Convergences et théorèmes limite

I Convergences en probabilité page 2

1. Inégalité de Markov et inégalité de Bienaymé-Tchebichev
2. Loi faible des grands nombres

II Convergence en loi page 3

1. Convergence en loi dans le cas discret
2. Convergence en loi dans le cas général : théorème central limit page 4

III Test de conformité sur la moyenne page 6

1. Deuxième forme du théorème central limit
2. Test de conformité sur la moyenne
3. Test de conformité sur une proportion page 7

I Convergences en probabilité

On considére une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$, toutes définies sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$.
On étudie le comportement de X_n lorsque n tend vers $+\infty$.

1. Inégalité de Markov et inégalité de Bienaymé-Tchebichev

Proposition 1 Inégalité de Markov
Soit X une variable aléatoire réelle, à valeurs positives, admettant une espérance $E(X)$. Alors $\forall a > 0,\ P(X \geqslant a) \leqslant \dfrac{E(X)}{a}$

Preuve •Si X est discrète alors $aP(X \geqslant a) = \sum\limits_{k\geqslant a} aP(X=k) \leqslant \sum\limits_{k\geqslant a} kP(X=k) \leqslant \sum\limits_{k\in X(\Omega)} kP(X=k)$ car $X \geqslant 0$.
•Si X est à densité f alors $aP(X \geqslant a) = \int_a^{+\infty} af(t)\,\mathrm{d}t \leqslant \int_a^{+\infty} tf(t)\,\mathrm{d}t \leqslant \int_0^{+\infty} tf(t)\,\mathrm{d}t$ car $X \geqslant 0$.

Proposition 2 Inégalité de Bienaymé-Tchebichev
Soit X une variable aléatoire réelle admettant une espérance $E(X)$ et une variance $V(X)$.
Alors $\forall \varepsilon > 0,\ P(|X - E(X)| \geqslant \varepsilon) \leqslant \dfrac{V(X)}{\varepsilon^2}$
ce qui équivalent à : $\forall \varepsilon > 0,\ P(|X - E(X)| < \varepsilon) \geqslant 1 - \dfrac{V(X)}{\varepsilon^2}$

Preuve On a $[|X - E(X)| \geqslant \varepsilon] = \left[(X - E(X))^2 \geqslant \varepsilon^2\right]$.
$(X - E(X))^2$ est une variable aléatoire à valeurs positives admettant une espérance car X admet une variance et $E\left[(X - E(X))^2\right] = V(X)$.
L'inégalité de Markov appliquée à $(X - E(X))^2$ pour $a = \varepsilon^2 > 0$ permet d'écrire
$P((X - E(X))^2 \geqslant \varepsilon^2) \leqslant \dfrac{V(X)}{\varepsilon^2}$.

Exercice 1 Une usine produit en moyenne 35 pièces par semaine.
On note X la variable aléatoire donnant le nombre de pièces produites par semaine.
Que peut-on dire de la probabilité que l'usine produise plus de 70 pièces par semaine ?

2. Loi faible des grands nombres

Définition 1 Soit (X_n) une suite de variables aléatoires réelles et X une variable aléatoire réelle.
La suite (X_n) converge en probabilité vers X si, et seulement si :

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

Exercice 2 Soit $X_n \hookrightarrow \mathcal{N}\left(0, \frac{1}{n}\right)$. Montrer que (X_n) converge en probabilité vers la variable certaine nulle.

Proposition 3 Soit $(X_n)_{n>0}$ une suite de variables aléatoires réelles, indépendantes et de même loi admettant toutes la même espérance μ et la même variance σ^2.
Alors la variable moyenne de $X_1, \cdots, X_n$, définie par : $M_n = \dfrac{1}{n}\sum\limits_{k=1}^{n} X_k$ admet pour espérance $E(M_n) = \mu$ et pour variance $V(M_n) = \dfrac{\sigma^2}{n}$.

Preuve Par linéarité de l'espérance, $E(M_n) = \dfrac{1}{n}\sum_{k=1}^{n} E(X_k) = \mu$ et, par indépendance des X_n,
$V(M_n) = \dfrac{1}{n^2}\sum_{k=1}^{n} V(X_k) = \dfrac{\sigma^2}{n}$.

Proposition 4 Loi faible des grands nombres

Soit $(X_n)_{n>0}$ une suite de variables aléatoires réelles.
On suppose que les variables X_n sont indépendantes et de même loi.
Soit μ leur espérance et σ^2 leur variance commune.
Alors la suite $(M_n)_{n>0}$ converge en probabilité vers la variable certaine égale à μ.

