PC 8 : Estimation statistique

1 Estimation de paramètres

Exercice 1 (Loi géométrique)**.** On observe $\mathbf{x} = (x_1, \dots, x_n)$ que l'on considère comme la réalisation du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$, où les X_i sont des variables aléatoires i.i.d. de la loi géométrique $\mathrm{Geo}(\theta)$ de paramètre inconnu $0 < \theta < 1$.

1. Calculer l'estimateur par la méthode des moments $\hat{\theta}_n^{\mathrm{MM}}$ de θ (s'il existe). On utilisera le moment d'ordre 1.
2. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\mathrm{MV}}$ de θ (s'il existe).
3. Vérifier si l'estimateur $\hat{\theta}_n^{\mathrm{MV}}$ est convergent.
4. Montrer que $\hat{\theta}_n^{\mathrm{MV}}$ est asymptotiquement normal et préciser sa vitesse de convergence.

Exercice 2 (Fonction caractéristique, TLC)**.** Soient X_i, $i = 1 \cdots, n$ variables aléatoires *i.i.d.* de loi de Poisson de paramètre $\lambda > 0$.
1. Calculer la loi de $S_n = n\overline{X}$.
2. Trouver deux suites $(a_n)_n$ et $(b_n)_n$ telles que $a_n\overline{X} + b_n$ converge en loi vers une variable de loi non dégénérée.

Exercice 3 (EMM, EMV, CV asymptotique)**.** Soient n variables aléatoires i.i.d. $X_1, \dots, X_n$, de densité de Pareto

$$f_\theta(x) = \frac{\theta}{x^{\theta+1}} \mathbb{1}_{\{x \geq 1\}},$$

où $\theta > 0$ est un paramètre inconnu que l'on souhaite estimer.
1. On suppose d'abord que l'ensemble des paramètres est $\Theta = \{\theta > 1\}$. Estimer θ par la méthode des moments. On utilisera le moment d'ordre 1.
2. On suppose maintenant que l'ensemble des paramètres est $\Theta = \{\theta > 0\}$. Peut-on utiliser le moment d'ordre 1 ?
3. $\Theta = \{\theta > 0\}$. Estimer θ par la méthode des moments avec un autre moment que le moment d'ordre 1.
4. $\Theta = \{\theta > 0\}$. Estimer θ par celle du maximum de vraisemblance.
5. Étudier la loi limite de l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$, c'est-à-dire la convergence en loi de $\sqrt{n}(\widehat{\theta}_n^{EMV} - \theta)$.

Exercice 4 (Loi de Cauchy)**.** On observe $\mathbf{x} = (x_1, \dots, x_n)$ que l'on considère comme la réalisation du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$, où les X_i sont des variables aléatoires i.i.d. de la loi de Cauchy avec un paramètre d'échelle dont la densité est donnée par

$$f_\theta(x) = \frac{\theta}{\pi(\theta^2 + x^2)},$$

pour $\theta > 0$ inconnu.

1. Calculer l'estimateur par la méthode des moments $\hat{\theta}_n^{\mathrm{MM}}$ de θ (s'il existe). On utilisera le moment d'ordre 1. Vérifier si l'estimateur est convergent et asymptotiquement normal, et calculer son risque quadratique moyen.
2. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\mathrm{MV}}$ de θ (s'il existe). Que dire des propriétés de cet estimateur ?

2 Modèle statistique

Exercice 5 (Méthode de capture-recapture)**.** On souhaite estimer le nombre N de poissons vivant dans un bassin. Pour cela, on en pêche k que l'on marque. Ensuite, on pêche et on relâche successivement n poissons et l'on compte le nombre X de poissons marqués parmi les n. Donner l'estimateur du maximum de vraisemblance $\hat{N}$ de N en fonction de X, n et k. Montrer que $\hat{N}$ est convergent et donner une approximation de sa loi quand n est grand. Comment la variance de $\hat{N}$ dépend du choix de k ?

