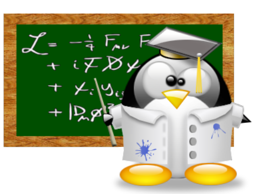
Stéphane PASQUET, 18 mai 2017

Sommaire



Prérequis

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

Propriété

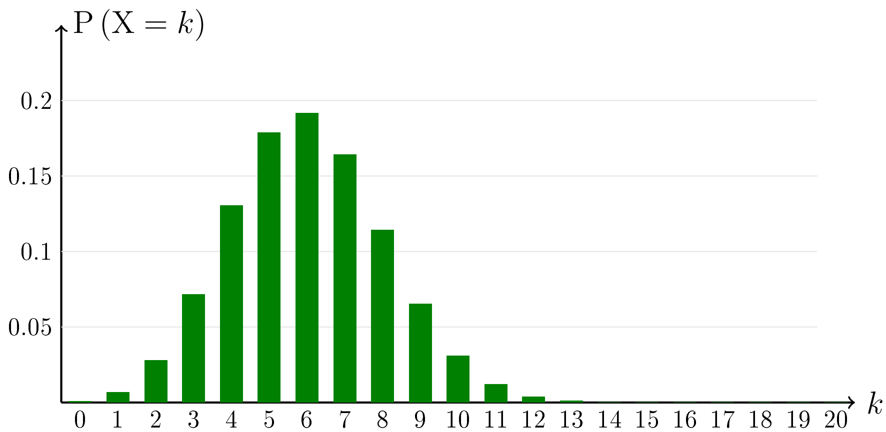
Loi binomiale (rappels)

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

Variable aléatoire centrée réduite

On dit d'une variable aléatoire qu'elle est `centrée réduite` quand son espérance mathématique est nulle et son écart-type égal à 1.

Propriété

Changement de variable aléatoire

Soit X une variable aléatoire d'espérance m et d'écart-type $\sigma \neq 0$.
Alors, la variable aléatoire Z définie par :

$$Z = \frac{X - m}{\sigma}$$

est une variable aléatoire centrée réduite.

Démonstration

Rappelons les formules :

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b \qquad ; \qquad \sigma(aX + b) = |a|\sigma(X).$$

Alors,

$$\begin{aligned}\mathbb{E}(Z) &= \mathbb{E}\left(\frac{1}{\sigma}X - \frac{m}{\sigma}\right)\\ &= \frac{1}{\sigma}\mathbb{E}(X) - \frac{m}{\sigma}\\ &= \frac{1}{\sigma} \times m - \frac{m}{\sigma}\\ &= 0\end{aligned}$$

et

$$\begin{aligned}\sigma(Z) &= \frac{1}{\sigma}\sigma\\ &= 1.\end{aligned}$$

Z est donc bien une variable aléatoire centrée réduite. ■

Théorème

Théorème de Moivre-Laplace

Soit $p \in]0\,;1[$ et soit, pour $n \in \mathbb{N}$, X_n une variable aléatoire suivant la loi binomiale $\mathscr{B}(n\,;p)$.
On pose alors :

$$Z_n = \frac{X_n - m}{\sigma_n}$$

où $m = \mathbb{E}(X_n)$ et σ_n l'écart-type de X_n.
Pour tous nombres réels a et b, $a < b$, on a :

$$\lim_{n \to +\infty} P(a \leqslant Z_n \leqslant b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx.$$

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. $\mathrm{P}(\mathrm{X} \leqslant 0) = \mathrm{P}(\mathrm{X} \geqslant 0) = \dfrac{1}{2}$.

Démonstration

1. $f(-x) = f(x)$ donc $\displaystyle\int_{-\infty}^{-u} f(x)\ \mathrm{d}x = \int_{u}^{+\infty} f(x)\ \mathrm{d}x$, d'où le résultat.
2. En prenant $u = 0$ dans le résultat précédent, on a $\mathrm{P}(\mathrm{X} \leqslant 0) = \mathrm{P}(\mathrm{X} \geqslant 0)$ et comme $\mathrm{P}(\mathrm{X} \leqslant 0) + \mathrm{P}(\mathrm{X} \geqslant 0) = 1$, on arrive au résultat souhaité. ■

Propriété *Espérance et écart-type de la loi normale centrée réduite*

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

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

Démonstration

$$\mathbb{E}(\mathrm{X}) = \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\ \mathrm{d}x.$$

Or, la fonction $f\ :\ x \mapsto \dfrac{x}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$ est impaire, ce qui signifie que :

$$\int_{-x}^{0} f(x)\ \mathrm{d}x \ = \ -\int_{0}^{x} f(x)\ \mathrm{d}x$$

et donc $\mathbb{E}(\mathrm{X}) = 0$.

