16 T.D. Théorèmes limites

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

Exercice 1 : ✶ Obtenir des inégalités grâce à bienaymé-Tchebychev

① Soit $X \hookrightarrow \exp(\lambda)$.

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

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

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

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

② Montrer que, $\forall x \in \mathbb{R}_+^*,\, \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-t^2/2} dt \geq \frac{1}{2}\left(1 - \frac{1}{x^2}\right)$

Exercice 2 : ✶

Une urne contient en proportion égale des boules blanches et des boules noires. On extrait successivement, avec remise, six boules de l'urne. On marque deux points pour une boule blanches et cinq points pour une boule noire. Soit S la variable aléatoire égale à la somme des points obtenus.

① Quelle est la loi de S ? Préciser son espérance μ et sa variance σ_S

② En utiliser l'inégalité de Bienaymé-Tchebychev, calculer la probabilité pour que $|S - \mu|$ soit strictement supérieur à 5 ou strictement supérieur à 10.

③ Effectuer le même calcul en utilisant la loi de probabilité de S. Comparer les résultats.

Exercice 3 : ✶ Savoir simuler une variable aléatoire à densité

Dans cet exercice, hormis les fonctions des bibliothèques `math` et `matplolib.pyplot`, la seule fonction utilisable est la fonction `rd.random()` de la bibliothèque `random`.

① Écrire une fonction Python `loiUniforme` d'arguments a et b, deux réels, et renvoyant une simulation de la loi uniforme sur $[a, b]$.

② **a.** Écrire, en vous aidant du théorème central limite sous sa première forme, une fonction `Zncr()` qui simule une loi normale centrée réduite.

b. Écrire une fonction Python permettant de comparer la densité de la loi normale centrée réduite avec un histogramme obtenu à l'aide de la fonction `Zncr`.

c. Généraliser la question précédente et écrire une fonction permettant de comparer une densité de la loi normale de paramètres m et σ^2 avec un histogramme issu de l'exécution répétée de `sig*Zncr()+m`.

Exercice 4 : ✶✶

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 *Application 1 :* Déterminer $\underset{n\to\infty}{lim} \left(e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \right)$.

④ *Application 2 :* Une région comporte dix structures hospitalières ayant chacune une capacité opératoire journalière de dix patients. Le nombre de personnes se présentant chaque jour pour être opérées dans chacune de ces structures suit une loi de Poisson de paramètre 8, ces nombres étant supposés indépendants d'un hôpital à l'autre.
On considère un jour donné.

a. Quelle est la probabilité qu'un hôpital donné soit obligé de refuser un patient ?
b. Quelle est la probabilité q que l'un au moins des hôpitaux soit obligé de refuser un patient ?
c. On suppose maintenant qu'un hôpital saturé a la possibilité de se « délester » sur un autre qui ne l'est pas. Quelle est la probabilité r pour qu'un patient ne puisse se faire opérer un jour donné ?

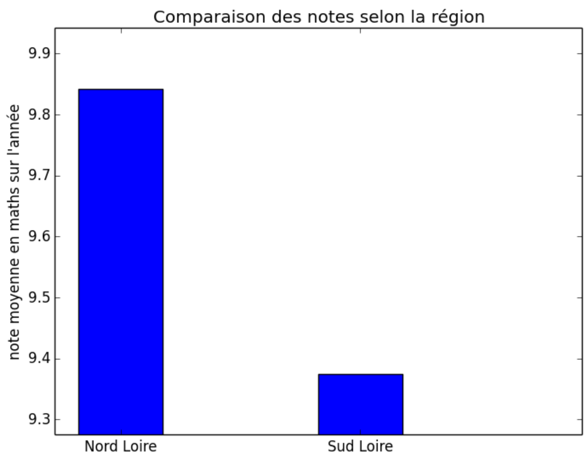
Exercice 5 : ✶ Intervalles de confiance

① On suppose qu'un échantillon de $n = 106$ températures corporelles prises à midi sur des individus en bonne santé donne une moyenne de 36.78^oC et un écart-type empirique de 0.34^oC. Peut-on considérer que $\mu = 37^o$C au seuil de risque de 5% ?

