Feuille 4 : correction

Exercice 1 (Distance en variation totale)

1. Posons $\Lambda = \{x \in E,\, p_x > q_x\}$ et soit A une partie quelconque de E. On a
$$P(A) - Q(A) \le P(A \cap \Lambda) - Q(A \cap \Lambda) \le P(\Lambda) - Q(\Lambda)\,,$$
et
$$Q(A) - P(A) \le Q(\Lambda^c) - P(\Lambda^c) = P(\Lambda) - Q(\Lambda).$$
Ainsi le supremum est atteint en Λ et
$$d_{\mathrm{VT}}(P, Q) = P(\Lambda) - Q(\Lambda) = \sum_{x \in E} (p_x - q_x)_+\,.$$
2. Il suffit d'écrire $(p_x - q_x)_+ = p_x - \min(p_x, q_x)$.
3. Soit (X, Y) un couplage de P et Q. Pour tout $A \subset E$, on a
$$\begin{aligned} P(A) - Q(A) &= \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \\ &\le \mathbb{P}(X \in A) - \mathbb{P}(Y \in A, X \in A) \\ &= \mathbb{P}(X \in A, Y \notin A) \le \mathbb{P}(X \ne Y)\,. \end{aligned}$$
Soit maintenant (X, Y) un couple de variables aléatoires dont la loi est donnée par : pour tout $(x, y) \in E^2$,
$$\mathbb{P}(X = x, Y = y) = \min(p_x, q_x)\mathbb{1}_{x=y} + \frac{(p_x - q_x)_+(q_y - p_y)_+}{d_{\mathrm{VT}}(P, Q)}\,.$$
Pour tout $x \in E$, on a
$$\mathbb{P}(X = x) = \sum_{y \in E} \mathbb{P}(X = x, Y = y) = \min(p_x, q_x) + (p_x - q_x)_+ = p_x\,.$$
Ainsi X est bien de loi P, et de même Y est de loi Q. De plus, $\mathbb{P}(X \ne Y) = d_{\mathrm{VT}}(P, Q)$.
4. Soient $(X_1, Y_1), \ldots, (X_n, Y_n)$ des couples i.i.d. selon le couplage optimal de P et Q (celui de la question précédente). Alors on a
$$\begin{aligned} d_{\mathrm{VT}}(P^{\otimes n}, Q^{\otimes n}) &\le \mathbb{P}\left((X_1, \ldots, X_n) \ne (Y_1, \ldots, Y_n)\right) \\ &\le \sum_{i=1}^n \mathbb{P}(X_i \ne Y_i) \\ &= n d_{\mathrm{VT}}(P, Q)\,. \end{aligned}$$
5. Soient $(X_1, Y_1), \ldots, (X_n, Y_n)$ des couples i.i.d. avec $X_i \sim \mathcal{B}\left(\frac{\lambda}{n}\right)$ et $Y_i \sim \mathcal{P}\left(\frac{\lambda}{n}\right)$, et (X_i, Y_i) distribué selon le couplage optimal entre $\mathcal{B}\left(\frac{\lambda}{n}\right)$ et $\mathcal{P}\left(\frac{\lambda}{n}\right)$. Alors $\sum_{i=1}^n X_i$ est de loi P et $\sum_{i=1}^n Y_i$ est de loi Q. On a
$$\mathbb{P}\left(\sum_{i=1}^n X_i \ne \sum_{i=1}^n Y_i\right) \le \sum_{i=1}^n \mathbb{P}(X_i \ne Y_i) = n d_{\mathrm{VT}}\left(\mathcal{B}\left(\frac{\lambda}{n}\right), \mathcal{P}\left(\frac{\lambda}{n}\right)\right)\,.$$
Or
$$d_{\mathrm{VT}}\left(\mathcal{B}\left(\frac{\lambda}{n}\right), \mathcal{P}\left(\frac{\lambda}{n}\right)\right) = \sum_{k \in \mathbb{N}} \left(\mathbb{P}(X_i = k) - \mathbb{P}(Y_i = k)\right)_+\,.$$
Tous les termes de la somme ci-dessus sont nuls sauf pour $k = 1$. Ainsi $\mathbb{P}(X_i \ne Y_i) = \frac{\lambda}{n}\left(1 - e^{-\lambda/n}\right)$.

Exercice 2 (Théorème de De Finetti)