La démonstration de la propriété sur l'écart-type est hors-programme. Cependant, voyons les points essentiels de la démonstration.

La formule de König-Huygens nous dit que $\mathbb{V}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - (\mathbb{E}(\mathrm{X}))^2 = \mathbb{E}(\mathrm{X}^2)$ car $\mathbb{E}(\mathrm{X}) = 0$.

$$\mathbb{E}\left(\mathrm{X}^2\right) = \int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^4}{2}}\,\mathrm{d}x = 2\int_{-\infty}^{0} \frac{x^2}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^4}{2}}\,\mathrm{d}x$$

À l'aide d'une intégration par parties, en posant $u(x) = \dfrac{x^2}{\sqrt{2\pi}}$ et $v(x) = \mathrm{e}^{-\frac{x^4}{2}}$, on a :

$$\int_{-\infty}^{0} \frac{x^2}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^4}{2}}\,\mathrm{d}x = \underbrace{\left[-\frac{x}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}\right]_{-\infty}^{0}}_{=0} + \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{2}.$$

On arrive donc bien à $\mathbb{V}(\mathrm{X}) = 1$. Or, $\sigma(\mathrm{X}) = \sqrt{\mathbb{V}(\mathrm{X})}$, donc $\sigma(\mathrm{X}) = 1$. ■

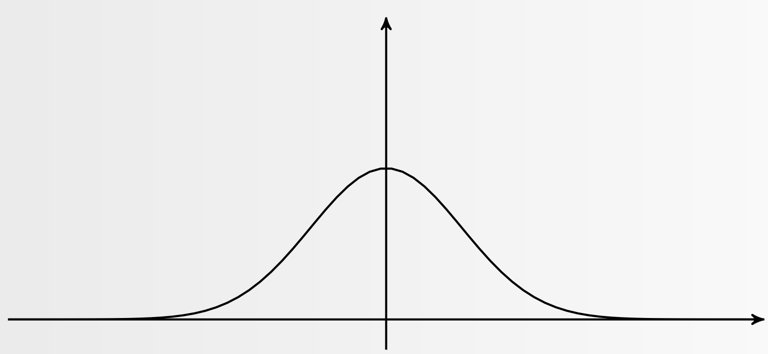
Théorème *Répartition des valeurs d'une loi normale centrée réduite*

Soit X une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(0\,;1)$.
Alors,

$$\forall\ \alpha \in]0\,;1[,\ \exists!\ u_\alpha \in \mathbb{R}^+ \mid \mathrm{P}\left(-u_\alpha \leqslant \mathrm{X} \leqslant u_\alpha\right) = 1 - \alpha.$$

Démonstration

La fonction de densité de la loi normale est représentée par la courbe suivante :



Par sa symétrie, on a :

$$\forall\ u \in \mathbb{R}^+,\ \mathrm{P}\left(-u \leqslant \mathrm{X} \leqslant u\right) = 2\mathrm{P}\left(0 \leqslant \mathrm{X} \leqslant u\right) = 2\int_0^u f(x)\,\mathrm{d}x = 2F(u),$$

où F est une primitive de f sur $\mathbb{R}$ qui s'annule en 0.
La fonction F est alors continue et strictement croissante sur $\mathbb{R}$ ($F(u)$ représente l'aire du domaine délimité par la courbe, l'axe des abscisses, l'axe des ordonnées et la droite d'équation $x = u$).

⋯

De plus, $2F(0) = 0$ et $\lim\limits_{u \to +\infty} 2F(u) = 1$ (par définition d'une fonction de densité) donc, d'après le théorème de la valeur intermédiaire, il existe un unique nombre u_α dans $]0\,;+\infty[$ tel que $2F(u_\alpha) = 1 - \alpha$ pour $\alpha \in]0\,;1[$. D'où le tableau de variations suivant :

u	0		u_α		$+\infty$
$2F(u)$	0	$\nearrow$	$1-\alpha$	$\nearrow$	1

Le résultat est alors démontré. ■

Valeurs particulières de u_α.

