Feuille 3 : correction

Exercice 1 (Tests dans le modèle gaussien)

1. *Test de Student d'une relation affine.*
   (a) Comme $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2({}^tXX)^{-1})$, on a ${}^tc\widehat{\beta} - {}^tc\beta \sim \mathcal{N}(0, \sigma^2\,{}^tc({}^tXX)^{-1}c)$. De plus, on sait par le théorème de Cochran que $\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$ et que que $\widehat{\beta}$ et $\widehat{\sigma}^2$ sont indépendantes. Donc la variable
   $$\frac{{}^tc\widehat{\beta} - {}^tc\beta}{\widehat{\sigma}\sqrt{{}^tc({}^tXX)^{-1}c}} = \frac{\frac{{}^tc\widehat{\beta} - {}^tc\beta}{\sigma\sqrt{{}^tc({}^tXX)^{-1}c}}}{\sqrt{\frac{(n-p)\widehat{\sigma}^2}{(n-p)\sigma^2}}}$$
   suit une loi de Student $\mathcal{T}(n-p)$.
   (b) Sous H_0, on sait que la statistique $T_n = \frac{{}^tc\widehat{\beta} - a}{\widehat{\sigma}\sqrt{{}^tc({}^tXX)^{-1}c}}$ suit une loi de Student $\mathcal{T}(n-p)$. En notant $t_{n-p}(1-\alpha/2)$ le quantile d'ordre $1-\alpha/2$ de la loi de Student $\mathcal{T}(n-p)$, le test qui rejette H_0 si et seulement si $|T_n| > t_{n-p}(1-\alpha/2)$ est donc un test de niveau α.
   (c) Par la question 1.(a), l'intervalle
   $$\left[{}^tc\widehat{\beta} - t_{n-p}(1-\alpha/2)\widehat{\sigma}\sqrt{{}^tc({}^tXX)^{-1}c}\,,\ {}^tc\widehat{\beta} + t_{n-p}(1-\alpha/2)\widehat{\sigma}\sqrt{{}^tc({}^tXX)^{-1}c}\right]$$
   est un intervalle de niveau $1-\alpha$ pour ${}^tc\beta$.
2. *Test de Fisher d'un sous-modèle.*
   (a) On décompose $\mathbb{R}^n$ en $\mathbb{R}^n = W \oplus W^{\perp_V} \oplus V^{\perp}$, où $W^{\perp_V}$ est le sous-espace orthogonal à W dans V. Par le théorème de Cochran, on a
   — $P_WY \sim \mathcal{N}(P_WX\beta, \sigma^2 P_W)$;
   — $P_{W^{\perp_V}}Y \sim \mathcal{N}(P_{W^{\perp_V}}X\beta, \sigma^2 P_{W^{\perp_V}})$;
   — $P_{V^\perp}Y \sim \mathcal{N}(0, \sigma^2 P_{V^\perp})$;
   — les variables P_WY, $P_{W^{\perp_V}}Y$ et $P_{V^\perp}Y$ sont indépendantes ;
   — $\sigma^{-2}\|P_W(Y - X\beta)\|^2 \sim \chi^2(q)$;
   — $\sigma^{-2}\|P_{W^{\perp_V}}(Y - X\beta)\|^2 \sim \chi^2(p-q)$;
   — $\sigma^{-2}\|P_{V^\perp}Y\|^2 \sim \chi^2(n-p)$;
   Remarquons que si $X\beta \in W$, alors $P_{W^{\perp_V}}X\beta = 0$ et $\sigma^{-2}\|P_{W^{\perp_V}}Y\|^2$ suit la loi $\chi^2(p-q)$. Donc, par indépendance de $P_{W^{\perp_V}}Y$ et $P_{V^\perp}Y$, si $X\beta \in W$, la statistique
   $$F = \frac{\|P_{W^{\perp_V}}Y\|^2/(p-q)}{\|P_{V^\perp}Y\|^2/(n-p)}$$
   suit une loi de Fisher $\mathcal{F}(p-q, n-p)$.
   (b) En notant $f_{p-q,n-p}(1-\alpha)$ le quantile d'ordre $1-\alpha$ de la loi de Fisher $\mathcal{F}(p-q, n-p)$, on déduit de la question précédente que le test qui rejette H_0 si et seulement si $F > f_{p-q,n-p}(1-\alpha)$ est un test de niveau α.
   (c) Si $q = p-1$, le sous-espace W est un hyperplan de V. Donc il existe $h \in V$ tel que
   $$X\beta \in W \iff {}^thX\beta = 0 \iff {}^tc\beta = 0\,,$$
   avec $c = {}^tXh$. Ainsi tester si $X\beta \in W$ revient à tester si ${}^tc\beta = a$ avec $c = {}^tXh$ et $a = 0$. Montrons maintenant que dans ce cas, les deux tests obtenus sont les mêmes. Plus précisément, montrons que $|T_n|^2 = F$ et $t_{n-p}(1-\alpha/2)^2 = f_{1,n-p}(1-\alpha)$. Notons d'abord que si $Z \sim \mathcal{T}(n-p)$, alors, $Z^2 \sim \mathcal{F}(1, n-p)$. On a donc bien $t_{n-p}(1-\alpha/2)^2 = f_{1,n-p}(1-\alpha)$. La statistique du test de la question 1 élevée au carré est :
   $$|T_n|^2 = \frac{({}^tc\widehat{\beta})^2}{\widehat{\sigma}^2\,{}^tc({}^tXX)^{-1}c} = \frac{({}^thX\widehat{\beta})^2}{\widehat{\sigma}^2\,{}^thP_Vh} = \frac{({}^thX\widehat{\beta})^2}{\widehat{\sigma}^2\|h\|^2}.$$