Autrement dit :

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

Preuve L'inégalité de Bienaymé-Tchebichev appliquée à M_n s'écrit :
$\forall \varepsilon > 0,\ P(|M_n - E(M_n)| \geqslant \varepsilon) \leqslant \dfrac{V(M_n)}{\varepsilon^2}$ d'où $P(|M_n - \mu| \geqslant \varepsilon) \leqslant \dfrac{\sigma^2}{n\varepsilon^2}$.
On a bien $\lim\limits_{n\to+\infty} P(|M_n - m| \geqslant \varepsilon) = 0$

Exercice 3 On lance indéfiniment un dé bien équilibré.
Quelle est la limite en probabilité de la fréquence d'apparition du 6 ?

Remarque Cela justifie les simulations Python pour trouver des valeurs approchées de l'espérance : on suppose qu'on dispose d'une fonction *simul_X* `()` qui simule une variable aléatoire X.

```
def Esp_X(x,N):
    s = 0
    for _ in range(N):
        s += simul_X(x)
    return s/N
```

ou

```
def Esp_X(x,N):
    return sum([simul_X(x) for _ in range(N)])/N
```

II Convergence en loi

1. Convergence en loi dans le cas d'une variable à valeurs dans $\mathbb{N}$

Définition 2 Soit (X_n) une suite de variable aléatoire discrètes à valeur dans $\mathbb{N}$.
La suite (X_n) converge en loi vers une variable X si, et seulement si :

$$\forall k \in \mathbb{N}, \lim_{n\to+\infty} P(X_n = k) = P(X = k).$$

Proposition 5 Approximation d'une loi binomiale par une loi de Poisson

Soit (p_n) une suite de nombres réels telle que :
•$\forall n \in \mathbb{N},\ p_n \in [0,1]$
• $\lim\limits_{n\to+\infty} np_n = \lambda > 0$ (par exemple $p_n = \dfrac{\lambda}{n}$)
Soit (X_n) une suite de variables aléatoires telle que :$\forall n \in \mathbb{N},\ X_n \hookrightarrow \mathcal{B}(n,p_n)$
Alors la suite (X_n) converge en loi vers une variable aléatoire de loi de Poisson $\mathcal{P}(\lambda)$
C'est-à-dire que : $\forall k \in \mathbb{N},\ \lim\limits_{n\to+\infty} P(X_n = k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$

Preuve Comme $X_n \hookrightarrow \mathcal{B}(n,p_n)$, $\forall k \in \mathbb{N}$, $P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$

Soit $\lambda_n = np_n$ et $\forall k \in \mathbb{N}$, $P(X_n = k) = \dfrac{n(n-1)\cdots(n-k+1)}{n^k} \dfrac{\lambda_n^k}{k!}\left(1-\dfrac{\lambda_n}{n}\right)^{n-k}$.

Or $\lim\limits_{n\to+\infty} \dfrac{n(n-1)\cdots(n-k+1)}{n^k} = \lim\limits_{n\to+\infty} \prod\limits_{i=1}^{k} \dfrac{n-i+1}{n} = 1$,

$\lim\limits_{n\to+\infty} \dfrac{\lambda_n^k}{k!} = \dfrac{\lambda^k}{k!}$ et $\left(1-\dfrac{\lambda_n}{n}\right)^{n-k} = \exp\left[(n-k)\ln\left(1-\dfrac{\lambda_n}{n}\right)\right]$

Comme $\ln\left(1-\dfrac{\lambda_n}{n}\right) = -\dfrac{\lambda_n}{n} + o\left(\dfrac{1}{n}\right)$, $(n-k)\ln\left(1-\dfrac{\lambda_n}{n}\right) = -\lambda_n + o(1)$

donc $\lim\limits_{n\to+\infty}\left(1-\dfrac{\lambda_n}{n}\right)^{n-k} = \lim\limits_{n\to+\infty} e^{-\lambda_n + o(1)} = e^{-\lambda}$. D'où $\forall k \in \mathbb{N}$, $\lim\limits_{n\to+\infty} P(X_n = k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$

Remarque Il est communément admis que la condition $n \geqslant 30$ et $p \leqslant 0,1$ est suffisante pour approximer une loi $\mathcal{B}(n,p)$ par la loi $\mathcal{P}(np)$.

