UQÀM

Devoir 2 - Éléments de correction

STT 1000 - Automne 2025

Devoir à rendre le Mercredi 26 Novembre 2025 au plus tard à 10h00
Pénalité de 5 points par heure de retard
Un rendu manuscrit par personne étudiante

Exercice 1 (Compromis biais-variance (**30%**))**.** Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon i.i.d. de loi populationnelle $\mathscr{U}([0, \theta])$ avec $\theta > 0$. On note $M = X_{(n)} = \max\{X_1, \ldots, X_n\}$ et on admettra que $M/\theta \sim \text{Beta}(n, 1)$

1. En utilisant le fait que pour une v.a. $Y \sim \text{Beta}(a, b)$

$$\mathbb{E}[Y] = \frac{a}{a+b}, \quad \text{et} \quad \mathbb{V}(Y) = \frac{ab}{(a+b)^2(a+b+1)},$$

   calculez $\mathbb{E}[M]$ et $\mathbb{V}(M)$.

2. Considérez l'estimateur de θ

$$\widehat{\theta}_{\text{SB}} = \frac{n+1}{n} M.$$

   (a) Cet estimateur est-il biaisé ?

   (b) Calculez sa variance et son écart quadratique moyen (EQM).

3. On considère maintenant la famille d'estimateurs

$$\widehat{\theta}_c = cM, \quad \text{où } c > 0$$

   (a) Exprimez le biais, la variance et l'EQM de $\widehat{\theta}_c$ en fonction de c, de n et de θ.

   (b) Dérivez l'EQM par rapport à c et trouver la valeur c^* qui minimise l'EQM. Vérifiez qu'il s'agit bien d'un minimum à l'aide de la dérivée seconde.

   (c) L'estimateur $\widehat{\theta}_{c^*}$ est sans biais?

   (d) Comparer l'EQM de $\widehat{\theta}_{c^*}$ à celui de $\widehat{\theta}_{\text{SB}}$.

4. Lequel de ces deux estimateurs préférez-vous au sens de l'EQM?
   Quelle conclusion générale peut-on tirer sur le compromis entre biais et variance ?

Solution

1. On a que $M = \theta \times M/\theta$ avec $M/\theta \sim \text{Beta}(n, 1)$

$$\mathbb{E}[M] = \theta \mathbb{E}[M/\theta] = \boxed{\theta \frac{n}{n+1}}$$

et

$$\mathbb{V}(M) = \theta^2 \mathbb{V}(M/\theta) = \boxed{\theta^2 \frac{n}{(n+1)^2(n+2)}}$$

2. Estimateur $\widehat{\theta}_{\text{SB}}$

   (a) **Biais:**

   $$\mathbb{E}\left[\widehat{\theta}_{\text{SB}}\right] = \frac{n+1}{n}\mathbb{E}[M] = \theta \times \frac{n+1}{n} \times \frac{n}{n+1} = \theta$$

   et donc

   $$\text{biais}_\theta(\widehat{\theta}_{\text{SB}}) = \mathbb{E}\left[\widehat{\theta}_{\text{SB}}\right] - \theta = 0.$$

   Donc $\widehat{\theta}_{\text{SB}}$ est sans biais

   (b) **Variance et EQM**:

   $$\mathbb{V}\left(\widehat{\theta}_{\text{SB}}\right) = \left(\frac{n+1}{n}\right)^2 \mathbb{V}(M) = \theta^2 \frac{(n+1)^2}{n^2} \frac{n}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}$$

   et comme $\widehat{\theta}_{\text{SB}}$ est sans biais, alors

   $$\text{EQM}_\theta(\widehat{\theta}_{\text{SB}}) = \mathbb{V}\left(\widehat{\theta}_{\text{SB}}\right) = \frac{\theta^2}{n(n+2)}$$

3. Estimateur $\widehat{\theta}_c$

   (a) **Biais, variance, EQM**

   $$\text{biais}_\theta\left(\widehat{\theta}_c\right) = \mathbb{E}[cM] - \theta = c\frac{n}{n+1}\theta - \theta = \theta\left(\frac{cn}{n+1} - 1\right)$$

   $$\mathbb{V}\left(\widehat{\theta}_c\right) = c^2 \mathbb{V}(M) = c^2 \frac{n\theta^2}{(n+1)^2(n+2)}$$

   $$\text{EQM}_\theta(\widehat{\theta}_c) = \mathbb{V}\left(\widehat{\theta}_c\right) + \text{biais}_\theta\left(\widehat{\theta}_c\right)^2 = \theta^2\left[\frac{c^2 n}{(n+1)^2(n+2)} + \left(\frac{cn}{n+1} - 1\right)^2\right]$$

