (1-z_1)^{\sum_{j=2}^K a_j-1} \int_{[0,1]^{K-2}} \prod_{i=2}^{K-1} u_i^{a_i-1} \left(1 - \sum_{j=2}^{K-1} u_j\right)^{a_{K-1}} \mathbb{1}_{\{\sum_{i=2}^{K-1} u_i \leq 1\}} du_2 \dots du_{K-1}$$

où le terme intégral a le bon goût de ne plus dépendre de z_1 , donc

$$f(z_1) \propto z_1^{a_1-1} (1-z_1)^{\sum_{j=2}^K a_j-1}$$

ce qui prouve que $Z_1 \sim \text{Beta}(a_1, \sum_{j=2}^K a_j)$.

(b) On se place dans le cadre bayésien :

$$p = (p_1, \dots, p_K) \sim \Pi = \text{Dir}(a), \quad \text{avec } a = (a_1, \dots, a_K) \in (\mathbb{R}_+^*)^K$$

$$\mathbf{X} | p \sim \left(\sum_{k=1}^K p_k \delta_k \right)^{\otimes n}.$$

Autrement dit, conditionnellement à p , les variables $X_1, \dots, X_n$ sont i.i.d. à valeurs dans $\{1, \dots, K\}$, de loi donnée par

$$\forall k \in \{1, \dots, K\}, \mathbb{P}(X_1 = k | p) = p_k.$$

La loi a posteriori s'écrit

$$\begin{aligned} \pi(p | \mathbf{X}) &\propto \prod_{k=1}^K p_k^{a_k-1} \mathbb{1}_{p \in \mathcal{S}_K} \prod_{i=1}^n \prod_{k=1}^K p_k^{\mathbb{1}_{X_i=k}} \\ &\propto \prod_{k=1}^K p_k^{a_k + N_k - 1} \mathbb{1}_{p \in \mathcal{S}_K}, \end{aligned}$$

où l'on a noté $N_k = \sum_{i=1}^n \mathbb{1}_{X_i=k}$ le nombre d'observations égales à k dans l'échantillon. Ainsi $\Pi[p | \mathbf{X}] = \text{Dir}(a_1 + N_1, \dots, a_K + N_K)$. La famille des lois de Dirichlet est donc bien conjuguée pour ce modèle.

2. (a) Pour tout $j \in \{1, \dots, K\}$, la variable N_j suit une loi binomiale $\mathcal{B}(n, p_j)$.

(b) Pour avoir $N_j = n_j$ pour tout j , il faut

- i. choisir une partition de $\{1, \dots, n\}$ en K sous-ensembles $U_1 \cup \dots \cup U_K$ de tailles respectives $n_1, \dots, n_K$. Il y a, par définition, $\binom{n}{n_1, \dots, n_K}$ façons de faire cela.
- ii. pour une partition fixée, la probabilité que les observations $(X_i)_{i \in U_k}$ soient toutes égales à k vaut $p_k^{n_k}$, pour tout $1 \leq k \leq K$.

On note enfin qu'il faut avoir $\sum_{j=1}^K n_j = n$, puisque

$$\sum_j \sum_i \mathbb{1}_{X_i=j} = \sum_i \sum_j \mathbb{1}_{X_i=j} = \sum_i 1 = n.$$

(c) Soit

$$p = (p_1, \dots, p_K) \sim \Pi = \text{Dir}(a), \quad \text{avec } a = (a_1, \dots, a_K) \in (\mathbb{R}_+^*)^K$$

$$N = (N_1, \dots, N_K) \mid p \sim \text{Mult}(n, p_1, \dots, p_K).$$

La loi a posteriori s'écrit

$$\begin{aligned} \pi(p \mid N) &\propto \prod_{k=1}^K p_k^{a_k-1} \mathbb{1}_{p \in \mathcal{S}_K} \prod_{k=1}^K p_k^{N_k} \\ &\propto \prod_{k=1}^K p_k^{a_k+N_k-1} \mathbb{1}_{p \in \mathcal{S}_K}, \end{aligned}$$

Ainsi $\Pi[p \mid N] = \text{Dir}(a_1 + N_1, \dots, a_K + N_K)$. La famille des lois de Dirichlet est donc bien conjuguée pour ce modèle.

