

Exercice

Un vol Marseille - Paris est assuré par un Airbus de 150 places; pour ce vol des estimations ont montré que la probabilité pour qu'une personne confirme son billet est $p = 0,75$. La compagnie vend n billets, $n > 150$. Soit X la variable aléatoire égale au nombre de personnes parmi les n possibles, ayant confirmé leur réservation pour ce vol.

1. Quelle est la loi exacte suivie par X ?
2. Quel est le nombre maximum de places que la compagnie peut vendre pour qu'elle soit sûre à (au moins) 95% que tout le monde puisse monter dans l'avion.

Corrigé

1. On vérifie sans problème que X suit la loi binomiale de paramètres n et p .
2. On cherche n tel que $\mathbb{P}(X \leq 150) \geq 0,95$. D'après le théorème de Moivre-Laplace, on peut approcher la loi de X par la loi normale d'espérance np et de variance $np(1-p)$ et de manière équivalente, la loi de $X^* = \frac{X - np}{\sqrt{np(1-p)}}$ par la loi normale centrée réduite. En notant Φ la fonction de répartition de la loi normale centrée réduite, remarquons que :

$$\begin{aligned} \mathbb{P}(X \leq 150) &= \mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{150 - np}{\sqrt{np(1-p)}}\right) \\ &= \mathbb{P}\left(X^* \leq \frac{150 - np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{150 - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

Puisque $\Phi(1,65) \geq 0,95$, on est ramené à résoudre l'inéquation :

$$\begin{aligned} \Phi\left(\frac{150 - np}{\sqrt{np(1-p)}}\right) \geq 0,95 &\Leftrightarrow \frac{150 - np}{\sqrt{np(1-p)}} \geq 1,65 \text{ par croissance de } \Phi \text{ sur } \mathbb{R} \\ &\Leftrightarrow np + 1,65\sqrt{np(1-p)} - 150 \leq 0 \\ &\Leftrightarrow 0,75x^2 + 1,65\sqrt{0,1875x} - 150 \leq 0 \text{ en posant } x = \sqrt{n} \\ &\Leftrightarrow x \in [0, x_1] \text{ où } x_1 \approx 13,61 \\ &\Leftrightarrow n \in [0, x_1^2]. \end{aligned}$$

Puisque $x_1^2 \approx 187,1$, la compagnie aérienne peut vendre au maximum 187 places pour rester sûre à (plus de) 95% que tout le monde puisse monter dans l'avion.

Exercice

Une usine fabrique des pièces dont une proportion inconnue p est défectueuse, et on souhaite trouver une valeur approchée de p . On effectue un prélèvement de n pièces. On suppose que le prélèvement se fait sur une population très grande, et donc qu'il peut s'apparenter à une suite de n tirages indépendants avec remise. On note X_n la variable aléatoire égale au nombre de pièces défectueuses.

1. Quelle est la loi de X_n ? Sa moyenne ? Sa variance ?
2. Majorer, pour tout $\varepsilon > 0$, $\mathbb{P}\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right)$.
3. En déduire une condition sur n pour que $\frac{X_n}{n}$ soit une valeur approchée de p à 10^{-2} près avec une probabilité supérieure ou égale à 95%.
4. Répondre à la question précédente en utilisant cette fois une loi approchant la loi de X_n . Qu'en pensez-vous?

Corrigé

1. On vérifie sans problème que X_n suit la loi binomiale de paramètres n et p . $\mathbb{E}(X_n) = np$ et $\mathbb{V}(X_n) = np(1-p)$.
2. Puisque X_n admet une espérance, $\frac{X_n}{n}$ aussi et $\mathbb{E}\left(\frac{X_n}{n}\right) = p$ et $\mathbb{V}\left(\frac{X_n}{n}\right) = \frac{p(1-p)}{n}$.