   (b) **Minimiser EQM**

   Comme $\theta^2 > 0$, alors minimiser $\text{EQM}_\theta(\widehat{\theta}_c)$ ou $\frac{\text{EQM}_\theta(\widehat{\theta}_c)}{\theta^2}$ est équivalent.
   Posons: $a = \frac{n}{(n+1)^2(n+2)}$ et $b = \frac{n}{n+1}$. Alors:

   $$\frac{\text{EQM}_\theta(\widehat{\theta}_c)}{\theta^2} = c^2 a + (cb-1)^2 = c^2(a+b^2) - 2bc + 1.$$

   Et donc:

   $$\frac{\partial}{\partial c}\frac{\text{EQM}_\theta(\widehat{\theta}_c)}{\theta^2} = 2(a+b)c - 2b$$

   Si on annule la dérivée, on a que:

   $$2(a+b)c - 2b = 0 \quad \Longleftrightarrow \quad \boxed{c^* = \frac{b}{a+b^2}}$$

Comme

$$a + b^2 = \frac{n}{(n+1)^2(n+2)} + \frac{n^2}{(n+1)^2} = \frac{n + n^2(n+2)}{(n+1)^2(n+2)} = \frac{n(n+1)^2}{(n+1)^2(n+2)} = \frac{n}{n+2}$$

et donc

$$\boxed{c^* = \frac{n/n+1}{n/n+2} = \frac{n+2}{n+1}}.$$

De plus,

$$\frac{\partial^2}{\partial c^2} \frac{\mathrm{EQM}_\theta(\widehat{\theta}_c)}{\theta^2} = 2(a + b^2) = 2\frac{n}{n+2} > 0$$

La fonction est donc convexe et il s'agit bien un d'un minimum.

(c) **Biais de $\widehat{\theta}_{c^*}$**

$$\mathrm{biais}_\theta\left(\widehat{\theta}_{c^*}\right) = \theta\left(\frac{n+2}{n+1}\frac{n}{n+1} - 1\right) = -\frac{\theta}{(n+1)^2}$$

Non, l'estimateur $\widehat{\theta}_{c^*}$ est biaisé.

(d) **Comparaison d'EQM**
On a que:

$$\mathrm{EQM}_\theta(\widehat{\theta}_{c^*}) = \frac{\theta^2}{(n+1)^2}$$

On peut calculer l'efficacité entre $\widehat{\theta}_{c^*}$ et $\widehat{\theta}_{\mathrm{SB}}$:

$$\mathrm{e}_\theta(\widehat{\theta}_{c^*}, \widehat{\theta}_{\mathrm{SB}}) = \frac{\frac{\theta^2}{n(n+2)}}{\frac{\theta^2}{(n+1)^2}} = \frac{(n+1)^2}{n(n+2)} = \frac{n^2+2n+1}{n^2+2n} = 1 + \frac{1}{n(n+2)} > 1$$

4. Comme $\mathrm{e}_\theta(\widehat{\theta}_{c^*}, \widehat{\theta}_{\mathrm{SB}}) > 1$, alors $\mathrm{EQM}_\theta(\widehat{\theta}_{c^*}) < \mathrm{EQM}_\theta(\widehat{\theta}_{\mathrm{SB}})$, on préférera toujours $\widehat{\theta}_{c^*}$ au sens de l'EQM, malgré le fait qu'il soit biaisé. On peut en conclure qu'au sens de l'EQM, un estimateur biaisé peut être préféré à un estimateur sans biais.

Exercice 2 (Estimation et intervalle de confiance (**40%**))**.** Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon i.i.d. de loi populationnelle ayant pour densité

$$f(t;\theta) = \theta^2 t \exp(-\theta t)\mathbb{1}_{[0,\infty)}(t).$$

On admettra que, pour $i = 1, \ldots, n$, $2\theta X_i \sim \chi^2(4)$.

1. En utilisant le fait que pour une v.a. $Y \sim \chi^2(4)$

$$\mathbb{E}[Y] = 4, \quad \text{et} \quad \mathbb{V}(Y) = 8,$$

calculez $\mathbb{E}[X_1]$ et $\mathbb{V}(X_1)$.

