Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

1er décembre 2023
Aucun document autorisé
Durée : 1 heure

Contrôle Continu

Rappels

- Il sera tenu compte du soin apporté à la présentation et à la rédaction.
- Si vous n'arrivez pas à démontrer un résultat, vous pouvez l'admettre pour la suite de l'exercice.
- Pour $\lambda > 0$ et $p > 0$, si $T \sim \Gamma(p, \lambda)$ alors $\mathbb{E}[T] = \frac{p}{\lambda}$ et $\mathrm{Var}(T) = \frac{p}{\lambda^2}$. De plus, si $p > 2$, alors

$$\mathbb{E}\left[\frac{1}{T}\right] = \frac{\lambda}{p-1}, \qquad \text{et} \qquad \mathbb{E}\left[\frac{1}{T^2}\right] = \frac{\lambda^2}{(p-1)(p-2)}.$$

Exercice 1 On part du modèle linéaire gaussien classique

$$Y = X\beta + \boldsymbol{\varepsilon},$$

où X est la matrice $(n \times p)$ du plan d'expérience, $\beta = [\beta_1, \dots, \beta_p]'$ un vecteur de $\mathbb{R}^p$, Y est le vecteur $(n \times 1)$ des observations, et $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$. Les paramètres β et σ^2 sont supposés inconnus et on souhaite les estimer. Néanmoins, on n'observe pas comme d'habitude les Y_i, mais des moyennes par classe. Plus précisément, les n données sont réparties en L classes $C_1, \dots, C_L$ d'effectifs respectifs **connus** $n_1, \dots, n_L$ et on a seulement accès aux moyennes par classe, à savoir pour tout $\ell \in \{1, \dots, L\}$ et pour tout $j \in \{1, \dots, p\}$:

$$\overline{Y}_\ell = \frac{1}{n_\ell} \sum_{i \in C_\ell} Y_i \qquad \text{et} \qquad \overline{x}_{\ell j} = \frac{1}{n_\ell} \sum_{i \in C_\ell} x_{ij},$$

avec $\overline{X} := [\overline{x}_{\ell j}]_{1 \leq \ell \leq L, 1 \leq j \leq p}$ supposée de rang p. En notant $\overline{\varepsilon}_\ell = \frac{1}{n_\ell} \sum_{i \in C_\ell} \varepsilon_i$, le modèle peut donc se mettre sous la forme

$$\overline{Y} = \overline{X}\beta + \overline{\boldsymbol{\varepsilon}}.$$

1. Donner la loi de $\overline{\boldsymbol{\varepsilon}}$.
   Correction. Les variables $\overline{\varepsilon}_\ell$ sont indépendantes de gaussiennes centrées et de variances respectives σ^2/n_ℓ donc $\overline{\boldsymbol{\varepsilon}} \sim \mathcal{N}(0, \sigma^2 \Delta^{-2})$, où Δ est la matrice diagonale dont l'élément (ℓ, ℓ) vaut $\sqrt{n_\ell} > 0$, de sorte que Δ^{-2} est la matrice diagonale dont l'élément (ℓ, ℓ) vaut $1/n_\ell$.
2. Déterminer une matrice diagonale Δ telle que si l'on définit $X^* = \Delta\overline{X}$, $Y^* = \Delta\overline{Y}$ et $\boldsymbol{\varepsilon}^* = \Delta\overline{\boldsymbol{\varepsilon}}$, on ait $Y^* = X^*\beta + \boldsymbol{\varepsilon}^*$, avec $\boldsymbol{\varepsilon}^* \sim \mathcal{N}(0, \sigma^2 I_L)$.
   Correction. Avec la matrice Δ définie ci-dessus, la propriété de stabilité des vecteurs gaussiens par transformation linéaire donne bien $\boldsymbol{\varepsilon}^* \sim \mathcal{N}(0, \sigma^2 I_L)$.
3. Justifier que la matrice X^* est de rang p.
   Correction. Puisque $X^* = \Delta\overline{X}$ avec $\overline{X}$ de rang p et Δ bijective, le rang de la matrice X^* est lui aussi égal à p.