- $u_{0,05} \approx 1,96$ d'où $\mathrm{P}(-1,96 \leqslant \mathrm{X} \leqslant 1,96) \approx 0,95$.
- $u_{0,01} \approx 2,58$ d'où $\mathrm{P}(-2,58 \leqslant \mathrm{X} \leqslant 2,58) \approx 0,99$.

Cela signifie que si une variable aléatoire X suit la loi $\mathcal{N}(0\,;1)$, 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 $\mathcal{N}(\mu\,;\sigma^2)$, 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 $\mathcal{N}(0\,;1)$.

Propriété

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

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

Démonstration

On a :

$$\begin{aligned} \mu-\sigma \leqslant \mathrm{X} \leqslant \mu+\sigma &\Leftrightarrow -\sigma \leqslant \mathrm{X}-\mu \leqslant \sigma\\ &\Leftrightarrow -1 \leqslant \frac{\mathrm{X}-\mu}{\sigma} \leqslant 1 \end{aligned}$$

…

donc

$$\begin{aligned}P(\mu-\sigma \leqslant X \leqslant \mu+\sigma) &= P\left(-1 \leqslant \frac{X-\mu}{\sigma} \leqslant 1\right)\\ &= P(-1 \leqslant Z \leqslant 1)\\ &\approx 0,68 \text{ (À l'aide de la calculatrice ou d'un logiciel de calcul)}\end{aligned}$$

Nous faisons de même pour les autres résultats. ■

Ce dernier résultat nous dit que si X est une variable aléatoire suivant la loi normale $\mathscr{N}(\mu\,;\sigma^2)$, 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 $[\mu-3\sigma\,;\mu+3\sigma]$.

Théorème *Intervalles de fluctuation asymptotiques*

Soit $p \in]0\,;1[$. On pose alors X_n une suite de variables aléatoires suivant la loi binomiale $\mathscr{B}(n\,;p)$. Alors,

$$\forall\ \alpha \in]0\,;1[,\ \lim_{n\to+\infty} P\left(\frac{X_n}{n} \in I_n\right) = 1-\alpha$$

où

$$I_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

et où u_α est l'unique réel tel que $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1-\alpha$, Z suivant la loi $\mathscr{N}(0\,;1)$.

Démonstration

$X_n \sim \mathscr{B}(n\,;p)$ donc $\mathbb{E}(X_n) = np$ et $\sigma(X_n) = \sqrt{np(1-p)}$.
Ainsi,

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}.$$

On a :

$$\begin{aligned}P(a \leqslant Z_n \leqslant b) &= P\left(a \leqslant \frac{X_n-np}{\sqrt{np(1-p)}} \leqslant b\right)\\ &= P(a\sqrt{np(1-p)} \leqslant X_n - np \leqslant b\sqrt{np(1-p)})\\ &= P(a\sqrt{np(1-p)} + np \leqslant X_n \leqslant b\sqrt{np(1-p)} + np)\\ &= P\left(a\frac{\sqrt{np(1-p)}}{n} + p \leqslant \frac{X_n}{n} \leqslant b\frac{\sqrt{np(1-p)}}{n} + p\right).\end{aligned}$$

D'après le théorème de Moivre-Laplace,

$$\forall\,(a\,;b) \in \mathbb{R}\times\mathbb{R},\ \lim_{n\to+\infty} P(a \leqslant Z_n \leqslant b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\ dx.$$

…

On en déduit alors :

$$\lim_{n\to+\infty} \mathrm{P}\left(a\frac{\sqrt{np(1-p)}}{n}+p \leqslant \frac{\mathrm{X}_n}{n} \leqslant b\frac{\sqrt{np(1-p)}}{n}+p\right) = \int_a^b \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x.$$

En prenant $a = -u_\alpha$ et $b = u_\alpha$, on obtient :

$$\lim_{n\to+\infty} \mathrm{P}\left(-u_\alpha\frac{\sqrt{np(1-p)}}{n}+p \leqslant \frac{\mathrm{X}_n}{n} \leqslant u_\alpha\frac{\sqrt{np(1-p)}}{n}+p\right) = \int_{-u_\alpha}^{u_\alpha} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = 1-\alpha.$$

Ainsi,

$$\lim_{n\to+\infty} \mathrm{P}\left(\frac{\mathrm{X}_n}{n} \in \mathrm{I}_n\right) = 1-\alpha. \qquad \blacksquare$$

Remarque

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

L'intervalle $\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est un intervalle de fluctuation asymptotique au seuil de 95% (p désigne la proportion dans la population).