D'après l'inégalité de Bienaymé-Tchebychev, on trouve que :

$$\forall \varepsilon > 0, \mathbb{P}\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}\left(\frac{X_n}{n}\right)}{\varepsilon^2} \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

3. On cherche n tel que $\mathbb{P}\left(\left|\frac{X_n}{n} - p\right| < 10^{-2}\right) \geq 0,95$, i.e. $\mathbb{P}\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq 0,05$.
D'après la question précédente, il suffit de choisir n tel que $\frac{10^4}{4n} \leq 0,05$, i.e. $n \geq 50000$.

4. D'après le théorème de Moivre-Laplace, on peut approcher la loi de $\frac{X_n - np}{\sqrt{np(1-p)}}$ par la loi normale centrée réduite. Notons Φ la fonction de répartition de la loi normale centrée réduite.

On cherche n tel que $\mathbb{P}\left(\left|\frac{X_n}{n} - p\right| < 10^{-2}\right) \geq 0,95$. Or :

$$\begin{aligned}\mathbb{P}\left(\left|\frac{X_n}{n} - p\right| < 10^{-2}\right) &\geq 0,95 \Leftrightarrow \mathbb{P}\left(\left|\frac{X_n - np}{\sqrt{np(1-p)}}\right| < \frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0,95 \\ &\Leftrightarrow \Phi\left(\frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0,95 \\ &\Leftrightarrow 2\Phi\left(\frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0,95 \\ &\Leftrightarrow \Phi\left(\frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0,975.\end{aligned}$$

Puisque $\Phi(1,96) = 0,96$ et puisque Φ est croissante sur $\mathbb{R}$, il suffit de choisir n tel que $\frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}} \geq 1,96$. Or :

$$\frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}} \geq 1,96 \Leftrightarrow n \geq 196^2 p(1-p).$$

Le paramètre p étant inconnu, on utilise la fameuse inégalité $p(1-p) \leq \frac{1}{4}$ (à redémontrer systématiquement). Il suffit donc de choisir $n \geq \frac{196^2}{4} = 9604$ pour être certain à (plus de) 95% que $\frac{X_n}{n}$ soit une approximation à 10^{-2} -près du paramètre p inconnu.

Exercice

On suppose que l'intervalle de temps entre deux voitures successives à un passage à niveau (peu fréquenté) suit une loi exponentielle de moyenne 30 minutes. On suppose de plus qu'il y a indépendance entre les intervalles de temps séparant les instants de passage de voitures. Calculer (une valeur approchée) de la probabilité qu'il y ait plus de 50 voitures qui empruntent le passage à niveau une journée donnée.

Corrigé

Pour tout $n \in \mathbb{N}^*$, notons X_n la variable aléatoire égale à la durée (en minutes) séparant le passage de la n -ème voiture de celui de la $(n+1)$ -ème (on suppose pour simplifier que l'expérience commence au passage de la première voiture).

D'après l'énoncé, $\mathbb{E}(X_n) = 30$ donc X_n suit la loi exponentielle de paramètre $\lambda = \frac{1}{30}$. On sait alors que $\mathbb{V}(X_n) = \frac{1}{\lambda^2} = 900$. Notons $m = \mathbb{E}(X_n) = 30$ et $\sigma = \sigma(X_n) = 30$.

La variable aléatoire $S_n = X_1 + \dots + X_n$ représente alors la durée séparant les passages de la première voiture et de la $(n+1)$ -ème. On cherche alors à calculer $\mathbb{P}(S_{49} \leq 24 \times 60)$. Or :

$$\mathbb{P}(S_{49} \leq 24 \times 60) = \mathbb{P}\left(\frac{\frac{S_{49}}{49} - m}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\frac{24 \times 60}{49} - m}{\frac{\sigma}{\sqrt{n}}}\right)$$