EXERCICE 8 (Estimateur de Laplace-Bayes)

1. S'il y a r boules rouges dans l'urne, alors, pour $x \in \{0, \dots, r\}$, on a

$$\mathbb{P}_r(X = x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}.$$

En effet, le nombre total de tirages possibles est $\binom{N}{n}$, et le nombre de tirages qui donnent x boules rouges est $\binom{r}{x} \binom{N-r}{n-x}$ (on choisit x boules parmi les r boules rouges et $n-x$ parmi les $N-r$ boules noires). La loi conditionnelle de X sachant r , notée P_r , est la loi hypergéométrique de paramètre (N, n, r) .

2. La loi conditionnelle de r sachant X est donnée par

$$\pi(r \mid X) = \frac{\frac{1}{N+1} \frac{\binom{r}{X} \binom{N-r}{n-X}}{\binom{N}{n}}}{\sum_{k=0}^N \frac{1}{N+1} \frac{\binom{k}{X} \binom{N-k}{n-X}}{\binom{N}{n}}} = \frac{\binom{r}{X} \binom{N-r}{n-X}}{\sum_{k=0}^N \binom{k}{X} \binom{N-k}{n-X}} = \frac{\binom{r}{X} \binom{N-r}{n-X}}{\binom{N+1}{n+1}}.$$

où l'on a utilisé la formule donnée en indication sous la forme : $\binom{N+1}{n+1} = \sum_{k=1}^{N+1} \binom{k-1}{X} \binom{N+1-k}{n-X}$. En effet, si on considère $N+1$ boules numérotées et on s'intéresse au nombre de façons de tirer (sans prendre en compte l'ordre) $n+1$ boules parmi ces $N+1$ boules. On met les boules ainsi tirées dans l'ordre et on regarde la $X+1$ -ème boules ainsi obtenue. Si celle-ci vaut k , cela veut dire qu'il y a X boules prenant leurs valeurs dans $\{1, \dots, k-1\}$ (et donc $\binom{k-1}{X}$ façons de les choisir) et $N-X$ boules prenant leurs valeurs dans $\{k+1, \dots, N+1\}$ (et donc $\binom{N+1-k}{N-X}$ façons de les choisir), ce qui nous fait $\binom{k-1}{X} \binom{N+1-k}{N-X}$ tirages tels que la $X+1$ -ème plus grande boule vaille k . On obtient ainsi (en sommant sur tous les k possibles)

$$\binom{N+1}{n+1} = \sum_{k=1}^{N+1} \binom{k-1}{X} \binom{N+1-k}{n-X}.$$

Une autre preuve, analytique, est la suivante : on commence par montrer par récurrence que pour tout $n \in \mathbb{N}^*$:

$$\frac{1}{(1-X)^n} = \sum_{j=0}^{\infty} \binom{n+j-1}{j} X^j,$$

puis on développe de deux manières $\frac{1}{(1-X)^{n+1}} \times \frac{1}{(1-X)^{m+1}}$.

3. Notons E l'événement « la $(n+1)^{\text{ième}}$ boule tirée est rouge ». Par la formule des probabilités totales appliquée, on a pour tout $x \leq N$

$$\mathbb{P}(E | X = x) = \sum_{r=x}^N \pi(r | X = x) \mathbb{P}(E | X = x, R = r).$$

Or

$$\mathbb{P}(E | X = x, R = r) = \frac{r-x}{N-n}.$$

Ainsi

$$\begin{aligned} \mathbb{P}(E | X = x) &= \sum_{r=x}^N \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N+1}{n+1}} \cdot \frac{r-x}{N-n} \\ &= \frac{(x+1)}{(N-n) \binom{N+1}{n+1}} \sum_{r=x}^N \binom{r}{x+1} \binom{N-r}{n-x} \\ &= \frac{x+1}{N-n} \cdot \frac{\binom{N+1}{n+2}}{\binom{N+1}{n+1}} \\ &= \frac{x+1}{n+2}. \end{aligned}$$

Puisque cette formule est valable pour tout $x \leq N$, on a donc

$$\mathbb{P}(E | X) = \frac{X+1}{n+2}.$$

EXERCICE 9 (Identifiabilité)

1. Soit $(\varepsilon_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. de loi de Bernoulli $\mathcal{B}(p)$. Alors la variable Y a la même loi que

$$\sum_{i=1}^X \varepsilon_i.$$

On a que $\mathcal{L}(X) = \mathcal{P}(\lambda)$ et X et les ε_i sont indépendantes.