Par définition, $\widehat{\sigma}^2 = \frac{\|P_{V^\perp} Y\|^2}{n-p}$. D'autre part, puisque W est l'hyperplan de V orthogonal au vecteur h de V, $W^{\perp_V}$ est l'espace vectoriel engendré par h, et $P_{W^{\perp_V}} Y = P_{W^{\perp_V}} X\widehat{\beta} = \langle h, X\widehat{\beta}\rangle \frac{h}{\|h\|^2}$. Ainsi

$$\|P_{W^{\perp_V}} Y\|^2 = \frac{({}^t hX\widehat{\beta})^2}{\|h\|^2}$$

et l'on a bien $F = |T_n|^2$.

3. *Test de Wald de plusieurs hypothèses affines.*
   (a) On a
   $$C\widehat{\beta} - C\beta \sim \mathcal{N}(0, \sigma^2 C({}^t XX)^{-1}\,{}^t C)$$
   (b) Clairement, la matrice Σ est symétrique. De plus, pour $x \in \mathbb{R}^k$, on a
   $${}^t x\Sigma x = {}^t({}^t Cx)({}^t XX)^{-1}\,{}^t Cx$$
   Si x est non nul, alors, comme ${}^t C$ est de rang k donc injective, le vecteur ${}^t Cx$ est lui aussi non nul. Et comme $({}^t XX)^{-1}$ est définie positive, on a ${}^t({}^t Cx)({}^t XX)^{-1}\,{}^t Cx > 0$.
   (c) On peut écrire $\Sigma = Q\Delta Q^{-1}$ avec Δ une matrice diagonale et Q une matrice orthogonale. On définit alors $\Sigma^{1/2} = Q\Delta^{1/2}Q^{-1}$, où $\Delta^{1/2}$ est la matrice diagonale dont les coefficients sont les racines carrées des coefficients de Δ. On a alors
   $$\frac{\Sigma^{-1/2}(C\widehat{\beta} - C\beta)}{\sigma} \sim \mathcal{N}(0, I_k)\,,$$
   et donc
   $$\frac{\|\Sigma^{-1/2}(C\widehat{\beta} - C\beta)\|^2}{\sigma^2} = \frac{1}{\sigma^2}\,{}^t(C\widehat{\beta} - C\beta)\Sigma^{-1}(C\widehat{\beta} - C\beta) \sim \chi^2(k)\,.$$
   De plus
   $$\frac{\|Y - X\widehat{\beta}\|^2}{\sigma^2} = \frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)\,,$$
   et $\widehat{\beta}$ et $\widehat{\sigma}^2$ sont indépendantes. Donc la variable
   $$\frac{{}^t(C\widehat{\beta} - C\beta)\Sigma^{-1}(C\widehat{\beta} - C\beta)/k}{\widehat{\sigma}^2}$$
   suit une loi de Fisher $\mathcal{F}(k, n-p)$.
   (d) Sous H_0, la statistique
   $$W = \frac{{}^t(C\widehat{\beta} - a)\Sigma^{-1}(C\widehat{\beta} - a)/k}{\widehat{\sigma}^2}$$
   suit une loi de Fisher $\mathcal{F}(k, n-p)$. Ainsi, en notant $f_{k,n-p}(1-\alpha)$ le quantile d'ordre $1-\alpha$ de la loi de Fisher $\mathcal{F}(k, n-p)$, le test qui rejette H_0 si et seulement si $W > f_{k,n-p}(1-\alpha)$ est un test de niveau $1-\alpha$.
   (e) Le sous-ensemble de $\mathbb{R}^k$ défini par
   $$\mathcal{E}_\alpha = \left\{ a \in \mathbb{R}^k,\ \frac{\|\Sigma^{-1/2}(C\widehat{\beta} - a)\|^2/k}{\widehat{\sigma}^2} \le f_{k,n-p}(1-\alpha) \right\}$$
   est une ellipsoïde de confiance de niveau $1-\alpha$ pour $C\beta$.