1. On utilise les résultats de l'exercice précédent sur la distance en variation totale, et notamment la question 3. Pour majorer $d_{\mathrm{VT}}(P,Q)$, on cherche à construire un couplage qui soit tel que les variables coïncident le plus souvent possible. Soit $(Y_i)_{i=1}^{+\infty}$ une suite i.i.d. de tirages uniformes dans $\{1,\ldots,N\}$ et soit τ_i le temps d'apparition de la $i^{\text{ème}}$ nouvelle valeur dans cette suite. Alors, $Y=(Y_1,\ldots,Y_n)$ est de loi Q et $X=(Y_{\tau_1},\ldots,Y_{\tau_n})$ est de loi P. De plus,
$$\mathbb{P}(X\neq Y)=\delta(n,N)\leq \mathbb{P}\left(\exists 1\leq i<j\leq n,\, Y_i=Y_j\right)\leq \sum_{i<j}\mathbb{P}(Y_i=Y_j)=\frac{1}{N}\binom{n}{2}.$$
Une autre façon de le voir est de remarquer que $P=Q_{|A}$, où $Q_{|A}$ est la loi Q conditionnée à l'évènement $A=\{$ tous les tirages sont distincts$\}$, et l'on montre facilement que $d_{\mathrm{VT}}(Q,Q_{|A})\leq Q(A^c)$.
2. Par échangeabilité, la loi conditionnelle de $(X_1,\ldots,X_n)$ sachant $\boldsymbol{\theta}_N$ (i.e. sachant le nombre de 1 apparus entre 1 et N) correspond à la loi de n tirages sans remise dans une urne contenant $N\boldsymbol{\theta}_N$ boules noires et $N(1-\boldsymbol{\theta}_N)$ boules blanches. Par la question précédente, la distance en variation totale entre un échantillon avec et sans remise est inférieure à $\frac{1}{N}\binom{n}{2}$.
3. Toujours par échangeabilité, la loi conditionnelle de $2^k\boldsymbol{\theta}_{2^k}$ sachant $\boldsymbol{\theta}_{2^{k+1}}$ est celle du nombre de boules noires tirées après n tirages sans remise dans une urne contenant $2^{k+1}\boldsymbol{\theta}_{2^{k+1}}$ boules noires et $2^{k+1}(1-\boldsymbol{\theta}_{2^{k+1}})$ boules blanches. Par l'exercice 3 du TD1 (avec $a_i=1$ si la boule i est noire, 0 si elle est blanche), l'inégalité de Hoeffding s'applique et l'on obtient
$$\mathbb{P}\left(|\boldsymbol{\theta}_{2^k}-\boldsymbol{\theta}_{2^{k+1}}|>\varepsilon \;\middle|\; \boldsymbol{\theta}_{2^{k+1}}\right)\leq 2e^{-\frac{2\cdot(2^k\varepsilon)^2}{2^k}}=2e^{-2^{k+1}\varepsilon^2}.$$
Il suffit alors de prendre l'espérance pour obtenir l'inégalité demandée. En prenant $\varepsilon=\frac{1}{k^2}$, on obtient alors
$$\sum_{k=1}^{+\infty}\mathbb{P}\left(|\boldsymbol{\theta}_{2^k}-\boldsymbol{\theta}_{2^{k+1}}|>\frac{1}{k^2}\right)<+\infty\,.$$
Par le lemme de Borel–Cantelli, avec probabilité 1, l'événement $|\boldsymbol{\theta}_{2^k}-\boldsymbol{\theta}_{2^{k+1}}|>\frac{1}{k^2}$ ne se réalise qu'un nombre fini de fois. En particulier,
$$\sum_{k=1}^{+\infty}|\boldsymbol{\theta}_{2^k}-\boldsymbol{\theta}_{2^{k+1}}|<+\infty \qquad \text{p.s.}\,,$$
ce qui implique que la suite $(\boldsymbol{\theta}_{2^k})_{k\geq 1}$ est p.s. de Cauchy, donc converge p.s. vers une variable aléatoire notée $\boldsymbol{\theta}$.
4. En prenant $N=2^k$ et en utilisant simplement la convergence en loi de $\boldsymbol{\theta}_N$ vers $\boldsymbol{\theta}$ quand $k\to+\infty$ et le fait que la fonction considérée est continue bornée sur $[0,1]$), on a
$$\mathbb{E}\left[\prod_{i=1}^n \boldsymbol{\theta}_N^{x_i}(1-\boldsymbol{\theta}_N)^{1-x_i}\right]\underset{k\to+\infty}{\to}\mathbb{E}\left[\prod_{i=1}^n \boldsymbol{\theta}^{x_i}(1-\boldsymbol{\theta})^{1-x_i}\right].$$
Remarquons maintenant que l'inégalité de la question 2 implique, par l'inégalité triangulaire, que
$$\left|\mathbb{P}\left(X_1=x_1,\ldots,X_n=x_n\right)-\mathbb{E}\left[\prod_{i=1}^n \boldsymbol{\theta}_N^{x_i}(1-\boldsymbol{\theta}_N)^{1-x_i}\right]\right|\leq\frac{1}{N}\binom{n}{2}.$$
En faisant tendre k vers $+\infty$ dans cette inégalité, on obtient
$$\mathbb{P}\left(X_1=x_1,\ldots,X_n=x_n\right)=\mathbb{E}\left[\prod_{i=1}^n \boldsymbol{\theta}^{x_i}(1-\boldsymbol{\theta})^{1-x_i}\right].$$
Il ne reste plus qu'à montrer que $\boldsymbol{\theta}_n \underset{n\to\infty}{\longrightarrow}\boldsymbol{\theta}$ p.s.. On a montré que la loi de $(X_1,\ldots,X_n)$ était $\mathcal{B}(\boldsymbol{\theta})^{\otimes n}$ avec $\boldsymbol{\theta}$ aléatoire. Ainsi, par l'inégalité de Hoeffding, pour tout $\varepsilon>0$,
$$\mathbb{P}\left(|\boldsymbol{\theta}_n-\boldsymbol{\theta}|>\varepsilon\right)=\mathbb{E}\left[\mathbb{P}\left(|\boldsymbol{\theta}_n-\boldsymbol{\theta}|>\varepsilon \;\middle|\; \boldsymbol{\theta}\right)\right]\leq 2e^{-2n\varepsilon^2}.$$
Par Borel–Cantelli, on a bien $\boldsymbol{\theta}_n \underset{n\to\infty}{\longrightarrow}\boldsymbol{\theta}$ p.s.

Exercice 3 (Urnes de Polya)

1. Soit $(i_1, \dots, i_n) \in \{0,1\}^n$ et notons $x_n = \sum_j i_j$ le nombre de 1 dans la suite. Lorsque l'on écrit la probabilité d'obtenir $(i_1, \dots, i_n)$ comme produit des probabilités de chaque tirage sachant les tirages précédents, le numérateur est égal à
$$(a+b)(a+b+1)\dots(a+b+n-1)$$
le $k^{\text{ième}}$ terme de ce produit correspondant au nombre total de boules dans l'urne à l'instant du $k^{\text{ième}}$ tirage. Quant au numérateur, quitte à réordonner les termes, il est égal à
$$a(a+1)\dots(a+x_n-1).b(b+1)\dots(b+n-x_n-1)\,.$$
En effet, au moment où l'on tire la $k^{\text{ième}}$ boule noire, il y a $a+k-1$ boules noires dans l'urne, et au moment où l'on tire la $k^{\text{ième}}$ boule blanche, il y a $b+k-1$ boules blanches dans l'urne. La loi de $(I_1, \dots, I_n)$ ne dépend donc pas de l'ordre dans lequel on tire les boules mais uniquement du nombre de boules noires tirées. Autrement dit, elle est échangeable.