2. Trouvez l'estimateur de θ par la méthode des moments et par la méthode du maximum de vraisemblance et montrez que:

$$\widehat{\theta}_{\mathrm{MM}}(\mathbf{X}) = \widehat{\theta}_{\mathrm{EMV}}(\mathbf{X}) = \frac{2n}{\sum_{i=1}^n X_i}.$$

3. En utilisant le fait que si $Y_1, \dots, Y_n$ i.i.d. de loi $\chi^2(4)$, alors

$$\sum_{i=1}^{n} Y_i \sim \chi^2(4n),$$

déterminez la loi de $2\theta \sum_{i=1}^{n} X_i$. Déduisez-en un intervalle de confiance bilatéral au niveau $1-\alpha$.

4. On observe l'échantillon suivant, pour $n = 10$:

$$0.192,\ 0.208,\ 0.114,\ 0.127,\ 0.473,\ 0.169,\ 0.098,\ 0.352,\ 0.083,\ 0.166.$$

Calculez l'intervalle de confiance observé au niveau de risque $\alpha = 0.05$

Solution

1. **Moments de X_1**
On sait que $Y = 2\theta X_1 \sim \chi^2(4)$, avec

$$\mathbb{E}[Y] = 4, \qquad \mathbb{V}(Y) = 8.$$

Comme $X_1 = \dfrac{Y}{2\theta}$, on a

$$\mathbb{E}[X_1] = \frac{1}{2\theta}\mathbb{E}[Y] = \boxed{\frac{2}{\theta}}, \qquad \mathbb{V}(X_1) = \frac{1}{(2\theta)^2}\mathbb{V}(Y) = \boxed{\frac{2}{\theta^2}}.$$

2. **Estimateurs de θ (méthode des moments et EMV)**

   (a) **Méthode des moments**
   La densité $f(t;\theta) = \theta^2 t e^{-\theta t}\mathbf{1}_{[0,\infty)}(t)$ est une gamma $\Gamma(k=2,\ \text{taux}=\theta)$, d'où $\mathbb{E}[X_1] = \dfrac{2}{\theta}$.
   En égalant $\overline{X}_n$ à $\mathbb{E}[X_1]$,

$$\overline{X}_n = \frac{2}{\theta} \quad \Longrightarrow \quad \widehat{\theta}_{\text{MM}} = \boxed{\frac{2}{\overline{X}_n}} = \boxed{\frac{2n}{\sum_{i=1}^{n} X_i}}.$$

   (b) **Estimateur du maximum de vraisemblance**
   La log-vraisemblance vaut, pour $\mathbf{X} = (X_1, \dots, X_n)$,

$$\ell(\theta) = \sum_{i=1}^{n}\Big(2\log\theta + \log X_i - \theta X_i\Big) = 2n\log\theta + \sum_{i=1}^{n}\log X_i - \theta\sum_{i=1}^{n} X_i.$$

   Dérivée première et condition du maximum :

$$\ell'(\theta) = \frac{2n}{\theta} - \sum_{i=1}^{n} X_i = 0 \quad \Longrightarrow \quad \widehat{\theta}_{\text{EMV}} = \boxed{\frac{2n}{\sum_{i=1}^{n} X_i}}.$$

   Dérivée seconde $\ell''(\theta) = -\dfrac{2n}{\theta^2} < 0$, donc il s'agit bien d'un maximum.

   Conclusion : $\boxed{\widehat{\theta}_{\text{MM}}(\mathbf{X}) = \widehat{\theta}_{\text{EMV}}(\mathbf{X}) = \dfrac{2n}{\sum_{i=1}^{n} X_i}}$.

3. **Loi de $2\theta \sum X_i$ et IC bilatéral pour θ**
On a $Y_i = 2\theta X_i \sim \chi^2(4)$ i.i.d., donc par additivité :

$$\sum_{i=1}^{n} Y_i \sim \chi^2(4n) \quad \Longleftrightarrow \quad 2\theta \sum_{i=1}^{n} X_i \sim \chi^2(4n).$$

Pour un niveau $1-\alpha$, en notant $c_{\alpha,\nu}$ le quantile d'ordre α d'une $\chi^2(\nu)$,

$$c_{\alpha/2,4n} \;\le\; 2\theta \sum_{i=1}^{n} X_i \;\le\; c_{1-\alpha/2,4n}.$$

En isolant θ :

$$\boxed{\left[\; \frac{c_{\alpha/2,4n}}{2\sum_{i=1}^{n} X_i} \;,\; \frac{c_{1-\alpha/2,4n}}{2\sum_{i=1}^{n} X_i} \;\right]} \quad \text{est un IC bilatéral de niveau } 1-\alpha \text{ pour } \theta.$$