Exercice 2 (Régression Ridge - Régularisation de Tikhonov)

1. Si $k > n$, le noyau de X n'est pas réduit à 0 et cela équivaut à dire que la matrice $X^T X$ n'est pas inversible. Le modèle n'est pas identifiable et il existe une infinité de solutions à l'équation $X^T X\theta = X^T Y$.
2. La fonction $f : \theta \mapsto \|Y - X\theta\|^2 + \lambda\|\theta\|^2$ est strictement convexe et différentiable. Le minimiseur est donc un point critique :
   $$\nabla f(\theta) = 2X^T(X\theta - Y) + 2\lambda\theta = 0 \;\Leftrightarrow (X^T X + \lambda I)\theta = X^T Y\,.$$
   Comme $\lambda > 0$, la matrice $X^T X + \lambda I$ est définie positive, donc inversible. L'unique solution est
   $$\widehat{\theta}_\lambda = (X^T X + \lambda I)^{-1} X^T Y\,.$$

3. On a
$$\mathbb{E}[\widehat{\theta}_\lambda] = (X^TX + \lambda I)^{-1}X^TX\theta \neq \theta\,,$$
et
$$\begin{aligned}\mathrm{Cov}(\widehat{\theta}_\lambda) &= (X^TX + \lambda I)^{-1}X^T\,\mathrm{Cov}(Y)X(X^TX + \lambda I)^{-1}\\ &= \sigma^2(X^TX + \lambda I)^{-1}X^TX(X^TX + \lambda I)^{-1}\\ &= \sigma^2(X^TX + \lambda I)^{-1}\left(I - \lambda(X^TX + \lambda I)^{-1}\right)\,.\end{aligned}$$

Exercice 3 (Maximum de gaussiennes corrélées)

1. On a $\max(X_1, X_2) = \frac{1}{2}\left(X_1 + X_2 + |X_1 - X_2|\right)$. Comme $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$, on obtient $\mathbb{E}\left[\max(X_1, X_2)\right] = \frac{1}{2}\mathbb{E}[|X_1 - X_2|]$. Or $X_1 - X_2 \sim \mathcal{N}(0, 2(1-\rho))$. Ainsi
$$\begin{aligned}\mathbb{E}\left[\max(X_1, X_2)\right] &= \int_0^{+\infty} \frac{1}{\sqrt{4\pi(1-\rho)}} x e^{-\frac{x^2}{4(1-\rho)}}\,dx\\ &= \sqrt{\frac{1-\rho}{\pi}}\left[-e^{-\frac{x^2}{4(1-\rho)}}\right]_0^{+\infty} = \sqrt{\frac{1-\rho}{\pi}}\,.\end{aligned}$$