4. En déduire un estimateur $\widehat{\beta}$ de β et préciser sa loi.
Correction. Comme X^* est de rang p, l'estimateur des moindres carrés de β dans le modèle $Y^* = X^*\beta + \boldsymbol{\varepsilon}^*$ est $\widehat{\beta} := ((X^*)'X^*)^{-1}(X^*)'Y^*$ et $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2((X^*)'X^*)^{-1})$.
5. Déterminer de même un estimateur sans biais $\widehat{\sigma}^2$ de σ^2 et préciser sa loi.
Correction. En notant $\widehat{Y}^* := X^*\widehat{\beta}$, un estimateur sans biais de σ^2 est $\widehat{\sigma}^2 := \|Y^* - \widehat{Y}^*\|^2/(L-p)$, avec $(L-p)\widehat{\sigma}^2/\sigma^2 \sim \chi^2_{L-p}$.
6. Quelle est la variance de $\widehat{\sigma}^2$? La comparer à celle de l'estimateur sans biais $\widetilde{\sigma}^2$ dans le modèle linéaire gaussien initial $Y = X\beta + \boldsymbol{\varepsilon}$ où l'on a accès à toutes les observations $(Y_i)_{1 \le i \le n}$. Interpréter le résultat.

   Correction. On commence par noter que le rang de $\overline{X}$ étant égal à p, il en va de même pour celui de X. Ceci peut se voir par l'absurde : si tel n'était pas le cas, il existerait un vecteur $\alpha \in \mathbb{R}^p \backslash \{0\}$ tel que $X\alpha = 0$, mais alors on aurait aussi $\overline{X}\alpha = 0$, ce qui est exclu. Dès lors, on sait que dans le modèle classique, $(n-p)\widetilde{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}$, donc

$$\mathrm{Var}(\widetilde{\sigma}^2) = \frac{2\sigma^4}{n-p} \le \mathrm{Var}(\widehat{\sigma}^2) = \frac{2\sigma^4}{L-p}.$$

   Ce résultat est intuitivement clair : en agrégeant des données, on a perdu de l'information donc on aboutit à un estimateur moins précis pour σ^2.
   Remarque : on a vu que si $\overline{X}$ est de rang p, alors X aussi. La réciproque est fausse, comme on peut s'en rendre compte en considérant les classes $C_1 = \{1\}$ et $C_2 = \{2, 3\}$ pour la matrice

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix}.$$

Exercice 2 Soient $\lambda > 0$ et $n \ge 5$ entier. On considère le modèle bayésien suivant :

$$\boldsymbol{\theta} \sim \Pi = \mathcal{E}(\lambda),$$
$$\mathbf{X} = (X_1, \ldots, X_n) \,\big|\, \boldsymbol{\theta} \sim \mathcal{N}\left(0, \frac{1}{\boldsymbol{\theta}}\right)^{\otimes n}.$$

1. Déterminer la loi a posteriori $\Pi\left[\cdot \,\middle|\, \mathbf{X}\right]$. On notera $S^2_{\mathbf{X}} = \sum_{i=1}^n X_i^2$.
Correction. Pour tout $\theta > 0$, on a

$$\pi(\theta \,|\, \mathbf{X}) \propto e^{-\lambda\theta}\theta^{n/2}e^{-\frac{1}{2}\theta\sum_{i=1}^n X_i^2}\mathbb{1}_{\theta>0} \propto \theta^{n/2}e^{-\theta\left(\lambda + \frac{1}{2}S^2_{\mathbf{X}}\right)}\mathbb{1}_{\theta>0}$$

   donc

$$\Pi\left[\cdot \,\middle|\, \mathbf{X}\right] = \Gamma\left(1 + \frac{n}{2}, \lambda + \frac{1}{2}S^2_{\mathbf{X}}\right).$$

2. Donner un estimateur de Bayes pour la perte quadratique.
Correction. Un estimateur de Bayes pour la perte quadratique est la moyenne a posteriori, i.e.

$$T^\star(\mathbf{X}) = \mathbb{E}[\boldsymbol{\theta} \,|\, \mathbf{X}] = \frac{n+2}{2\lambda + S^2_{\mathbf{X}}}.$$

3. On considère maintenant l'estimateur $\widehat{\theta}_n(\mathbf{X}) = \frac{n}{S^2_{\mathbf{X}}}$. L'objectif de cette question est de calculer son risque de Bayes pour Π et la perte quadratique.

