Contrôle

La calculatrice et les documents sont autorisés. Toutes vos réponses doivent être soigneusement rédigées. Le barème est donné à titre indicatif

Exercice 1 (3 points)

Un joueur lance un dé. S'il obtient un 6 il gagne 8 euros. Sinon, il perd 2 euros.

1. Calculer l'espérance de gain du joueur.

$$E = 8 \times \frac{1}{6} + (-2) \times \frac{5}{6} = -\frac{1}{3}.$$

On propose au joueur d'acheter pour n euros un deuxième dé avant de jouer. Il peut dans ce cas lancer les deux dés. Les gains sont les mêmes : 8 euros s'il obtient au moins un 6, et -2 euros sinon.

2. Pour quelles valeurs de n cet achat rend-il le jeu plus avantageux ?

Calculons l'espérance de gain de ce second jeu : il a 11 chances sur 36 d'obtenir au moins un 6.

$$E' = 8 \times \frac{11}{36} + (-2) \times \frac{25}{36} = \frac{38}{36}.$$

Mais pour comparer avec le premier jeu, il faut tenir compte de l'achat du dé. Ce jeu est plus favorable si $\frac{38}{36} - n \geq -\frac{1}{3}$. Le jeu est donc plus avantageux pour $n < \frac{50}{36} \approx 1,39$ euros.

Exercice 2 : loi de Pareto (10 points)

Soit $k > 1$. La loi de Pareto de paramètre k, notée $\mathcal{P}ar(k)$, est donnée par la densité de probabilité f nulle sur $]-\infty, 1[$ et définie sur $[1, +\infty]$ par

$$f(x) = \frac{k}{x^{k+1}}.$$

Soit $X \sim \mathcal{P}ar(k)$.

1. (a) Vérifier que f est bien une densité de probabilité.
   f est une fonction positive et $\int_{-\infty}^{+\infty} f(x)dx = \int_1^{+\infty} \frac{k}{x^{k+1}}dx = [-\frac{1}{x^k}]_1^{+\infty} = 1$.

   (b) Calculer l'espérance de X.

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_1^{+\infty} \frac{k}{x^k}dx = [-\frac{k}{(k-1)x^{k-1}}]_1^{+\infty} = \frac{k}{k-1}.$$

(On rappelle que $k > 1$.)

(c) Soit $a > 1$. Calculer $\mathbf{P}(X > a)$.

$$\mathbf{P}(X > a) = \int_a^{+\infty} \frac{k}{x^{k+1}} dx = [-\frac{1}{x^k}]_a^{+\infty} = \frac{1}{a^k}.$$

(d) Déterminer la médiane de X, c'est-à-dire la valeur m telle que $\mathbf{P}(X < m) = \mathbf{P}(X > m)$.

On cherche donc m tel que $\mathbf{P}(X > a) = \frac{1}{2}$. Donc $\frac{1}{m^k} = \frac{1}{2}$ et $m = 2^{\frac{1}{k}}$.

(e) Soit $b > 1$. Calculer $\mathbf{P}(X > a + b|X > a)$. Montrer que cette probabilité croît avec a.

$$\mathbf{P}(X > a+b|X > a) = \frac{\mathbf{P}(X > a + b \cap X > a)}{\mathbf{P}(X > a)} = \frac{\mathbf{P}(X > a + b)}{\mathbf{P}(X > a)} = \frac{a^k}{(a + b)^k} = (\frac{a}{a + b})^k.$$

La fonction $a \mapsto \frac{a}{a+b}$ est croissante puisque sa dérivée $\frac{b}{(a+b)^2}$ est positive. Donc comme $k > 1$, $(\frac{a}{a+b})^k$ est croissante avec a.

(f) Si X représente une durée de vie, comment peut-on interpréter ce résultat ?

On a calculé la probabilité qu'ayant vécu un temps a, on vive un temps b supplémentaire. la croissance de cette probabilité avec a signifie donc que plus on a vécu longtemps, plus la probabilité de vivre un temps b supplémentaire est élevée. Autrement dit, le temps passé ne fait qu'augmenter le temps qu'on peut espérer vivre encore. On peut dire que le système rajeunit.

2. (a) Posons $Y = \ln(X)$. Montrer que Y suit la loi exponentielle de paramètre k.