Exercice 6 (Ampoules défaillantes)**.** Un statisticien observe pendant n jours le nombre d'ampoules défaillantes par jour à la sortie d'une chaîne de fabrication, noté $x_1, \dots, x_n$. Il considère que $x_1, \dots, x_n$ sont des réalisations i.i.d. d'une loi $\mathbb{P}$, et il souhaite estimer la probabilité p de n'avoir aucune ampoule défaillante par jour ($p = \mathbb{P}(X = 0)$).

1. Dans un premier temps, il compte le nombre de jours où aucune ampoule n'est défaillante, noté N_n, qui est alors le nombre de X_i, $i = 1, \dots, n$, égaux à 0. Il propose d'estimer la probabilité p par
$$\hat{p}_1 = \frac{1}{n} N_n.$$
Montrer que l'estimateur est convergent et sans biais. Calculer son risque quadratique, et donner sa loi limite.
2. Désormais le statisticien suppose que X_i suivent une loi de Poisson $\mathrm{Poi}(\lambda)$ de paramètre $\lambda > 0$ inconnu. Il propose comme estimateur de p
$$\hat{p}_2 = \mathrm{e}^{-\bar{X}_n},$$
où $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
   (i) Expliquer sa démarche. Montrer que $\hat{p}_2$ est convergent. Calculer sa variance et son biais. Déterminer des équivalents asymptotiques des quantités précédentes.
   (ii) Montrer que l'on peut choisir t_n tel que $\hat{p}_3 = \mathrm{e}^{-t_n \bar{X}_n}$ soit sans biais.
   (iii) Lequel de $\hat{p}_1$, $\hat{p}_2$ et $\hat{p}_3$ choisiriez-vous pour estimer $\mathrm{e}^{-\lambda}$?

3 Estimation statistique dans le modèle linéaire

Exercice 7 (Régression linéaire)**.** Soit $X \in M_{n,p}(\mathbb{R})$, $\beta \in \mathbb{R}^p$ avec $p < n$. On définit
$$Y = X\beta + \varepsilon,$$
où $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$. La matrice X est connue et on observe Y. L'objectif est d'estimer β. On supposera que X est de rang maximal, ce qui implique en particulier que $X'X$ est inversible.

1. Quelle est la loi du vecteur $Y = (Y_1, \dots, Y_n)'$? Les cordonnées Y_i sont-elles indépendantes et identiquement distribuées ?
2. La définition de l'estimateur du maximum de vraisemblance se généralise à des observations non i.i.d.. La fonction de vraisemblance est alors définie comme la densité du vecteur d'observations. Ainsi, l'estimateur du maximum de vraisemblance $\hat{\beta}$ de β est défini comme le point où la densité de Y, notée f_β, est maximale (on maximise $\beta \mapsto f_\beta(Y)$).
Montrer que $\hat{\beta}$ minimise $\beta \mapsto \|Y - X\beta\|_2^2$. En déduire que
$$\hat{\beta} = (X'X)^{-1}X'Y.$$
Donner la loi de $\hat{\beta}$.
3. On définit un estimateur de σ^2 par
$$\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|_2^2}{n-p}.$$
Montrer que $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants.

4. Donner la loi de $\hat{\sigma}^2$ et en déduire qu'il s'agit d'un estimateur sans biais qui converge en probabilité vers σ^2.
Rappel du Théorème de Cochran : Soient $(V_i)_{1\leq i\leq n}$ une décomposition orthogonale en somme directe de $\mathbb{R}^n$ et $(\Pi_{V_i})_{1\leq i\leq n}$ les matrices de projection orthogonale associées. Pour un vecteur gaussien $Z \sim \mathcal{N}_n(0,\sigma^2 I_n)$, on a
— $\Pi_{V_1}Z, \cdots, \Pi_{V_k}Z$ *sont des vecteurs aléatoires indépendants de loi* $\mathcal{N}_n(0,\sigma^2\Pi_{V_i})$.
— *Pour tout* $1 \leq i \leq n$, $\|\Pi_{V_j}Z\|_2^2/\sigma^2$ *suit la loi* $\chi^2_{p_i}$ *où* p_i *est la dimension de* V_i.

