229) Suites de variables aléatoires indépendantes de même loi de Bernoulli. Variables aléatoires de loi binomiale et approximations de la loi binomiale.

1 Loi(s) de Bernoulli

Définition:
Une variable aléatoire X suit une loi de Bernoulli de paramètre $p \in [0,1]$
$\iff X \sim \mathcal{B}(p)$
$\iff \begin{cases} \mathbb{P}(X=1) = p \\ \mathbb{P}(X=0) = 1-p \end{cases}$

Preuve: $\mathbb{E}[X] = 1 \times \mathbb{P}(X=1) + 0 \times \mathbb{P}(X=0) = p$.
$Var(X) = 1^2 \times \mathbb{P}(X^2=1) + 0^2 \times \mathbb{P}(X^2=0) - p^2 = p - p^2$
$g_X(s) = s^0 \times \mathbb{P}(X=0) + s^1 \times \mathbb{P}(X=1) = 1-p+sp$.

Exemple: jeu de pile ou face (biaisé si $p \neq 0{,}5$), plus généralement, tte expérience à 2 issues ("échec" ou "succès").

Propriétés:
Si $X \sim \mathcal{B}(p)$ alors:
$\begin{cases} \mathbb{E}[X] = p \\ Var(X) = p(1-p) \\ g_X(s) = 1-p+ps \quad, s \in [-1,1] \end{cases}$

Rappels:

Définition: Variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ tte application $X : \Omega \longrightarrow \mathbb{R}$ vérifiant :
$$\forall B \in \mathcal{B}(\mathbb{R}),\ X^{-1}(B) \in \mathcal{A}$$

→ $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé, ie un espace mesuré de masse totale égale à 1.
→ Ω est l'ensemble des résultats possibles de l'expérience aléatoire
→ $\mathcal{A}$ est l'ensemble des évènements, c'est une tribu.
→ $\mathbb{P}$ est une probabilité, c'est-à-dire une mesure positive sur $(\Omega, \mathcal{A})$ telle que $\mathbb{P}(\Omega) = 1$.
→ $\mathcal{B}(\mathbb{R})$ est la tribu borélienne réelle, qui contient tous les intervalles réels.

Définition: Fonction de répartition de la v.a.r. X l'application:
$$F_X(x) = \mathbb{P}(X \leq x) \quad, \forall x \in \mathbb{R}.$$

Thm:
$F_X = F_Y \implies \underbrace{\mathbb{P}_X}_{\text{loi de } X} = \mathbb{P}_Y$

Thm:
$F : \mathbb{R} \to \mathbb{R}$ est une fct° de répartition ssi :
$\begin{cases} F \text{ est croissante} \\ F \text{ est continue à droite} \\ \lim_{t \to -\infty} F(t) = 0 \\ \lim_{t \to +\infty} F(t) = 1 \end{cases}$

Définition : Soit X une v.a.r. discrète tq $X(\Omega)=\{x_0,\dots,x_n,\dots\}$

* **Espérance :** $\mathbb{E}[X]=\sum_{n=0}^{+\infty} x_n \, \mathbb{P}(X=x_n)$
* **Variance :** $Var(X)=\mathbb{E}\left[(X-\mathbb{E}[X])^2\right]$
$= \mathbb{E}[X^2]-(\mathbb{E}[X])^2$
* **Covariance :** $Cov(X,Y)=\mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right]$

Thm :

→ $\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]$
→ $\forall a\in\mathbb{R}$, $\mathbb{E}[aX]=a\cdot\mathbb{E}[X]$.
→ $\forall a,b\in\mathbb{R}$, $Var(aX+b)=a^2\,Var(X)$
→ $Cov(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$
→ $Var(X+Y)=Var(X)+Var(Y)+2\,Cov(X,Y)$.

$$\left.\begin{array}{l} \to \mathbb{E}[X_1X_2\cdots X_n]=\prod_{i=1}^{n}\mathbb{E}[X_i] \\ \to Var\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}Var(X_i) \end{array}\right\} X_i \perp\!\!\!\perp \,!$$

Définition : Fonction génératrice de la v.a.r. X à valeurs dans $\mathbb{N}$:

$\forall s\in[-1;1]$:
$$g_X(s)=\mathbb{E}[s^X]=\sum_{n=0}^{+\infty}\mathbb{P}(X=n)\cdot s^n$$

→ Rmq : c'est une série entière.