(a) Les variables aléatoires X et Y sont à valeurs dans $\mathbb{N}$, avec par construction $Y \leq X$. De plus, pour tous $0 \leq y \leq x$,

$$\begin{aligned} \mathbb{P}(Y = y | X = x) &= \frac{\mathbb{P}(Y = y \text{ et } X = x)}{\mathbb{P}(X = x)} \\ &= \frac{\mathbb{P}(\sum_{i=1}^x \varepsilon_i = y \text{ et } X = x)}{\mathbb{P}(X = x)} \\ &= \frac{\mathbb{P}(\sum_{i=1}^x \varepsilon_i = y) \mathbb{P}(X = x)}{\mathbb{P}(X = x)} = \mathbb{P}\left(\sum_{i=1}^x \varepsilon_i = y\right), \end{aligned}$$

où l'on a utilisé l'indépendance de X et des ε_i . Or $\sum_{i=1}^x \varepsilon_i$ suit une loi binomiale $\mathcal{B}(x, p)$, donc la loi $\mathcal{L}(Y | X = x)$ est une $\mathcal{B}(x, p)$, i.e.

$$\mathbb{P}(Y = y | X = x) = \mathbb{P}(\mathcal{B}(x, p) = y) = \binom{x}{y} p^y (1-p)^{x-y}.$$

- (b) Le modèle statistique s'écrit $\mathcal{P} = \{P_{\lambda,p}, (\lambda,p) \in \mathbb{R}_*^+ \times]0,1[\}$, où $P_{\lambda,p}$ est la loi du couple (X,Y) . Pour déterminer cette loi, il suffit de se donner $\mathbb{P}(X = x, Y = y)$ pour tous $x, y \in \mathbb{N}$. On a

$$\begin{aligned} \mathbb{P}(X = x, Y = y) &= \mathbb{P}(X = x)\mathbb{P}(Y = y \mid X = x) \\ &= \mathbb{P}(\mathcal{P}(\lambda) = x)\mathbb{P}(\mathcal{B}(x,p) = y) \\ &= \binom{x}{y} p^y (1-p)^{x-y} e^{-\lambda} \frac{\lambda^x}{x!}. \end{aligned}$$

On peut aussi noter que se donner la loi jointe de (X,Y) est équivalent à se donner les lois $\mathcal{L}(Y \mid X)$ et $\mathcal{L}(X)$, ce qui revient à la première identité de la série d'égalités précédente.

Montrons que ce modèle est identifiable. Supposons $P_{\lambda_1,p_1} = P_{\lambda_2,p_2}$. Alors si X, Y suivent cette même loi, on a $\mathbb{E}_{\lambda_1,p_1}[X] = \mathbb{E}[\mathcal{P}(\lambda_1)] = \lambda_1$ et de même $\mathbb{E}_{\lambda_2,p_2}[X] = \mathbb{E}[\mathcal{P}(\lambda_2)] = \lambda_2$. Les deux lois étant les mêmes, les espérances sont égales soit $\lambda_1 = \lambda_2$. Puis notons que

$$\begin{aligned} \mathbb{E}_{\lambda_1,p_1}[Y] &= \mathbb{E}_{\lambda_1,p_1}[\mathbb{E}_{\lambda_1,p_1}[Y \mid X]] \\ &= \mathbb{E}_{\lambda_1,p_1}[\mathbb{E}_{\lambda_1,p_1}[\mathcal{B}(X,p_1) \mid X]] = \mathbb{E}_{\lambda_1,p_1}[Xp_1] = \lambda_1 p_1. \end{aligned}$$

En écrivant ceci sous P_{λ_2,p_2} et en utilisant le fait que les espérances sont égales on obtient $\lambda_1 p_1 = \lambda_2 p_2$. Donc $p_1 = p_2$ car $\lambda_1 = \lambda_2 > 0$. Le modèle est donc identifiable.

- (c) Puisque $\mathbb{E}_{\lambda,p}[X] = \lambda$ et $\mathbb{E}_{\lambda,p}[Y] = \lambda p$, la méthode des moments incite à proposer

$$\begin{aligned} \hat{\lambda}_n &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n \\ \hat{p}_n &= \frac{\bar{Y}_n}{\bar{X}_n} \mathbb{1}_{\bar{X}_n > 0}. \end{aligned}$$

En effet la LFGN implique que $\hat{\lambda}_n$ converge p.s. vers $\mathbb{E}_{\lambda,p}[X] = \lambda > 0$ et que $\bar{Y}_n$ converge p.s. vers $\mathbb{E}_{\lambda,p}[Y] = \lambda p$. Ceci assure que $\hat{p}_n$ converge p.s. vers $\lambda p / \lambda = p$.