② On calcule sur 27 étudiants de BCPST la moyenne annuelle des notes en maths selon qu'il est originaire du nord de la Loire ou du sud Loire. Les notes étant les suivantes :

$$NL = [12, 8, 16, 16, 3, 11, 9, 12, 4, 3, 12, 3, 13, 13, 11, 16, 7, 12, 6] \text{ et } SL = [10, 13, 8, 4, 16, 7, 12, 5]$$

on obtient les moyennes suivantes : Que pouvez-vous conclure ?



Exercice 6 : ✶✶ Estimation du risque par simulation

Max et Jojo disposent d'une urne contenant une boule blanche et une boule noire. Max effectue des tirages selon le protocole suivant : Il commence par tirer au hasard une boule de l'urne.

- Si elle est noire, il arrête de tirer et la partie est terminée.
- Si elle est blanche, il la remet dans l'urne et il y rajoute une boule noire ; puis il poursuit les tirages selon la même règle.

On admet que toute partie s'arrête avec une probabilité égale à 1 et qu'on peut donc définir la variable X suivante :
On désigne par X la variable aléatoire égale au nombre de tirages effectués par Max lors de l'arrêt de la partie.
Si X est impair et vaut n, Max donne n euros à Jojo. Le gain en euros de Jojo vaut $G = n$.
Si X est pair et vaut n, Jojo donne n euros à Max. Le gain en euros de Jojo vaut $G = -n$.

① Écrire une fonction `tirage()` qui simule une partie et retourne la valeur prise par X.

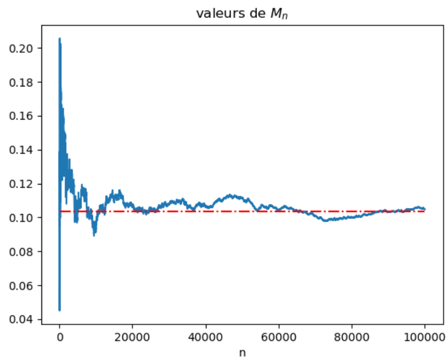
② Déterminer la loi de X.

③ Exprimer la variable aléatoire G simplement en fonction de X. Justifier que G admet une espérance (✐ *Le calcul effectif de l'espérance est traité à la question suivante*).

④ **a.** Déterminer trois réels a, b et c tels que pour tout entier naturel n : $n^2 = an(n+1) + b(n+1) + c$.
b. En déduire la valeur de $\mathbb{E}(G)$.

⑤ Max et Jojo décident de jouer un grand nombre de parties.

On désigne par G_n le gain en euros de Jojo à la n-ième partie. On représente ci-contre les moyennes successives M_n des gains au cours des 100000 parties où $M_n = \dfrac{G_1 + \cdots + G_n}{n}$.
Commenter le graphique.
Celui-ci vous semble-t-il cohérent ?



⑥ On admet que les variables G_n ont une variance non nulle. On note m l'espérance de G_n et σ l'écart-type de G_n.
On note pour tout $n \in \mathbb{N}^*$, $M_n^* = \dfrac{M_n - m}{\sigma/\sqrt{n}}$

a. Donner pour $\alpha \in \mathbb{R}_+^*$, l'approximation de la probabilité $\mathbb{P}(-\alpha < M_n^* < \alpha)$ donnée par le théorème central limite.
b. En déduire une valeur de n à partir de laquelle l'intervalle $[M_n - 0.01; M_n + 0.01]$ est un intervalle de confiance de m au seuil de risque de 95%. ✐ *On admettra que* $\sigma \approx 1.93$.

⑦ Changeons les règles en modifiant le protocole. Max commence par tirer au hasard une boule de l'urne.

- Si elle est noire, il arrête de tirer et la partie est terminée.
- Si elle est blanche, il la remet dans l'urne et il y rajoute une boule blanche, puis il poursuit les tirages selon les mêmes règles.

On désigne par X' la variable aléatoire égale au nombre de tirages effectués par Max lors de l'arrêt de la partie et par G' le gain. Déterminer la loi de X', puis à l'aide d'un programme Python, représenter les moyennes successives M'_n des gains au cours de 50000 parties où, pour tout $n \in \mathbb{N}^*$, $M'_n = \dfrac{G'_1 + \cdots + G'_n}{n}$

Qu'observez-vous en faisant plusieurs simulations ? Proposer une explication.

