1. **Estimateurs de variables de Bernoulli.** Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi $\mathcal{B}(p)$. On considère les estimateurs classiques de la moyenne et de la variance
$$\widehat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad ; \quad \widehat{v} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{m})^2.$$
   (a) Calculer le biais et l'erreur moyenne au carré de $\widehat{m}$, vu comme estimateur de p (indication : quelle est la loi de la variable $n\widehat{m}$?).
   (b) Montrer que $\widehat{m}$ donne une suite consistante d'estimateurs de p.
   (c) Calculer le biais et l'erreur moyenne au carré de $\widehat{v}$, vu comme estimateur du paramètre $p(1-p)$. On pourra remarquer que $(X_i)^k = X_i$ pour tout $k \geq 1$.

2. **Meilleur estimateur de la variance d'une gaussienne.** Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes gaussiennes, de loi $\mathcal{N}(m, \sigma^2)$. On pose
$$T = c\sum_{i=1}^{n}\left(X_i - \frac{X_1 + \cdots + X_n}{n}\right)^2,$$
avec $c > 0$.
   (a) Rappeler quelle est la loi de la variable $\frac{1}{c\sigma}T$.
   (b) Calculer en fonction de c le biais, puis l'erreur moyenne au carré de T en tant qu'estimateur de la variance σ^2.
   (c) Quelle valeur de c donne un estimateur sans biais ?
   (d) Quelle valeur de c donne l'estimateur avec la plus petite erreur moyenne au carré ?

3. **Estimateur du maximum de vraisemblance : cas de la loi de Poisson.** Soit $X_1, \ldots, X_n$ des variables indépendantes de loi $\mathcal{P}(\lambda)$. On souhaite estimer le paramètre λ. Étant donné un paramètre $t > 0$ et des valeurs entières $x_1, \ldots, x_n$, on note
$$h(x_1, \ldots, x_n; t) = \log\left(\prod_{i=1}^{n}\mathcal{P}(t)[x_i]\right)$$
avec $\mathcal{P}(t)[x] = \mathrm{e}^{-t}\,\frac{t^x}{x!}$ probabilité de la loi de Poisson $\mathcal{P}(t)$.
   (a) Montrer que
$$h(x_1, \ldots, x_n; t) = (x_1 + \cdots + x_n)\log t - nt - \sum_{i=1}^{n}\log(x_i)!\,.$$
   (b) Les nombres entiers $x_1, \ldots, x_n$ étant fixés, calculer la valeur de t qui maximise la quantité $h(x_1, \ldots, x_n; t)$ ci-dessus. On note cette valeur $t = \mathrm{EMV}(x_1, \ldots, x_n)$, et on dit que t est l'*estimateur du maximum de vraisemblance*. Commenter cette terminologie.

(c) On pose maintenant $T_n = \mathrm{EMV}(X_1, \ldots, X_n)$. Quel estimateur retrouve-t-on ? Montrer que $(T_n)_{n\in\mathbb{N}}$ est une suite d'estimateurs sans biais du paramètre λ, qui est consistante.

4. **Estimateur du maximum de vraisemblance : cas de la loi normale.** Soit $X_1, \ldots, X_n$ des variables indépendantes de loi $\mathcal{N}(m, \sigma^2)$. On souhaite estimer les deux paramètres m et σ^2. Étant donnés des paramètres $u \in \mathbb{R}$ et $v > 0$ et des valeurs réelles $x_1, \ldots, x_n$, on note

$$h(x_1, \ldots, x_n; u, v) = \log\left(\prod_{i=1}^{n} f_{\mathcal{N}(u,v)}(x_i)\right)$$

avec $f_{\mathcal{N}(u,v)}(x) = \frac{1}{\sqrt{2\pi v}}\,\mathrm{e}^{-\frac{(x-u)^2}{2v}}$ densité de la loi gaussienne $\mathcal{N}(u, v)$.

(a) On fixe v et les nombres réels $x_1, \ldots, x_n$. Montrer que la valeur u qui maximise la quantité $h(x_1, \ldots, x_n; u, v)$ est $u = \frac{x_1+\cdots+x_n}{n}$.

(b) Les nombres réels $x_1, \ldots, x_n$ étant fixés, trouver les deux paramètres u et v qui maximise la quantité $h(x_1, \ldots, x_n; u, v)$. On pose $(u, v) = \mathrm{EMV}(x_1, \ldots, x_n)$.

