Modélisation statistique (STA221)
Examen (2h)

Olivier Rioul

5 octobre 2023

Contrôle avec ***polycopié et notes de cours manuscrites, UNIQUEMENT.****.*
Calculettes autorisées.
On détaillera avec précision tous les raisonnements.
Rédiger lisiblement, ***en faisant des phrases*** *en langue humaine SVP.*
Le soin et la présentation comptent pour la note finale.

I Factorisation de Neyman-Fisher

ON CONSIDÈRE un modèle statistique $\underline{X} = (X_1, \dots, X_N) \sim p_\theta(\underline{x})$ (pas nécessairement régulier) et une fonction S qui ne dépend pas de θ. On suppose que $S = S(\underline{X})$ est une *statistique suffisante* pour θ, ce qui signifie que si $s = S(\underline{x})$, la distribution conditionnelle

$$p_\theta(\underline{x}|s) = p(\underline{x}|s)$$

ne dépend pas de θ. (Autrement dit les données $\underline{X}$ ne dépendent de θ que par l'intermédiaire de la statistique $S = S(\underline{X})$).

1. Si $s = S(\underline{x})$, en écrivant $p_\theta(\underline{x}) = p_\theta(\underline{x}, s)$ et en développant, montrer qu'on peut factoriser le modèle $p_\theta(\underline{x})$ sous la forme

$$\boxed{p_\theta(\underline{x}) = g_\theta(S(\underline{x})) \cdot h(x)}$$

   où la fonction g_θ dépend de θ mais la fonction h ne dépend pas de θ.

2. Réciproquement, si on a la factorisation $p_\theta(\underline{x}) = g_\theta(S(\underline{x})) \cdot h(x)$, appliquer la formule de Bayes à $p_\theta(\underline{x}|s)$ pour montrer que $S = S(\underline{X})$ est une statistique suffisante.
3. Application : On considère le modèle i.i.d. Bernoulli $X \sim \mathscr{B}(\theta)$. Montrer que la somme $S = \sum_{i=1}^N X_i$ est une statistique suffisante.
4. Même question pour le modèle gaussien $X \sim \mathcal{N}(\theta, \sigma^2)$
5. On considère le modèle i.i.d. uniforme $X \sim \mathscr{U}[0, \theta]$. Ce modèle est-il régulier ? On notant

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

   les échantillons $X = (X_1, X_2, \dots, X_N)$ *réordonnés par ordre croissant*, montrer que $S = X_{(N)}$ est une statistique suffisante.

II Théorème de Rao-Blackwell

On se donne un estimateur quelconque $\hat{\theta} = \hat{\theta}(\underline{X})$ (pas nécessairement non biaisé) et $S = S(\underline{X})$ une *statistique suffisante* pour θ (voir l'exercice précédent). Soit l'espérance conditionnelle

$$\hat{\theta}^* = \mathbf{E}(\hat{\theta}|S).$$

1. Montrer que c'est bien une fonction de $\underline{X}$ qui ne dépend pas de θ (c'est bien un estimateur).
2. En appliquant la formule de l'espérance totale[1], comparer les biais de $\hat{\theta}$ et de $\hat{\theta}^*$.
3. En appliquant la formule de la variance totale[2], comparer les variances de $\hat{\theta}$ et de $\hat{\theta}^*$. Conclure!
4. Application : On considère le modèle i.i.d. Bernoulli $X \sim \mathscr{B}(\theta)$ avec $N \geqslant 2$ observations et $S = \sum_{i=1}^N X_i$. En prenant l'estimateur inadmissible $\hat{\theta}(\underline{X}) = X_1$, trouver l'estimateur $\hat{\theta}^*$ de Rao-Blackwell.
   (Indication : justifier que $\mathbf{E}(X_1|S) = \mathbf{E}(X_i|S)$ pour tout $i = 1, \dots, N$ et utiliser la linéarité de l'espérance).
5. Même question pour le modèle gaussien $X \sim \mathcal{N}(\theta, \sigma^2)$ avec toujours $S = \sum_{i=1}^N X_i$.
6. *Plus difficile* On considère le modèle i.i.d. uniforme $X \sim \mathscr{U}[0, \theta]$ avec $S = X_{(N)}$. Justifier que l'estimateur inadmissible $\hat{\theta} = 2X_1$ est non biaisé, et trouver l'estimateur $\hat{\theta}^*$ de Rao-Blackwell.

