

Dans tout le chapitre, $(\Omega, \mathcal{T}, \mathbf{P})$ est un espace probabilisé.

On étudie une suite de variables aléatoires réelles $(X_n)_{n \in \mathbf{N}}$ définies sur Ω .

Ces variables aléatoires sont indifféremment finies, dénombrables, ou à densité.

I Inégalités usuelles

1 Inégalité de Markov

PROPOSITION

Soit X une VAR positive ($X(\Omega) \subset \mathbf{R}_+$), et admettant une espérance $\mathbf{E}(X)$. Alors :

$$\forall a > 0, \mathbf{P}(X \geq a) \leq \frac{\mathbf{E}(X)}{a}$$

2 Inégalité de Bienaymé-Tchebychev

PROPOSITION

Soit X une VAR admettant un moment d'ordre 2. Alors :

$$\forall \varepsilon > 0, \mathbf{P}(|X - \mathbf{E}(X)| \geq \varepsilon) \leq \frac{\mathbf{V}(X)}{\varepsilon^2}$$

Remarque : ce résultat équivaut à : $\mathbf{P}(|X - \mathbf{E}(X)| < \varepsilon) \geq 1 - \frac{\mathbf{V}(X)}{\varepsilon^2}$

Exercice 1 : Soit X suivant une loi exponentielle : $X \hookrightarrow \mathcal{E}(\lambda)$. Soit $\varepsilon > 0$.

1. on suppose ici que $\varepsilon \leq \frac{1}{\lambda}$. Que penser dans ce cas de l'inégalité de Bienaymé-Tchebychev ?

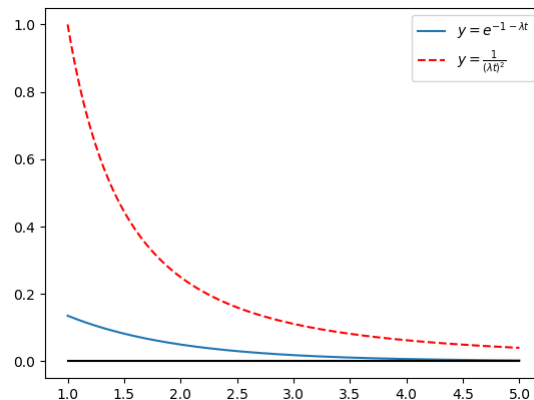
2. On suppose dans la suite que $\varepsilon > \frac{1}{\lambda}$.

a) Calculer : $\mathbf{P}(|X - \mathbf{E}(X)| \geq \varepsilon)$.

b) Le graphique ci-contre montre la représentation graphique des fonctions

$$f(t) = e^{-1-\lambda t} \text{ et } g(t) = \frac{1}{(\lambda t)^2}$$

pour $\lambda = 1$. Que penser de la précision de l'inégalité de Bienaymé-Tchebychev ?



II Loi faible des grands nombres

1 Moyenne empirique

DÉFINITION

Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de VAR. La suite des **moyennes empiriques** des X_n est la suite de VAR définie par :

$$\forall n \geq 1, M_n = \frac{1}{n} \sum_{k=1}^n X_k$$

PROPOSITION

Si $(X_n)_{n \in \mathbf{N}^*}$ est une suite de VAR mutuellement indépendantes, de même loi, admettant une espérance μ et une variance σ^2 , alors : $\mathbf{E}(M_n) = \mu$ et $\mathbf{V}(M_n) = \frac{\sigma^2}{n}$.

2 Loi faible des grands nombres

THÉORÈME

Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de VAR mutuellement indépendantes, de même loi, admettant une espérance μ et une variance σ^2 . Alors la suite $(M_n)_{n \geq 1}$ des moyennes empiriques vérifie :

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbf{P}(|M_n - \mu| > \varepsilon) = 0$$

Rq : on dit que la suite $(M_n)_{n \geq 1}$ converge en probabilité vers une VAR certaine de valeur μ .

Exercice 2 : Soit $n \in \mathbf{N}^*$. On jette n fois un dé, et on note A_n l'événement :

A_n : « au cours des n lancers, on a obtenu 6 moins d'une fois sur dix ».

