

CORRIGÉ 10

Exercice 1. (i). (a) Pour $0 \leq y \leq \theta$, la fonction de répartition de $M_n = \max\{Y_1, \dots, Y_n\}$ est

$$\begin{aligned} F_{M_n}(y) &= \Pr_\theta(M_n \leq y) \\ &= \Pr_\theta(Y_1 \leq y, Y_2 \leq y, \dots, Y_n \leq y) \\ &= \Pr_\theta(Y_1 \leq y) \cdots \Pr_\theta(Y_n \leq y), \quad \text{car ils sont indépendants,} \\ &= \left(\frac{y}{\theta}\right)^n. \end{aligned}$$

D'autre part, $F_{M_n}(y) = 0$ pour $y \leq 0$ et $F_{M_n}(y) = 1$ pour $y > \theta$. En dérivant $F_{M_n}(y)$ on trouve la fonction de densité

$$f_{M_n}(y) = \frac{n}{\theta} \left(\frac{y}{\theta}\right)^{n-1}, \quad y \in [0, \theta],$$

et $f_{M_n}(y) = 0$ sinon.

(b) Le biais de l'estimateur M_n est égal à $\mathbb{E}_\theta[M_n] - \theta$. Donc pour calculer le biais b_n , on commence par calculer l'espérance de M_n ,

$$\mathbb{E}_\theta[M_n] = \int_{-\infty}^{\infty} y f_{M_n}(y) dy = \int_0^\theta \frac{n}{\theta^n} (y)^n dy = \frac{n}{n+1} \theta.$$

Le biais est donc $\mathbb{E}_\theta[M_n] - \theta = -\theta/(n+1) \neq 0$ et M_n est donc un estimateur biaisé (en particulier M_n sous-estime θ).

(c) Pour obtenir un estimateur non-biaisé, on doit corriger M_n et poser $\tilde{\theta}_1 = \frac{n+1}{n} M_n$. Pour ce nouvel estimateur, on a bien $\mathbb{E}_\theta[\tilde{\theta}_1] = \theta$ et donc son biais est nul pour tout $\theta > 0$. La variance est $\text{var}_\theta(\tilde{\theta}_1) = \left(\frac{n+1}{n}\right)^2 \text{Var}(M_n)$. On obtient la variance de M_n à l'aide de la fonction de densité trouvée en (a) :

$$\mathbb{E}_\theta[M_n^2] = \int_{-\infty}^{\infty} x^2 f_{M_n}(x) dx = \frac{n}{n+2} \theta^2,$$

et donc $\text{var}_\theta(M_n) = \mathbb{E}_\theta[M_n^2] - (\mathbb{E}_\theta M_n)^2 = \frac{n}{(n+1)^2(n+2)} \theta^2$. Finalement, $\text{var}_\theta(\tilde{\theta}_1) = \frac{1}{n(n+2)} \theta^2$.

(ii). (a) Par définition, le premier moment empirique de l'échantillon $Y_1, Y_2, \dots, Y_n$ est

$$\frac{1}{n} \sum_{i=1}^n Y_i$$

et le premier moment théorique est

$$\mu_{(1)} = \mathbb{E}_\theta[Y_1] = \theta/2,$$

(comme tous les Y_i suivent la même distribution uniforme sur $[0, \theta]$). L'estimateur des moments, qu'on dénote $\tilde{\theta}_2$, doit résoudre l'équation

$$\begin{aligned}\mu_{(1)} &= \frac{1}{n} \sum_{i=1}^n Y_i \\ \Leftrightarrow \quad \frac{\tilde{\theta}_2}{2} &= \frac{1}{n} \sum_{i=1}^n Y_i, \\ \Leftrightarrow \quad \tilde{\theta}_2 &= 2 \frac{1}{n} \sum_{i=1}^n Y_i\end{aligned}$$

On vérifie que $\mathbb{E}_\theta[\tilde{\theta}_2] = \theta$ et donc l'estimateur est non-biaisé.

(b) Sa variance est $\text{var}_\theta(\tilde{\theta}_2) = \theta^2/(3n)$.

