18 Statistiques inférentielles.

« *Exemples de capacités :* Trouver un intervalle de confiance de la moyenne ;
faire un test de conformité sur la moyenne.

Exercice 1 : ✶ Utiliser l'inégalité de Bienaymé-Tchebychev

Soit $X \hookrightarrow \exp(\lambda)$.

① Écrire l'inégalité de Bienaymé-Tchebychev et montrer que :

$$\forall \varepsilon > 0,\ \mathbb{P}\left(X - \frac{1}{\lambda} \geq \varepsilon\right) \leq \frac{1}{\lambda^2 \varepsilon^2}$$

② En déduire que : $\forall t \in \mathbb{R}_+^*,\ t^2 e^{-t} \leq e$

③ Vérifier par une étude de fonction la qualité de cette inégalité.

Exercice 2 : ✶

Un dé équilibré est lancé 900 fois. Estimer la probabilité d'obtenir entre 140 et 160 fois l'as en utilisant le théorème de Moivre-Laplace et l'inégalité de Bienaymé-Tchebychev. Faire de même avec la probabilité d'obtenir entre 130 fois et 170 fois l'as.

Exercice 3 : ✶✶✶

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et T une variable aléatoire définie sur Ω suivant une loi normale centrée réduite dont la fonction de répartition sera notée ϕ.

① Montrer que : $\forall x > 0,\ 0 < 1 - \phi(x) \leq \dfrac{1}{2x^2}$.

② En déduire l'existence de l'intégrale $\displaystyle\int_0^{+\infty} (1 - \phi(x))\, dx$ puis la calculer.

Exercice 4 : ✶✶

Un cinéma comporte deux salles de n places ($n \geq 30$).
N personnes se présentent à l'entrée et on admet que les choix des spectateurs sont indépendants.

① Soit S le nombre de spectateurs choisissant la première salle. Déterminer la loi de S et de $N - S$.

② Exprimer en fonction de S, N et n la probabilité p que tous les spectateurs ne puissent pas voir le film de leur choix.

③ Si $N = 1000$, comment choisir n si on souhaite avoir $p < 0.01$?

Exercice 5 : ✶✶

Soit $(X_k)_{k \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes telle que, pour tout $k \in \mathbb{N}^*$, la variable X_k suit une loi de Poisson de paramètre $\lambda > 0$.

① Soit $S_n = \sum_{k=1}^{n} X_k$ pour tout $n \in \mathbb{N}^*$. Déterminer la loi de S_n et préciser son espérance et sa variance.

② Appliquer le théorème central limite pour la suite $(X_k)_{k\in\mathbb{N}^*}$ et montrer que, sous certaines conditions, une loi de Poisson peut être approchée par une loi normale dont on précisera les paramètres.

③ Le nombre de clients d'une épicerie de quartier un jour quelconque de l'année suit une loi de Poisson de paramètre 12. On admet que les variables correspondant à des jours différents sont indépendantes. Estimer la probabilité d'avoir au moins 250 clients au cours d'un mois comportant 22 jours ouvrables.

Exercice 6 : ✶ Intervalles de confiance

On a mesuré la tension artérielle systolique, au repos de 30 étudiants en BCPST. Les résultats obtenus ont été consignés dans le tableau suivant :

t_i	13.7	11.9	10.8	12.9	12.5	13.1	10.2	7.5	12.1	13.0	12.9	12.4	12.7	11.6	11.7
	15.8	11.5	11.5	12.4	15.2	14.3	11.4	9.6	14.1	13.6	10.5	12.4	13.7	14.0	10.5

① Calculer l'espérance et la variance de cette série statistique.

② On considère dix classes de même amplitude. Tracer l'histogramme des fréquences et des fréquences cumulées de cette série.

③ Donner des arguments permettant d'admettre que les observations proviennent d'une loi normale.

④ Donner pour μ l'intervalle de confiance au niveau de confiance de 95%.

Exercice 7 : ✶✶ : Simulation et comparaison d'intervalles de confiance - Agro 2015

Dans une population d'individus amateurs de café, un proportion p (inconnue) préfère le robusta à l'arabica. On interroge n individus de cette population et on note $X_i = 1$ si l'individu i préfère le robusta et $X_i = 0$ sinon. On suppose que les X_i sont indépendantes. Soit Z_n le nombre d'individus interrogés préférant le robusta à l'arabica.

① Quelle est la loi suivie par Z_n ? Donner son espérance et sa variance.

② On se propose de simuler informatiquement le tirage des X_i ainsi que les informations statistiques qu'on peut en tirer ; on rappelle à cet effet que la fonction `random()` de la bibliothèque Python `random` renvoie un nombre pseudo-aléatoire qu'on peut supposer uniformément distribué dans 0 et 1.

a. Proposer une fonction Python `observation()` qui, pour n et p donnés en entrée, renvoie une liste de 0 et de 1 correspondant aux valeurs prises par les X_i pour une observation d'un échantillon aléatoire de taille n répondant au schéma de Bernoulli de paramètre p.

b. Proposer une fonction Python `moyempir()` fournissant la moyenne empirique (c'est-à-dire, la fréquence des 1) à partir de la donnée d'un échantillon sous la forme d'une liste de 0 et de 1.

c. Proposer une fonction Python `varempir()` fournissant la variance empirique à partir de la donnée d'un échantillon sous la forme d'une liste de 0 et de 1.

Dans la suite, on souhaite proposer (par diverses méthodes) un intervalle de confiance pour p de niveau de confiance 99%, c'est-à-dire un intervalle que l'on peut calculer à partir des observations dont on dispose (c'est-à-dire Z_n) et auquel p appartient pour plus de 99% des échantillons utilisés.

③ Montrer que pour n assez grand, il existe un nombre u pour lequel :

$$\mathbb{P}\left(\frac{Z_n}{n} - u\sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{Z_n}{n} + u\sqrt{\frac{p(1-p)}{n}}\right) \approx 0.99$$

Pour accéder concrètement au nombre u, on pourra faire appel à la bibliothèque `scipy.stats` qui fournit les fonctions suivantes :

- `norm.cdf()` qui donne la fonction de répartition d'une loi normale centrée réduite.
- `norm.ppf()` qui donne la fonction réciproque de la précédente (également nommée *fonction des quantiles*)

④ Montrer que pour tout $p \in [0,1]$, $p(1-p) \leq 1/4$. En déduire un premier intervalle de confiance à 99% pour p qui sera note I_1 dans la suite.

⑤ Proposer une fonction Python `ic1()` fournissant (sous forme d'une liste de deux valeurs) un intervalle de confiance pour p au niveau 99%, à partir d'un échantillon fourni sous la forme d'une liste de 0 et de 1.

⑥ En recourant à la seconde forme du TCL, proposer un autre intervalle de confiance de niveau 99% por p, qui sera noté I_2 dans la suite.

⑦ Comparer les intervalles de confiance I_1 et I_2 (lorsque n est suffisamment grand).

⑧ Le principe des intervalles de confiance est que, si l'on dispose d'un grand nombre d'échantillons issus d'un tirage de Bernoulli de paramètre p et de longueur n (n grand), alors p doit appartenir, dans 99% des cas, aux intervalles de confiance I_1 et I_2. Vérifier ce faite par simulation.