2 Loi(s) binomiale(s)

Définition : Une v.a.r. X suit une loi binomiale de paramètres $n\in\mathbb{N}^*$ et $p\in[0,1]$
$\iff X\sim \mathcal{B}in(n,p)$
$\iff \forall k\in[\![0;n]\!]$, $\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

Exemple : Nombre de piles obtenus si on lance n fois une pièce (biaisée si $p\neq 0{,}5$). Plus généralement, toute répétition d'une exp. de Bernoulli indépendantes.

Propriétés :
Si $X\sim\mathcal{B}(n,p)$ alors :
$$\begin{cases} \mathbb{E}[X]=np \\ Var(X)=np(1-p) \\ g_X(s)=(1-p+ps)^n \end{cases}$$

Preuve : Découle directement du résultat suivant :

Thm :
$X_1,\dots,X_n$ indépendantes de même loi $\mathcal{B}(p)$ (iid)
$\Rightarrow \sum_{i=1}^{n}X_i\sim\mathcal{B}(n,p)$

Thm (corollaire) :

$$\left.\begin{array}{l} Y_1 \sim \mathcal{B}(n_1, p) \\ Y_2 \sim \mathcal{B}(n_2, p) \end{array}\right\} \Rightarrow Y_1 + Y_2 \sim \mathcal{B}(n_1+n_2, p)$$

Preuve : pas si simple...

3 Théorèmes de convergence.

Thm : Inégalité de Bienaymé-Tchebychev

X une v.a.r. admettant un moment d'ordre 2.

Alors $\forall a > 0$:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var(X)}{a^2}$$

Rmq : Loin d'être optimale !

Preuve :

$a > 0$ donc

$|X - \mathbb{E}[X]| \geq a \iff (X - \mathbb{E}[X])^2 \geq a^2$

Or $(X - \mathbb{E}[X])^2 \geq 0$.

On applique l'inégalité de Markov (voir ci-dessous.) à la v.a.r. $(X - \mathbb{E}[X])^2$ et au réel a^2 :

$$\Rightarrow \mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2) \leq \frac{\overbrace{\mathbb{E}[(X - \mathbb{E}[X])^2]}^{Var(X)}}{a^2}$$

D'où le résultat recherché. ▨

Thm : Inégalité de Markov

X une v.a.r. ≥ 0 qui admet un moment d'ordre 1.

Alors $\forall a > 0$:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Rmq : Permet de trouver un majorant mais pas forcém^T un sup.

Preuve :

Notons $(x_i)_{i \in [\![1;n]\!]}$ les valeurs prises par X.

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \cdot \mathbb{P}(X = x_i)$$

$$= \sum_{x_i \geq a} x_i \cdot \mathbb{P}(X = x_i) + \underbrace{\sum_{x_i < a} \overbrace{x_i}^{\geq 0} \cdot \overbrace{\mathbb{P}(X = x_i)}^{\geq 0}}_{\geq 0}$$

$$\geq \sum_{x_i \geq a} \underbrace{x_i}_{\geq a} \cdot \mathbb{P}(X = x_i) \geq a \sum_{x_i \geq a} \mathbb{P}(X = x_i) = a \cdot \mathbb{P}(X \geq a)$$

D'où le résultat recherché. ▨

indépendantes ($\perp\!\!\!\perp$) et identiquement distribuées.

Thm: Loi faible des grands nombres.

Soient $(X_i)_{i\in[\![1;n]\!]}$ des v.a.r. iid de loi $\mathcal{B}(p)$.

Alors :

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{\mathbb{P}} p.$$

Preuve :

Lemme : Inégalité de concentration.

X une v.a.r. d'espérance $\mathbb{E}[X]$ et de variance $Var(X)$.

$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ avec X_i iid (de même loi que X).

Alors :

$$\forall a>0, \quad \mathbb{P}(|\overline{X_n}-\mathbb{E}[X]|\geq a) \leq \frac{Var(X)}{na^2}$$

Preuve :

- $\mathbb{E}[\overline{X_n}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\times n\,\mathbb{E}[X] = \mathbb{E}[X].$

- $Var(\overline{X_n}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$ ← X_i $\perp\!\!\!\perp$.

