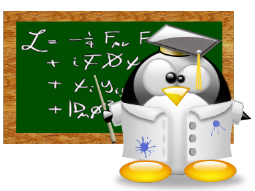
Stéphane PASQUET, 17 août 2018

Sommaire



Prérequis

- Variables aléatoires discrètes (avec espérance et écart-type)
- Loi binomiale

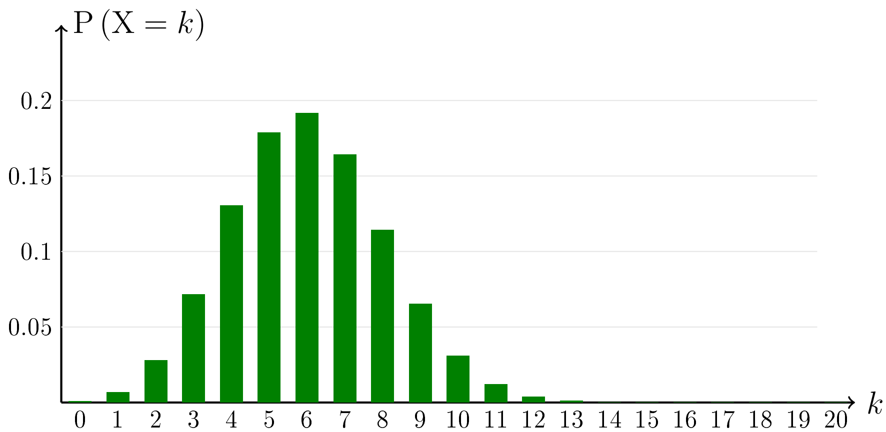
Propriété — *Loi binomiale (rappels)*

Soit X une variable aléatoire discrète désignant le nombre de succès obtenus après n répétitions d'une même expérience à deux issues. Si p représente la probabilité d'obtenir un succès lors d'une expérience, X suit la loi binomiale de paramètres n et p, et on note :

$$\mathrm{X} \sim \mathscr{B}\,(n\,;p)\,.$$

Dans ce cas, on a :

$$\forall\ k \in [\![0\,;n]\!]\,,\ \mathrm{P}\,(\mathrm{X}=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad , \quad \mathbb{E}\,(\mathrm{X}) = np \quad \text{et} \quad \sigma(X) = \sqrt{np(1-p)}.$$



Cas ou $n = 20$ et $p = 0,3$

Définition — *Intervalle de fluctuation au seuil de 95 %*

Un `intervalle de fluctuation au seuil de 95 %` relatif à un échantillon de taille n est un intervalle dans lequel est la fréquence observée dans un échantillon de même taille n avec une probabilité de 95 %.

Remarque

En classe de Seconde, on a vu que $\left[p - \dfrac{1}{\sqrt{n}}\,;p + \dfrac{1}{\sqrt{n}}\right]$ était un intervalle de fluctuation *approché* au seuil de 95 % relatif aux échantillon de taille n.
Quelques fois, la probabilité que la fréquence appartienne à cet intervalle est très proche de 0,95 mais en étant inférieure (d'où le « *approché* »).
Dans la pratique, on utilise cet intervalle lorsque $p \in [0,2\,;0,8]$ et $n \geqslant 25$.

Définition

Loi normale centrée réduite

Une variable aléatoire X à densité f sur I suit une `loi normale centrée réduite` si $f(x) = \dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$ et $\mathrm{I} = [a\,;b]$, a et b étant deux réels, $a < b$.
On note alors :

$$\mathrm{X} \sim \mathscr{N}(0\,;1)\,.$$

Propriétés

Soit $\mathrm{X} \sim \mathscr{N}(0\,;1)$. Alors,

1. $\forall\ u \in \mathbb{R}^+,\ \mathrm{P}(\mathrm{X} \leqslant -u) = \mathrm{P}(\mathrm{X} \geqslant u)$,
2. $\mathrm{P}(\mathrm{X} \leqslant 0) = \mathrm{P}(\mathrm{X} \geqslant 0) = \dfrac{1}{2}$.