2. Par la question précédente, on a
$$\begin{aligned}\mathbb{P}(X_n = k) &= \binom{n}{k}\frac{a(a+1)\dots(a+k-1).b(b+1)\dots(b+n-k-1)}{(a+b)(a+b+1)\dots(a+b+n-1)} \\ &= \binom{n}{k}\frac{\Gamma(a+k)\Gamma(b+n-k)\Gamma(a+b)}{\Gamma(a)\Gamma(b)\Gamma(a+b+n)} \\ &= \binom{n}{k}\frac{\mathrm{B}(a+k, b+n-k)}{\mathrm{B}(a,b)}\,.\end{aligned}$$

3. Dans le scénario bayésien $\boldsymbol{\theta} \sim \mathrm{Beta}(a,b)$ et $Z_n \mid \boldsymbol{\theta} \sim \mathrm{Bin}(n, \boldsymbol{\theta})$, 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.

4. Par le théorème de De Finetti, on sait que $\frac{X_n}{n}$ converge presque sûrement vers une variable aléatoire $\boldsymbol{\theta}$ et la question précédente montre que cette variable limite est de loi $\mathrm{Beta}(a,b)$. Ainsi
$$\boldsymbol{\theta}_n = \frac{X_n + a}{n+a+b} = \frac{n}{n+a+b}\cdot\frac{X_n}{n} + \frac{a}{n+a+b} \xrightarrow{\text{p.s.}} \boldsymbol{\theta}\,.$$

Exercice 4 (Lois conjuguées)

1. Soit $\boldsymbol{\theta} \sim \Pi = \mathcal{N}(\mu, \sigma^2)$ et $\mathbf{X} = (X_1, \dots, X_n) \mid \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}, 1)^{\otimes n}$. Montrons que la loi a posteriori $\Pi[\cdot \mid \mathbf{X}]$ est de la forme $\mathcal{N}(\mu', \sigma'^2)$. On a
$$d\Pi(\theta) = \pi(\theta)d\theta, \qquad \pi(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(\theta-\mu)^2}.$$
$$dP_\theta(x) = p_\theta(x)dx, \qquad p_\theta(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2}.$$

La densité a posteriori s'écrit

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

Ainsi $\Pi[\cdot \mid \mathbf{X}] = \mathcal{N}\left(\frac{\mu\sigma^{-2}+n\overline{X}_n}{\sigma^{-2}+n}, \frac{1}{\sigma^{-2}+n}\right)$. Donc la famille considérée est conjuguée. La moyenne a posteriori vaut $\mathbb{E}[\boldsymbol{\theta} \mid \mathbf{X}] = \frac{\mu\sigma^{-2}+n\overline{X}_n}{\sigma^{-2}+n}$.

2. On part de $\Pi = \mathrm{Beta}(a,b)$. Dans le modèle de Bernoulli $\mathcal{B}(\theta)^{\otimes n}$, la loi a posteriori $\Pi[\cdot \mid \mathbf{X}]$ a pour densité

$$
\begin{aligned}
\pi(\theta \mid \mathbf{X}) &\propto \pi(\theta) \prod_{i=1}^{n} p_\theta(X_i) \\
&\propto \theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{0\le\theta\le1} \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} \\
&\propto \theta^{a+n\overline{X}_n-1}(1-\theta)^{b+n-n\overline{X}_n-1}\mathbb{1}_{0\le\theta\le1}
\end{aligned}
$$

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

3. Soit $\boldsymbol{\theta} \sim \Pi = \mathrm{Gamma}(a,b)$ et $\mathbf{X} = (X_1,\dots,X_n) \mid \boldsymbol{\theta} \sim \mathrm{Gamma}(p,\boldsymbol{\theta})^{\otimes n}$, pour $p>0$ fixé. La loi a posteriori $\Pi[\cdot \mid \mathbf{X}]$ a pour densité

$$
\begin{aligned}
\pi(\theta \mid \mathbf{X}) &\propto \pi(\theta) \prod_{i=1}^{n} p_\theta(X_i) \\
&\propto \theta^{a-1}\mathrm{e}^{-b\theta}\mathbb{1}_{\theta\ge0} \prod_{i=1}^{n} \theta^{p} e^{-\theta X_i}\mathbb{1}_{X_i\ge0} \\
&\propto \theta^{np+a-1}\mathrm{e}^{-(b+n\overline{X}_n)\theta}\mathbb{1}_{\theta\ge0}.
\end{aligned}
$$

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

4. Soit $\boldsymbol{\theta} \sim \Pi = \mathrm{Gamma}(a,b)$ et $\mathbf{X} = (X_1,\dots,X_n) \mid \boldsymbol{\theta} \sim \mathcal{P}(\boldsymbol{\theta})^{\otimes n}$. La loi a posteriori $\Pi[\cdot \mid \mathbf{X}]$ a pour densité

$$
\begin{aligned}
\pi(\theta \mid \mathbf{X}) &\propto \pi(\theta) \prod_{i=1}^{n} p_\theta(X_i) \\
&\propto \theta^{a-1}\mathrm{e}^{-b\theta}\mathbb{1}_{\theta\ge0} \prod_{i=1}^{n} e^{-\theta}\theta^{X_i} \\
&\propto \theta^{a+n\overline{X}_n-1}\mathrm{e}^{-(b+n)\theta}\mathbb{1}_{\theta\ge0}.
\end{aligned}
$$

Ainsi $\Pi[\cdot \mid \mathbf{X}] = \mathrm{Gamma}(a+n\overline{X}_n, b+n)$. La moyenne a posteriori vaut $\frac{a+n\overline{X}_n}{b+n}$.