2. On peut remarquer que
$$(X_1, \dots, X_n) \sim (\sqrt{\rho}Y + \sqrt{1-\rho}Y_1, \dots, \sqrt{\rho}Y + \sqrt{1-\rho}Y_n)\,,$$
où $Y, Y_1, \dots, Y_n$ sont i.i.d. de loi $\mathcal{N}(0,1)$. En effet, il s'agit bien d'un vecteur gaussien, d'espérance nulle, et l'on a, pour tout $i \in [\![1, n]\!]$,
$$\mathrm{Var}(\sqrt{\rho}Y + \sqrt{1-\rho}Y_i) = (\sqrt{\rho})^2 + (\sqrt{1-\rho})^2 = 1\,,$$
et pour $i \neq j$, par indépendance de Y, Y_i et Y_j,
$$\mathrm{Cov}(\sqrt{\rho}Y + \sqrt{1-\rho}Y_i, \sqrt{\rho}Y + \sqrt{1-\rho}Y_j) = \mathrm{Cov}(\sqrt{\rho}Y, \sqrt{\rho}Y) = \rho\,.$$
Avec cette représentation, on voit que
$$a_n(\rho) = \mathbb{E}\left[\sqrt{\rho}Y + \sqrt{1-\rho}\max(Y_1, \dots, Y_n)\right] = \sqrt{1-\rho}\,a_n(0)\,.$$

3. En utilisant l'inégalité de Jensen puis le fait que le maximum de variables positives est inférieur à la somme, on a, pour tout $\lambda \in \mathbb{R}_+$,
$$\begin{aligned}e^{\lambda \mathbb{E}[\max_{1\le i\le n}(X_i)]} &\le \mathbb{E}\left[e^{\lambda \max_{1\le i\le n}(X_i)}\right]\\ &= \mathbb{E}\left[\max_{1\le i\le n} e^{\lambda X_i}\right]\\ &\le \sum_{i=1}^n \mathbb{E}\left[e^{\lambda X_i}\right] = n\mathbb{E}\left[e^{\lambda X_1}\right] = ne^{\lambda^2/2}\,,\end{aligned}$$
où l'on a utilisé que la transformée de Laplace d'une loi $\mathcal{N}(0, \sigma^2)$ est donnée par $\lambda \mapsto e^{\frac{\lambda^2\sigma^2}{2}}$. Ainsi pour tout $\lambda \in \mathbb{R}_+^*$,
$$a_n(0) \le \frac{1}{\lambda}\left(\ln(n) + \frac{\lambda^2}{2}\right)$$
En optimisant sur $\lambda > 0$, on obtient, pour $\lambda = \sqrt{2\ln(n)}$,
$$a_n(0) \le \sqrt{2\ln(n)}\,.$$

Exercice 4 (Maximum de variables de Poisson indépendantes)

1. On utilise la même méthode que dans l'exercice précédent. Pour tout $\lambda \in \mathbb{R}_+$, par l'inégalité de Jensen,

$$e^{\lambda \mathbb{E}[\max_{1\le i\le n} X_i]} \le \mathbb{E}\left[\max_{1\le i\le n} e^{\lambda X_i}\right] \le n\mathbb{E}\left[e^{\lambda X_1}\right] = ne^{e^\lambda - 1}\,.$$

Ainsi pour tout $\lambda \in \mathbb{R}_+^*$,

$$\mathbb{E}\left[\max_{1\le i\le n} X_i\right] \le \frac{1}{\lambda}\left(\ln(n) + e^\lambda - 1\right)\,.$$

En choisissant λ tel que $e^\lambda - 1 = \ln(n)$, soit $\lambda = \ln(1+\ln(n))$, on obtient

$$\mathbb{E}\left[\max_{1\le i\le n} X_i\right] \le \frac{2\ln(n)}{\ln(1+\ln(n))}\,.$$

2. Pour tout $k \in \mathbb{N}^*$, on a, par l'inégalité de Markov,

$$\begin{aligned}
\mathbb{E}\left[\max_{1\le i\le n} X_i\right] &\ge k\mathbb{P}\left(\max_{1\le i\le n} X_i \ge k\right)\\
&= k\left(1 - \mathbb{P}\left(\max_{1\le i\le n} X_i < k\right)\right)\\
&= k\left(1 - \mathbb{P}(X_1 < k)^n\right)\\
&= k\left(1 - (1-\mathbb{P}(X_1 \ge k))^n\right)\\
&\ge k\left(1 - (1-\mathbb{P}(X_1 = k))^n\right)\,.
\end{aligned}$$