Exercice 5 On suppose que la probabilité de rupture d'un stock de médicaments est $p = 10\%$ par mois.
Quelle est la probabilité pour qu'il y ait exactement quatre défaillances sur cinq ans ?
Comparer la valeur exacte avec la valeur obtenue par approximation.

2. Convergence en loi dans le cas général : théorème central limite

Définition 3 Soit (X_n) une suite de variables aléatoires réelles et X une variable aléatoire réelle.
Pour tout n de $\mathbb{N}$, on note F_{X_n} la fonction de répartition de X_n.
Soit F_X la fonction de répartition de X et D l'ensemble des points de discontinuité de F.
Alors, la suite (X_n) <u>converge en loi</u> vers X si, et seulement si : $\forall x \in \mathbb{R}\backslash D$, $\lim\limits_{n\to+\infty} F_n(x) = F(x)$.

Remarque Si (X_n) converge en loi vers X alors
$\forall (a,b) \in (\mathbb{R}\backslash D)^2 \mid a < b$, $\lim\limits_{n\to+\infty} P(a < X_n \leqslant b) = P(a < X \leqslant b)$.

Proposition 6 <u>Théorème central limite (première forme)</u>
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes, de même loi, admettant toutes la même espérance μ et la même variance non nulle σ^2.

Soit, pour tout $n > 0$, $M_n = \dfrac{1}{n}\sum\limits_{k=1}^{n} X_k$ et $M_n^* = \dfrac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}$.

Alors la suite $(M_n^*)_{n\in\mathbb{N}^*}$ converge en loi vers une variable de loi $\mathcal{N}(0,1)$.

Preuve Théorème admis

Remarques • Si n est assez grand ($n \geqslant 30$), on peut approximer la loi de $M_n^* = \dfrac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ par $\mathcal{N}(0,1)$:

$\forall (a,b) \in \overline{\mathbb{R}}^2 \mid a < b$, $\lim\limits_{n\to+\infty} P(a < M_n^* \leqslant b) = \Phi(b) - \Phi(a) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_a^b e^{-t^2/2}\,\mathrm{d}t$.

• soit la somme $S_n = \sum\limits_{k=1}^{n} X_k = nM_n$ alors $E(S_n) = \sum\limits_{k=1}^{n} E(X_k) = n\mu$ et $V(S_n) = \sum\limits_{k=1}^{n} V(X_k) = n\sigma^2$

donc $E(M_n) = \dfrac{1}{n}E(S_n) = \mu$ et $V(M_n) = \dfrac{1}{n^2}V(S_n) = \dfrac{\sigma^2}{n}$.

Ainsi $S_n^* = \dfrac{S_n - E(S_n)}{\sqrt{n\sigma^2}} = \dfrac{S_n - n\mu}{\sigma\sqrt{n}} = \dfrac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}} = M_n^*$.

On a aussi la suite $(S_n^*)_{n\in\mathbb{N}^*}$ converge en loi vers une variable de loi $\mathcal{N}(0,1)$.

Proposition 7 Théorème de Moivre-Laplace : approximation d'une loi binomiale par une loi normale

Soit $p \in]0,1[$.
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires telle que : $\forall n \in \mathbb{N}^*,\ X_n \hookrightarrow \mathcal{B}(n,p)$.
Pour tout n de $\mathbb{N}^*$, on pose $X_n^* = \dfrac{X_n - np}{\sqrt{n(1-p)}}$.
Alors la suite $(X_n^*)_{n\in\mathbb{N}^*}$ converge en loi vers une variable de loi normale $\mathcal{N}(0,1)$.

Autrement dit : $\forall (a,b) \in \overline{\mathbb{R}}^2 \ \mid\ a < b,\ \lim\limits_{n\to+\infty} P(a < X_n^* \leqslant b) = \Phi(b) - \Phi(a) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_a^b e^{-t^2/2}\,\mathrm{d}t$.

Remarque Les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont généralement considérées comme suffisantes pour approximer la loi de $X_n^* = \dfrac{X_n - np}{\sqrt{npq}}$ par la loi $\mathcal{N}(0,1)$.

Preuve On pose $Y_1, \cdots, Y_n$ n variables aléatoires indépendantes toutes de même loi de Bernoulli de paramètre p.
Alors $\forall k \in [\![1,n]\!],\ Y_k$ admet une espérance $E(Y_k) = p$ et une variance $V(Y_k) = p(1-p)$.
On a alors $\displaystyle\sum_{k=1}^{n} Y_k = X_n \hookrightarrow \mathcal{B}(n,p)$
et $X_n^* = \dfrac{X_n - np}{\sqrt{np(1-p)}}$.
D'après le théorème central limite, $(X_n^*)_{n\in\mathbb{N}^*}$ converge en loi vers une variable de loi normale $\mathcal{N}(0,1)$.