Or $\frac{S_{49}}{49}$ est la moyenne empirique d'un échantillon de 49 variables aléatoires indépendantes de même loi, d'espérance $m = 30$ et de variance $\sigma^2 = 900$. Le théorème central limite assure alors qu'on peut approcher la loi de $\frac{\frac{S_{49}}{49} - m}{\frac{\sigma}{\sqrt{n}}}$ par loi normale centrée réduite. Ainsi, en notant Φ la fonction de répartition de la loi normale centrée réduite, on a :

$$\mathbb{P}(S_{49} \leq 24 \times 60) \approx \Phi\left(\frac{\frac{24 \times 60}{49} - 30}{\frac{30}{\sqrt{49}}}\right) = \Phi\left(-\frac{1}{7}\right) = 1 - \Phi\left(\frac{1}{7}\right) \approx 1 - 0,5557 = 0,4443.$$

Exercice

Le taux moyen de personnes à soigner pour un problème de cholestérol élevé est de 7,5% dans la population française. Un échantillon de 10000 personnes sur une population étant donné, donner un intervalle de confiance à 95% du nombre de personnes à soigner.

Corrigé

Notons X la variable égale au nombre de personnes à soigner dans l'échantillon. La variable X suit la loi hypergéométrique $\mathcal{H}(N, n, p)$ où :

- N est la taille de la population française,
- $n = 10000$ et $p = 0,075$.

Puisque N est très grand devant n , on peut approcher la loi de X par la loi binomiale $\mathcal{B}(n, p)$ (c'est ce qui a été effectué dans l'énoncé de l'exercice sur les pièces d'usine). Le théorème de Moivre-Laplace assure qu'on peut approcher cette loi par la loi normale d'espérance np et de variance $np(1-p)$. Sous cette approximation, la variable aléatoire $Z = \frac{X - np}{\sqrt{np(1-p)}}$ suit la loi normale centrée réduite.

On sait que $\mathbb{P}(-1,96 \leq Z \leq 1,96) = 0,95$ d'après une table de la loi normale centrée réduite. Or :

$$\begin{aligned} [-1,96 \leq Z \leq 1,96] &= \left[-1,96 \leq \frac{X - np}{\sqrt{np(1-p)}} \leq 1,96 \right] \\ &= \left[np - 1,96\sqrt{np(1-p)} \leq X \leq np + 1,96\sqrt{np(1-p)} \right] \\ &= \left[np - 1,96\sqrt{np(1-p)} \leq X \leq np + 1,96\sqrt{np(1-p)} \right] \end{aligned}$$

Or $np - 1,96\sqrt{np(1-p)} \approx 698,4$ et $np + 1,96\sqrt{np(1-p)} \approx 801,6$. Ainsi :

$$\mathbb{P}(X \in [698; 802]) \geq \left(\mathbb{P}(X \in [np - 1,96\sqrt{np(1-p)}, np + 1,96\sqrt{np(1-p)}]) \right) = 0,95.$$

L'intervalle de confiance au niveau de confiance 95% du nombre de personnes à soigner est donc $[698; 802]$.

Exercice

Sur 12000 individus d'une espèce, on a dénombré 13 albinos. Estimer la proportion d'albinos dans l'espèce.

Corrigé

Notons $n = 12000$ le nombre d'individus de l'échantillon, p la proportion d'albinos à estimer et X_n la variable aléatoire égale au nombre d'individus atteints d'albinisme. En supposant que cette maladie atteint des individus de manière indépendante, il vient que X_n suit la loi binomiale $\mathcal{B}(n, p)$.

- Méthode 1 : théorème de Moivre-Laplace

D'après le théorème de Moivre-Laplace, on peut approcher la loi de la variable aléatoire $Z = \frac{X_n - np}{\sqrt{np(1-p)}}$ par la loi normale centrée réduite. On sait, d'après une table de la loi normale centrée réduite que $\mathbb{P}(-1,96 \leq Z \leq 1,96) = 0,95$. Ainsi :

$$\begin{aligned} 0,95 &= \mathbb{P}\left(-1,96 \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq 1,96\right) \\ &= \mathbb{P}\left(-1,96 \leq \frac{np - X_n}{\sqrt{np(1-p)}} \leq 1,96\right) \\ &= \mathbb{P}\left(\frac{X_n}{n} - 1,96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{X_n}{n} + 1,96\sqrt{\frac{p(1-p)}{n}}\right) \end{aligned}$$

Puisque le paramètre p est inconnu, on cherche un intervalle contenant l'intervalle $\left[\frac{13}{12000} - 1,96\sqrt{\frac{p(1-p)}{12000}}, \frac{13}{12000} + 1,96\sqrt{\frac{p(1-p)}{12000}} \right]$ ne dépendant pas de p .