  $= \frac{1}{n^2}\times n\,Var(X) = \frac{Var(X)}{n}.$

- D'après l'inégalité de B-T :

  $\forall a>0, \quad \mathbb{P}(|\overline{X_n}-\mathbb{E}[\overline{X_n}]|\geq a) \leq \frac{Var(\overline{X_n})}{a^2}$

  soit d'après les calculs précédents :

  $\mathbb{P}(|\overline{X_n}-\mathbb{E}[X]|\geq a) \leq \frac{Var(X)}{na^2}.$ ∎

D'après l'inégalité de concentration :

$$0 \leq \mathbb{P}(|\overline{X_n}-\mathbb{E}[X]|\geq a) \leq \frac{Var(X)}{na^2}$$

Or $\lim_{n\to+\infty}\frac{Var(X)}{na^2} = 0$ donc d'après le thm des gendarmes :

$$\lim_{n\to+\infty}\mathbb{P}(|\overline{X_n}-\underbrace{\mathbb{E}[X]}_{p}|\geq a) = 0, \quad \forall a>0$$ ∎

Définition : Convergence en probas :

$(X_n)_{n\in\mathbb{N}}$ suite de v.a.r. converge en probabilité vers une v.a.r. X

$\iff X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$

$\iff \forall\varepsilon>0, \lim_{n\to+\infty}\mathbb{P}(|X_n-X|>\varepsilon)=0$

Interprétation : $\forall\varepsilon>0$ fixé, Plus n est grand, il est de moins en moins probable d'observer un écart supérieur à ε entre X_n et X.

4 Approximations de la loi binomiale.

a) Par la loi de Poisson.

Définition : Une v.a.r. X suit une loi de Poisson de paramètre $\lambda > 0$
$\iff X \sim \mathcal{P}(\lambda)$
$\iff \forall k \in \mathbb{N},\ \mathbb{P}(X=k) = e^{-\lambda} \times \frac{\lambda^k}{k!}$

Propriété : $\lambda > 0$, $(p_n) \in [0;1]^{\mathbb{N}}$ tq $\lim_{n \to +\infty} n \cdot p_n = \lambda$.
Alors si $\forall n \in \mathbb{N}^*, S_n \sim \mathcal{B}(n, p_n)$:
$\forall k \in \mathbb{N},\ \mathbb{P}(S_n = k) \xrightarrow[n \to +\infty]{} e^{-\lambda} \frac{\lambda^k}{k!}$

i.e. la suite de lois $(\mathcal{B}(n, p_n))$ cv vers $\mathcal{P}(\lambda)$.

Preuve : $S_n \sim \mathcal{B}(n, p_n)$ avec $np_n \xrightarrow[n \to +\infty]{} \lambda$
$\forall k \in \mathbb{N}$:

$$\mathbb{P}(S_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} p_n^k (1-p_n)^{n-k}$$

$$\overset{n \text{ en facteur}}{=} \frac{n\, n^{k-1}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{k!} p_n^k (1-p_n)^{n-k}$$

$$= \frac{(np_n)^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \cdot (1-p_n)^{n-k}$$

$$\overset{p = \frac{np}{n}}{=} \frac{(np)^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \cdot \left(1-\frac{np_n}{n}\right)^n \times \left(1-\frac{np_n}{n}\right)^{-k}$$

$$\overset{x^n = e^{n \ln x}}{=} \frac{(np_n)^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \cdot e^{n\ln\left(1-\frac{np_n}{n}\right)} \times e^{-k\ln\left(1-\frac{np_n}{n}\right)}$$

Or :

* $\lim_{n \to +\infty} np_n = \lambda$ d'où :

$$\ln\left(1-\frac{np_n}{n}\right) \underset{n \to +\infty}{\sim} \frac{-\lambda}{n} \Rightarrow \begin{cases} n\ln\left(1-\frac{np_n}{n}\right) \underset{n \to +\infty}{\sim} -\lambda \\ -k\ln\left(1-\frac{np_n}{n}\right) \underset{n \to +\infty}{\sim} k \times \frac{\lambda}{n} \end{cases}$$

$$\Rightarrow \begin{cases} \lim_{n \to +\infty} e^{n\ln\left(1-\frac{np_n}{n}\right)} = e^{-\lambda} \\ \lim_{n \to +\infty} e^{-k\left(1-\frac{np_n}{n}\right)} = e^0 = 1 \end{cases}$$

* $\lim_{n \to +\infty} \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) = 1$

* $\lim_{n \to +\infty} \frac{(np_n)^k}{k!} = \frac{\lambda^k}{k!}$ car $np_n \xrightarrow[n \to +\infty]{} \lambda$

D'où au final : $\lim_{n \to +\infty} \mathbb{P}(S_n = k) = \frac{\lambda^k}{k!} \times 1 \times e^{-\lambda} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$

<u>Règle empirique</u>:

Si $n \geqslant 30$ et $p \leq 0{,}1$, alors $\mathcal{B}(n,p) \approx \mathcal{P}(\lambda)$.