4 Exercices corrigés

Exercice 8 (Stabilisation de la variance)**.** On dispose d'un échantillon $X_1,\ldots,X_n$ de v.a. *i.i.d.* de même loi de Bernoulli de paramètre $p \in]0,1[$.
1. On pose $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ (moyenne empirique). Que nous dit la loi des grands nombres ? Et le théorème de la limite centrale ?
2. Trouver une fonction g, qui ne dépend pas de p, telle que $\sqrt{n}\left(g(\overline{X}_n) - g(p)\right)$ converge en loi vers une loi gaussienne centrée réduite.
Indication : utiliser la méthode Delta.

Solution. 1. La loi (forte) des grands nombres nous dit que $\overline{X}_n \xrightarrow{\text{p.s.}} \mathbb{E}_\theta(X_1) = p$ et le TLC nous dit que $\sqrt{n}(\overline{X}_n - p) \xrightarrow{\text{loi}} \mathcal{N}(0,p(1-p))$.

2. On utilise la Delta-méthode : si g est une fonction dérivable en p telle que $g'(p) \neq 0$ alors

$$\sqrt{n}\left(g(\overline{X}_n) - g(p)\right) \xrightarrow{\text{loi}} \mathcal{N}\big(0, p(1-p)(g'(p))^2\big).$$

On cherche donc g telle que $x(1-x)(g'(x))^2 = 1$, c.-à-d. $g'(x) = \frac{1}{\sqrt{x(1-x)}}$, $x \in]0,1[$.

À une constante additive près, on a $g(x) = 2\arcsin(\sqrt{x}\,)$, car $\arcsin'(x) = \left(1-x^2\right)^{-1/2}$.[1] Donc

$$\sqrt{n}\left(\arcsin\left(\sqrt{\overline{X}_n}\,\right) - \arcsin\left(\sqrt{p}\right)\right) \xrightarrow{\text{loi}} \mathcal{N}(0,1)\,.$$

Exercice 9 (Comparaison de deux estimateurs selon le risque quadratique moyen)**.** Soit $U_1,\ldots,U_n$ des v.a. *i.i.d.* de même loi uniforme sur $[0,\theta]$ où $\theta > 0$ est inconnu. Soit

$$T_n^{(1)} = 2\,\frac{U_1+\cdots+U_n}{n} \quad \text{et} \quad T_n^{(2)} = \frac{n+1}{n}\,\max\{U_1,\ldots,U_n\}.$$

1. Montrer que $T_n^{(1)} \xrightarrow{\text{p.s.}} \theta$ et $T_n^{(2)} \xrightarrow{\text{p.s.}} \theta$. *Indication : on commencera par montrer que $T_n^{(2)}$ converge en probabilité*
2. Calculer l'espérance et la variance de $T_n^{(1)}$, puis celles de $T_n^{(2)}$.
Indication : commencer par trouver la loi de $Y = \max\{U_1,\ldots,U_n\}$.
3. Comparer les deux estimateurs du point de vue du risque quadratique moyen.

Solution. On a $\mathbb{E}_\theta(U_i) = \theta/2$ et $\mathbb{V}ar_\theta(U_i) = \theta^2/12$.

1. Par la LFGN on a $T_n^{(1)} \xrightarrow{\text{p.s.}} \theta$ car $\frac{U_1+\cdots+U_n}{n} \xrightarrow{\text{p.s.}} \mathbb{E}_\theta(U_1) = \frac{\theta}{2}$. Pour l'autre estimateur, il suffit de montrer que
$$M_n := \max\{U_1,\ldots,U_n\} \xrightarrow{\text{p.s.}} \theta.$$
Pour tout $\epsilon \in]0,\theta[$, on a
$$\mathbb{P}(|M_n-\theta| > \epsilon) = \mathbb{P}(M_n < \theta-\epsilon) = \big(\mathbb{P}(U_1 < \theta-\epsilon)\big)^n = \left(\frac{\theta-\epsilon}{\theta}\right)^n.$$
On a donc convergence en probabilité de M_n vers θ et par le lemme de Borel-Cantelli on déduit que cette convergence est en fait presque sûre. Donc $T_n^{(2)} \xrightarrow{\text{p.s.}} \theta$.