Cherchons la fonction de répartition de Y. Soit $x > 0$.

$$F_Y(x) = \mathbf{P}(Y \leq x) = \mathbf{P}(\ln(X) \leq x) = \mathbf{P}(X \leq e^x) = 1 - \frac{1}{(e^x)^k} = 1 - e^{-kx}.$$

Sa densité est donc donnée par la dérivée de cette fonction :

$$f_Y(x) = F'_Y(x) = ke^{-kx}.$$

On reconnaît la densité de la loi exponentielle de paramètre k.

(b) Soit $n \geq 30$. Écrire le théorème central limite pour n tirages Y_j indépendants de la variable Y. L'espérance de Y est $1/k$ et sa variance $1/k^2$. On peut donc écrire le théorème central limite :

$$\frac{\sum_{j=1}^n Y_j - n/k}{\sqrt{n}/k} = \frac{1}{\sqrt{n}}(k \sum_{j=1}^n Y_j - n) \sim \mathcal{N}(0, 1).$$

(c) On dispose d'un échantillon de 50 valeurs de la variable X tirées indépendamment. On souhaite estimer la valeur du paramètre k. On sait simplement que le produit des 50 valeurs vaut approximativement 500000.

Donner une estimation de k à l'aide de la variable Y.

Notons $x_1, \ldots, x_{50}$ les résultats des tirages. On sait que la moyenne de Y est $1/k$. Or $\ln(x_1), \ldots, \ln(x_{50})$ constituent des tirages indépendants de la variable Y. Leur moyenne empirique fournit donc une estimation de $1/k$:

$$\frac{\ln(x_1) + \cdots + \ln(x_{50}}{50} = \frac{\ln(x_1 x_2 \cdots x_{50})}{50} \approx \frac{1}{k}.$$

On peut ainsi estimer k par

$$k \approx \frac{50}{\ln(500000)} \approx 3,81.$$

(d) Donner un intervalle de confiance à 95% pour k.

On utilise pour cela le théorème central limite. On sait que

$$\mathbf{P}(-1,96 < \frac{1}{\sqrt{50}}(k\sum_{j=1}^{50} Y_j - 50) < 1,96) \approx 0,95.$$

On dispose ici de $\sum_{j=1}^{50} Y_j = \ln(500000) \approx 13,12$. Donc k est situé dans l'intervalle de confiance à 95%

$$[(50 - \sqrt{50}1,96)/13,12; (50 + \sqrt{50}1,96)/13,12] = [2,75; 4.87].$$

Exercice 3 : Valeur moyenne d'un signal bruité (7 points)

On reçoit un signal f entre les instants $t = 0$ et $t = 1$ et on souhaite calculer sa valeur moyenne donnée par $\int_0^1 f(t)dt$. Le résultat ne sera pas exact pour deux raisons. Le signal a été échantilloné, l'intégrale sera donc approchée par la moyenne des valeurs observées et de plus le signal émis a été perturbé par la présence d'un bruit. Le but de cet exercice est de mesurer l'impact qu'a ce bruit sur l'exactitude du résultat obtenu.

On note f le signal émis et ε le bruit. Ces fonctions sont définies sur $[0,1]$. Le signal reçu est ainsi donné par $g = f + \varepsilon$. Le signal a été échantilloné avec un pas de $\frac{1}{n}$ avec $n \in \mathbf{N}^*$. On on a donc reçu les valeurs $g(\frac{k}{n})$ pour $k = 1 \dots n$. La valeur moyenne de f sera donc estimée par $\frac{1}{n}\sum_{k=1}^{n} g(\frac{k}{n})$. En l'absence de bruit, on sait que l'erreur commise pour la fonction f considérée aurait été donnée par $|\int_0^1 f(t)dt - \frac{1}{n}\sum_{k=1}^{n} f(\frac{k}{n})| < \frac{10}{n}$.

Le bruit ε est modélisé à l'aide de la loi normale. On suppose que les variables $X_k = \varepsilon(\frac{k}{n})$ pour $k = 1 \dots n$ sont des variables indépendantes de même loi $\mathcal{N}(0,1)$.

On prendra $n = 100$ comme fréquence d'échantillonage (hormis dans les questions 3 et 4).