Exercice 5 (Famille gaussienne à moyenne et variance inconnues)

1. La ≪ densité ≫ jointe de $(\mu, \sigma^2, \mathbf{X})$ 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 $s_{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2$. L'intégrale de la quantité précédente par rapport à (μ, σ^2) est finie. En effet

$$\begin{aligned} &\int_{\mathbb{R}_+^*} (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}} \int_{\mathbb{R}} e^{-\frac{n}{2\sigma^2}(\mu-\overline{X}_n)^2} d\mu d\sigma^2 \\ &\quad = \sqrt{\frac{2\pi}{n}} \int_{\mathbb{R}_+^*} (\sigma^2)^{-\frac{n+1}{2}} e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}} d\sigma^2 \\ &\quad = \sqrt{\frac{2\pi}{n}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\left(\frac{ns_{\mathbf{X}}}{2}\right)^{\frac{n-1}{2}}}, \end{aligned}$$

où l'on a reconnu la densité d'une variable IG $\left(\frac{n-1}{2}, \frac{ns_{\mathbf{X}}}{2}\right)$ (il faut $n \geq 2$ pour que cela soit bien défini).

2. La densité de la loi $\mathcal{L}(\sigma^2 \mid \mathbf{X})$ est proportionnelle à

$$\begin{aligned} \int_{\mathbb{R}} p_{\mu,\sigma^2}(\mathbf{X})\sigma^{-2} d\mu &\propto (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}} \sqrt{2\pi\frac{\sigma^2}{n}} \\ &\propto (\sigma^2)^{-\frac{n+1}{2}} e^{-\frac{ns_{\mathbf{X}}}{2\sigma^2}}. \end{aligned}$$

On reconnait la densité d'une loi IG$(\frac{n-1}{2}, \frac{ns_{\mathbf{X}}}{2})$.

3. La densité de la loi $\mathcal{L}(\mu \mid \sigma^2, \mathbf{X})$ est la densité en μ proportionnelle à la quantité

$$p_{\mu,\sigma^2}(\mathbf{X})\sigma^{-2} \propto \exp\left\{-\frac{n}{2\sigma^2}(\mu-\overline{X}_n)^2\right\}.$$

On en déduit $\mathcal{L}(\mu \mid \sigma^2, \mathbf{X}) = \mathcal{N}(\overline{X}_n, \sigma^2/n)$, et on a vu plus haut que $\mathcal{L}(\sigma^2 \mid \mathbf{X}) = \mathrm{IG}(\frac{n-1}{2}, \frac{ns_{\mathbf{X}}}{2})$.

4. Pour déterminer une région de crédibilité pour μ, il suffit de déterminer la loi $\mathcal{L}(\mu \mid \mathbf{X})$ et de prendre une région déduite, par exemple, des quantiles a posteriori. Pour déterminer $\mathcal{L}(\mu \mid \mathbf{X})$, on procède comme pour celle de $\mathcal{L}(\sigma^2 \mid \mathbf{X})$: 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

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

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 $b \in \mathbb{R}^*$ sont deux constantes, alors $\frac{Z-a}{b}$ a pour densité $u \mapsto bg(bu+a)$, on obtient

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

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 le quantile d'ordre $\alpha/2$ puisque la loi de Student est symétrique), on en conclut que

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

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 $1-\alpha$.

Remarque. Cet intervalle de crédibilité coïncide avec l'intervalle de confiance utilisé en statistiques fréquentistes. En effet, si $X_1,\ldots,X_n \sim \mathcal{N}(\mu_0,\sigma_0^2)^{\otimes n}$, par définition de la loi de Student, $\frac{\sqrt{n-1}(\overline{X}_n-\mu_0)}{\sqrt{s_{\mathbf{X}}}}$ suit une loi de Student à $n-1$ degrés de liberté et donc

$$\mathbb{P}_{\mu_0,\sigma_0}\left(\mu_0 \in \left[\overline{X}_n \pm \sqrt{\frac{s_{\mathbf{X}}}{n-1}}z_\alpha\right]\right) = 1-\alpha,$$

donc $I(\mathbf{X})$ est aussi un intervalle de confiance pour μ au niveau $1-\alpha$.

Exercice 6 (De Bayes de risque constant implique minimax)

1. Soit T un estimateur de Bayes pour Π et une fonction de perte ℓ, de risque constant (i.e. pour tout $\theta\in\Theta$, $\mathbf{R}(\theta,T) = \mathbf{R}_{\max}(T)$). Si T n'était pas minimax, on pourrait trouver un estimateur T' tel que

$$\mathbf{R}_B(\Pi,T') \leq \mathbf{R}_{\max}(T') < \mathbf{R}_{\max}(T).$$

   Mais comme T est de risque constant par hypothèse, $\mathbf{R}_{\max}(T) = \mathbf{R}_B(\Pi,T)$, ce qui implique avec l'inégalité ci-dessus que T ne peut pas être de Bayes pour Π. Contradiction : T est donc minimax.