1. On retrouve ce résultat facilement si on utilise que pour tout $u \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ on a $\arcsin(\sin u) = u$ et qu'on dérive.

2. On a
$$\mathbb{E}_\theta(T_n^{(1)}) = \theta \quad \text{et} \quad \mathbb{V}ar_\theta(T_n^{(1)}) = \frac{4}{n}\,\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$
On suppose que $n \geq 2$. La f.r. de $Y = \max\{U_1, \ldots, U_n\}$ est
$$F_Y(y) = \begin{cases} 0 & \text{si} \quad y < \theta \\ \big(\mathbb{P}(U_1 \leq y)\big)^n = \big(F_{U_1}(y)\big)^n = \left(\frac{y}{\theta}\right)^n & \text{si} \quad y \in [0, \theta[\\ 1 & \text{si} \quad y \geq \theta. \end{cases}$$
Donc la densité de Y est $f_Y(y) = \frac{n}{\theta^n}\, y^{n-1}\, \mathbb{1}_{[0,\theta]}(y)$. On a
$$\mathbb{E}_\theta(Y) = \int_0^\theta y\, \frac{n}{\theta^n}\, y^{n-1}\, \mathrm{d}y = \frac{n}{n+1}\,\theta$$
donc $\mathbb{E}_\theta(T_n^{(2)}) = \frac{n+1}{n}\mathbb{E}_\theta(Y) = \theta$. Calculons la variance de Y.
$$\mathbb{E}_\theta(Y^2) = \int_0^\theta y^2\, \frac{n}{\theta^n}\, y^{n-1}\, \mathrm{d}y = \frac{n}{n+2}\,\theta^2.$$
Donc
$$\mathbb{V}ar_\theta(Y) = \frac{n}{n+2}\,\theta^2 - \frac{n^2}{(n+1)^2}\,\theta^2 = \frac{n}{(n+2)(n+1)^2}\,\theta^2$$
et
$$\mathbb{V}ar_\theta(T_n^{(2)}) = \frac{(n+1)^2}{n^2}\,\mathbb{V}ar_\theta(Y) = \frac{1}{n(n+2)}\,\theta^2.$$
3. Les deux estimateurs sont sans biais, donc il suffit de comparer leur variance :
$$\mathbb{V}ar_\theta(T_n^{(1)}) - \mathbb{V}ar_\theta(T_n^{(2)}) = \left(\frac{1}{3n} - \frac{1}{n(n+2)}\right)\theta^2 = \frac{n-1}{3n(n+2)}\,\theta^2.$$
C'est une quantité strictement positive dès que $n \geq 2$, quel que soit $\theta > 0$, donc $T_n^{(2)}$ est meilleur que $T_n^{(1)}$ selon le critère du risque quadratique moyen.

Exercice 10 (Loi de Poisson)**.** On observe $\mathbf{x} = (x_1, \ldots, x_n)$ que l'on considère comme la réalisation du vecteur aléatoire $\mathbf{X} = (X_1, \ldots, X_n)$, où les X_i sont des variables aléatoires i.i.d. de la loi de Poisson $\mathcal{P}(\theta)$ de paramètre inconnu $\theta > 0$.