Utiliser l'inégalité de Bienaymé-Tchebychev pour donner une majoration de $\mathbf{P}(A_n)$ lorsque :

$$* n = 10 \quad * n = 100 \quad * n = 1000$$

III Convergence en loi

1 Cas de VAR entières

DÉFINITION

Soit $(X_n)_{n \in \mathbf{N}}$ une suite de VAR entières ($\forall n \in \mathbf{N}, X_n(\Omega) \subset \mathbf{N}$), et soit X une VAR entière. Alors on dit que $(X_n)_{n \in \mathbf{N}}$ **converge en loi** vers X si et seulement si :

$$\forall k \in \mathbf{N}, \lim_{n \rightarrow +\infty} \mathbf{P}(X_n = k) = \mathbf{P}(X = k)$$

PROPRIÉTÉ ** Approximation d'une loi binomiale par une loi de Poisson **

Soit $(X_n)_{n \in \mathbf{N}}$ une suite de VAR telle que : $\forall n \in \mathbf{N}, X_n \hookrightarrow \mathcal{B}(n, p_n)$.

Si $\lim_{n \rightarrow +\infty} n \times p_n = \lambda \in \mathbf{R}$, alors $(X_n)_n$ converge en loi vers une VAR X telle que : $X \hookrightarrow \mathcal{P}(\lambda)$.

Usage : on admet communément que lorsque $n \geq 30$ et $p \leq 0,1$, on peut approximer la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$.

Exercice 3 : Soit $X \hookrightarrow \mathcal{B}\left(1000, \frac{1}{40}\right)$. Donner une estimation de : $\mathbf{P}(X = 25)$.

2 Cas général

DÉFINITION

Soit $(X_n)_{n \in \mathbf{N}}$ une suite de VAR, et soit X une VAR.

On note F_{X_n} et F_X respectivement les fonctions de répartition de X_n et de X .

Soit $\mathcal{D}$ l'ensemble des points de discontinuité de F_X .

Alors on dit que $(X_n)_{n \in \mathbf{N}}$ **converge en loi** vers X si et seulement si :

$$\forall x \in \mathbf{R} \setminus \mathcal{D}, \lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

Rappel : $\overline{\mathbf{R}}$ désigne $\mathbf{R} \cup \{-\infty, +\infty\}$.

PROPOSITION

Si $(X_n)_{n \in \mathbf{N}}$ converge en loi vers X , alors :

$$\forall a, b \in \overline{\mathbf{R}} \setminus \mathcal{D} \text{ (avec } a < b), \lim_{n \rightarrow +\infty} \mathbf{P}(a < X_n \leq b) = \mathbf{P}(a < X \leq b)$$

3 Théorème central-limite première forme

Rappels : • Soit X une VAR admettant une variance non nulle. Alors on note X^* la VAR centrée réduite

$$\text{associée à } X : X^* = \frac{X - \mathbf{E}(X)}{\sqrt{\mathbf{V}(X)}}.$$

• Φ désigne la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$:

$$\forall x \in \mathbf{R}, \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

THÉORÈME

Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de VAR mutuellement indépendantes, de même loi, admettant une espérance μ et une variance $\sigma^2 \neq 0$. Alors la suite $(M_n^*)_{n \geq 1}$ converge en loi vers une VAR suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$:

$$\forall a, b \in \overline{\mathbf{R}} \text{ (avec } a < b), \lim_{n \rightarrow +\infty} \mathbf{P}(a < M_n^* \leq b) = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt$$

Remarque : Soit $S_n = \sum_{k=1}^n X_k = n \times M_n$. Alors $S_n^* = M_n^*$ donc $(S_n^*)_n$ converge aussi en loi vers $\mathcal{N}(0, 1)$.

PROPRIÉTÉ ** Approximation d'une loi binomiale par une loi normale **

Soit $p \in]0, 1[$ fixé, et soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de VAR telle que : $\forall n \in \mathbf{N}^*, X_n \hookrightarrow \mathcal{B}(n, p)$.

Alors $(X_n^*)_{n \geq 1}$ converge en loi vers $\mathcal{N}(0, 1)$:

$$\forall a, b \in \overline{\mathbf{R}} \text{ (avec } a < b), \lim_{n \rightarrow +\infty} \mathbf{P}\left(a < \frac{X_n - np}{\sqrt{np(1-p)}} \leq b\right) = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt$$

Usage : on admet communément que lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, on peut approximer la loi $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np, npq)$.