Exercice 7 : ✶✶ Estimation du risque par simulation

Soit X une variable aléatoire.
Pour n entier naturel non nul, on souhaite générer 100000 n-échantillons associés à la variable aléatoire X et, pour différentes lois, différentes valeurs de n, valider le théorème central limite en considérant des intervalles de confiance au seuil de confiance de 95%.
On considérera successivement que X suit la loi normale centrée-réduite, une loi symétrique comme la loi uniforme sur $[-1;1]$ et une loi asymétrique comme la loi exponentielle de paramètre 1

① En supposant que M_n suit une loi normale d'espérance μ et de variance inconnue, écrire les bornes de l'intervalle de confiance de la moyenne en fonction de la taille de l'échantillon.

② Pour chacune des lois ci-dessous, générer 1000 n-échantillons de la variable aléatoire X pour des tailles d'échantillon $n = 10$, 30 et 100 et retourner le pourcentage de fois où l'espérance μ de X n'est pas dans l'intervalle de confiance au niveau de confiance de 95% en considérant que la variance est connue puis qu'elle est inconnue.

a. $X \hookrightarrow \mathcal{N}(0,1)$
b. $X \hookrightarrow \mathcal{U}_{[-1,1]}$
c. $X \hookrightarrow \mathcal{E}(\lambda)$

Exercice 8 : ✶✶ Estimation du risque par simulation

On souhaite estimer un paramètre $p \in]0,1[$. on note $q = 1-p$.
Soit un entier $n \geq 1$ fixé. On considère $X_1, X_2, \cdot, X_n$ des variables aléatoires, indépendantes, suivant une même loi de Bernoulli de paramètre p et définies sur un même espace probabilisé.
On note $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$.

① **a.** Justifier que $p(1-p) \leq \dfrac{1}{4}$.

b. En utilisant l'inégalité de Bienaymé-Tchebychev, montrer que l'intervalle $\left[\overline{X_n} - \sqrt{\dfrac{5}{n}}, \overline{X_n} + \sqrt{\dfrac{5}{n}}\right]$ est un intervalle de confiance de p au niveau de confiance 0.95.

② Écrire une fonction Python `test(n,p,a,b)` qui prend en arguments un entier n, une probabilité p, deux flottants a et b, simule une réalisation de $\overline{X_n}$ et qui retourne 1 si $\overline{X_n}$ appartient à $[a,b]$ et 0 sinon.
Utiliser cette fonction pour valider la réponse obtenue en 1.b).
On cherche par la suite un intervalle de confiance de p au niveau de confiance 0.95 d'amplitude plus petite.

③ On fixe un réel strictement positif t quelconque et ε un réel strictement positif quelconque.

a. Établir l'égalité : $\mathbb{P}\left(\overline{X_n} - p \geq \varepsilon\right) = \mathbb{P}\left(e^{nt\overline{X_n}} \geq e^{nt(p+\varepsilon)}\right)$.

b. En utilisant l'inégalité de Markov, établir l'inégalité suivante : $\mathbb{P}\left(\overline{X_n} - p \geq \varepsilon\right) \leq e^{n(\ln(pe^t+q)-t(p+\varepsilon))}$.

④ On admet l'inégalité : $\ln(pe^t + q) - tp \leq \dfrac{t^2}{8}$. Ainsi, on a l'inégalité suivante :

$$\forall t \in \mathbb{R}_+^*,\ \mathbb{P}(\overline{X_n} - p \geq \varepsilon) \leq e^{n\left(\frac{t^2}{8} - t\varepsilon\right)}$$

En déduire l'inégalité : $\mathbb{P}(\overline{X_n} - p \geq \varepsilon) \leq e^{-2n\varepsilon^2}$

⑤ Déduire des questions 3.c) et 4. l'inégalité :

$$\forall \varepsilon \in \mathbb{R}_+^*,\ \mathbb{P}\left(|\overline{X_n} - p| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$$

⑥ Comment choisir ε pour obtenir un intervalle de confiance de p au niveau de confiance 0.95 ? L'amplitude de l'intervalle de confiance est-elle plus réduite que celle obtenue à la question 1.b) ?