On a

$$\mathbb{P}(X_1 = k) = \frac{e^{-1}}{k!} \ge \frac{e^{-1}}{k^k}\,.$$

Pour $k = \left\lfloor \frac{\ln(n)}{\ln\ln(n)} \right\rfloor$, on a $k^k \le n$. Ainsi

$$\mathbb{E}\left[\max_{1\le i\le n} X_i\right] \ge \left\lfloor \frac{\ln(n)}{\ln\ln(n)} \right\rfloor \left(1 - \left(1 - \frac{e^{-1}}{n}\right)^n\right) \ge \left\lfloor \frac{\ln(n)}{\ln\ln(n)} \right\rfloor \left(1 - e^{-e^{-1}}\right)\,.$$

Exercice 5 (Conditionnement linéaire gaussien)
Voir le poly de S. Boucheron, section 2.5.

Exercice 6 (Statistique exhaustive, statistique complète)

1. (a) Posons $\theta = e^{-\lambda} \in]0,1]$. La densité de X s'écrit, pour tout $x \in \mathbb{N}^n$,

$$f_\theta(x) = \prod_{i=1}^n \frac{\theta(\ln(1/\theta))^{x_i}}{x_i!} = \frac{1}{\prod_{i=1}^n x_i!}\theta^n(\ln(1/\theta))^{\sum_{i=1}^n x_i}\,.$$

Ainsi, la propriété de factorisation est vérifiée avec $T(x) = \sum_{i=1}^n x_i$, $h(x) = \left(\prod_{i=1}^n x_i!\right)^{-1}$, et $g(\theta, T(x)) = \theta^n(\ln(1/\theta))^{T(x)}$. On peut aussi vérifier que la loi de X sachant $S_n = \sum_{i=1}^n X_i$ ne dépend pas de θ. En effet, la variable S_n suit une loi de Poisson de paramètre $n\lambda$. Ainsi

$$\mathbb{P}(S_n = k) = \frac{e^{-\lambda n}(n\lambda)^k}{k!} = \frac{\theta^n(n\ln(1/\theta))^k}{k!}\,,$$

et pour $x = (x_1, \dots, x_n)$ tel que $\sum x_i = k$, on a

$$\mathbb{P}\left(X = x \mid S_n = k\right) = \frac{\mathbb{P}(X=x)}{\mathbb{P}(S_n = k)} = \frac{\theta^n(\ln(1/\theta))^k}{\prod_{i=1}^n x_i!} \cdot \frac{k!}{\theta^n(n\ln(1/\theta))^k} = \frac{\binom{k}{x_1,\dots,x_n}}{n^k}\,,$$

où $\binom{k}{x_1,\dots,x_n}$ est le nombre de mots de longueur k que l'ont peut former avec un alphabet de taille n en utilisant x_1 fois la lettre 1, …, x_n fois la lettre n (coefficient multinomial). La loi de X sachant $S_n = k$ est appelée loi multinomiale de paramètres k et $(1/n, \dots, 1/n)$. En particulier, la loi de X sachant S_n est indépendante de θ.

(b) Notons $x \mapsto g(x \mid T(X))$ la densité conditionnelle de X sachant $T(X)$. Comme T est exhaustive, cette fonction ne dépend pas de θ. On a

$$\theta^*(X) = \int_E \widehat{\theta}(x) g(x \mid T(X)) d\mu(x) .$$

La variable $\theta^*(X)$ est donc une fonction mesurable de X ne dépendant pas de θ. C'est bien un estimateur. Le théorème de Rao–Blackwell découle simplement de l'inégalité de Jensen conditionnelle. En effet,

$$\begin{aligned}
\mathbb{E}\left[\|\theta^*(X) - \theta\|^2\right] &= \sum_{j=1}^d \mathbb{E}\left[(\theta^*(X)_j - \theta_j)^2\right] \\
&= \sum_{j=1}^d \mathbb{E}\left[\left(\mathbb{E}\left[\widehat{\theta}(X)_j \mid T(X)\right] - \theta_j\right)^2\right] \\
&\leq \sum_{j=1}^d \mathbb{E}\left[\mathbb{E}\left[\left(\widehat{\theta}(X)_j - \theta_j\right)^2 \mid T(X)\right]\right] \\
&= \sum_{j=1}^d \mathbb{E}\left[\left(\widehat{\theta}(X)_j - \theta_j\right)^2\right] \\
&= \mathbb{E}\left[\left\|\widehat{\theta}(X) - \theta\right\|^2\right] .
\end{aligned}$$