Remarque Si on veut obtenir une approximation de $P(X_n = k)$, prendre $a = b = k$ n'est pas envisageable puisque l'on obtiendrait alors 0, on effectue alors une correction de continuité.
On écrit $P(X_n^* = k) = P(k - \frac{1}{2} \leqslant X_n^* \leqslant k + \frac{1}{2})$ et $\lim\limits_{n\to+\infty} P(X_n^* = k) = \Phi\left(k + \dfrac{1}{2}\right) - \Phi\left(k - \dfrac{1}{2}\right)$.

Exercice 6 Déterminer la probabilité que, en 36 lancers d'une pièce bien équilibrée, on obtienne exactement le même nombre de "Pile" que de "Face".

Proposition 8 Approximation d'une loi de Poisson par une loi normale

Soit $\alpha > 0$, $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles telle que : $\forall n \in \mathbb{N},\ X_n \hookrightarrow \mathcal{P}(n\alpha)$.
Pour tout n de $\mathbb{N}$, on pose $X_n^* = \dfrac{X_n - n\alpha}{\sqrt{n\alpha}}$.
Alors la suite $(X_n^*)_{n\in\mathbb{N}^*}$ converge en loi vers une variable de loi normale $\mathcal{N}(0,1)$.
C'est-à-dire que $\forall (a,b) \in \overline{\mathbb{R}}^2 \ \mid\ a < b,\ \lim\limits_{n\to+\infty} P(a < X_n^* \leqslant b) = \Phi(b) - \Phi(a) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_a^b e^{-t^2/2}\,\mathrm{d}t$.

Preuve X_n peut être considérée comme la somme de n variables de Poisson indépendantes de même paramètre α, d'espérance α et de variance $\sigma^2 = \alpha$.
Le théorème central limite permet de conclure.

Remarques •Là encore, on effectue une correction de continuité.
•La condition $\lambda \geqslant 15$ est généralement considérée comme suffisante pour approximer la loi de $X_n^* = \dfrac{X_n - n\alpha}{\sqrt{n\alpha}}$ par la loi $\mathcal{N}(0,1)$.

Exercice 7 En appliquant le théorème central limit à une suite de variables aléatoires réelles indépendantes $(X_n)_{n\in\mathbb{N}}$ suivant toutes une loi $\mathcal{P}(1)$, calculer : $\lim\limits_{n\to+\infty} e^{-n} \displaystyle\sum_{k=0}^{n} \dfrac{n^k}{k!}$.

III Test de conformité sur la moyenne

1. Deuxième forme du théorème central limite

Proposition 9 Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, indépendantes, de même loi, admettant toutes une espérance non nulle μ et une variance inconnue non nulle σ^2.

On définit la moyenne empirique $M_n = \frac{1}{n}\sum_{k=1}^{n} X_k$

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

Soit $Y_n = \frac{M_n - \mu}{\frac{S_n}{\sqrt{n}}}$ alors la suite $(Y_n)_{n\in\mathbb{N}^*}$ converge en loi vers la loi $\mathcal{N}(0,1)$.

C'est-à-dire que : $\forall (a,b) \in \mathbb{R}^2 \;\mid\; a < b,\ \lim\limits_{n\to+\infty} P(a < Y_n \leqslant b) = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}\,\mathrm{d}t.$

Preuve Théorème admis

Remarques • l'intérêt de cette deuxième forme provient du fait que lorsqu'on fait une analyse statistique, on ne connait pas nécessairement à l'avance la variance σ^2.

On la remplace donc par la variance empirique $S_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - M_n)^2$.

• Ce théorème exprime que si n est assez grand (en pratique $n \geqslant 30$), il est possible d'approximer la loi de $\frac{M_n - \mu}{\frac{S_n}{\sqrt{n}}}$ par la loi $\mathcal{N}(0,1)$.

• Une autre version de ce théorème utilise l'écart-type empirique corrigé S'_n où

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

2. Test de conformité sur la moyenne

Le but de ce test est de comparer une moyenne observée μ et une moyenne théorique μ_0.

Proposition 10 Soit X une variable aléatoire d'espérance μ et de variance σ^2.

Soit $X_1, \cdots, X_n$ n variables aléatoires indépendantes toutes de même loi que X.

