Devoir 1 de Statistique

Durée : une heure. Les calculatrices sont autorisées ; tout autre matériel électronique (comme le téléphone portable) est interdit.

Exercice 1. (5 pts)

On dispose d'un dé équilibré à n faces numérotées de 8 à $7+n$.

1. Que valent l'espérance et la variance d'une loi uniforme discrète entre 1 et n ?

   Si X est une variable aléatoire uniforme dont les valeurs sont les entiers compris entre 1 en n, son espérance est $E(X) = (1+n)/2$ et sa variance $V(X) = (n^2 - 1)/12$.

2. Déduisez-en l'espérance et la variance d'une loi uniforme discrète entre 8 et $7+n$.

   Si Y est uniforme entre 8 et $7+n$, alors $X = Y - 7$ est uniforme entre 1 et n, ce dont on déduit : $E(Y-7) = (1+n)/2$ et $V(Y-7) = (n^2-1)/12$. Or $E(Y-7) = E(Y) - 7$ et $V(Y-7) = V(Y)$. Donc : $E(Y) = 7 + (1+n)/2$ et $V(Y) = (n^2-1)/12$.

3. Combien le dé a-t-il de faces si la variance est égale à l'espérance des faces qui apparaissent ?

   Lorsqu'on lance un dé équilibré à n faces numérotées de 8 à $7+n$, la face qui apparaît est une variable aléatoire Y uniforme entre 8 et $7+n$. Si son espérance est égale à sa variance, alors : $7 + (1+n)/2 = (n^2-1)/12$. Cette équation du second degré en n est équivalente à : $n^2 - 6n - 91 = 0$; elle admet une unique solution positive : $n = 7$. Le dé a donc sept faces.

Exercice 2. (5 pts)

20% de la population possède le gène A ; α désigne la probabilité d'observer de dix à douze sujets porteurs de A dans un échantillon de trente personnes.

Déterminez la valeur exacte de α, puis une valeur approchée de α basée sur l'approximation de la loi binomiale par une loi normale[a].

Le nombre X de sujets porteurs de A dans un échantillon de trente personnes est distribué selon une loi binômiale de paramètres $n = 30$ et $p = 0,2$: $X \sim \mathscr{B}(30; 0,2)$.

On en déduit : $\alpha = \mathbb{P}(10 \leq X \leq 12) = \sum_{k=10}^{12} \mathbb{P}(X = k) = \sum_{k=10}^{12} \binom{30}{k} 0,2^k 0,8^{30-k} \approx 0,058$ ou en approchant la loi binomiale $\mathscr{B}(0,2; 30)$ par la loi normale de moyenne $\mu = 30 \times 0,2 = 6$ et de variance $\sigma^2 = 30 \times 0,2 \times 0,8 = 4,8$: $\mathbb{P}(10 \leq X \leq 12) = \mathbb{P}(10 \leq \mathscr{B}(30; 0,2) \leq 12) \underset{\text{cont.}}{\approx} \mathbb{P}(9,5 \leq \mathscr{N}(6; 4,8) \leq 12,5) = \phi((12,5-6)/\sqrt{4,8}) - \phi((9,5-6)/\sqrt{4,8}) \approx 0,54$.

Exercice 3. (5 pts)

M. Li fabrique des filets de deux citrons et trois oranges. La masse, en gramme, des citrons produits est distribuée selon une loi normale de moyenne 100 et de variance 100 et la masse des oranges selon une loi normale de moyenne 120 et de variance 150.

M. Li envisage deux types de filets : dans chaque filet de type A, la masse des deux citrons est la même et celle des oranges est libre ; dans les filets de type B, la masse des citrons, comme celle des oranges, est libre.

Déterminez dans chacun des cas la probabilité de produire un filet dont la masse est inférieure à 500 g.

- Filets de type A.

La masse X du premier citron est distribuée selon $\mathscr{N}(100, 100)$; puisque les deux citrons ont la même masse, la masse des deux citrons réunis est : $G = 2X \sim \mathscr{N}(2 \times 100 = 200; 2^2 \times 100 = 400)$.

Pour $i = 1, 2, 3$, la variable Y_i représentant la masse de l'orange i est distribuée selon $\mathscr{N}(120; 150)$; en supposant Y_1, Y_2, Y_3 indépendantes, la masse des trois oranges réunies est $H = \sum_{i=1}^{3} Y_i \sim \mathscr{N}(3 \times 120 = 360; 3 \times 150 = 450)$.

En admettant que la masse globales des citrons G est indépendante de celles des oranges H, la masse du filet $F = G + H$ est distribuée selon $\mathscr{N}(200 + 360 = 560; 400 + 450 = 850)$ et $\mathbb{P}(X \leq 500) = \mathbb{P}(\mathscr{N}(560; 850) \leq 500) = \phi((500-560)/\sqrt{850}) = \phi(-2,06) = 1 - \phi(2,06) \approx 0,2$.

[a] $\binom{30}{10} = 30045015$, $\binom{30}{11} = 54627300$, $\binom{30}{12} = 86493225$

• Filets de type B.

La masse H des trois oranges réunies est toujours distribuée selon $\mathcal{N}(360, 450)$, mais dans les filets B, la masse X_1 du citron 1 est indépendante de la masse X_2 du citron 2 ; ainsi, la masse globale des deux citrons est : $G = X_1 + X_2 \sim \mathcal{N}(100+100 = 200; 100+100 = 200)$. La masse du filet $F = G + H$ est distribuée selon $\mathcal{N}(200 + 360 = 560; 200 + 450 = 650)$ et $\mathbb{P}(X \leq 500) = \mathbb{P}(\mathcal{N}(560; 650) \leq 500) = \phi((500 - 560)/\sqrt{650}) = \phi(-2,35) = 1 - \phi(2,35) \approx 0,009$.

Exercice 4. (5 pts)

Combien de fois faut-il lancer une pièce de monnaie équilibrée[b] pour que la fréquence de Pile ait 95% de chances d'être comprise entre $0,48$ et $0,52$?

n désigne le nombre de lancers et X_i une variable aléatoire prenant la valeur 1 si Pile apparait au lancer i et 0 sinon. Les variables X_i ($i = 1, \ldots, n$) sont distribuées selon une loi de Bernoulli de paramètre $1/2$: leur espérance vaut $0,5$ et leur variance $0,25$. En supposant les variables X_i indépendantes et d'après le théorème *central limit*, la fréquence de Pile : $F = \sum_{i=1}^{n} X_i/n$ est approximativement distribuée selon une loi normale de moyenne $0,5$ et de variance $0,25/n$. Ainsi, $\mathbb{P}(0,48 \leq F \leq 0,52) \approx \mathbb{P}(0,48 \leq \mathcal{N}(0,5; 0,25/n) \leq 0,52) = \phi((0,52 - 0,5)/\sqrt{0,25/n}) - \phi((0,48 - 0,5)/\sqrt{0,25/n}) = \phi(0,04\sqrt{n}) - \phi(-0,04\sqrt{n}) = 2\phi(0,04\sqrt{n}) - 1$. On cherche donc n de sorte que : $2\phi(0,04\sqrt{n}) - 1 = 0,95$. Ainsi : $\phi(0,04\sqrt{n}) = 0,975$; $0,04\sqrt{n} = 1,96$ et $n = (1,96/0,04)^2 \approx 2401$.

Il faut 2401 lancers d'une pièce équilibrée pour que la fréquence de Pile soit comprise entre $0,48$ et $0,52$ avec 95% de chances.

---

[b] la pièce est équilibrée, la monnaie ne l'est pas forcément