TD N° 2 : Estimation ponctuelle

Exercice 1. (QQ-plot)
Ci-dessous vous trouverez les quantiles associés aux probabilités $0.05, 0.10, \ldots, 0.95$ de la durée de la grossesses (en jours) pour les mères de l'étude CHDS. Tracez 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.

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 de 2 individus.

a) Calculez la distribution exacte de la moyenne de x sur l'échantillon.
b) Utilisez cette distribution exacte pour calculer l'espérance et la variance de cet estimateur.

Exercice 3. (Quantiles gaussiens)
Supposez que les quantiles y_p d'une loi $\mathcal{N}(\mu, \sigma^2)$ sont tracés en fonctions 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. (Déterministe vs. aléatoire)
En gardant les notations du cours, dîtes quels sont les éléments aléatoires ou non et expliquez pourquoi.

$$x_1, x_{i_2}, \bar{x}_n, N, \mu, i_1, n$$

Exercice 5. (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 $N = 500000$ comprimés. On a ainsi obtenu après collecte des 500 mesures :

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

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

Exercice 6. (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 $i \in [\![1, N]\!]$ et que la proportion de 1 dans la population est p. Calculez $\mathbb{E}[x_{i_1} x_{i_2}]$ et en déduire la covariance entre x_{i_1} et x_{i_2} pour ce cas particulier.

Exercice 7. (Moyenne empirique et optimisation)
Montrez que $\bar{x}_n$ (moyenne empirique des $x_1, \ldots, x_n$) est la valeur qui minimise la fonction f définie pour tout $x \in \mathbb{R}$ par

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

Aide : on pourra forcer l'introduction du terme $\bar{x}_n$ et développer le carré.

Exercice 8. (Biais de la variance empirique)
On suppose que le $x_1, \ldots, x_n$ sont *i.i.d.*et ont comme espérance μ et comme variance σ^2. Montrer que pour tout réel x, la relation suivante est vraie :

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

En appliquant cette relation pour le chois $x = \mu$, 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 $\mathbb{B}(s_n(\mathbf{x})) = \mathbb{E}(s_n(\mathbf{x})) - \sigma^2$. Proposer une modification de l'estimateur précédent pour le rendre non biaisé.