Estimation de $\mathbf{P}(X_n = k)$ par correction de continuité

Pour $k \in \llbracket 0, n \rrbracket$, on veut une approximation de : $\mathbf{P}(X_n = k) = \mathbf{P}\left(X_n^* = \frac{k - np}{\sqrt{npq}}\right)$.

Prendre $a = b = \frac{k - np}{\sqrt{npq}}$ donnerait 0... On effectue alors une **correction de continuité** :

$$\mathbf{P}(X_n = k) = \mathbf{P}\left(k - \frac{1}{2} < X_n \leq k + \frac{1}{2}\right) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt \quad \text{avec} \quad \begin{cases} a = \frac{k - \frac{1}{2} - np}{\sqrt{npq}} \\ b = \frac{k + \frac{1}{2} - np}{\sqrt{npq}} \end{cases}$$

Exercice 4 : Soit $X \hookrightarrow \mathcal{B}\left(1000, \frac{1}{40}\right)$.

- 1) Estimer la probabilité de l'événement : $[20 < X \leq 30]$.
- 2) Estimer la probabilité de l'événement : $[X = 25]$. Comparer avec le résultat de l'exercice 3.

IV Tests statistiques

1 Écart-type empirique

DÉFINITION

Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de VAR. On définit pour tout $n \geq 1$ la **variance empirique** de $X_1, \dots, X_n$ par :

$$S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - M_n)^2$$

L'écart-type empirique de $X_1, \dots, X_n$ est : $S_n = \sqrt{S_n^2}$.

2 Théorème central-limite seconde forme

THÉORÈME

Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de VAR mutuellement indépendantes, de même loi, admettant une espérance μ et une variance $\sigma^2 \neq 0$. Alors la suite $\left(\frac{M_n - \mu}{\frac{S_n}{\sqrt{n}}}\right)_{n \geq 1}$ converge en loi vers la loi normale centrée réduite.

Usage : on se sert de cette deuxième forme lorsqu'on ne connaît pas l'écart-type commun des VAR X_n .

3 Test de conformité à la moyenne

Situation : on peut expérimentalement effectuer n mesures d'un phénomène, et on souhaite tester statistiquement l'hypothèse : $\ll H_0 : \text{la valeur moyenne du phénomène vaut } \mu_0 \gg$.

La valeur de μ_0 peut être obtenue par calcul selon une modélisation mathématique, ou bien être issue d'une bibliographie... On l'appelle ici *moyenne théorique*.

On pose H_1 l'hypothèse alternative : $\ll H_1 : \text{la valeur moyenne du phénomène est différente de } \mu_0 \gg$.

On souhaite prendre la décision : accepte-t-on H_0 , ou bien doit-on la rejeter ?

On peut se tromper dans deux cas :

* on rejette H_0 , alors qu'elle est vraie. C'est le risque de *première espèce*.

* on accepte H_0 , alors qu'elle est fautive. C'est le risque de *seconde espèce*.

Même en multipliant les expériences, on ne pourra jamais être certain d'avoir pris la bonne décision.

On fixe le *risque d'erreur de première espèce*, noté α , égal à : $\alpha = \mathbf{P}_{[H_0 \text{ est vraie}]}(\text{on rejette } H_0)$.

On choisit la plupart du temps $\alpha = 0,05$ ou $0,01$ (risque d'erreur à 5% ou à 1%).

Le risque d'erreur de seconde espèce n'est pas étudié ici.

DÉFINITION

Soit X une VAR admettant une moyenne μ_0 et un écart-type $\sigma_0 \neq 0$. Soient $n \in \mathbf{N}^*$ et $\alpha \in]0, 1[$.

* On appelle $(X_1, \dots, X_n)$ un **n -échantillon** de X tout vecteur de VAR mutuellement indépendantes et suivant la même loi que X . Sa moyenne empirique est notée M_n .

* On appelle **marge d'erreur** au seuil $1 - \alpha$ le réel : $Err = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \times \frac{S_n}{\sqrt{n}}$ où S_n est l'écart-type empirique issu d'un n -échantillon de X .