Puisque $p(1-p) \leq \frac{1}{4}$, on considère l'intervalle :

$$\left[\frac{13}{12000} - \frac{1,96}{2\sqrt{12000}}, \frac{13}{12000} + \frac{1,96}{2\sqrt{12000}} \right] = [-7,9 \times 10^{-3}; 1 \times 10^{-2}].$$

Il y a donc 95% de chances que la proportion d'albinos dans l'espèce soit inférieur à 1%.

Remarque : bien que ce résultat soit correct, la proportion d'albinos est largement surestimée car $p(1-p)$ est très probablement très inférieur à $\frac{1}{4}$. La seconde approche décrite

ci-dessous est alors plus précise, mais n'est absolument pas un attendu du programme; je vous la donne uniquement pour aller plus loin.

- Méthode 2 (pour aller plus loin) : théorème central limite, version variance inconnue

En notant S_n^2 la variance empirique, et en remarquant que $\frac{X_n}{n}$ est la moyenne empirique d'un n -échantillon de variables aléatoires indépendantes identiquement distribuées de variables de Bernoulli égales à 1 si un individu est albinos, le théorème central limite assure qu'on peut approcher la loi de la variable aléatoire $Z = \frac{\frac{X_n}{n} - p}{\frac{S_n}{\sqrt{n}}}$ par la loi normale centrée réduite.

On sait, d'après une table de la loi normale centrée réduite que $\mathbb{P}(-1,96 \leq Z \leq 1,96) = 0,95$. Ainsi :

$$\begin{aligned} 0,95 &= \mathbb{P}\left(-1,96 \leq \frac{\frac{X_n}{n} - p}{\frac{S_n}{\sqrt{n}}} \leq 1,96\right) \\ &= \mathbb{P}\left(-1,96 \leq \frac{p - \frac{X_n}{n}}{\frac{S_n}{\sqrt{n}}} \leq 1,96\right) \\ &= \mathbb{P}\left(\frac{X_n}{n} - 1,96 \frac{S_n}{\sqrt{n}} \leq p \leq \frac{X_n}{n} + 1,96 \frac{S_n}{\sqrt{n}}\right) \end{aligned}$$

Puisque la variance de la loi commune est $p(1-p)$, on peut approcher la réalisation de S_n^2 par $f(1-f)$ où $f = \frac{13}{12000}$ (car $\frac{X_n}{n}$ est un estimateur sans biais de l'espérance et S_n^2 est un estimateur asymptotiquement sans biais de la variance). On considère alors l'intervalle :

$$\left[f - 1,96 \sqrt{\frac{f(1-f)}{12000}}, \frac{13}{12000} + 1,96 \sqrt{\frac{f(1-f)}{12000}} \right] = [4,9 \times 10^{-4}; 1,7 \times 10^{-3}].$$

Il y a donc 95% de chances que la proportion d'albinos dans l'espèce soit comprise entre $4,9 \times 10^{-2}\%$ et $1,7 \times 10^{-1}\%$.

Exercice

Soient $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées et admettant un moment d'ordre 2. On suppose que leur loi a la propriété suivante : $X_1 + X_2$ a même loi $\sqrt{2}X_1$.