4. **Application numérique ($n = 10$, $\alpha = 0.05$)**
Données : 0.192, 0.208, 0.114, 0.127, 0.473, 0.169, 0.098, 0.352, 0.083, 0.166.
Somme observée :

$$\sum_{i=1}^{10} X_i = 1.982, \qquad 2\sum_{i=1}^{n} X_i = 3.964.$$

L'estimateur commun (MM et EMV) :

$$\widehat{\theta} = \frac{2n}{\sum X_i} = \boxed{\frac{20}{1.982} \approx 10.091}.$$

Pour $\nu = 4n = 40$ et $\alpha = 0.05$ (donc $\alpha/2 = 0.025$), on prend:

$$c_{0.025,40} \approx 24.423, \qquad c_{0.975,40} \approx 59.345.$$

L'IC bilatéral au niveau 95% :

$$\left[\; \frac{24.423}{3.964} \;,\; \frac{59.345}{3.964} \;\right] = \boxed{[\,6.161\;;\;14.971\,]}.$$

Exercice 3 (Intervalles de confiance (**30%**))**.** Pour chacun des 5 études de cas suivantes, justifiez votre choix d'intervalle de confiance, puis calculez-les à l'aide des observations et estimations qui vous sont fournies. Arrondissez vos résultats à deux chiffres après la virgule.

Etude de cas 1 (Contrôle de mise en bouteille) Une entreprise embouteille un jus pressé artisanalement en bouteilles de 1 L. Pour garantir la constance du processus, on veut estimer la moyenne du volume réel des bouteilles. On suppose que les volumes (en mL) suivent une loi normale de variance $\sigma^2 = 2$.

Un échantillon de 12 bouteilles donne:

1002.1, 998.4, 1000.7, 1001.3, 999.2, 1000.1, 1003.6, 1001.8, 999.9, 1002.0, 1001.4, 1000.6

Construisez les intervalles de confiance unilatéraux à 95% pour la moyenne du volume.

Etude de cas 2 (Analyse nutritionnelle) Un laboratoire souhaite vérifier la teneur moyenne en sodium (en mg) d'un type de soupe en conserve. On prélève un échantillon de taille 308 et on obtient $\overline{x}_n = 423.32$ mg et $s^2 = 8.6$

Construisez un intervalle de confiance bilatéral à 99% pour la moyenne de la teneur en sodium.

Etude de cas 3 (Comparaison de deux machines) Une usine possède deux machines (**A** et **B**) produisant des vis métalliques. On veut savoir si les deux machines produisent des vis de même longueur moyenne. Les longueurs (en mm) sont supposées suivre des lois normales, et la variance est égale entre les deux machines.

On observe:

- **Machine A:** $n_1 = 10$, $\overline{x}_1 = 49.82$, $s_1^2 = 0.025$
- **Machine B:** $n_2 = 12$, $\overline{x}_2 = 49.90$, $s_2^2 = 0.014$

Construisez un intervalle de confiance bilatéral à 95% pour répondre à la question.

Etude de cas 4 (Efficacité d'un programme de formation) Une compagnie souhaite évaluer l'effet d'un programme de formation sur la productivité de ses employé·e·s. On mesure la production (en unités/heure) avant et après la formation pour les mêmes 8 employé·e·s. On suppose que ces mesures de production suivent une loi normale.

Employé	Avant la formation	Après la formation
1	12.1	12.6
2	11.8	12.4
3	13.0	13.5
4	11.5	12.1
5	12.3	12.8
6	12.7	13.1
7	11.9	12.5
8	12.4	12.9

Construisez un intervalle de confiance à 99% qui permettra d'aider la compagnie à évaluer ce programme.

Etude de cas 5 (Comparaison des performances sportives) Un journaliste spécialiste de hockey souhaite comparer la constance des tirs au filet entre les Canadiens de Montréal et les Maple Leafs de Toronto. On mesure la vitesse (en km/h) des tirs au filet lors d'une séance d'entraînement des deux équipes. On suppose que, pour chaque équipe, les vitesses suivent une loi normale et que les deux échantillons sont indépendants.

On observe:

- **Canadiens de Montréal:** $n_1 = 10$, $\overline{x}_1 = 154.8$, $s_1^2 = 10.5$
- **Maple Leafs de Toronto:** $n_2 = 12$, $\overline{x}_2 = 156.1$, $s_2^2 = 12.4$

Calculez un intervalle de confiance bilatéral à 95% pour comparer les variances de tir.