(iii). On se retrouve donc avec trois candidats potentiels $(M_n, \tilde{\theta}_1, \tilde{\theta}_2)$ pour estimer le paramètre inconnu θ . On utilise l'erreur quadratique moyenne $EQM_\theta(\tilde{\theta}) = b(\tilde{\theta})^2 + \text{Var}(\tilde{\theta})$ pour les comparer.

$$\begin{aligned}EQM_\theta(M_n) &= [-\theta/(n+1)]^2 + n\theta^2/[(n+1)^2(n+2)] \\ EQM_\theta(\tilde{\theta}_1) &= 0 + \frac{1}{n(n+2)}\theta^2 \\ EQM_\theta(\tilde{\theta}_2) &= 0 + \theta^2/(3n)\end{aligned}$$

Pour tout $n > 1$, $EQM(\tilde{\theta}_1)$ est le plus petit des trois et l'estimateur $\tilde{\theta}_1$ est donc préférable.

(iv). Notons tout d'abord que les bornes de l'intervalle ne dépendent pas de θ mais uniquement des données $Y_1, \dots, Y_n$.

Comme $M_n \leq \theta$ forcément, on a pour tout $\theta > 0$

$$\Pr_\theta(M_n \leq \theta \leq M_n \alpha^{-1/n}) = \Pr_\theta(M_n \geq \theta \alpha^{1/n}) = 1 - \Pr_\theta(M_n \leq \theta \alpha^{1/n}) = 1 - \frac{(\theta \alpha^{1/n})^n}{\theta^n} = 1 - \alpha.$$

On a donc bien un intervalle de confiance de niveau $1 - \alpha$. Il est construit à l'aide du pivot de la diapositive 180.

Exercice 2. (i). $E_p(\hat{p}_1) = p$

$$\text{Var}_p(\hat{p}_1) = \frac{p(1-p)}{m}$$

(ii). $E_p(\hat{p}_2) = E_p\left(\frac{\sum_i Y_i + 1/2}{m+1}\right) = \frac{mp + 1/2}{m+1} = p + \frac{1/2-p}{m+1}$.

$$\text{Var}_p(\hat{p}_2) = \text{var}_p\left(\frac{\sum_i Y_i + 1/2}{m+1}\right) = \frac{m}{(m+1)^2} p(1-p).$$

(iii). L'erreur quadratique moyenne (en anglais mean squared error, MSE) d'un estimateur est égale à la somme de sa variance et de son biais au carré.

L'erreur quadratique moyenne de $\hat{p}_2$ et $\hat{p}_1$ sont

$$\begin{aligned}EQM_p(\hat{p}_1) &= \text{Var}_p(\hat{p}_1) + (\mathbb{E}_p(\hat{p}_1) - p)^2 = \frac{p(1-p)}{m} + (p-p)^2 = \frac{p(1-p)}{m}, \\ EQM_p(\hat{p}_2) &= \text{Var}_p(\hat{p}_2) + (\mathbb{E}_p(\hat{p}_2) - p)^2 = \frac{mp(1-p)}{(m+1)^2} + \frac{(1/2-p)^2}{(m+1)^2}, \\ EQM_p(\hat{p}_2) - EQM_p(\hat{p}_1) &= \frac{mp(1-p)}{(m+1)^2} + \frac{(1/2-p)^2}{(m+1)^2} - \frac{p(1-p)}{m} = \frac{1}{m(m+1)^2} (p^2(3m+1) - p(3m+1))\end{aligned}$$

On a donc l'équivalence suivante

$$\text{EQM}_p(\hat{p}_2) < \text{EQM}_p(\hat{p}_1) \iff p^2(3m+1) - p(3m+1) + \frac{m}{4} < 0 \iff p \in \left[\frac{1}{2} - \frac{1}{2}\sqrt{\frac{2m+1}{3m+1}}, \frac{1}{2} + \frac{1}{2}\sqrt{\frac{2m+1}{3m+1}} \right]$$

Donc le fait que $\hat{p}_2$ soit un meilleur estimateur que $\hat{p}_1$ dépend de la vraie valeur du paramètre p (et de la taille de l'échantillon, m). $\hat{p}_2$ est meilleur si p est proche de $1/2$.

Exercice 3. (i). On a calculé la semaine passée $\ell''(\theta) = -\frac{n}{\theta^2}$. Dans ce cas, ce n'est pas une variable aléatoire. L'information de Fisher est $-\mathbb{E}_\theta[\ell''(\theta)] = n/\theta^2$. donc $\hat{\theta}_{ML} \overset{app}{\approx} \mathcal{N}(\theta, 1/I_n(\theta)) = \mathcal{N}(\theta, \theta^2/n)$.