1. Montrer que, pour tout $n \in \mathbb{N}^*$, les variables aléatoires $X_1 + X_2 + \dots + X_{2^n}$ et $\sqrt{2^n} X_1$ ont même loi.
2. Montrer que l'espérance de toutes les variables de la suite $(X_n)_{n \geq 1}$ est nulle.
3. À l'aide du théorème central limite, déterminer la loi commune à toutes les variables de la suite $(X_n)_{n \geq 1}$.

Corrigé

1. Démontrons le résultat par récurrence sur $n \in \mathbb{N}^*$.

Le résultat est vrai par hypothèse pour $n = 1$.

Soit $n \in \mathbb{N}^*$. Supposons que $X_1 + X_2 + \dots + X_{2^n}$ et $\sqrt{2^n} X_1$ ont même loi. Or :

- $X_1 + X_2 + \dots + X_{2^{n+1}} = (X_1 + X_2 + \dots + X_{2^n}) + (X_{2^n+1} + X_{2^n+2} + \dots + X_{2^{n+1}})$.
- $X_1 + X_2 + \dots + X_{2^n}$ suit la même loi que $\sqrt{2^n} X_1$.
- $X_{2^n+1} + X_{2^n+2} + \dots + X_{2^{n+1}}$ suit la même loi que $X_1 + X_2 + \dots + X_{2^n}$ donc que $\sqrt{2^n} X_1$ ou $\sqrt{2^n} X_2$.

Par indépendance de $X_1, X_2, \dots, X_{2^{n+1}}$, les variables $X_1 + X_2 + \dots + X_{2^n}$ et $X_{2^n+1} + X_{2^n+2} + \dots + X_{2^{n+1}}$ sont indépendantes par le lemme des coalitions. Ainsi, $X_1 + X_2 + \dots + X_{2^{n+1}}$ suit la même loi que $\sqrt{2^n} X_1 + \sqrt{2^n} X_2 = \sqrt{2^n} (X_1 + X_2)$. On en déduit enfin que $X_1 + X_2 + \dots + X_{2^{n+1}}$ suit la même loi que $\sqrt{2^{n+1}} X_1$, ce qui conclut la récurrence.

2. La loi admet un moment d'ordre 2 donc une espérance, qu'on notera m . D'après la question précédente, $\mathbb{E}(X_1 + X_2 + \dots + X_{2^n}) = \sqrt{2^n} \mathbb{E}(X_1)$, i.e. $2^n m = \sqrt{2^n} m$. On en déduit que $m = 0$.
3. Notons σ l'écart-type de la loi commune à toutes les variables de la suite $(X_n)_{n \geq 1}$. Puisque (X_n) forme une suite de variables aléatoires indépendantes de même loi et admettant une variance, le théorème central limite assure que :

$$\forall x \in \mathbb{R}, \lim_{n \rightarrow +\infty} \mathbb{P}\left(\frac{\frac{X_1 + \dots + X_{2^n}}{2^n} - m}{\frac{\sigma}{\sqrt{2^n}}} \leq x\right) = \Phi(x).$$

Or, pour tout $n \in \mathbb{N}^*$, la variable aléatoire $\frac{\frac{X_1 + \dots + X_{2^n} - m}{\frac{\sigma}{\sqrt{2^n}}}}{\frac{\sigma}{\sqrt{2^n}}} = \frac{X_1 + \dots + X_{2^n}}{\sigma\sqrt{2^n}}$ suit la

même loi que $\frac{X_1}{\sigma}$. Ainsi :

$$\forall x \in \mathbb{R}, \mathbb{P}\left(\frac{X_1}{\sigma} \leq x\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\frac{\frac{X_1 + \dots + X_{2^n} - m}{\frac{\sigma}{\sqrt{2^n}}}}{\frac{\sigma}{\sqrt{2^n}}} \leq x\right) = \Phi(x).$$

On en déduit que la fonction de répartition de $\frac{X_1}{\sigma}$ est celle de la loi normale centrée réduite

donc que $\frac{X_1}{\sigma}$ suit cette loi. On en déduit enfin que X_1 suit la loi normale d'espérance nulle et de variance (inconnue) σ^2 .