Remarque

Pour que toutes les approximations précédentes soient réalisées, on prendra pour habitude d'avoir :

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

On utilise les intervalles de fluctuation pour déterminer si la fréquence f observée dans une population est compatible ou pas avec un modèle de Bernoulli, en regardant si la fréquence observée est dans l'intervalle de fluctuation asymptotique. La borne inférieure de l'intervalle devrait être arrondie par excès et la borne supérieure, par défaut (en effet, il est préférable d'avoir un intervalle J inclus dans l'intervalle de fluctuation plutôt qu'un intervalle plus grand) mais dans la pratique, très peu nombreuses sont les personnes qui le font car cela ne change pas grand-chose dans 99,99% des situations. Mais si l'intervalle est $I = [0,5298\,;0,6877]$, qu'on l'arrondi à $I' = [0,53\,;0,688]$ et que $f = 0,6879$, remarquez que $f \in I'$ mais $f \notin I$; ainsi, en théorie, f « ne convient pas » alors qu'avec notre intervalle arrondi, f « convient »...

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.

Remarque

Pour tout réel $p \in \,]0\,;1[$, la fonction $p \mapsto p(1-p)$ admet un maximum pour $p = \dfrac{1}{2}$ et vaut $\dfrac{1}{2}\left(1 - \dfrac{1}{2}\right) = \dfrac{1}{4}$. Donc :$p(1-p) \leqslant \dfrac{1}{4}$.
On en déduit alors :
$$1,96\sqrt{p(1-p)} \leqslant 1,96\sqrt{\frac{1}{4}} \leqslant 1\ , \qquad \text{et donc :} \qquad 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{1}{\sqrt{n}}.$$
Ainsi,
$$\mathrm{I}_n = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subseteq \left[p - \frac{1}{\sqrt{n}}\,;p - \frac{1}{\sqrt{n}}\right] = \mathrm{J}_n$$
et donc :
$$\mathrm{P}\left(\frac{\mathrm{X}_n}{n} \in \mathrm{I}_n\right) \leqslant \mathrm{P}\left(\frac{\mathrm{X}_n}{n} \in \mathrm{J}_n\right).$$
Ceci légitime l'utilisation de l'intervalle J_n en classe de Seconde.

Théorème

Soit $p \in]0\,;1[$ un réel. Soit alors (X_n) une suite de variables aléatoires suivant la loi binomiale $\mathscr{B}(n\,;p)$.

$$\exists\, n_0 \in \mathbb{N},\ n \geqslant n_0 \Rightarrow \mathrm{P}\left(\frac{\mathrm{X}_n}{n} \in \left[p - \frac{1}{\sqrt{n}}\,;p + \frac{1}{\sqrt{n}}\right]\right) \geqslant 0,95.$$

Démonstration

- Dans un premier temps, on démontre que :

$$\mathrm{P}\left(p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{\mathrm{X}_n}{n} \leqslant p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \leqslant \mathrm{P}\left(p - 2\frac{1}{\sqrt{n}} \leqslant \frac{\mathrm{X}_n}{n} \leqslant p + 2\frac{1}{\sqrt{n}}\right)$$

en utilisant le même raisonnement que dans la remarque précédente.