(a) Soit $\theta > 0$. Sachant $\boldsymbol{\theta} = \theta$, quelle loi suit $\theta S^2_{\mathbf{X}}$?
Correction. Sachant $\boldsymbol{\theta} = \theta$, les variables $\sqrt{\theta} X_i$ sont i.i.d. gaussiennes standards donc

$$\theta S^2_{\mathbf{X}} = \sum_{i=1}^{n} (\sqrt{\theta} X_i)^2 \sim \chi^2_n.$$

(b) On rappelle que, pour tout entier non-nul d, la loi χ^2_d est identique à la loi $\Gamma\left(\frac{d}{2}, \frac{1}{2}\right)$. En déduire, pour tout $\theta > 0$, $\mathbb{E}_\theta\left[\frac{1}{S^2_{\mathbf{X}}}\right]$ et $\mathbb{E}_\theta\left[\frac{1}{S^4_{\mathbf{X}}}\right]$.
Correction. On a

$$\mathbb{E}_\theta\left[\frac{1}{S^2_{\mathbf{X}}}\right] = \theta \mathbb{E}\left[\frac{1}{\theta S^2_{\mathbf{X}}}\right] = \frac{\theta}{n-2} \qquad \text{et} \qquad \mathbb{E}_\theta\left[\frac{1}{S^4_{\mathbf{X}}}\right] = \frac{\theta^2}{(n-2)(n-4)}.$$

(c) En déduire que, pour tout $\theta > 0$, le risque ponctuel vaut

$$\mathbf{R}(\theta, \widehat{\theta}_n(\mathbf{X})) = \frac{2(n+4)}{(n-2)(n-4)} \theta^2.$$

Correction. On a

$$\mathbb{E}_\theta\left[\left(\theta - \frac{n}{S^2_{\mathbf{X}}}\right)^2\right] = \theta^2 - 2n\theta \mathbb{E}_\theta\left[\frac{1}{S^2_{\mathbf{X}}}\right] + n^2 \mathbb{E}_\theta\left[\frac{1}{S^4_{\mathbf{X}}}\right] = \frac{2(n+4)}{(n-2)(n-4)} \theta^2.$$

(d) En déduire le risque de Bayes pour Π de l'estimateur $\widehat{\theta}_n(\mathbf{X})$.
Correction. Puisque $\boldsymbol{\theta} \sim \mathcal{E}(\lambda)$, on sait que $\mathbb{E}[\boldsymbol{\theta}^2] = 2/\lambda^2$ donc

$$\mathbf{R}_B(\Pi, \widehat{\theta}_n) = \frac{2(n+4)}{(n-2)(n-4)} \mathbb{E}[\boldsymbol{\theta}^2] = \frac{4(n+4)}{(n-2)(n-4)} \times \frac{1}{\lambda^2}.$$

4. (a) L'objectif est maintenant d'utiliser la méthode du Bayésien empirique pour calibrer λ. Montrer que, pour tout $\lambda > 0$, la densité marginale f_λ de la loi de $\mathbf{X}$ vérifie :

$$f_\lambda(\mathbf{X}) \propto \frac{\lambda}{\left(\lambda + \frac{S^2_{\mathbf{X}}}{2}\right)^{\frac{n}{2}+1}}$$

et déterminer l'estimateur du maximum de vraisemblance marginale de λ.
Correction. Pour tout $\lambda > 0$, on a

$$f_\lambda(\mathbf{X}) \propto \int_0^{+\infty} \lambda \theta^{n/2} e^{-\theta\left(\lambda + \frac{1}{2} S^2_{\mathbf{X}}\right)} d\theta \propto \frac{\lambda}{\left(\lambda + \frac{1}{2} S^2_{\mathbf{X}}\right)^{\frac{n}{2}+1}}.$$

En passant à la log-vraisemblance, on a

$$\ell_\lambda(\mathbf{X}) = c + \log(\lambda) - (n/2+1) \log\left(\lambda + \frac{1}{2} S^2_{\mathbf{X}}\right),$$

puis

$$\ell'_\lambda(\mathbf{X}) = \frac{1}{\lambda} - (n/2+1) \frac{1}{\lambda + \frac{1}{2} S^2_{\mathbf{X}}} > 0 \iff \lambda < \frac{1}{n} S^2_{\mathbf{X}}.$$

Le maximum est donc atteint en $\widehat{\lambda}(\mathbf{X}) = \frac{1}{n} S^2_{\mathbf{X}}$.

(b) Donner la pseudo-loi a posteriori obtenue par la méthode du Bayésien empirique dans ce cas. Que retrouve-t-on en considérant la pseudo-moyenne a posteriori associée ?
Correction. La pseudo-loi a posteriori obtenue par bayésien empirique est alors la loi $\Gamma\left(\frac{n+2}{2}, \frac{n+2}{2n} S^2_{\mathbf{X}}\right)$, qui, sachant $\mathbf{X}$, est d'espérance $\widehat{\theta}_n(\mathbf{X}) = \frac{n}{S^2_{\mathbf{X}}}$.