* On appelle **intervalle de confiance** au seuil $1 - \alpha$ l'intervalle réel : $IC_n = [M_n - Err, M_n + Err]$.

PROPOSITION

$$\lim_{n \rightarrow +\infty} \mathbf{P}(\mu_0 \in IC_n) = 1 - \alpha$$

Conduite du test :

1. on fixe arbitrairement $\alpha \in]0, 1[$.
2. on détermine $u_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ grâce à une table de la loi normale.
3. On effectue n simulations de X et on calcule moyenne et écart-type empiriques M_n et S_n .
4. On explicite l'intervalle de confiance $IC_n = \left[M_n - u_\alpha \times \frac{S_n}{\sqrt{n}}, M_n + u_\alpha \times \frac{S_n}{\sqrt{n}} \right]$.
5. On décide de rejeter H_0 si $\mu_0 \notin IC_n$. Sinon, on conclut que le test ne permet pas de rejeter H_0 au seuil d'erreur α .

On a alors asymptotiquement (si n est grand) un risque d'erreur α .

En pratique, on admet la validité du test lorsque $n \geq 30$.

Exercice 5 : Une usine agro-alimentaire conditionne des barquettes de fraises de 250 grammes.

En pratique, les barquettes ne font bien sûr pas toutes exactement 250 g : cette valeur doit être comprise comme une masse moyenne. Un contrôle est effectué pour savoir si l'usine respecte cette moyenne de 250 g. On prélève au hasard $n = 100$ barquettes issues de cette usine, et on calcule la masse moyenne M_n et l'écart-type S_n de cet échantillon. On trouve : $M_n = 247$ g et $S_n = 5$ g.

Au risque d'erreur de 5%, peut-on considérer que l'usine respecte la moyenne de 250 g ?

4 Test de conformité sur une proportion

Situation : on connaît la moyenne théorique μ d'un phénomène, et on souhaite tester si des valeurs peuvent être issues de ce phénomène, en comparant la moyenne empirique M_n à la moyenne théorique μ .

DÉFINITION

Soit un phénomène apparaissant avec une fréquence μ dans une population.

Soit $\alpha \in]0, 1[$. On appelle **intervalle de fluctuation** au seuil $1 - \alpha$ l'intervalle réel :

$$IF_n = \left[\mu - \frac{a}{\sqrt{n}}, \mu + \frac{a}{\sqrt{n}} \right]$$

où $a = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \times \sqrt{\mu(1 - \mu)}$.

DÉFINITION SIMPLIFIÉE : pour $\alpha = 0,05$ et en majorant $\mu(1 - \mu)$ par $\frac{1}{4}$, on obtient : $IF_n \approx \left[\mu - \frac{1}{\sqrt{n}}, \mu + \frac{1}{\sqrt{n}} \right]$.

PROPOSITION

On choisit au hasard et indépendamment $n \geq 1$ éléments de cette population, et on note M_n la moyenne empirique obtenue. Alors : $\lim_{n \rightarrow +\infty} \mathbf{P}(M_n \in IF_n) = 1 - \alpha$.

Conduite du test au risque d'erreur de 5% :

1. on pose $\alpha = 0,05$ (si autre choix, adapter les valeurs suivantes).
2. On explicite l'intervalle de fluctuation simplifié $IF_n = \left[\mu - \frac{1}{\sqrt{n}}, \mu + \frac{1}{\sqrt{n}} \right]$.
3. On calcule M_n .
4. On décide que les valeurs ne sont pas issues du phénomène étudié (elles ne sont pas représentatives de ce phénomène) si $M_n \notin IF_n$. Sinon, on conclut que le test ne permet pas d'exclure que ces valeurs soient issues du phénomène.

Comme précédemment, on convient que $n \geq 30$ est suffisant pour assurer la validité du test.

Exercice 6 : En 2024, on compte 51,5% de femmes en France. Une étude montre que parmi les 80 postes de direction (président ou directeur général) des entreprises du CAC40, on compte 5 femmes.

Peut-on affirmer au risque d'erreur de 5%, que les femmes sont bien représentées dans les directions des plus grandes entreprises françaises ?