2. (a) La formule de Bayes donne, pour tous $a,b>0$,

$$\pi_{a,b}(\theta \mid \mathbf{X}) \propto \theta^{a-1}(1-\theta)^{b-1}\theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n} \propto \theta^{a+n\overline{X}_n-1}(1-\theta)^{b+n-n\overline{X}_n-1}.$$

   La loi a posteriori est une loi $\mathrm{Beta}(a+n\overline{X}_n, b+n-n\overline{X}_n)$. Un estimateur de Bayes $\widehat{\theta}_{a,b}(\mathbf{X})$ pour $\Pi_{a,b}$ et la perte quadratique est la moyenne a posteriori $\int \theta d\Pi_{a,b}(\theta \mid \mathbf{X})$. En utilisant que l'espérance d'une loi $\mathrm{Beta}(a,b)$ est égale à $\frac{a}{a+b}$, on obtient

$$\widehat{\theta}_{a,b}(\mathbf{X}) = \frac{a+n\overline{X}_n}{a+b+n}.$$

   (b) Par la question 1., un estimateur de Bayes de risque constant est minimax. Il suffit donc de trouver, parmi les estimateurs $\left(\widehat{\theta}_{a,b}(\mathbf{X})\right)_{a,b>0}$ (qui sont de Bayes), un estimateur qui soit de risque constant. L'énoncé suggère de poser $a=b$, on calcule donc le risque quadratique de l'estimateur $\widehat{\theta}_{a,a}(\mathbf{X})$,

$$\begin{aligned}\mathbb{E}_\theta\left[(\widehat{\theta}_{a,a}(\mathbf{X})-\theta)^2\right] &= \mathbb{E}_\theta\left[\left(\frac{n(\overline{X}_n-\theta)}{2a+n} - \frac{2a\theta-a}{2a+n}\right)^2\right] \\ &= \frac{n^2\,\mathrm{Var}_\theta(\overline{X}_n)}{(2a+n)^2} + \frac{a^2(2\theta-1)^2}{(2a+n)^2} \\ &= (2a+n)^{-2}\left(n\theta(1-\theta)+a^2(2\theta-1)^2\right) \\ &= (2a+n)^{-2}\left((4a^2-n)\theta^2+(n-4a^2)\theta+a^2\right).\end{aligned}$$

   On en conclut que si l'on choisit a tel que $4a^2=n$, soit

$$a_n = \frac{\sqrt{n}}{2},$$

   l'estimateur $\widehat{\theta}_{a_n,a_n}(\mathbf{X})$ est de risque constant et de Bayes donc minimax.

(c) Pour savoir si T est minimax, il suffit de comparer $\mathbf{R}_{\max}(T)$ au risque minimax $\mathbf{R}_M$. Or on peut calculer ce dernier car on connait un estimateur minimax, $\widehat{\theta}_{a_n,a_n}(\mathbf{X})$. Le risque maximal de ce dernier est, d'après la question précédente, puisqu'il est de risque constant, égal à

$$\mathbf{R}_{\max}(\widehat{\theta}_{a_n,a_n}(\mathbf{X})) = \frac{a_n^2}{(2a_n+n)^2} = \frac{n}{4(n+\sqrt{n})^2}.$$

Or, le risque maximal de T vaut

$$\mathbf{R}_{\max}(T) = \sup_{\theta\in(0,1)} \mathbf{R}(\theta,T) = \sup_{\theta\in(0,1)} \mathrm{Var}_\theta(\overline{X}_n) = \sup_{\theta\in(0,1)} \frac{\theta(1-\theta)}{n} = \frac{1}{4n},$$

On en conclut $\mathbf{R}_{\max}(T) > \mathbf{R}_{\max}(\widehat{\theta}_{a_n,a_n}(\mathbf{X})) = \mathbf{R}_M$, donc T n'est pas minimax.

Exercice 7 (De Bayes et « unique » implique admissible)

1. (a) L'espérance a posteriori $T^\star(\mathbf{X}) = \int \theta d\Pi(\theta \mid \mathbf{X})$ est un estimateur de Bayes pour Π.

(b) Supposons que pour tout $\theta \in \Theta$, $P_\theta \ll Q$. Le risque a posteriori de T s'écrit

$$\begin{aligned}
\boldsymbol{\rho}(\Pi, T \mid \mathbf{X}) &= \mathbb{E}\left[(T(\mathbf{X}) - \boldsymbol{\theta})^2 \mid \mathbf{X}\right] \\
&= (T(\mathbf{X}) - T^\star(\mathbf{X}))^2 + \mathbb{E}\left[(T^\star(\mathbf{X}) - \boldsymbol{\theta})^2 \mid \mathbf{X}\right] + 2(T(\mathbf{X}) - T^\star(\mathbf{X}))\mathbb{E}\left[T^\star(\mathbf{X}) - \boldsymbol{\theta} \mid \mathbf{X}\right] \\
&= (T(\mathbf{X}) - T^\star(\mathbf{X}))^2 + \boldsymbol{\rho}(\Pi, T^\star \mid \mathbf{X}),
\end{aligned}$$

le terme croisé étant nul par définition de $T^\star(\mathbf{X}) = \mathbb{E}[\boldsymbol{\theta} \mid \mathbf{X}]$. En prenant l'espérance selon la loi marginale de $\mathbf{X}$, on obtient

$$\mathbf{R}_B(\Pi, T) = \mathbb{E}\left[(T(\mathbf{X}) - T^\star(\mathbf{X}))^2\right] + \mathbf{R}_B(\Pi, T^\star).$$

Or $\mathbf{R}_B(\Pi, T) = \mathbf{R}_B(\Pi, T^\star) = \mathbf{R}_B(\Pi)$ puisque T et $T^\star$ sont de Bayes. Ainsi

$$\mathbb{E}\left[(T(\mathbf{X}) - T^\star(\mathbf{X}))^2\right] = 0,$$

et donc $T(\mathbf{X}) = T^\star(\mathbf{X})$, Q-presque sûrement. Mais comme Q domine toutes les lois P_θ, on a $T(\mathbf{X}) = T^\star(\mathbf{X})$, P_θ-presque sûrement, pour tout $\theta \in \Theta$. On a donc, pour tout $\theta \in \Theta$,

$$\mathbf{R}(\theta, T) = \int_E (T(x) - \theta)^2 \dot{\mathrm{P}}_\theta(x) = \int_E (T^\star(x) - \theta)^2 \dot{\mathrm{P}}_\theta(x) = \mathbf{R}(\theta, T^\star).$$

