

TD N° 2 : Estimation ponctuelle

EXERCICE 1. (QQ-plot)

Ci-dessous vous trouverez les quantiles associés aux probabilités 0.05, 0.10, ..., 0.95 de la durée de la grossesse (en jours) pour les mères de l'étude CHDS vue en cours. Tracez (idéalement utiliser un notebook pour cela) ces quantiles en fonction de ceux d'une loi $\mathcal{U}(0, 1)$ uniforme sur le segment $[0, 1]$. Décrire la forme de la distribution de la durée de la grossesse par rapport à la loi uniforme.

250, 262, 267, 270, 272, 274, 275, 277, 278, 280, 282, 283, 284, 286, 288, 290, 292, 295, 302.

 Les quantiles peuvent être obtenus depuis la base de données `babies23.data` en utilisant les lignes de codes disponibles dans `StatsDescriptives.ipynb`.

EXERCICE 2. (Espérance et aléatoire)

On considère une population de 6 individus sur lesquels une variable x vaut

$$x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 4, x_5 = 4, x_6 = 5 .$$

Dans la suite, on travaille sur un échantillon aléatoire simple (tirage sans remise) de 2 individus :

- Donner la loi de distribution de la moyenne sur l'échantillon.
- Utilisez cette distribution pour calculer l'espérance et la variance de cet estimateur.

EXERCICE 3. (Quantiles gaussiens)

Supposez que les quantiles y_p (pour tout $p \in \mathbb{R}$) d'une loi $\mathcal{N}(\mu, \sigma^2)$ sont tracés en fonction des quantiles z_p d'une loi $\mathcal{N}(0, 1)$. Montrez que la pente et l'ordonnée à l'origine de la droite des points sont σ et μ respectivement.

EXERCICE 4. (Covariance et échantillon aléatoire)

On considère un échantillon de taille $n = 2$ issu d'un échantillonnage aléatoire simple sur une population de taille $N = 100$. Ainsi on tire un couple (x_{i_1}, x_{i_2}) de manière uniforme parmi tous les couples possibles. Supposons que $x_i = 0$ ou 1 pour tout $i \in \llbracket 1, N \rrbracket$ et que la proportion de 1 dans la population soit (en espérance) p . Calculez $\mathbb{E}[x_{i_1} x_{i_2}]$ et en déduire la covariance entre x_{i_1} et x_{i_2} .

EXERCICE 5. (Moyenne empirique et optimisation)

Montrez que \bar{x}_n (moyenne empirique des x_1, \dots, x_n) est la valeur qui minimise la fonction f définie pour tout $x \in \mathbb{R}$ par

$$f(x) = \frac{1}{n} \sum_{i=1}^n (x_i - x)^2 .$$

Aide : Montrer que pour tout réel x , la relation suivante est vraie :

$$\frac{1}{n} \sum_{i=1}^n (x_i - x)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 + (\bar{x}_n - x)^2 . \quad (1)$$

En déduire le résultat en étudiant la formulation quadratique ainsi donnée. Que vaut la fonction f quand elle atteint son minimum ?

EXERCICE 6. (Biais de la variance empirique)

On suppose que le x_1, \dots, x_n sont *i.i.d.* et ont comme espérance μ et comme variance σ^2 . En déduire la valeur du biais de la variance empirique

$$s_n^2(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 ,$$

c'est-à-dire la valeur de $\mathbb{B}(s_n^2(\mathbf{x})) = \mathbb{E}(s_n^2(\mathbf{x})) - \sigma^2$. Proposer une modification de l'estimateur précédent pour le rendre non biaisé.

Aide : utiliser la relation (1) de l'exercice précédent.

EXERCICE 7. (Intervalle de confiance et théorème central limite)

Lors d'un contrôle de qualité dans une firme pharmaceutique, la quantité d'acide acétylsalicylique x dans un comprimé d'aspirine a été mesurée pour $n = 500$ comprimés prélevés dans une production de $N = 500\,000$ comprimés. On a ainsi obtenu après collecte des 500 mesures :

$$\bar{x}_n = \frac{1}{500} \sum_{i=1}^{500} x_i = 0.496 \text{ mg} .$$

On suppose connue la variance (théorique) des mesures, qui vaut $\sigma^2 = 0.004 \text{ mg}^2$. Construire un intervalle de confiance au niveau 95% de la quantité d'acide acétylsalicylique contenue dans un comprimé. On utilisera pour cela une approximation donnée par le théorème central limite.