Propriété

Espérance et écart-type de la loi normale centrée réduite

Soit $\mathrm{X} \sim \mathscr{N}(0\,;1)$. Alors,

$$\mathbb{E}(\mathrm{X}) = 0 \qquad \text{et} \qquad \sigma(\mathrm{X}) = 1.$$

- $\mathrm{P}(-1,96 \leqslant \mathrm{X} \leqslant 1,96) \approx 0,95$.
- $\mathrm{P}(-2,58 \leqslant \mathrm{X} \leqslant 2,58) \approx 0,99$.

Cela signifie que si une variable aléatoire X suit la loi $\mathscr{N}(0\,;1)$, alors 95 % de ses valeurs se trouvent entre $-1,96$ et $1,96$, et 99 % entre $-2,58$ et $2,58$.

Définition

Loi normale

Une variable aléatoire X suit une loi normale $\mathscr{N}(\mu\,;\sigma^2)$, d'espérance μ et d'écart-type σ, si la variable aléatoire $\mathrm{Z} = \dfrac{\mathrm{X}-\mu}{\sigma}$ suit la loi normale centrée réduite $\mathscr{N}(0\,;1)$.

Propriété

Soit $\mathrm{X} \sim \mathscr{N}(\mu\,;\sigma^2)$.
Alors,

$$\begin{array}{rl}
\mathrm{P}(\mu-\sigma \leqslant \mathrm{X} \leqslant \mu+\sigma) \approx 0,68 & \quad (\text{à } 10^{-2} \text{ près}) \\
\mathrm{P}(\mu-2\sigma \leqslant \mathrm{X} \leqslant \mu+2\sigma) \approx 0,95 & \quad (\text{à } 10^{-2} \text{ près}) \\
\mathrm{P}(\mu-3\sigma \leqslant \mathrm{X} \leqslant \mu+3\sigma) \approx 0,997 & \quad (\text{à } 10^{-3} \text{ près})
\end{array}$$

Ce dernier résultat nous dit que si X est une variable aléatoire suivant la loi normale $\mathcal{N}(\mu\,;\sigma^2)$, alors

- $68\,\%$ de ses valeurs se trouvent dans $[\mu-\sigma\,;\mu+\sigma]$;
- $95\,\%$ de ses valeurs se trouvent dans $[\mu-2\sigma\,;\mu+2\sigma]$;
- $99,7\,\%$ de ses valeurs se trouvent dans $[\mu-3\sigma\,;\mu+3\sigma]$.

Définition

Intervalle de fluctuation asymptotique au seuil de 95 %

Si :

$$n \geqslant 30 \quad ; \quad np \geqslant 5 \quad ; \quad n(1-p) \geqslant 5.$$

alors l'intervalle $\mathrm{I}_n = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est appelé un `intervalle de fluctuation asymptotique` au seuil de $1-\alpha$ de la variable aléatoire $\mathrm{F}_n = \dfrac{\mathrm{X}_n}{n}$.

Exemple

Exemple d'application (d'après les ressources de l'E.N.)

Le responsable de la maintenance des machines à sous d'un casino doit vérifier qu'un certain type de machines est bien réglé sur une fréquence de succès de 0,06.
Il décide de régler chaque machine pour laquelle il aura observé, dans l'historique des jeux, une fréquence de succès se situant en dehors d'un intervalle de fluctuation au seuil de 95 %.
Lors du contrôle d'une machine, le technicien constate qu'elle a fourni 9 succès sur 85 jeux.

1. Déterminer la fréquence observée f de succès de cette machine.
2. Déterminer un intervalle de fluctuation asymptotique au seuil de 95 % de la fréquence de succès.
3. Le technicien va-t-il modifier le réglage de la machine ?
4. Quelle aurait été sa décision s'il y avait eu 21 succès sur 200 jeux ?