- De plus,

$$\begin{aligned} & p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{\mathrm{X}_n}{n} \leqslant p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \\ \Leftrightarrow\ & np - 2\sqrt{np(1-p)} \leqslant \mathrm{X}_n \leqslant np + 2\sqrt{np(1-p)} \\ \Leftrightarrow\ & -2 \leqslant \frac{\mathrm{X}_n - np}{\sqrt{np(1-p)}} \leqslant 2. \end{aligned}$$

Donc,

$$\mathrm{P}\left(p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{\mathrm{X}_n}{n} \leqslant p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = \mathrm{P}\left(-2 \leqslant \frac{\mathrm{X}_n - np}{\sqrt{np(1-p)}} \leqslant 2\right).$$

Posons :

$$a_n = \mathrm{P}\left(-2 \leqslant \frac{\mathrm{X}_n - np}{\sqrt{np(1-p)}} \leqslant 2\right).$$

D'après le théorème de Moivre-Laplace,

$$\lim_{n \to +\infty} a_n = \int_{-2}^{2} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \approx 0,9545. \quad \text{(calculatrice)}$$

Ainsi, (a_n) converge vers un nombre fini supérieur à 0,95.
Il existe donc bien un entier n_0 à partir duquel $a_n \geqslant 0,95$. D'où le théorème. ■

Remarque

L'entier n_0 dépend de p. Son maximum est atteint pour $p = \dfrac{1}{2}$ et vaut $n_0 = 529$. (Résultat obtenu à l'aide d'un algorithme)

Propriété

Intervalle de confiance

Soit $p \in]0\,;1[$ et soit (X_n) une suite de variables aléatoires suivant la loi binomiale $\mathscr{B}(n\,;p)$. On pose alors $\mathrm{F}_n = \dfrac{\mathrm{X}_n}{n}$ la variable aléatoire représentant la fréquence du nombre de « succès ».
Alors, pour n assez grand, $p \in \left[\mathrm{F}_n - \dfrac{1}{\sqrt{n}}\,;\mathrm{F}_n + \dfrac{1}{\sqrt{n}}\right]$ avec une probabilité de 95%.

Remarque

Intervalle aléatoire

L'intervalle $\left[\mathrm{F}_n - \dfrac{1}{\sqrt{n}}\,;\mathrm{F}_n + \dfrac{1}{\sqrt{n}}\right]$ est appelé un `intervalle aléatoire` (intervalle dont les bornes sont des variables aléatoires).

Démonstration

D'après le théorème précédent,

$$\exists\, n_0 \in \mathbb{N},\ n \geqslant n_0 \Rightarrow \mathrm{P}\left(\frac{\mathrm{X}_n}{n} \in \left[p - \frac{1}{\sqrt{n}}\,;p + \frac{1}{\sqrt{n}}\right]\right) \geqslant 0,95.$$

Or,

$$\frac{\mathrm{X}_n}{n} \in \left[p - \frac{1}{\sqrt{n}}\,;p + \frac{1}{\sqrt{n}}\right] \Leftrightarrow \mathrm{F}_n - \frac{1}{\sqrt{n}} \leqslant p \leqslant \mathrm{F}_n + \frac{1}{\sqrt{n}}.$$

D'où le résultat souhaité. ■

Si $n \geqslant 30$, $nf \geqslant 5$ et $n(1-f) \geqslant 5$, on convient que f est une estimation de p et que $\left[f - \dfrac{1}{\sqrt{n}}\,;f + \dfrac{1}{\sqrt{n}}\right]$ est un **intervalle de confiance** au niveau de 95% pour la proportion p.
Cet intervalle est parfois appelé **fourchette de sondage**.

La précision de l'estimation est alors donnée par l'amplitude de l'intervalle $\left[f - \dfrac{1}{\sqrt{n}}\,;f + \dfrac{1}{\sqrt{n}}\right]$, soit $\dfrac{2}{\sqrt{n}}$. En notant ε la précision souhaitée, on peut donc calculer la taille n de l'échantillon pour que $\dfrac{2}{\sqrt{n}} \leqslant \varepsilon$. On a alors $n \geqslant \dfrac{4}{\varepsilon^2}$.

Par exemple, pour avoir une précision de 10^{-2}, on prendra un échantillon de taille $n = 40\,000$.

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 $\left[0,525-\frac{1}{\sqrt{1\,112}}\,;0,525+\frac{1}{\sqrt{1\,112}}\right]$, soit $[0,4951\,;0,5549]$ pour le candidat 1.

L'intervalle de confiance pour l'institut B est $\left[0,505-\frac{1}{\sqrt{1\,112}}\,;0,505+\frac{1}{\sqrt{1\,112}}\right]$, 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 : $p=\frac{1}{49}\approx 2{,}04\%$.
Un intervalle de fluctuation au seuil de 95% est :

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

Exemple (suite)

Contrôle du Loto

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.