(c) Dans le cas poissonien, on a vu que la somme S_n était une statistique exhaustive. La version Rao–Blackwellisée de W est alors

$$\mathbb{E}[W \mid S_n] = \mathbb{E}[\mathbb{1}_{X_1=0} \mid S_n] = \mathbb{P}(X_1 = 0 \mid S_n) .$$

Pour $k \in \mathbb{N}$, on a

$$\mathbb{P}(X_1 = 0 \mid S_n = k) = \frac{\mathbb{P}(X_1 = 0, S_n = k)}{\mathbb{P}(S_n = k)} = \frac{\mathbb{P}(X_1 = 0)\mathbb{P}(S_{n-1} = k)}{\mathbb{P}(S_n = k)} ,$$

où l'on a utilisé que l'événement $\{X_1 = k\} \cup \{S_n = k\}$ est égal à l'événement $\{X_1 = 0\} \cup \{X_2 + \cdots + X_n = k\}$. Or ces deux événement sont indépendants, et $X_2 + \cdots + X_n$ a la même loi que S_{n-1}. En utilisant la formule pour $\mathbb{P}(S_n = k)$ donnée plus haut et le fait que $\mathbb{P}(X_1 = 0) = \theta$, on a

$$\mathbb{P}(X_1 = 0 \mid S_n = k) = \theta \, \frac{\theta^{n-1}(n-1)^k}{\theta^n n^k} = \left(1 - \frac{1}{n}\right)^k .$$

Ainsi

$$\mathbb{E}[W \mid S_n] = \left(1 - \frac{1}{n}\right)^{S_n} .$$

2. (a) On se place dans le cas poissonien avec $S_n = \sum X_i$. Soit $g : \mathbb{N} \to R$ telle que pour tout $\theta \in \Theta$, $\mathbb{E}_\theta\left[g(S_n)\right] = 0$. Ainsi pour tout $\lambda \geq 0$,

$$\sum_{k \geq 0} g(k) \frac{(n\lambda)^k}{k!} = 0 .$$

Pour $\lambda = 0$, on obtient $g(0) = 0$. Puis, pour $j \geq 1$, en dérivant j fois par rapport à λ, on a

$$\sum_{k \geq j} g(k) n^k \frac{\lambda^{k-j}}{(k-j)!} = 0 ,$$

et en évaluant cette égalité en $\lambda = 0$, on obtient $g(j) = 0$. Ainsi g est la fonction nulle et l'on a bien, pour tout $\theta \in]0, 1]$, $\mathbb{P}_\theta\left(g(S_n) = 0\right) = 1$. La statistique S_n est donc bien une statistique complète.

(b) Tout d'abord, montrons que si T^* est un estimateur sans biais de risque quadratique minimal, alors il est unique p.s. Soit T un estimateur sans biais de θ avec $\mathbb{E}[\|T-\theta\|^2] = \mathbb{E}[\|T^*-\theta\|^2] = R_{min}$, où R_{min} est le risque minimal. Considérons l'estimateur $T' = \frac{T^*+T}{2}$. Alors T' est lui aussi sans biais. Et son risque vérifie :

$$\begin{aligned}\mathbb{E}\left[\|T'-\theta\|^2\right] &= \frac{1}{4}\mathbb{E}\left[\|T^*-\theta+T-\theta\|^2\right] \\ &= \frac{1}{4}\left(\mathbb{E}\left[\|T^*-\theta\|^2\right] + \mathbb{E}\left[\|T-\theta\|^2\right] + 2\mathbb{E}\left[\langle T^*-\theta\,,\,T-\theta\rangle\right]\right) \\ &\frac{1}{2}\left(R_{min} + \mathbb{E}\left[\langle T^*-\theta\,,\,T-\theta\rangle\right]\right)\,.\end{aligned}$$

Comme $\mathbb{E}\left[\|T'-\theta\|^2\right] \geq R_{min}$ par définition du risque minimal, on doit avoir

$$\mathbb{E}\left[\langle T^*-\theta\,,\,T-\theta\rangle\right] \geq R_{min}\,.$$