1. Appliquer le théorème central limite aux variables X_k.
   Les X_k ont pour loi $\mathcal{N}(0,1)$. Elles sotn donc d'espérance nulle et de variance 1. Comme elles sont de plus indépendantes et que $n = 100 > 30$, on peut appliquer le théorème central limite :
   $$\frac{\sum_{k=1}^{n} X_k - 0}{1 \times \sqrt{n}} \text{ suit à peu près la loi } \mathcal{N}(0,1).$$
2. Déterminer la probabilité p que $|\frac{1}{n}\sum_{k=1}^{n} g(\frac{k}{n}) - \frac{1}{n}\sum_{k=1}^{n} f(\frac{k}{n})| < 0,1$.

   $|\frac{1}{n}\sum_{k=1}^{n} g(\frac{k}{n}) - \frac{1}{n}\sum_{k=1}^{n} f(\frac{k}{n})| = |\frac{1}{n}\sum_{k=1}^{n} \varepsilon(\frac{k}{n})| = |\frac{\sum_{k=1}^{n} X_k}{n}|$. Ainsi

   $$p = \mathbf{P}(|\frac{\sum_{k=1}^{n} X_k}{n}| < 0,1) = \mathbf{P}(-0,1\sqrt{n} < \frac{\sum_{k=1}^{n} X_k}{\sqrt{n}} < 0,1\sqrt{n}) = \mathbf{P}(-1 < \frac{\sum_{k=1}^{n} X_k}{\sqrt{n}} < 1)$$

   car $n = 100$. En utilisant le théorème central limite, on obtient

   $$p \approx F_{\mathcal{N}}(1) - F_{\mathcal{N}}(-1) = 2F_{\mathcal{N}}(1) - 1 \approx 0,68.$$

3. Déterminer une valeur de n, c'est-à-dire une fréquence d'échantillonage, telle que cette probabilité p soit de l'ordre de 0,99.

   On prend n quelconque supérieur à 30. On a encore

   $$p = \mathbf{P}(-0,1\sqrt{n} < \frac{\sum_{k=1}^{n} X_k}{\sqrt{n}} < 0,1\sqrt{n}) = 2F_{\mathcal{N}}(0,1\sqrt{n}) - 1.$$

   On souhaite avoir $p = 0,99$. Il faut donc avoir $F_{\mathcal{N}}(0,1\sqrt{n}) = 0,995$. Ainsi, d'après les tables, $0,1\sqrt{n} = 2,58$. Donc $n = 665,64$. Il faut donc prendre au moins $n = 666$ pour que la probabilité que l'erreur soit supérieure à 10% soit inférieure à 1%.

4. Que peut-on alors dire de la précision de la valeur moyenne de f obtenue ?

   Il y a deux erreurs commises dans le calcul de la valeur moyenne de f. On peut majorer l'erreur totale ainsi :

   $$|\int_0^1 f - \frac{1}{n}\sum_{k=0}^{n} g(\frac{k}{n})| \leq |\int_0^1 f - \frac{1}{n}\sum_{k=0}^{n} f(\frac{k}{n})| + |\frac{1}{n}\sum_{k=0}^{n} f(\frac{k}{n}) - \frac{1}{n}\sum_{k=0}^{n} g(\frac{k}{n})|$$

   Pour $n = 666$, le premier terme est majoré par $10/666 \approx 0,002$ et le second par $0,1$ avec un risque de 1%. L'erreur totale est donc majorée par $0,102$. On constate que c'est le bruit qui est la principale cause d'erreur dans le calcul.

5. Si on ne peut pas augmenter la fréquence d'échantillonage $n = 100$, on peut essayer de diminuer le bruit. Déterminer une valeur de l'écart-type σ telle que si le bruit est décrit par une loi normale $\mathcal{N}(0, \sigma^2)$, alors la probabilité p sera de l'ordre de $0,99$.

   On réécrit le théorème central limite pour $n = 100$:

   $$\frac{\sum_{k=1}^{100} X_k}{10\sigma} \text{ suit à peu près la loi } \mathcal{N}(0,1).$$

   Ainsi $p = \mathbf{P}(-0,1 < \frac{\sum_{k=1}^{100} X_k}{100} < 0,1) = \mathbf{P}(-\frac{1}{\sigma} < \frac{\sum_{k=1}^{100} X_k}{10\sigma} < \frac{1}{\sigma})$. Pour avoir $p = 0,99$, il faut donc avoir $2F_{\mathcal{N}}(\frac{1}{\sigma}) - 1 = 0,99$. Donc $F_{\mathcal{N}}(\frac{1}{\sigma}) = 0,995$ et $\frac{1}{\sigma} \approx 2,58$. Il faut donc avoir un écart-type au plus égal à $\sigma = 0,39$.