Solution

1. **Contrôle de mise en bouteille**
   Cadre de travail. Nous sommes dans le cadre de travail suivant:
   - Échantillon Gaussien, variance connue $\sigma^2 = 2$
   - $\alpha = 0.05$
   - $\overline{x}_n = 1000.92$
   - $n = 12$

   Pivot. Dans ce cadre là, le pivot est donné par:
   $$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \sim \mathcal{N}(0,1).$$
   Intervalle de confiance. Ainsi, on peut utiliser les IC unilatéraux au niveau 95%:
   - Droit: $(-\infty\ ;\ \overline{x}_n + z_{0.95}\frac{\sigma}{\sqrt{n}}]$
   - Gauche: $[\overline{x}_n - z_{0.95}\frac{\sigma}{\sqrt{n}}\ ;\ \infty)$

2. **Analyse nutritionnelle**
   Cadre de travail. Nous sommes dans le cadre de travail suivant:
   - Échantillon de loi inconnue
   - $\alpha = 0.01$
   - $\overline{x}_n = 423.32$ et $s^2 = 8.6$
   - $n = 308$

   Pivot. Dans ce cadre là, comme $n > 30$, nous sommes dans le régime asymptotique et acceptons l'approximation asymptotique suivante:
   $$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{S}\right) \sim \mathcal{N}(0,1)$$
   d'après le Théorème de la Limite Centrale.
   Intervalle de confiance. Ainsi, on peut utiliser l'IC asymptotique bilatéral suivant au niveau 99%:
   $$\left[\overline{x}_n - z_{0.995}\frac{s}{\sqrt{n}}\ ;\ \overline{x}_n - z_{0.995}\frac{s}{\sqrt{n}}\right]$$

3. **Comparaison de deux machines**
   Cadre de travail. Nous sommes dans le cadre de travail suivant:
   - Deux échantillons indépendants Gaussiens, de variances inconnues mais égales
   - $\alpha = 0.05$

- $\overline{x}_1 = 49.82$ et $s_1^2 = 0.025$
- $\overline{x}_2 = 49.9$ et $s_2^2 = 0.014$
- $n_1 = 10$, $n_2 = 12$

Pivot. Dans ce cadre là, le pivot est donné par:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}} \sim \mathscr{T}(n_1 + n_2 - 2).$$

Intervalle de confiance. Ainsi, on peut utiliser l'IC bilatéral suivant au niveau 95%:

$$\left[(\overline{x}_1 - \overline{x}_2) - t_{0.975,n_1+n_2-2}\, s_p\sqrt{1/n_1 + 1/n_2}\ ;\ (\overline{x}_1 - \overline{x}_2) + t_{0.975,n_1+n_2-2}\, s_p\sqrt{1/n_1 + 1/n_2}\right]$$

où $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

4. **Efficacité d'un programme de formation**
   Cadre de travail. Nous sommes dans le cadre de travail suivant:

   - Deux échantillons appariés Gaussiens, de variances inconnues mais égales
   - $\alpha = 0.01$
   - $\overline{d}_n = 0.525$, $s_D = 0.071$
   - $n = 8$

   Pivot. Dans ce cadre là, le pivot est donné par:

$$\sqrt{n}\left(\frac{\overline{D}_n - \mu_D}{S_D}\right) \sim \mathscr{T}(n-1).$$

   Intervalle de confiance. Ainsi, on peut utiliser l'IC bilatéral suivant au niveau 99%:

$$\left[\overline{d}_n - t_{0.995,n-1}\frac{s_D}{\sqrt{n}}\ ;\ \overline{d}_n + t_{0.995,n-1}\frac{s_D}{\sqrt{n}}\right].$$

5. **Comparaison des performances sportives (variances)**
   Cadre de travail. Nous sommes dans le cadre de travail suivant:

   - Deux échantillons indépendants Gaussiens, moyennes et variances inconnues
   - $\alpha = 0.05$
   - $\overline{x}_1 = 154.8$ et $s_1^2 = 10.5$
   - $\overline{x}_2 = 156.1$ et $s_2^2 = 12.4$
   - $n_1 = 10$, $n_2 = 12$

   Pivot. Dans ce cadre là, le pivot est donné par:

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim \mathscr{F}(n_1 - 1, n_2 - 1).$$

   Intervalle de confiance. Ainsi, on peut utiliser l'IC bilatéral suivant au niveau 95%:

$$\left[\frac{s_1^2/s_2^2}{F_{0.975,n_1-1,n_2-1}}\ ;\ \frac{s_1^2/s_2^2}{F_{0.025,n_1-1,n_2-1}}\right].$$