D'autre part, en appliquant deux fois l'inégalité de Cauchy–Schwarz, on a

$$\mathbb{E}\left[\langle T^*-\theta\,,\,T-\theta\rangle\right] \leq \mathbb{E}\left[\|T^*-\theta\|\cdot\|T-\theta\|\right] \leq \sqrt{\mathbb{E}\left[\|T^*-\theta\|^2\right]\mathbb{E}\left[\|T-\theta\|^2\right]} = R_{min}\,.$$

On en déduit que

$$\mathbb{E}\left[\|T^*-\theta\|\cdot\|T-\theta\| - \langle T^*-\theta\,,\,T-\theta\rangle\right] = 0\,.$$

Comme cette variable est positive, on obtient $\langle T^*-\theta\,,\,T-\theta\rangle = \|T^*-\theta\|\cdot\|T-\theta\|$ p.s. Il existe donc $\lambda > 0$ tel que $T^*-\theta = \lambda(T-\lambda)$ p.s. Mais comme les risque de T^* et de T sont les mêmes, $\lambda = 1$ et $T^* = T$ p.s.

Montrons maintenant que l'estimateur $\theta^*(X) = \mathbb{E}\left[\widehat{\theta}(X) \bigm| T(X)\right]$, avec $\widehat{\theta}(X)$ un estimateur sans biais et $T(X)$ une statistique exhaustive et complète, est de risque minimal (il est clairement sans biais puisque $\mathbb{E}[\theta^*(X)] = \mathbb{E}[\widehat{\theta}(X)] = \theta$). Soit $T'(X)$ un estimateur sans biais de θ. On a

$$\begin{aligned}\mathbb{E}\left[\|T'(X)-\theta\|^2\right] &= \mathbb{E}\left[\|T'(X)-\theta^*(X)\|^2\right] + \mathbb{E}\left[\|\theta^*(X)-\theta\|^2\right] + 2\mathbb{E}\left[\langle T'(X)-\theta^*(X)\,,\,\theta^*(X)-\theta\rangle\right] \\ &= \mathbb{E}\left[\|T'(X)-\theta^*(X)\|^2\right] + \mathbb{E}\left[\|\theta^*(X)-\theta\|^2\right] \\ &\quad + 2\mathbb{E}\left[\left\langle \mathbb{E}[T'(X) \mid T(X)] - \mathbb{E}[\widehat{\theta}(X) \mid T(X)]\,,\,\mathbb{E}[\widehat{\theta}(X) \mid T(X)] - \theta\right\rangle\right]\,.\end{aligned}$$

Notons $g(T(X)) = \mathbb{E}[T'(X) \mid T(X)] - \mathbb{E}[\widehat{\theta}(X) \mid T(X)]$. Comme T' et $\widehat{\theta}$ sont sans biais, on a, pour tout $\theta \in \Theta$, $\mathbb{E}_\theta\left[g(T(X))\right] = 0$. Ainsi, comme T est complète, pur tout $\theta \in \Theta$, $\mathbb{P}_\theta\left(g(T(X)) = 0\right) = 1$. Donc l'espérance du produit scalaire est nulle et $\mathbb{E}\left[\|T'(X)-\theta\|^2\right] \geq \mathbb{E}\left[\|\theta^*(X)-\theta\|^2\right]$, ce qu'il fallait démontrer.

(c) La statistique est exhaustive et complète, donc, comme W est sans biais, $\mathbb{E}[W \mid S_n] = \left(1-\frac{1}{n}\right)^{S_n}$ est l'unique estimateur sans biais de variance minimale. On a

$$\mathbb{E}\left[\mathbb{E}[W \mid S_n]^2\right] = \sum_{k\geq 0} \frac{e^{-\lambda n}(\lambda n)^k}{k!}\left(1-\frac{1}{n}\right)^2 k = e^{-\lambda n+\lambda n\left(1-\frac{1}{n}\right)^2} = e^{-\lambda\left(2-\frac{1}{n}\right)}\,.$$

Ainsi le risque minimal d'un estimateur vaut $e^{-\lambda\left(2-\frac{1}{n}\right)} - e^{-2\lambda} = e^{-2\lambda}\left(e^{\lambda/n}-1\right)$.