<u>Applications</u>: Évènements rares :

* Nombre de clients entrant dans un magasin
* Nombre de coquilles dans une page de journal
* Nombre de raisins dans une tranche de cake

<u>Apparté</u>:

→ Calculer $\mathbb{E}[X]$ et $Var(X)$ quand $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$:

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}^*} k \cdot \mathbb{P}(X=k) = \sum_{k \in \mathbb{N}^*} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k \in \mathbb{N}^*} \frac{\lambda^k}{(k-1)!}$$

si $k=0$, c'est 0...

$$= \lambda \cdot e^{-\lambda} \cdot \sum_{k \in \mathbb{N}^*} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot e^{-\lambda} \cdot \sum_{k \in \mathbb{N}} \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda}$$

$$= \lambda.$$

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sum_{k \in \mathbb{N}} k^2 \cdot \mathbb{P}(X=k) - \lambda^2$$

Or :

$$\sum_{k \in \mathbb{N}} k^2 \, \mathbb{P}(X=k) = \sum_{k \in \mathbb{N}} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} \overset{k^2 = k + k(k-1)}{=} \sum_{k \in \mathbb{N}^*} \frac{k}{k!} \lambda^k e^{-\lambda} + \sum_{k \in \mathbb{N}^*} \frac{k(k-1)}{k!} \lambda^k e^{-\lambda}$$

$$= \sum_{k \geqslant 1} \frac{\lambda^k}{(k-1)!} e^{-\lambda} + \sum_{k \geqslant 2} \frac{\lambda^k}{(k-2)!} e^{-\lambda}$$

$$= \lambda e^{-\lambda} \underbrace{\sum_{k \geqslant 1} \frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} + \lambda^2 e^{-\lambda} \cdot \underbrace{\sum_{k \geqslant 2} \frac{\lambda^{k-2}}{(k-2)!}}_{e^{\lambda}}$$

$$= \lambda \underbrace{e^{-\lambda} \cdot e^{\lambda}}_{1} + \lambda^2 \underbrace{e^{-\lambda} e^{\lambda}}_{1} = \lambda + \lambda^2$$

D'où $Var(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$.

b) Par la loi de Gauss.

Thm : Moivre-Laplace

Soit $p \in]0;1[$, $(S_n)_{n\in\mathbb{N}}$ tq $S_n \sim \mathcal{B}(n,p)$.

On pose :
$$\tilde{S}_n = \frac{S_n - np}{\sqrt{np(1-p)}} \quad \forall n \in \mathbb{N}^*.$$

Alors $\tilde{S}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$

Définition : Convergence en loi : ou "convergence faible"

$(X_n)_{n\in\mathbb{N}}$ suite de v.a.r. converge en loi vers une v.a.r. X

$\Leftrightarrow X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$

$\Leftrightarrow \forall f$ fonction à valeurs réelles, continue et bornée :

$$\lim_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

→ Dans le cas discret : $\lim_{n\to+\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X=k)$

Dans le cas du thm ci-dessus :

$\forall a,b \in \mathbb{R}$:

$$\mathbb{P}(a \leq \tilde{S}_n \leq b) \xrightarrow[n\to+\infty]{} \int_a^b \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

Rmq : CV en Probas $\Rightarrow$ CV en loi.

Règle empirique :

Si $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$, alors :
$$\mathcal{B}(n,p) \approx \mathcal{N}(np, np(1-p))$$

Preuve : ⚠ Utilise les fct's caractéristiques, pas au prgm !

Soit $(S_n)_{n\in\mathbb{N}}$ tq $S_n \sim \mathcal{B}(n,p)$. $q = 1-p$.

Définition (hors prgm) :

Soit X un v.a.r.