(c) On pose maintenant $(U_n, V_n) = \mathrm{EMV}(X_1, \ldots, X_n)$. Quels estimateurs classiques retrouve-t-on ? Montrer que $(U_n)_{n\in\mathbb{N}}$ et $(V_n)_{n\in\mathbb{N}}$ sont des suites consistantes d'estimateurs des paramètres m et σ^2.

5. **Intervalles de confiance pour des variables de Bernoulli.** Soit $X_1, \ldots, X_n$ des variables indépendantes de loi $\mathcal{B}(p)$. On note $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) Montrer que

$$\lim_{n\to\infty} \mathbb{P}\left[M_n - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq M_n + 1.96\sqrt{\frac{p(1-p)}{n}}\right] = 95\%.$$

(b) Quelle est la valeur maximale possible pour $\sqrt{p(1-p)}$? En déduire que pour n assez grand, $[M_n - \frac{1}{\sqrt{n}}, M_n + \frac{1}{\sqrt{n}}]$ est un intervalle de confiance pour le paramètre p de niveau supérieur à 95%.

(c) Que peut-on dire de la quantité $M_n(1 - M_n)$ lorsque n tend vers l'infini ? En déduire que

$$\left[M_n - \frac{1.96\sqrt{M_n(1-M_n)}}{\sqrt{n}}\ ,\ M_n + \frac{1.96\sqrt{M_n(1-M_n)}}{\sqrt{n}}\right]$$

est un intervalle de confiance asymptotiquement de niveau 95% pour le paramètre p.

6. **Pièce équilibrée.** Un joueur lance une pièce au hasard, et note $X_1, \ldots, X_n$ les résultats pile ou face de ces lancers, le pile étant noté 1 et le face étant noté 0. On suppose donc que $X_1, \ldots, X_n$ sont des variables de Bernoulli $\mathcal{B}(p)$, et l'on souhaite savoir si la pièce est équilibrée, c'est-à-dire si l'hypothèse

$$H_0\ :\ p = \frac{1}{2}$$

est vérifiée.

(a) Quelle hypothèse alternative H_1 est raisonnable ?

(b) On utilise le résultat de l'exercice précédent, qui affirme que l'intervalle

$$\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1.96\sqrt{p(1-p)}}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1.96\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

est un intervalle de confiance pour le paramètre p de niveau asymptotique 95%. En déduire un test pour la paire d'hypothèses (H_0, H_1), qui est asymptotiquement de niveau 95%. Peut-on évaluer sa puissance ?

7. **Tout-terrain.** Un fabriquant de pédaliers de vélo veut produire des boîtiers dont le diamètre est de 41mm, avec une tolérance autorisée égale à $+0.00$mm$/-0.10$mm (standard Pressfit GXP). Autrement dit, pour pouvoir être utilisées, les pièces qu'il produit doivent avoir un diamètre au moins égal à 40.9mm, et au plus égal à 41mm. Il étudie deux procédés de fabrication, qui donnent des pièces dont le diamètre X ou Y suit :

— dans le premier cas, une loi de densité

$$f_X(x) = \begin{cases} 40\,\mathrm{e}^{-40(41-x)} & \text{si } x \leq 41, \\ 0 & \text{sinon.} \end{cases}$$

— dans le second cas, une loi de densité

$$f_Y(y) = \frac{40}{\sqrt{2\pi}}\,\mathrm{e}^{-800(y-40.95)^2}.$$

On ne fera pas attention au fait que dans ces modèles, le diamètre X ou Y pourrait être négatif (ceci arrive avec une probabilité quasiment nulle).

(a) Identifier la loi de $41 - X$, et celle de Y.

(b) Quelle sont les espérance et variance de X ? celles de Y ? Commenter.

(c) Pour $t \leq 41$, que vaut $\mathbb{P}[X \leq t]$? On donne $\mathrm{e}^{-4} \leq 1.9\%$. Construire à partir de la variable X un intervalle de confiance de niveau au moins 98% pour le paramètre $m_X = \mathbb{E}[X]$.

(d) On donne $\int_2^{\infty} \mathrm{e}^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \geq 2\%$. Construire à partir de la variable Y un intervalle de confiance de niveau au plus 96% pour le paramètre $m_Y = \mathbb{E}[Y]$.

(e) Quel procédé permet de produire avec moins de 2% de pertes des boîtiers de pédalier utilisables ?