1. La fréquence observée est : $f = \dfrac{9}{85} \approx 0,106$.
2. Ici, $n = 85 \geqslant 30$, $p = 0,06$ donc $np = 5,1 \geqslant 5$ et $n(1-p) = 79,9 \geqslant 5$.
   Un intervalle de fluctuation au seuil de 95 % est donc :
   $$\mathrm{I} = \left[0,06 - 1,96\frac{\sqrt{0,06\times 0,94}}{\sqrt{85}}\,;0,06 + 1,96\frac{\sqrt{0,06\times 0,94}}{\sqrt{85}}\right] \approx [0,0096\,;0,1104].$$
3. $f \in \mathrm{I}$ donc le réglage de la machine n'est pas à modifier.
4. Dans cette question, $f \approx 0,105$ et $\mathrm{I} \approx [0,0271\,;0,0929]$ donc $f \notin \mathrm{I}$. Il faut donc régler la machine.

Définition

Intervalle de confiance

Soit f une fréquence observée sur une population de taille $n \geqslant 30$ telle que $nf \geqslant 5$ et $n(1-f) \geqslant 5$.
L'intervalle $\left[f - \dfrac{1}{\sqrt{n}}\,;f + \dfrac{1}{\sqrt{n}}\right]$ est appelé un `intervalle de confiance` au seuil de 95 % de la probabilité de l'événement observé.

Exemple

Exemple d'application : sondages et élections

Pour le 2e tour d'une élection présidentielle, l'institut de sondages A réalise un sondage qui montre que $52,5\,\%$ des intentions de votes sont pour le candidat 1, et donc $47,5\,\%$ pour le candidat 2 (en ne tenant pas compte ni de l'abstention ni des votes blancs).
L'institut B réalise son propre sondage et constate que $50,5\,\%$ des intentions de votes sont pour le candidat 1.
Les deux sondages ont été faits sur un échantillon de 1 112 personnes.
L'intervalle de confiance pour l'institut A est $\left[0,525 - \dfrac{1}{\sqrt{1\,112}}\,;0,525 + \dfrac{1}{\sqrt{1\,112}}\right]$, soit $[0,4951\,;0,5549]$ pour le candidat 1.
L'intervalle de confiance pour l'institut B est $\left[0,505 - \dfrac{1}{\sqrt{1\,112}}\,;0,505 + \dfrac{1}{\sqrt{1\,112}}\right]$, soit $[0,4751\,;0,5349]$ pour le candidat 1.
L'intersection des deux intervalles n'étant pas vide, les résultats des deux sondages ne sont pas en contradiction.
De plus, ces intervalles contiennent des valeurs inférieures à 0,5 ; ainsi, le candidat 1 n'est pas totalement sûr de gagner les élections.

Remarques

1. Remarquez l'importance de connaître la taille de l'échantillon sur lequel le sondage est effectué. Sans ce nombre, impossible de faire des conclusions probantes.
2. Comme pour les intervalles de fluctuation, il existe différents intervalles de confiance. Par exemple, l'intervalle :

$$\left[f - 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\,;f + 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right].$$

Exemple

Contrôle du Loto

Le site http://www.lototest.com offre certaines statistiques concernant le Loto, et notamment concernant le nouveau Loto (depuis 2008).
On y trouve, au 16 août 2013, que sur 763 tirages, la fréquence maximale enregistrée pour un numéro est de $2,49\,\%$ (le numéro 23) et la fréquence minimale est de $1,6\,\%$ (les numéros 42 et 19).

La question que l'on est en droit de se poser est si ces fréquences sont acceptables au seuil de $95\,\%$?

La probabilité d'un numéro quelconque est : $p = \dfrac{1}{49} \approx 2,04\,\%$.
Un intervalle de fluctuation au seuil de $95\,\%$ est :

$$\left[0,0204 - 1,96\frac{\sqrt{0,0204\times(1-0,0204)}}{\sqrt{763}}\,;0,0204 - 1,96\frac{\sqrt{0,0204\times(1-0,0204)}}{\sqrt{763}}\right] \approx [0,0104\,;0,031].$$

On constate donc que la fréquence minimale observée ainsi que la fréquence maximale observée sont dans cet intervalle.
On peut donc conclure qu'il n'y a aucune anomalie concernant les tirages observés.