(c) Supposons que $T^\star$ est unique à équivalence près, c'est-à-dire que tout estimateur de Bayes pour Π est équivalent à $T^\star$. Supposons $T^\star$ inadmissible. Alors il existerait T avec

$$\begin{aligned}
\forall\, \theta \in \Theta, &\quad \mathbf{R}(\theta, T) \le \mathbf{R}(\theta, T^\star) \\
\exists\, \theta_0 \in \Theta, &\quad \mathbf{R}(\theta_0, T) < \mathbf{R}(\theta_0, T^\star).
\end{aligned}$$

On intègre la première identité par rapport à Π, ce qui donne $\mathbf{R}_B(\Pi, T) \le \mathbf{R}_B(\Pi, T^\star) = \mathbf{R}_B(\Pi)$ car $T^\star$ est de Bayes pour Π, et donc T aussi. Par unicité à équivalence près, on en déduit $R(\theta, T) = R(\theta, T^\star)$ pour tout $\theta \in \Theta$, ce qui contredit l'inégalité ci-dessus si $\theta = \theta_0$.

2. (a) La moyenne a posteriori est de Bayes et vaut ici

$$\widehat{\theta}_{a,\sigma^2} = \frac{n\overline{X}_n + a}{n + \sigma^{-2}}.$$

(b) On peut noter que $\mathbf{X} = (X_1, \ldots, X_n)$ a même loi que le vecteur Y dont les coordonnées sont données par

$$Y_i = \boldsymbol{\theta} + \varepsilon_i,$$

où les ε_i sont i.i.d. $\mathcal{N}(0,1)$ et $\boldsymbol{\theta}$ est une variable aléatoire indépendante des ε_i et de loi $\mathcal{N}(a, \sigma^2)$. Le vecteur Y est gaussien, de moyenne a et de matrice de variance-covariance

$$\Sigma = \begin{bmatrix} 1+\sigma^2 & \sigma^2 & \ldots & \sigma^2 \\ \sigma^2 & 1+\sigma^2 & \ldots & \sigma^2 \\ \vdots & & & \\ \sigma^2 & \ldots & & 1+\sigma^2 \end{bmatrix}.$$

On en déduit que $\mathbf{X} \sim \mathcal{N}(a, \Sigma)$. Notons que Σ est inversible car définie positive puisque $y^T\Sigma y = \|y\|_2^2 + \sigma^2(\sum_i y_i)^2 > 0$ dès que $y \neq 0$.

(c) La loi Q_n a une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ donnée par

$$x \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{{}^t(x-a)\Sigma^{-1}(x-a)}{2}\right).$$

Cette densité est strictement positive sur tout $\mathbb{R}^n$. On en déduit que si $Q(A) = 0$ pour A mesurable, alors A est de mesure de Lebesgue sur $\mathbb{R}^n$ nulle, et donc $P_\theta^{\otimes n}(A) = 0$ pour tout θ. Ainsi Q_n domine toutes les lois $P_\theta^{\otimes n}$.

(d) Notons qu'un estimateur de Bayes pour la loi $\Pi = \mathcal{N}(a, \sigma^2)$ s'écrit, d'après 1., comme $\alpha\overline{X}_n + \beta$, avec $\alpha = \frac{n}{n+\sigma^{-2}}$ et $\beta = \frac{a}{n+\sigma^{-2}}$. De plus, d'après 3. et la première partie de l'exercice, il est unique à équivalence près et donc admissible. En faisant varier a dans $\mathbb{R}$ et σ^2 dans $(0, +\infty)$, on obtient tous les estimateurs $\alpha\overline{X}_n + \beta$ avec $\alpha \in (0, 1)$ et $\beta \in \mathbb{R}$. Le cas $\alpha = 0$ s'obtient par ailleurs en remarquant que l'estimateur constant égal à β est admissible, puisque la loi P_β domine toutes les P_θ, voir la proposition du début du chapitre 3 correspondante.

(e) Les estimateurs $\overline{X}_n + \beta$, pour $\beta \neq 0$, ne sont pas admissibles, car leur risque quadratique est toujours strictement supérieur à celui de l'estimateur $\overline{X}_n$ puisque

$$\mathbf{R}(\theta, \overline{X}_n + \beta) = \mathbb{E}_\theta\left[(\overline{X}_n - \theta + \beta)^2\right] = \frac{1}{n} + \beta^2 > \frac{1}{n} = \mathbf{R}(\theta, \overline{X}_n).$$

Exercice 8 (« Presque de Bayes » implique admissible)

1. On raisonne par l'absurde. Si $T^\star$ était inadmissible, il existerait T tel que

$$\begin{aligned} &\forall\, \theta \in \Theta\,, \quad \mathbf{R}(\theta, T) \leq \mathbf{R}(\theta, T^\star) \\ &\exists\, \theta_0 \in \Theta\,, \quad \mathbf{R}(\theta_0, T) < \mathbf{R}(\theta_0, T^\star)\,. \end{aligned}$$

Notons que T est alors de risque fini car $T^\star$ l'est. Par ailleurs, par continuité des fonctions $\theta \mapsto \mathbf{R}(\theta, T')$ pour $T' = T$ et $T' = T^\star$, il existe $\varepsilon > 0$ et un voisinage U de θ_0 tels que pour tout $\theta \in U$,

$$\mathbf{R}(\theta, T) \leq \mathbf{R}(\theta, T^\star) - \varepsilon. \tag{1}$$