(ii). Maintenant, on calcule

$$\Pr_\theta(-\log(X_1) \leq y) = \Pr_\theta(X_1 \geq e^{-y}) = 1 - \Pr_\theta(X_1 \leq e^{-y})$$

Si $y \leq 0$, alors $e^{-y} \geq 1$ et la probabilité est égale à 1. Si $y > 0$, alors $e^{-y} \in (0, 1)$ et

$$1 - \Pr_\theta(X_1 \leq e^{-y}) = 1 - \int_{-\infty}^{e^{-y}} f(x; \theta) dx = 1 - \int_0^{e^{-y}} \theta x^{\theta-1} dx = 1 - [(e^{-y})^\theta - 0^\theta] = 1 - e^{-y\theta}.$$

Ainsi la fonction de densité de chaque Y_i est $\theta e^{-y\theta}$ pour $y > 0$ et 0 pour $y \leq 0$. Donc $Y_i \sim \exp(\theta)$ qui a une espérance $1/\theta$ et variance $1/\theta^2$. Le théorème centrale limite donne $\bar{Y}_n \overset{app}{\approx} \mathcal{N}(1/\theta, 1/[n\theta^2])$.

Maintenant, remarquons que $\hat{\theta}_{ML} = 1/\bar{Y}$. Soit la fonction $g(y) = 1/y$ dont la dérivée est $g'(y) = -1/y^2$. En évaluant cette dérivée en $1/\theta$, on obtient avec la méthode delta que

$$\hat{\theta}_{ML} = g(\bar{Y}_n) \overset{app}{\approx} \mathcal{N}(g(1/\theta), [g'(1/\theta)]^2/[n\theta^2]) = \mathcal{N}(\theta, \theta^4/[n\theta^2]) = \mathcal{N}(\theta, \theta^2/n),$$

donc le même résultat obtenu avec l'information de Fisher.

(iii). On a

$$\mathbb{E}_\theta[X_i] = \int_0^1 x \theta x^{\theta-1} dx = \int_0^1 \theta x^\theta dx = \frac{\theta}{\theta+1}, \quad \mathbb{E}_\theta[X_i^2] = \int_0^1 \theta x^{\theta+1} dx = \frac{\theta}{\theta+2}$$

$$\text{var}_\theta(X_i) = \mathbb{E}_\theta[X_i^2] - (\mathbb{E}_\theta[X_i])^2 = \frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2} = \frac{\theta}{(\theta+2)(\theta+1)^2}.$$

D'après le théorème centrale limite

$$\bar{X}_n \overset{app}{\approx} \mathcal{N}\left(\frac{\theta}{\theta+1}, \frac{\theta}{(\theta+2)(\theta+1)^2 n}\right).$$

Soit $g(y) = y/(1-y)$ de sorte que $\hat{\theta}_{MOM} = g(\bar{X}_n)$ et $g'(y) = 1/(1-y)^2$. La méthode delta donne

$$\hat{\theta}_{MOM} = g(\bar{X}_n) \overset{app}{\approx} \mathcal{N}\left(g\left(\frac{\theta}{\theta+1}\right), g'\left(\frac{\theta}{\theta+1}\right)^2 \frac{\theta}{(\theta+2)(\theta+1)^2 n}\right) = \mathcal{N}\left(\theta, [(\theta+1)^2]^2 \frac{\theta}{(\theta+2)(\theta+1)^2 n}\right)$$

$$= \mathcal{N}\left(\theta, \frac{\theta(\theta+1)^2}{(\theta+2)n}\right)$$

Comme pour tout $\theta > 0$

$$\frac{\theta(\theta+1)^2}{(\theta+2)} > \theta^2,$$

l'estimateur de maximum de vraisemblance est préférable.

Exercice 4. Dans le cas binomial on a vu que $\hat{p}_{ML} = \bar{X}_n$ et que

$$\ell''(p) = -\frac{\sum_{i=1}^n X_i}{p^2} - \frac{n - \sum_{i=1}^n X_i}{(1-p)^2}.$$

Puisque l'espérance de chaque X_i est p , l'information de Fisher est

$$I_n(p) = -\mathbb{E}_p(\ell''(p)) = \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p(1-p)}.$$

L'information observée est $-\ell''(\hat{p}_{ML}) = n/[\hat{p}_{ML}(1-\hat{p}_{ML})] = n/[\bar{X}_n(1-\bar{X}_n)]$. La variance approximative de $\hat{p}_{ML}$ est $1/I_n(p)$ (dans ce cas, c'est la variance exacte, mais en générale c'est seulement une approximation) et si n est grand, elle est proche à $1/J_n(\hat{p}_{ML})$.