1. Calculer l'estimateur par la méthode des moments $\hat{\theta}_n^{\mathrm{MM}}$ de θ (s'il existe). On utilisera le moment d'ordre 1.
2. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\mathrm{MV}}$ de θ (s'il existe).
3. Vérifier si l'estimateur $\hat{\theta}_n^{\mathrm{MV}}$ est convergent, et déterminer son risque quadratique moyen $\mathrm{RQM}_\theta(\hat{\theta}_n^{\mathrm{MV}})$.
4. Montrer que $\hat{\theta}_n^{\mathrm{MV}}$ est asymptotiquement normal et préciser sa vitesse de convergence.

Solution. 1. Soit $X \sim \mathcal{P}(\theta)$ avec $\theta > 0$. On a $\mathbb{E}_\theta[X] = \theta$. Selon la méthode des moments, on approche $\mathbb{E}_\theta[X]$ par la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. On obtient alors que l'estimateur par la méthode des moments (EMM) est donné par $\hat{\theta}^{\mathrm{MM}} = \bar{X}_n$. Mais attention, l'espace de paramètre Θ est $\mathbb{R}_+$ et $\bar{X}_n = 0$ se produit avec probabilité non nulle. En effet,
$$\mathbb{P}_\theta(\bar{X}_n = 0) = \mathbb{P}_\theta(X_i = 0, i = 1, \ldots, n) = (\mathbb{P}_\theta(X_1 = 0))^n = \mathrm{e}^{-n\theta} > 0.$$
Donc, lorsque $\bar{x}_n = 0$, l'EMM n'est pas défini. Néanmoins, $\mathbb{P}_\theta(\bar{X}_n = 0) \to 0$ lorsque $n \to \infty$ pour tout $\theta > 0$. Donc, si n est suffisamment grand, le problème ne se pose pas.

2. La fonction de vraisemblance associée à $\mathbf{x} = (x_1, \ldots, x_n)$ est
$$\mathcal{L}(\mathbf{x}; \theta) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^n \frac{\theta^{x_i}}{x_i!}\mathrm{e}^{-\theta} = \mathrm{e}^{-n\theta}\theta^{\sum_{i=1}^n x_i}\,\frac{1}{\prod_{i=1}^n (x_i!)}.$$

On peut passer à la fonction de log-vraisemblance

$$\ell(\theta) = \log \mathcal{L}(\mathbf{x}; \theta) = -n\theta + \log \theta \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!).$$

Cette fonction est deux fois dérivable avec

$$\ell'(\theta) = -n + \frac{1}{\theta} \sum_{i=1}^{n} x_i, \qquad \ell''(\theta) = -\frac{1}{\theta^2} \sum_{i=1}^{n} x_i.$$

Comme $\ell''(\theta) \leq 0$ pour tout $\theta > 0$, la fonction $\ell(\theta)$ est concave. Pour la maximiser il suffit alors de trouver un point critique :

$$\ell'(\theta) = 0 \Longleftrightarrow -n + \frac{1}{\theta} \sum_{i=1}^{n} x_i = 0 \Longleftrightarrow \theta = \bar{x}_n.$$

L'EMV est alors $\hat{\theta}_n^{\mathrm{MV}} = \hat{\theta}_n^{\mathrm{MM}} = \bar{X}_n$.
La même remarque que pour l'EMM s'applique : si $\bar{x}_n = 0$, alors l'EMV n'existe pas.

3. La LFGN implique la convergence de $\hat{\theta}_n^{\mathrm{MV}}$. Le risque quadratique moyen est donné par

$$\mathrm{RQM}_\theta(\hat{\theta}^{\mathrm{MV}}) = (\mathbb{E}_\theta[\hat{\theta}^{\mathrm{MV}}] - \theta)^2 + \mathrm{Var}(\hat{\theta}^{\mathrm{MV}}) = 0 + \frac{1}{n}\mathrm{Var}(X_1) = \frac{1}{n}.$$

4. Par le TCL, on a $\sqrt{n}(\hat{\theta}_n^{\mathrm{MV}} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta)$ $(n \to \infty)$. Donc, l'EMV est bien asymptotiquement normal et sa vitesse de convergence est de $n^{-1/2}$.