Par hypothèse, pour ce choix de $\varepsilon > 0$ et de U, il existe une loi a priori $\Pi = \Pi_{U,\varepsilon}$ telle que

$$\mathbf{R}_B(\Pi, T^\star) < \mathbf{R}_B(\Pi) + \varepsilon\Pi(U). \tag{2}$$

En intégrant le risque de T contre cette loi Π, on a

$$\begin{aligned} \mathbf{R}_B(\Pi, T) &= \int_U \mathbf{R}(\theta, T) d\Pi(\theta) + \int_{U^c} \mathbf{R}(\theta, T) d\Pi(\theta) \\ &\leq \int_U \mathbf{R}(\theta, T^\star) d\Pi(\theta) - \varepsilon\Pi(U) + \int_{U^c} \mathbf{R}(\theta, T) d\Pi(\theta) \qquad \text{par (1)} \\ &\leq \mathbf{R}_B(\Pi, T^\star) - \varepsilon\Pi(U) \qquad \text{car pour tout } \theta \in \Theta,\ \mathbf{R}(\theta, T) \leq \mathbf{R}(\theta, T^\star). \end{aligned}$$

Par définition du risque bayésien, $\mathbf{R}_B(\Pi) \leq \mathbf{R}_B(\Pi, T)$, donc

$$\mathbf{R}_B(\Pi) \leq \mathbf{R}_B(\Pi, T^\star) - \varepsilon\Pi(U),$$

ce qui contredit l'hypothèse de l'énoncé. Donc $T^\star$ est admissible.

2. Soit Π une loi a priori telle que $\Pi(U) > 0$ pour tout ouvert non vide U de Θ et soit $T^\star$ un estimateur de Bayes pour Π. Comme $\mathbf{R}_B(\Pi, T^\star) = \mathbf{R}_B(\Pi)$, l'inégalité de la question précédente est vérifiée pour tout ouvert non vide U et tout $\varepsilon > 0$ en prenant comme loi a priori Π (celle pour laquelle $T^\star$ est de Bayes). Donc $T^\star$ est admissible.

3. (a) Un estimateur de Bayes pour Π et la perte quadratique est donné par la moyenne a posteriori $T^\star(\mathbf{X}) = \int \theta d\Pi(\theta \mid \mathbf{X})$. Ici, comme

$$\Pi[\cdot \mid \mathbf{X}] = \mathcal{N}\left(\frac{a\sigma^{-2} + n\overline{X}_n}{n + \sigma^{-2}}, \frac{1}{n + \sigma^{-2}}\right),$$

on a

$$T^\star(\mathbf{X}) = \frac{a\sigma^{-2} + n\overline{X}_n}{n+\sigma^{-2}}.$$

Le risque de Bayes pour la perte quadratique est alors égal à l'espérance de la variance a posteriori. Ici la variance a posteriori est égale à $\frac{1}{n+\sigma^{-2}}$ et ne dépend donc pas de $\mathbf{X}$. On a donc

$$\mathbf{R}_B(\Pi) = \frac{1}{n+\sigma^{-2}}\,.$$

(b) On applique la question 1. à $T^\star$. Il suffit de vérifier que $\Pi(\mathcal{V}) > 0$ pour tout ouvert non vide $\mathcal{V}$. C'est le cas car Π a une densité strictement positive par rapport à la mesure de Lebesgue sur $\mathbb{R}$ donc tout ouvert a une masse strictement positive sous Π, donc $T^\star$ est admissible.

(c) On applique le critère donné en début d'énoncé. Soit $\mathcal{V}$ un ouvert non vide fixé de $\mathbb{R}$. Il contient un intervalle $[c,d]$ avec $c<d$. D'après ce qui précède, comme $\mathbf{R}_B(\Pi) = \mathbf{R}_B(\Pi, T^\star)$, et comme $\mathbf{R}(\Pi, \overline{X}_n) = \frac{1}{n}$ (le risque de $\overline{X}_n$ étant constant égal à $1/n$), on a

$$\mathbf{R}_B(\Pi, \overline{X}_n) - \mathbf{R}_B(\Pi) = \frac{\sigma^{-2}}{n(n+\sigma^{-2})}.$$

Par ailleurs,

$$\Pi([c,d]) = \int_c^d \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}\,\mathrm{d}x \sim \frac{d-c}{\sqrt{2\pi}\sigma} \quad (\sigma\to\infty).$$

En effet, le théorème des valeurs intermédiaires implique qu'il existe $v\in[c,d]$ tel que

$$\int_c^d e^{-\frac{(x-a)^2}{2\sigma^2}}\,\mathrm{d}x = (d-c)e^{-\frac{(v-a)^2}{2\sigma^2}},$$

et cette quantité est équivalente à $d-c$ quand $\sigma\to\infty$. Comme $\sigma^{-2} = o(\sigma^{-1})$ quand $\sigma\to\infty$, on a donc

$$\mathbf{R}_B(\Pi, \overline{X}_n) - \mathbf{R}_B(\Pi) < \varepsilon\,\Pi([c,d]) \le \varepsilon\Pi[\mathcal{V}]$$

pour σ choisi assez grand. Donc $\overline{X}_n$ est admissible.