On pose : $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - M_n^2$.

Soit $\alpha \in [0,1]$ le risque et $u = U_{1-\frac{\alpha}{2}}$ c'est-à-dire $u = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

Sous l'hypothèse $H : \mu = \mu_0$, on a $\lim\limits_{n\to+\infty} P\left(M_n - \frac{S_n}{\sqrt{n}}u \leqslant \mu_0 \leqslant M_n + \frac{S_n}{\sqrt{n}}u\right) = 1 - \alpha$.

Preuve $\lim\limits_{n\to+\infty} P\left(M_n - \frac{S_n}{\sqrt{n}}u \leqslant \mu_0 \leqslant M_n + \frac{S_n}{\sqrt{n}}u\right) = \lim\limits_{n\to+\infty} P\left(\left|\frac{M_n - \mu}{\frac{S_n}{\sqrt{n}}}\right| \leqslant u\right) = 2\Phi(u) - 1 = 1 - \alpha$

Remarques • Le test est asymptotique , on l'utilise pour des grandes valeurs de n (en pratique $n \geqslant 30$).

• **En pratique**, on effectue les mesures, on calcule la moyenne M_n et la variance empirique S_n^2.

Si $M_n - \frac{S_n}{\sqrt{n}}u \leqslant \mu_0 \leqslant M_n + \frac{S_n}{\sqrt{n}}u$ on accepte l'hypothèse H au risque α sinon, on la rejette.

Exercice 8 De larges études ont montré que le taux de glycémie moyen chez l'homme est de $0,85$ gramme par litre de sang. Dans un hôpital, on mesure la glycémie de 49 volontaires. La moyenne observée est de $0,92\ g/L$ avec un écart-type de $0,1\ g/L$.

Peut-on affirmer avec une marge d'erreur de 1% que l'échantillon de volontaires est représentatif de la population ?

3. Test de conformité sur une proportion

Cas particulier du test de conformité sur la moyenne :
On souhaite étudier la proportion inconnue p d'individus présentant un caractère c dans une population.
On appelle X la variable aléatoire de loi de Bernoulli de paramètre p.
On a alors $\mu = p$ et $\sigma^2 = p(1-p)$.
Or $\forall p \in [0,1]$, $p(1-p) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leqslant \frac{1}{4}$ donc $\sigma \leqslant \frac{1}{2}$.
Soit $X_1, \cdots, X_n$ n variables aléatoires indépendantes toutes de même loi que X.

Ainsi, $\forall i \in [\![1,n]\!]$, $X_i = \begin{cases} 1 \text{ si le } i\text{-ième individu possède le caractère } c \\ 0 \text{ sinon} \end{cases}$.

On pose : $F_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ la fréquence d'appartion du caractère c et α le risque et $u = U_{1-\frac{\alpha}{2}}$.
Le théorème central limit (première forme) permet d'écrire que :
sous l'hypothèse $H : p = p_0$, $\lim\limits_{n\to+\infty} P\left(F_n - \frac{u}{2\sqrt{n}} \leqslant p_0 \leqslant F_n + \frac{u}{2\sqrt{n}}\right) = 1 - \alpha$.

Remarques • Le test est asymptotique , on l'utilise pour des grandes valeurs de n (en pratique $n \geqslant 30$).
• **En pratique**, on effectue les mesures, on calcule la moyenne M_n (proportion empirique d'individus présentant le caractère c).
Si $F_n - \frac{u}{2\sqrt{n}} \leqslant p_0 \leqslant F_n + \frac{u}{2\sqrt{n}}$ on peut dire, au risque α, que la proportion d'individus présentant le caractère c est p_0.

Exercice 9 On dit qu'un citron est de calibre C si son diamètre est compris entre $6,5$ et $7,3$ centimètres.
Un producteur affirme que la majorité des citrons sortant de sa production sont de calibre C.
Pour vérifier cette affirmation, un contrôleur extrait au hasard 40 citrons sortant de cette production et mesure leur diamètre.
Les résultats, en centimètres, sont :

6,71	6,43	7,74	6,51	6,79	7,32	6,65	6,78	7,22	6,42
5,82	6,18	6,45	5,95	7,86	7,11	7,02	6,80	6,93	6,90
6,16	6,79	6,82	6,13	5,93	6,95	6,63	6,44	6,18	7,03
7,48	7,21	6,05	6,66	6,77	6,08	6,75	6,74	6,25	7,24

Est-ce que le contrôleur peut dire, au risque 1 %, que le producteur a raison ?