Sa fonction caractéristique est la fct° à valeurs dans $\mathbb{C}$ définie sur $\mathbb{R}$ suivante :

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\,\mathbb{E}[\sin(tX)]$$

* X à densité : $\varphi_X(t) = \int_{\mathbb{R}} f_X(x) e^{itx} dx$

* X à val. ds $\mathbb{N}$: $\varphi_X(t) = \sum_{k=0}^{+\infty} \mathbb{P}(X=k) e^{itk}$

$$\varphi_{S_n}(t) = (pe^{it} + q)^n \quad *$$

$$\varphi_{\tilde{S}_n}(t) = \left(pe^{\frac{it}{\sqrt{npq}}} + q\right)^n \cdot e^{-\frac{itnp}{\sqrt{npq}}} \quad *$$

$$\ln(\varphi_{\tilde{S}_n}(t)) = n\ln\left[\left(pe^{it/\sqrt{npq}} + q\right)\right] - \frac{itnp}{\sqrt{npq}}$$

$$= n\ln\left[p\left(e^{\frac{it}{\sqrt{npq}}} - 1\right) + 1\right] - \frac{itnp}{\sqrt{npq}}$$

Or $e^x = 1 + x + \frac{x^2}{2} + o(x^2)$ d'où :

$$\ln\left(\varphi_{\tilde{S}_n}(t)\right) \approx n \cdot \ln\left[1 + p^*\left(1 + \frac{it}{\sqrt{npq}} + \frac{i^2t^2}{npq} - 1\right)\right] - \frac{itnp}{\sqrt{npq}}$$

$$\approx n \cdot \ln\left[1 + p^*\left(\frac{it}{\sqrt{npq}} - \frac{t^2}{npq}\right)\right] - \frac{itnp}{\sqrt{npq}}$$

On $\ln(x) \approx x - \frac{x^2}{2} + o(x^2)$ d'où :

$$\ln\left(\varphi_{\tilde{S}_n}(t)\right) \approx n \times \left[\frac{ipt}{\sqrt{npq}} - \frac{pt^2}{2npq} + \frac{p^2t^2}{2npq}\right] - \frac{itnp}{\sqrt{npq}}$$

$$\approx \cancel{\frac{itnp}{\sqrt{npq}}} - \frac{\cancel{np}t^2}{2\cancel{np}q} + \frac{\cancel{n}p^{\cancel{2}}t^2}{2\cancel{n}\cancel{p}q} - \cancel{\frac{itnp}{\sqrt{npq}}}$$

$$\approx -\frac{t^2}{2q} + \frac{pt^2}{2q} = \frac{t^2}{2q}\underbrace{(p-1)}_{=-(1-p)=-q} = \frac{t^2}{2q} \times \cancel{-q}$$

$\Rightarrow \ln\left(\varphi_{\tilde{S}_n}(t)\right) \approx \frac{-t^2}{2} \Rightarrow \varphi_{\tilde{S}_n}(t) \approx e^{-t^2/2}$, qui est la fonction caractéristique d'une $\mathcal{N}(0;1)$*. ▨

* Voir p. 18

Application : Intervalle de confiance/de fluctuation :

Lemme :
$\forall \gamma \in [0;1[, \ \exists! \, z_\gamma \geq 0$ tq $\int_{-z_\gamma}^{z_\gamma} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \gamma$

Preuve (idée) :
On sait que l'intégrale de Gauss converge :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1.$$

De plus : $\int_0^0 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0$.

La fonction $F(t) = \int_{-t}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ définie sur $\mathbb{R}_+$ est continue, (et strict ↗) avec $F(0) = 0$ et $F(t) \xrightarrow[t \to +\infty]{} 1$.
Donc $\forall \gamma \in [0;1[$, par le thm des valeurs intermédiaires, $\exists! \, z_\gamma \in \mathbb{R}_+$ tq $F(z_\gamma) = \gamma$. ▨

Propriété :

X_i iid $\mathcal{B}(p)$. Notons $\overline{X_n} = \frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$.

Alors :

$$\forall \gamma \in [0;1[, \quad \mathbb{P}\Bigg(\overline{X_n} \in \underbrace{\left[p - \frac{z_\gamma \sqrt{p(1-p)}}{\sqrt{n}} ; p + \frac{z_\gamma\sqrt{p(1-p)}}{\sqrt{n}}\right]}\Bigg) \xrightarrow[n\to+\infty]{} \gamma$$

Intervalle de fluctuation pour la variable $\overline{X_n}$ au seuil asymptotique γ.

En pratique : $\gamma = 0,95 \Rightarrow z_\gamma \approx 1,96$

Preuve :

Thm de Moivre-Laplace + lemme ci-dessus. ☒

Propriété :

X_i iid $\mathcal{B}(p)$, avec $p \in]0;1[$ inconnue. Alors :

$\forall \gamma \in [0;1[$:

$$\lim_{n\to+\infty} \mathbb{P}\Bigg(p \in \underbrace{\left[\overline{X_n} - \frac{z_\gamma}{2\sqrt{n}} ; \overline{X_n} + \frac{z_\gamma}{2\sqrt{n}}\right]}\Bigg) \geq \gamma$$

Intervalle de confiance pour le paramètre p, de niveau asymptotique γ

Application :

Meyre (T1) Ex. 3.7 p. 131.

