

Statistiques - Fiche d'exercices n°3

À rendre le lundi 04/05 avant votre TD, à l'adresse : gaspardabel@gmail.com

-> si vous n'arrivez pas à faire tous les exercices, aucun problème, mais envoyez vos pistes de réflexion (au propre si possible).

N'hésitez pas à me contacter par mail avant si vous avez des questions.

Exercice 1:

Un échantillon aléatoire $X_1, X_2, X_3, \dots, X_{100}$ est donné à partir d'une distribution dont la variance $V(X_i) = 16$ est connue. Pour l'échantillon observé, la moyenne de l'échantillon est $\bar{X} = 23,5$.

- 1) Trouvez un intervalle de confiance approximatif à 95% pour l'espérance $\mu = \mathbb{E}[X_i]$.
- 2) Supposons qu'on ait travaillé avec un échantillon de taille $n = 1024$ avec la même moyenne et variance. Calculez le nouvel intervalle de confiance. Interprétez le résultat.

Exercice 2:

Lors d'un sondage précédant les élections présidentielles, 500 personnes ont été interrogées. Bien que ce ne soit pas le cas en pratique, on suppose pour simplifier les calculs que les 500 personnes constituent un échantillon indépendant et identiquement distribué de la population française. Sur les 500 personnes, 150 ont répondu vouloir voter pour le candidat C1, et 140 pour le candidat C2.

1. Donner une estimation ponctuelle des intentions de votes pour chaque candidat, sous la forme d'un pourcentage f_i . Donner également la variance empirique s_i^2 pour chaque candidat.
2. Donner un intervalle de confiance $I_{i,\gamma}$ au seuil γ pour chacun des deux intentions de votes. Application pour $\gamma = 95\%$. Ici, on pourra approximer la variance de la loi $V(X_i)$ par la variance empirique s_i^2 .
3. Dessiner les deux intervalles de confiance $I_{i,\gamma}$. Peut-on prédire l'élection d'un candidat ?
4. Quelle est la procédure si on veut donner un intervalle de confiance sans faire l'approximation de la variance par sa valeur empirique ?
5. Donner la valeur des intervalles de confiance « absolus » et les dessiner sur un nouveau schéma.
6. Supposons maintenant qu'on cherche un intervalle de confiance à 99.7% de largeur $\pm 2\%$ pour le candidat 1. D'après la formule d'intervalle de confiance absolu, quelle est la taille de l'échantillon nécessaire pour obtenir ce seuil de confiance ?

Statistiques - Fiche d'exercices n°3

Exercice 3:

Un échantillon aléatoire $X_1, X_2, X_3, \dots, X_{16}$ est donné à partir d'une distribution normale avec une moyenne inconnue $\mu = \mathbb{E}[X_i]$ et une variance inconnue $V(X_i) = \sigma^2$. Pour l'échantillon observé, la moyenne de l'échantillon est $\bar{X} = 16,7$, et la variance de l'échantillon est $S_n^2 = 7,5$. On donne $t_{15,0.05} \approx 2.13$

- 1) Citer le théorème à utiliser pour trouver un intervalle de confiance pour μ , et l'écrire dans notre cas.
- 2) Trouvez un intervalle de confiance à 95 % pour μ

Exercice 4:

Soient $n \in \mathbb{N}^*$ et $(x_1, \dots, x_n)$ un échantillon pouvant être observé par un ensemble de variables aléatoires $X_1, \dots, X_n$ suivant la loi de Poisson $\mathcal{P}(\lambda)$, $\lambda > 0$. La loi est définie de manière suivante:

$$\forall k \in \mathbb{N}, \mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$$

Le paramètre λ est à estimer.

A/ Performance de deux estimateurs

1. Définir pour le paramètre λ la fonction de vraisemblance $\mathcal{L}(\lambda)$, avec l'échantillon de données $(x_1, \dots, x_n)$.
2. Écrire la fonction de log-vraisemblance $l(\lambda) = \ln \mathcal{L}(\lambda)$ pour $(x_1, \dots, x_n)$.
3. Trouver le maximum de $l(\lambda)$ qu'on notera λ^* .
4. Calculer $\mathbb{E}[\lambda^*]$. λ^* est-il sans biais?
5. Calculer $V(\lambda^*)$ et sa limite lorsque $n \rightarrow \infty$. Interpréter le résultat.
6. Calculer l'erreur quadratique moyenne de λ^* , $EQM(\lambda^*)$.
7. Soit un nouvel estimateur : $\hat{\lambda} = \frac{1}{3}(2X_1 + X_2)$. Calculer $\mathbb{E}[\hat{\lambda}]$. $\hat{\lambda}$ est-il sans biais?
8. Calculer $V(\hat{\lambda})$ puis $EQM(\hat{\lambda})$. Interpréter le résultat.
9. À partir de quelle valeur de n a-t-on $EQM(\lambda^*) < EQM(\hat{\lambda})$?

B/ Estimation ensembliste de λ

On souhaite construire un intervalle de confiance pour λ à partir de l'estimation ponctuelle donnée par notre EMV λ^* . On suppose que $n = 350$ et que $\lambda^* = 3.50$.

1. Écrire la formule de l'intervalle de confiance pour le paramètre λ , en approximant l'écart-type σ^2 par l'écart-type empirique S_X^2 . On prendra soin de justifier quelles valeurs sont prises pour la moyenne empirique $\bar{X}$ et la variance empirique S_X^2 pour la loi de Poisson.
2. Pour un seuil de confiance $\gamma = 95\%$, calculer cet intervalle.

Statistiques - Fiche d'exercices n°3

Exercice 5:

On dispose d'un échantillon de données $\vec{x} = x_1, \dots, x_n \in \mathbb{R}^+$. Ces données, recueillies auprès de n foyers, sont des données de durée de vie d'un grille pain de la marque *Stroumpf* et il est commun de modéliser ce type de phénomène de durée de vie par une loi de Weibull. On les modélise donc par un ensemble de variables aléatoires $X_1, \dots, X_n$, i.i.d. provenant de la loi de Weibull à deux paramètres $\theta = (\beta, \eta)$, dont la densité est, pour $\beta, \eta > 0$:

$$\forall x \in \mathbb{R}^+, f(x|\theta) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\eta}\right)^\beta\right)$$

1. En partant des valeurs de $x \rightarrow 0^+$ et $x \rightarrow +\infty$, tracer approximativement les lois pour $\beta = 1$, puis pour $\beta = 2$.
2. Écrire la fonction de vraisemblance $\mathcal{L}(\theta) = \prod_{i=1}^n f(x_i|\theta)$ pour ce modèle et la simplifier.

On fait maintenant un changement de variable en écrivant notre loi de Weibull à deux

paramètres avec $\theta' = (\beta, \gamma)$ ainsi : $\forall x \in \mathbb{R}^+, f(x|\theta') = \frac{\beta}{\gamma} x^{\beta-1} \exp\left(-\frac{x^\beta}{\gamma}\right)$.

3. Expliquer comment sont liés les vecteurs θ et θ' .
4. Écrire la log-vraisemblance $l(\theta')$ et calculer les dérivées par rapport à chacun des paramètres: $\frac{d}{d\beta} l(\theta)$ et $\frac{d}{d\gamma} l(\theta)$. Mais ne pas chercher à les résoudre. Pourquoi passe-t-on par le logarithme de la vraisemblance ?
5. On suppose β connu, fixé. Trouver l'estimateur maximum de vraisemblance pour le paramètre γ par rapport à nos données $x_1, \dots, x_n$.