Remarque : la méthode du maximum de vraisemblance mène aux mêmes estimateurs. En effet, la vraisemblance de l'échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ en (λ, p) vaut

$$L_n(\lambda, p) = \prod_{i=1}^n \binom{X_i}{Y_i} p^{Y_i} (1-p)^{X_i - Y_i} e^{-\lambda} \frac{\lambda^{X_i}}{X_i!} = C p^{n\bar{Y}_n} (1-p)^{n(\bar{X}_n - \bar{Y}_n)} e^{-n\lambda} \lambda^{n\bar{X}_n},$$

où C ne dépend ni de λ ni de p . La log-vraisemblance s'en déduit :

$$\ell_n(\lambda, p) = \log C + n\bar{Y}_n \log p + n(\bar{X}_n - \bar{Y}_n) \log(1-p) - n\lambda + n\bar{X}_n \log \lambda,$$

de dérivées partielles respectives

$$\frac{\partial \ell_n}{\partial \lambda}(\lambda, p) = n \left(\frac{\bar{X}_n}{\lambda} - 1 \right),$$

et

$$\frac{\partial \ell_n}{\partial p}(\lambda, p) = n \frac{\bar{Y}_n - p\bar{X}_n}{p(1-p)}.$$

On peut noter que si $\bar{X}_n = 0$, alors $\bar{Y}_n = 0$ (par construction du modèle, puisqu'on a toujours $0 \leq \bar{Y}_n \leq \bar{X}_n$), auquel cas la log-vraisemblance vaut $\log C - n\lambda$, maximale en $(\hat{\lambda}_n, \hat{p}_n) = (0, p)$ pour tout $0 < p < 1$. Noter que, en toute rigueur, ce n'est pas un estimateur du maximum de vraisemblance car 0 n'est pas une valeur possible de λ pour une loi de Poisson. Si $\bar{X}_n > 0$ (ce qui arrive p.s. asymptotiquement par la LFGN), l'unique maximum est obtenu en annulant les dérivées partielles ci-dessus, et l'on retrouve le couple $(\hat{\lambda}_n, \hat{p}_n)$ de la méthode des moments.

2. (a) Pour tout $y \in \mathbb{N}$, on a

$$\begin{aligned}
 \mathbb{P}(Y = y) &= \sum_{x=y}^{\infty} \mathbb{P}(X = x, Y = y) = \sum_{x=y}^{\infty} \mathbb{P}(X = x) \mathbb{P}(Y = y \mid X = x) \\
 &= \sum_{x=y}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} e^{-\lambda} \frac{\lambda^x}{x!} \\
 &= e^{-\lambda} \frac{(\lambda p)^y}{y!} \sum_{x=y}^{\infty} \frac{(1-p)^{x-y} \lambda^{x-y}}{(x-y)!} \\
 &= e^{-\lambda} \frac{(\lambda p)^y}{y!} \sum_{k=0}^{\infty} \frac{(1-p)^k \lambda^k}{k!} \\
 &= e^{-\lambda} \frac{(\lambda p)^y}{y!} e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^y}{y!} = \mathbb{P}(\mathcal{P}(\lambda p) = y).
 \end{aligned}$$

Ainsi la loi de Y est une loi de Poisson $\mathcal{P}(\lambda p)$.

(b) Si le paramètre d'intérêt est toujours le couple (λ, p) , le modèle statistique associé s'écrit $\mathcal{P} = \{Q_{\lambda, p}, (\lambda, p) \in \mathbb{R}_*^+ \times]0, 1[\}$, où $Q_{\lambda, p}$ est la loi de Y décrite ci-dessus, c'est-à-dire la loi de Poisson de paramètre λp . Ce modèle n'est pas identifiable. En effet, il suffit de choisir $\lambda_1, \lambda_2, p_1, p_2$ avec $\lambda_1 \neq \lambda_2$ mais $\lambda_1 p_1 = \lambda_2 p_2$ (par exemple $\lambda_1 = 2, p_1 = 1/4$ et $\lambda_2 = 1, p_2 = 1/2$), car alors la loi de Y sous chacune de ces lois est toujours $\mathcal{P}(\lambda_1 p_1) = \mathcal{P}(\lambda_2 p_2)$.

EXERCICE 10 (Lois conditionnelles) Une solution consiste à passer par les fonctions caractéristiques. Puisque X et Y sont indépendantes, on a en effet, pour tout réel u ,

$$\Phi_S(u) = \Phi_X(u) \Phi_Y(u).$$

Il "suffit" donc de connaître les fonctions caractéristiques des lois de X et de Y et de reconnaître celle qui en découle pour S . Dans ce qui suit, nous faisons les calculs de deux manières : via les fonctions caractéristiques d'une part, et d'autre part en calculant directement le produit de convolution des lois de X et de Y .