Sondage à la sortie des urnes : sur un échantillon de 900 personnes, 45 déclarent avoir voté pour C.

Hypothèse : La proportion de voix favorables pour C est $p \in [0,01 ; 0,2]$.

Quelle fourchette peut-on en déduire pour p avec un niveau de confiance d'environ 95% ?

C'est un tirage sans remise de 900 personnes dans la population, mais 900 << population totale d'électeurs donc on peut l'assimiler à un tirage avec remise.

Le nb de sondés ayant voté pour C est modélisé par la v.a.r. :

$S_{900} \sim \mathcal{B}(900, p)$

$np \geq 9 \geq 5$ et $n(1-p) \geq 720$ donc avec $z_{0,95} \approx 1,96$:

$\mathbb{P}\left(\frac{|S_{900} - 900p|}{\sqrt{900p(1-p)}} \leq 1,96\right) \approx 0,95$ ou si $\overline{X_{900}} = \frac{S_{900}}{900}$:

$\mathbb{P}\left(|\overline{X_{900}} - p| \leq \frac{1,96\sqrt{p(1-p)}}{30}\right) \approx 0,95$ Or $\sqrt{p(1-p)} \leq 0,4$ d'où l'intervalle de confiance : $\left[\overline{x_{900}} \pm \frac{1,96 \times 0,4}{30}\right] = [0,0239 ; 0,0761]$.

Remarques sur la fonction caractéristique :

$S_n \sim \mathcal{B}(n,p)$; $\widetilde{S_n} = \frac{S_n - np}{\sqrt{npq}}$; $q = 1-p$.

$$* \ \varphi_{S_n}(t) = \mathbb{E}[e^{itS_n}] = \sum_{k=0}^{+\infty} \mathbb{P}(S_n = k) e^{itk}$$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{itk} = \sum_{k=0}^{n} \binom{n}{k} (pe^{it})^k q^{n-k}$$

$$= (pe^{it} + q)^n.$$

$$* \ \varphi_{\widetilde{S_n}}(t) = \mathbb{E}\left[e^{it\frac{S_n - np}{\sqrt{npq}}}\right] = \mathbb{E}\left[e^{it\frac{S_n}{\sqrt{npq}}}\right] \cdot e^{-\frac{itnp}{\sqrt{npq}}}$$

$$= \left[\sum_{k=0}^{+\infty} \mathbb{P}(S_n = k) \cdot e^{it\frac{k}{\sqrt{npq}}}\right] e^{-\frac{itnp}{\sqrt{npq}}}$$

$$= \left[\sum_{k=0}^{n} \binom{n}{k} \left(pe^{i\frac{t}{\sqrt{npq}}}\right)^k q^{n-k}\right] \cdot e^{-\frac{itnp}{\sqrt{npq}}}$$

$$= \left(p \cdot e^{\frac{it}{\sqrt{npq}}} + q\right)^n \times e^{-\frac{itnp}{\sqrt{npq}}}.$$

* Soit $X \sim \mathcal{N}(0,1)$:

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{+\infty} f_X(x) \cdot e^{itx}\, dx$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot e^{itx}\, dx = \frac{e^{-t^2/2}}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-it)^2} dx}_{F(t)}$$

$$-\frac{1}{2}(x-it)^2 = -\frac{x^2}{2} + itx + t^2/2.$$

Mq F est constante :

$$F'(t) = \int_{-\infty}^{+\infty} \frac{\partial e^{-\frac{1}{2}(x-it)^2}}{\partial t} dx = \int_{-\infty}^{+\infty} -\frac{1}{2} \times 2 \times (-i) \times (x-it) \cdot e^{-\frac{1}{2}(x-it)^2} dx$$

thm. de dérivat° des intégrales à paramètres (⚠ conditions!)

$$= \int_{-\infty}^{+\infty} -i \times \frac{\partial e^{-\frac{1}{2}(x-it)^2}}{\partial x} dx = -i \left[e^{-\frac{1}{2}(x-it)^2}\right]_{-\infty}^{+\infty}$$

$$= 0 \Rightarrow \forall t \in \mathbb{R},\ F(t) = F(0) = \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

$$\Rightarrow \varphi_X(t) = e^{-t^2/2}$$