III Théorème de Lehmann-Scheffé

On se donne un estimateur *non biaisé* $\hat{\theta} = \hat{\theta}(\underline{X})$ et on note (voir l'exercice précédent) $\hat{\theta}^* = \mathbf{E}(\hat{\theta}|S)$ où $S = S(\underline{X})$ une *statistique suffisante "complète"*, c'est-à-dire qu'elle vérifie la propriété suivante pout toute fonction f :

$$(\forall \theta) \quad \mathbf{E}_\theta(f(S)) = 0 \quad \Longrightarrow \quad f(S) = 0 \text{ p.s.}$$

1. Soit $\hat{\theta}'$ un autre estimateur non biaisé. Montrer (en prenant la différence) que $\mathbf{E}(\hat{\theta}|S) = \mathbf{E}(\hat{\theta}'|S)$ p.s.
2. En déduire que l'estimateur $\hat{\theta}^*$ de Rao-Blackwell est MVU.
3. En particulier, si $\mathbf{E}_\theta(S) = 0$, montrer que S est elle-même un estimateur MVU.
4. Application : On considère le modèle i.i.d. Bernoulli $X \sim \mathscr{B}(\theta)$. (Re)trouver l'estimateur MVU.
5. Même question pour le modèle i.i.d. gaussien $X \sim \mathcal{N}(\theta, \sigma^2)$.
6. Même question pour le modèle i.i.d. uniforme $X \sim \mathscr{U}[0, \theta]$.

---

1. $\mathbf{E}(X) = \mathbf{E}(\mathbf{E}(X|Y))$. (On rappelle par ailleurs que $\mathbf{E}(X|X) = X$.)
2. $\mathbf{V}(X) = \mathbf{E}(\mathbf{V}(X|Y)) + \mathbf{V}(\mathbf{E}(X|Y))$.

Solutions

I-1) $p_\theta(x) = p_\theta(\underline{x}, s) = p_\theta(x|s)p_\theta(s = S(x)) = p(x|s)p_\theta(s = S(x))$ de la forme voulue avec $h(x) = p(x|s)$ qui ne dépend pas de θ et $g_\theta(s) = p_\theta(s)$ qui dépend de $S(x)$.

I-2) Bayes : $p_\theta(x|s) = \frac{p(s|x)p_\theta(x)}{\int p(s|x)p_\theta(x)d\mu(x)} = \frac{g_\theta(s)h(x)p(s|x)}{g_\theta(s)\int p(s|x)h(x)d\mu(x)} = \frac{h(x)p(s|x)}{\int p(s|x)h(x)d\mu(x)}$ ne dépend pas de θ.

I-3) $p_\theta(\underline{x}) = \prod_{i=1}^N \theta^{x_i}(1-\theta)^{1-x_i} = (\frac{\theta}{1-\theta})^{\sum_i x_i}(1-\theta)$ de la forme voulue avec $h \equiv 1$ et $S(\underline{x}) = \sum_i x_i$.

I-4) $p_\theta(\underline{x}) = \prod_{i=1}^N (2\pi\sigma^2)^{-N/2}\exp(-\frac{1}{2\sigma^2}\sum_i(x_i-\theta)^2) = \underbrace{(2\pi\sigma^2)^{-N/2}\exp(-\frac{1}{2\sigma^2}\sum_i x_i^2)}_{h(\underline{x})}\cdot\underbrace{\exp(\frac{\theta}{\sigma^2}\sum_i x_i - \frac{N\theta^2}{2\sigma^2})}_{g_\theta(\sum_i x_i)}$

de la forme voulue avec toujours $S(\underline{x}) = \sum_i x_i$.

I-5) $p_\theta(\underline{x}) = \prod_{i=1}^N \frac{1}{\theta}\mathbb{1}_{0\le x_i\leqslant\theta} = \frac{1}{\theta^N}\mathbb{1}_{0\le \max_i x_i\leqslant\theta}$ de la forme voulue avec $S(\underline{x}) = \max_i x_i = x_{(N)}$.

II-1) Par définition de l'espérance conditionnelle, $E(\hat\theta|S)$ est une fonction de $S = S(\underline{X})$, donc de $\underline{X}$, qui s'écrit $E(\hat\theta|S) = \int\hat\theta(x)p(x|S)\,\mathrm{d}\mu(x)$ où par définition d'une statistique suffisante, $p(x|S)$ ne dépend pas de θ. (Comme $\hat\theta(x)$ est un estimateur, la fonction $\hat\theta()$ ne dépend pas de θ non plus.)

II-2) $\mathbf{E}(\hat\theta^*) = \mathbf{E}\mathbf{E}(\hat\theta|S) = \mathbf{E}(\hat\theta) = \theta$. Donc les deux biais $\mathbf{E}(\hat\theta^*) - \theta$ et $\mathbf{E}(\hat\theta) - \theta$ sont égaux.