1. Par les fonctions caractéristiques : pour tout réel u ,

$$\Phi_S(u) = \Phi_X(u) \Phi_Y(u) = \exp(\lambda(e^{iu} - 1)) \exp(\mu(e^{iu} - 1)) = \exp((\lambda + \mu)(e^{iu} - 1))$$

donc $S \sim \mathcal{P}(\lambda + \mu)$.

Produit de convolution : pour tout $s \in \mathbb{N}$, on a

$$\begin{aligned}
 \mathbb{P}(S = s) &= \sum_{k=0}^s \mathbb{P}(X = k) \mathbb{P}(Y = s - k) \\
 &= \sum_{k=0}^s \frac{e^{-(\lambda+\mu)} \lambda^k \mu^{s-k}}{k!(s-k)!} = \frac{e^{-(\lambda+\mu)}}{s!} \sum_{k=0}^s \binom{s}{k} \lambda^k \mu^{s-k} = \frac{e^{-(\lambda+\mu)} (\lambda + \mu)^s}{s!}.
 \end{aligned}$$

Ainsi $S \sim \mathcal{P}(\lambda + \mu)$.

Pour $x, s \in \mathbb{N}$, on a

$$\begin{aligned}
 \mathbb{P}(X = x \mid S = s) &= \frac{\mathbb{P}(X = x) \mathbb{P}(S = s \mid X = x)}{\mathbb{P}(S = s)} = \frac{\mathbb{P}(X = x) \mathbb{P}(Y = s - x)}{\mathbb{P}(S = s)} \\
 &= \binom{s}{x} \frac{\lambda^x \mu^{s-x}}{(\lambda + \mu)^s}.
 \end{aligned}$$

Ainsi $\mathcal{L}(X | S) = \mathcal{B}\left(S, \frac{\lambda}{\lambda + \mu}\right)$ et $\mathbb{E}[X | S] = \frac{\lambda}{\lambda + \mu}S$.

2. Par les fonctions caractéristiques : pour tout réel u

$$\Phi_S(u) = \Phi_X(u)\Phi_Y(u) = (1 - iu/\lambda)^{-r}(1 - iu/\lambda)^{-t} = (1 - iu/\lambda)^{-(r+t)}$$

donc $S \sim \gamma(r+t, \lambda)$.

Produit de convolution : notons respectivement f_X , f_Y et f_S les densités en jeu. Puisque $S \geq 0$, le produit de convolution des densités s'écrit, pour tout $s \geq 0$,

$$f_S(s) = \int_0^s f_X(x)f_Y(s-x) dx = \int_0^s \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} \frac{\lambda^t (s-x)^{t-1} e^{-\lambda(s-x)}}{\Gamma(t)} dx$$

donc, via le changement de variable $u = x/s$,

$$f_S(s) = \frac{\lambda^{r+t} s^{r+t-1} e^{-\lambda s}}{\Gamma(r)\Gamma(t)} \int_0^1 u^{r-1} (1-u)^{t-1} du = \frac{\lambda^{r+t} s^{r+t-1} e^{-\lambda s}}{\Gamma(r+t)}$$

car on a reconnu dans l'intégrale la densité non normalisée d'une loi Beta(r, t). Ainsi $S \sim \gamma(r+t, \lambda)$.

Déterminons maintenant la densité de $S|X$. Soit φ une fonction borélienne bornée. Pour tout x ,

$$\begin{aligned} \mathbb{E}[\varphi(S) | X = x] &= \mathbb{E}[\varphi(x+Y) | X = x] \\ &= \mathbb{E}[\varphi(x+Y)] \\ &= \int_{\mathbb{R}} \varphi(x+y) f_Y(y) dy \\ &= \int_{\mathbb{R}} \varphi(s) f_Y(s-x) ds \end{aligned}$$

où l'on a utilisé l'indépendance de X et Y . Ainsi, $S|X = x$ a pour densité $s \mapsto f_Y(s-x)$.