Exercice 9 (Test bayésien)

1. Pour $\theta = p,q$, notons $p_\theta(\mathbf{X}) = \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n}$ la vraisemblance de $\mathbf{X}$ sous $\mathcal{B}(\theta)^{\otimes n}$. Par la formule de Bayes, la loi a posteriori est donnée par

$$\Pi\left[\{p\}\;\middle|\;\mathbf{X}\right] = \frac{\frac{1}{2}\cdot p_p(\mathbf{X})}{\frac{1}{2}\cdot p_p(\mathbf{X}) + \frac{1}{2}\cdot p_q(\mathbf{X})} = \frac{p_p(\mathbf{X})}{p_p(\mathbf{X}) + p_q(\mathbf{X})}\,,$$

et $\Pi\left[\{q\}\;\middle|\;\mathbf{X}\right] = 1 - \Pi\left[\{p\}\;\middle|\;\mathbf{X}\right]$. Ainsi le test de Bayes pour Π et la perte équilibré est donné par

$$\varphi^*(\mathbf{X}) = \mathbb{1}_{\left\{\Pi\left[\{p\}\,\middle|\,\mathbf{X}\right]\ge\Pi\left[\{q\}\,\middle|\,\mathbf{X}\right]\right\}} = \mathbb{1}_{\{p_p(\mathbf{X})\ge p_q(\mathbf{X})\}}\,.$$

Or

$$\begin{aligned}
p_p(\mathbf{X}) \ge p_q(\mathbf{X}) &\Leftrightarrow p^{n\overline{X}_n}(1-p)^{n-n\overline{X}_n} \ge q^{n\overline{X}_n}(1-q)^{n-n\overline{X}_n}\\
&\Leftrightarrow \left(\frac{p}{q}\right)^{n\overline{X}_n} \ge \left(\frac{1-q}{1-p}\right)^{n-n\overline{X}_n}\\
&\Leftrightarrow n\overline{X}_n \ln\left(\frac{p}{q}\right) \ge (n-n\overline{X}_n)\ln\left(\frac{1-q}{1-p}\right)\\
&\Leftrightarrow \overline{X}_n \ge \frac{\ln\left(\frac{1-q}{1-p}\right)}{\ln\left(\frac{p}{q}\right) + \ln\left(\frac{1-q}{1-p}\right)}\,.
\end{aligned}$$

Ainsi $\varphi^*(\mathbf{X}) = \mathbb{1}_{\left\{\overline{X}_n\ge c_{p,q}\right\}}$ avec

$$c_{p,q} = \frac{\ln\left(\frac{1-q}{1-p}\right)}{\ln\left(\frac{p}{q}\right) + \ln\left(\frac{1-q}{1-p}\right)}\,.$$

2. Le risque de Bayes pour Π et la perte équilibrée s'écrit

$$\begin{aligned}
\mathbf{R}_B(\Pi) &= \mathbb{E}\left[\mathbb{1}_{\boldsymbol{\theta}=q,\varphi^*(\mathbf{X})=1} + \mathbb{1}_{\boldsymbol{\theta}=p,\varphi^*(\mathbf{X})=0}\right] \\
&= \frac{1}{2}\mathbb{P}_q\left(\varphi^*(\mathbf{X})=1\right) + \frac{1}{2}\mathbb{P}_p\left(\varphi^*(\mathbf{X})=0\right) \\
&= \frac{1}{2}\Big(\mathbb{P}_q\left(p_p(\mathbf{X}) \geq p_q(\mathbf{X})\right) + \mathbb{P}_p\left(p_p(\mathbf{X}) < p_q(\mathbf{X})\right)\Big) \\
&= \frac{1}{2}\Big(1 - \Big\{\mathbb{P}_p\left(p_p(\mathbf{X}) \geq p_q(\mathbf{X})\right) - \mathbb{P}_q\left(p_p(\mathbf{X}) \geq p_q(\mathbf{X})\right)\Big\}\Big)\,.
\end{aligned}$$

Or

$$\begin{aligned}
\mathbb{P}_p\left(p_p(\mathbf{X}) \geq p_q(\mathbf{X})\right) - \mathbb{P}_q\left(p_p(\mathbf{X}) \geq p_q(\mathbf{X})\right) &= \sum_{x\in\{0,1\}^n} p_p(x)\mathbb{1}_{p_p(x)\geq p_q(x)} - \sum_{x\in\{0,1\}^n} p_q(x)\mathbb{1}_{p_p(x)\geq p_q(x)} \\
&= \sum_{x\in\{0,1\}^n} \left(p_p(x)-p_q(x)\right)_+ \\
&= d_{\mathrm{VT}}\left(\mathcal{B}(p)^{\otimes n}\,,\,\mathcal{B}(q)^{\otimes n}\right)\,.
\end{aligned}$$

3. On a

$$\mathbf{R}_B(\Pi,\varphi) = \frac{1}{2}\left(\mathbb{P}_p\left(\sum_{i=1}^n X_i < \frac{p+q}{2}n\right) + \mathbb{P}_q\left(\sum_{i=1}^n X_i \geq \frac{p+q}{2}n\right)\right)\,.$$

En soustrayant par le bon recentrage dans chacune des deux probabilités, on a d'une part

$$\mathbb{P}_p\left(\sum_{i=1}^n X_i < \frac{p+q}{2}n\right) = \mathbb{P}_p\left(\sum_{i=1}^n X_i - np < -\frac{p-q}{2}n\right)\,,$$

et

$$\mathbb{P}_q\left(\sum_{i=1}^n X_i \geq \frac{p+q}{2}n\right)\mathbb{P}_q\left(\sum_{i=1}^n X_i - nq \geq \frac{p-q}{2}n\right)\,.$$

Par l'inégalité de Hoeffding, comme $0 \leq X_i \leq 1$, chacune de ces deux probabilités est inférieure à $e^{-\frac{2(p-q)^2n^2}{4n}} = e^{-\frac{(p-q)^2n}{2}}$. On obtient

$$\mathbf{R}_B(\Pi,\varphi) \leq e^{-\frac{(p-q)^2n}{2}}\,.$$

En particulier, $\sum_{n\geq 1}\mathbb{P}_q(\varphi(\mathbf{X})=1) < \infty$ donc, par Borel–Cantelli, sous $\mathcal{B}(q)$, on a $\varphi(\mathbf{X}) \xrightarrow{\text{p.s.}} 0$, et de même, sous $\mathcal{B}(p)$, on a $\varphi(\mathbf{X}) \xrightarrow{\text{p.s.}} 1$.