Dans le cas Poisson on a

$$L(\lambda) = \prod_{i=1}^n \frac{X_i^\lambda}{X_i!} e^{-\lambda} = e^{-n\lambda} \frac{\lambda^{n\bar{X}_n}}{\prod_{i=1}^n X_i!}$$

$$\ell(\lambda) = \log L(\lambda) = n\bar{X}_n \log \lambda - n\lambda - \sum_{i=1}^n \log(X_i!)$$

$$\ell'(\lambda) = \frac{n\bar{X}_n}{\lambda} - n$$

$$\ell''(\lambda) = -\frac{n\bar{X}_n}{\lambda^2} < 0.$$

Donc $\hat{\lambda}_{ML} = \bar{X}_n$ correspond bien à un maximum. L'information de Fisher est

$$I_n(\lambda) = -\mathbb{E}[\ell''(\lambda)] = \frac{n}{\lambda} \mathbb{E}[\bar{X}_n] = \frac{n}{\lambda},$$

puisque l'espérance de chaque X_i est λ . Ainsi la variance asymptotique de $\hat{\lambda}_{ML}$ est λ/n . L'information observée est

$$-\ell''(\bar{X}_n) = \frac{n\bar{X}_n}{\bar{X}_n^2} = \frac{n}{\bar{X}_n}$$

(infinie si $\bar{X}_n = 0$, un événement rare de probabilité $e^{-n\lambda}$).

Exercice 5. (a) Si $X_1, \dots, X_n$ sont iid $\mathcal{N}(\mu, \sigma^2)$, alors

$$Z_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

On sait que

$$\Pr(-z < Z_n < z) = 2\Phi(z) - 1,$$

pour toute constante $z > 0$. Donc si on choisit $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ (le $(1-\alpha/2)$ -quantile de la loi $N(0, 1)$), on obtient que

$$\begin{aligned} 1 - \alpha &= \Pr_{\mu}(-z_{1-\alpha/2} < Z_n < z_{1-\alpha/2}) \\ &= \Pr_{\mu}\left(-z_{1-\alpha/2} < \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < z_{1-\alpha/2}\right) \\ &= \Pr_{\mu}\left(\bar{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma < \mu < \bar{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma\right). \end{aligned}$$

L'intervalle

$$\left(\bar{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma, \bar{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma\right)$$

couvre donc la vraie valeur de μ avec la probabilité $1 - \alpha$.

Dans notre cas, $\sigma^2 = 3.5$, $n = 12$, $\bar{x}_{12} = 13.31$, et $\alpha = 0.05$. On peut trouver dans le tableau de la fonction de répartition de la loi normale que $\Phi^{-1}(0.975) = 1.96$. L'intervalle cherché est donc $(12.25, 14.37)$.

- (b) Si $X_1, \dots, X_n$ sont iid $\mathcal{N}(\mu, \sigma^2)$, alors

$$T_{n-1} = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1},$$

où $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, et t_{ν} est la loi de Student avec ν degrés de liberté. La loi de Student est symétrique autour de zéro de la même manière que la loi $\mathcal{N}(0, 1)$. On a donc

$$\Pr_{\mu, \sigma^2}(|T| > t_{n-1, 1-\alpha/2}) = \alpha,$$

où $t_{n-1, 1-\alpha/2}$ est le $(1 - \alpha/2)$ -quantile de la loi t_{n-1} .

De la même manière que dans la partie (a) on obtient l'intervalle de confiance sous la forme

$$\left(\bar{X}_n - \frac{t_{n-1, 1-\alpha/2}}{\sqrt{n}} S_n, \bar{X}_n + \frac{t_{n-1, 1-\alpha/2}}{\sqrt{n}} S_n\right).$$

Dans notre situation on a $n = 12$, $\alpha = 0.05$, $s^2 = 3.69$, et on peut trouver dans le tableau que $t_{11, 0.975} = 2.201$. L'intervalle cherché est $(12.09, 14.53)$. On note que cet intervalle est plus large que celui de la partie (a). Cela vient du fait que nous avons maintenant deux paramètres à estimer et donc que l'incertitude est plus grande que dans le cas où un seul paramètre est à estimer.

- (c) On trouve que $t_{11}(0.95) = 1.796$, et que l'intervalle cherché est $(12.31, 14.31)$. Cet intervalle est plus étroit que celui calculé dans la partie (b), car son seuil de confiance est plus petit. Plus on veut être confiant qu'un intervalle couvre la vraie valeur de μ , plus cet intervalle doit être large (et vice-versa).
- (d) Récolter plus de données. Plus on a de données, plus l'incertitude est faible.