Autre méthode : pour déterminer la densité de $S|X$, on commence par celle du couple (X, S) . Soit φ une fonction borélienne bornée de $\mathbb{R}^2$ dans $\mathbb{R}$, alors l'indépendance de X et Y permet d'écrire, via un changement de variable,

$$\mathbb{E}[\varphi(X, S)] = \mathbb{E}[\varphi(X, X+Y)] = \iint \varphi(x, x+y) f_X(x) f_Y(y) dx dy$$

et le changement de variable $s = x + y$ donne

$$\mathbb{E}[\varphi(X, S)] = \iint \varphi(x, s) f_X(x) f_Y(s-x) dx ds,$$

ce qui montre que la densité jointe du couple (X, S) est $f_X(x)f_Y(s-x)$.

À présent, pour $s > 0$, la densité conditionnelle $f_{X|S=s}$ de X sachant $S = s$ est donnée, pour $x > 0$, par :

$$f_{X|S=s}(x) \propto_x f_X(x) f_Y(s-x) \propto_x x^{r-1} (s-x)^{t-1} \mathbf{1}_{0 \leq x \leq s} \propto_x \left(\frac{x}{s}\right)^{r-1} \left(1 - \frac{x}{s}\right)^{t-1} \mathbf{1}_{0 \leq \frac{x}{s} \leq 1}.$$

Cela correspond à la densité d'une variable s Beta(r, t). Ainsi $X | S \sim S \text{Beta}(r, t)$ et $\mathbb{E}[X | S] = \frac{r}{r+t}S$.

3. Par les fonctions caractéristiques : si $X \sim \mathcal{B}(n, p)$, alors $\Phi_X(u) = ((1-p) + pe^{iu})^n$ pour tout réel u , d'où

$$\Phi_S(u) = \Phi_X(u)\Phi_Y(u) = ((1-p) + pe^{iu})^n ((1-p) + pe^{iu})^m = ((1-p) + pe^{iu})^{n+m}$$

donc $S \sim \mathcal{B}(n + m, p)$.

Produit de convolution : pour tout $s \in \llbracket 0, n + m \rrbracket$, on a

$$\begin{aligned} \mathbb{P}(S = s) &= \sum_{k=0}^s \mathbb{P}(X = k) \mathbb{P}(Y = s - k) = \sum_{k=0}^s \binom{n}{k} \binom{m}{s-k} p^k (1-p)^{n-k} p^{s-k} (1-p)^{m-s+k} \\ &= \sum_{k=0}^s \binom{n}{k} \binom{m}{s-k} p^s (1-p)^{n+m-s} = \binom{n+m}{s} p^s (1-p)^{n+m-s} \end{aligned}$$

où l'on a utilisé l'identité de Vandermonde, à savoir que

$$\sum_{k=0}^s \binom{n}{k} \binom{m}{s-k} = \binom{n+m}{s},$$

laquelle peut se déduire des deux façons d'obtenir le coefficient de X^s dans le polynôme

$$(1 + X)^{n+m} = (1 + X)^n (1 + X)^m.$$

Ainsi $S \sim \mathcal{B}(n + m, p)$.

Pour arriver au même résultat, une méthode plus rapide consiste à considérer des variables $(B_i)_{1 \leq i \leq n+m}$ i.i.d. selon la loi de Bernoulli $\mathcal{B}(p)$ et à voir que, en loi, on a $X = B_1 + \dots + B_n$ et $Y = B_{n+1} + \dots + B_{n+m}$, de sorte que $S = X + Y = \sum_{i=1}^{n+m} B_i \sim \mathcal{B}(n + m, p)$. Pour $x, s \in \llbracket 0, n + m \rrbracket$, on a

$$\mathbb{P}(X = x \mid S = s) = \frac{\mathbb{P}(X = x) \mathbb{P}(Y = s - x)}{\mathbb{P}(S = s)} = \frac{\binom{n}{x} \binom{m}{s-x}}{\binom{n+m}{s}}.$$

On reconnaît la loi hypergéométrique de paramètres $(N, p, n) = (n + m, n/(n + m), s)$. On a donc par l'indication de l'énoncé : $\mathbb{E}[X \mid S] = \frac{n}{n+m} S$. Ce résultat peut se retrouver comme suit :

$$\mathbb{E}[X \mid S] = \sum_{x=0}^S \frac{x \binom{n}{x} \binom{m}{S-x}}{\binom{n+m}{S}} = \frac{1}{\binom{n+m}{S}} \sum_{x=0}^{S-1} n \binom{n-1}{x} \binom{m}{S-x-1} = \frac{n \binom{n+m-1}{S-1}}{\binom{n+m}{S}} = \frac{n}{n+m} S$$

où l'on a à nouveau appliqué l'identité de Vandermonde.