II-3) $\mathbf{V}(\hat\theta) = \mathbf{V}(\mathbf{E}(\hat\theta|S)) + \mathbf{E}(\mathbf{V}(\hat\theta|S)) \geqslant \mathbf{V}(\mathbf{E}(\hat\theta|S)) = \mathbf{V}(\hat{\theta^*})$. Donc (même biais et variance plus faible), l'estimateur $\hat\theta^*$ de Rao-Blackwell est meilleur (risque quadratique plus faible) que $\hat\theta$.

II-4) $\hat\theta^* = \mathbf{E}(X_1|S) = \mathbf{E}(X_i|S)$ pour tout $i = 1,\dots,N$ car les données sont i.i.d. Donc par linéarité, $\hat\theta^* = \frac{1}{N}\sum_{i=1}^N \mathbf{E}(X_i|S) = \frac{1}{N}\mathbf{E}(S|S) = \frac{S}{N} = \bar{X}_N$.

II-5) Exactement le même calcul que dans la question précédente.

II-6) $\hat\theta^* = \mathbf{E}(2X_1|X_{(N)} = \frac{2}{N}\mathbf{E}(\sum_i X_i|X_{(N)}$ comme ci-dessus. D'une part $\mathbf{E}(X_{(N)}|X_{(N)}) = X_{(N)}$ et d'autre part, $X_i|X_{(N)} = m$ suit la loi uniforme $\mathscr{U}[0,m]$ lorsque $i < N$ d'où $\mathbf{E}(X_{(i)}|X_{(N)}) = \frac{X_{(N)}}{2}$. Finalement $\hat\theta^* = \frac{2}{N}((N-1)X_{(N)}/2 + X_{(N)}) = \frac{N+1}{N}X_{(N)}$ qui est bien non biaisé comme $2X_1$ (exercice fait en classe).

III-1) $\mathbf{E}(\hat\theta|S) - \mathbf{E}(\hat\theta'|S) = \mathbf{E}(\hat\theta - \hat\theta'|S)$. Or par la loi de l'espérance totale, $\mathbf{E}_\theta\mathbf{E}(\hat\theta - \hat\theta'|S) = \mathbf{E}(\hat\theta - \hat\theta') = \theta - \theta = 0$ donc comme la statistique suffisante est *complète*, $\mathbf{E}(\hat\theta - \hat\theta'|S) = 0$ p.s.

III-2) D'après la question précédente, $\hat{\theta}^* = \mathbf{E}(\hat{\theta}|S) = \mathbf{E}(\hat{\theta}'|S)$, dont on sait (exercice précédent) qu'il est meilleur (de variance plus faible) que $\hat{\theta}'$. Or comme $\hat{\theta}'$ est quelconque, $\hat{\theta}^*$ non biaisé est de variance plus faible que tout estimateur, c'est-à-dire MVU.

III-3) Si $\mathbf{E}_\theta(S) = 0$, $S = S(\underline{X})$ est un estimateur non biaisé et $\mathbf{E}(S|S) = S$ est MVU.

III-4) Il s'agit de montrer que $S = \bar{X}_N$ est une statistique complète. Comme $\sum_i X_i$ suit la loi binomiale $\mathscr{B}(N,\theta)$, $\mathbf{E}(f(\bar{X}_N)) = \sum_{k=0}^N f(k/N)\binom{N}{k}\theta^k(1-\theta)^{N-k} = 0$ pour tout $\theta \in [0,1]$ implique $f(k/N) = 0$ pour tout k, c'est à dire $f(\bar{X}_N) = 0$. On retrouve que $\bar{X}_N$ est MVU.

III-5) Il s'agit de montrer que $S = \bar{X}_N$ est une statistique complète. Comme $\sum_i X_i$ suit la loi normale $\mathcal{N}(\theta,\sigma^2/N)$, $\mathbf{E}(f(\bar{X}_N)) \propto \int f(x)\exp(-\frac{(x-\theta)^2}{2\sigma^2/N})\,\mathrm{d}x \propto \int f(x)\exp(-\frac{\theta x}{2\sigma^2/N}) = 0$ implique $f = 0$ p.p. (transformée de Laplace) donc $f(\bar{X}_N) = 0$ p.s.

III-6) Il s'agit de montrer que $S = X_{(N)}$ est une statistique complète. D'après la loi suivie par $X_{(N)}$ (vu en exercice en classe), on a $\mathbf{E}(f(X_{(N)})) = \int_0^\theta f(x)\frac{Nx^{N-1}}{\theta^N}\,\mathrm{d}x$. Si cette quantité $= 0$ pour tout θ, on obtient par dérivation p.p. : $f = 0$ p.p. donc $f(X_{(N)}) = 0$ p.s.