Chapitre 17

Convergence de variables aléatoires.

Compétences attendues.

- ✓ Prouver une convergence en probabilité via la loi faible des grands nombres, l'inégalité de Markov ou de Bienaymé-Tchebychev, ou par un calcul direct.
- ✓ Étudier la convergence en loi d'une suite (X_n) de variables aléatoires.
- ✓ Connaitre et utiliser le Théorème Limite Central.
- ✓ Approximer une loi par une autre, en effectuant si nécessaire une correction de continuité.

Mathieu Mansuy - Professeur de Mathématiques en ECS2 au Lycée Louis Pergaud (Besançon)
`mansuy.mathieu@hotmail.fr`

1 Convergence en probabilités

Toutes les variables aléatoires considérées dans ce chapitre sont des variables discrètes ou à densité supposées définies sur un même espace probabilisé $(\Omega, \mathscr{A}, P)$.

1.1 Inégalités de concentration

Propriété 1 (*Inégalité de Markov* (1856 - 1922))

Si X est une variable aléatoire **positive** admettant une espérance, alors on a :

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

Preuve. Soit $a > 0$. Considérons la variable aléatoire $Y = a\mathbb{1}_{[X \geq a]}$. On a $Y \leq X$: en effet pour tout $\omega \in \Omega$,

- soit $X(\omega) \geq a$, et alors $Y(\omega) = a \leq X(\omega)$;
- soit $X(\omega) < a$, et dans ce cas $Y(\omega) = 0 \leq X(\omega)$ car X est positive.

Par croissance de l'espérance, on en déduit que :

$$E(X) \geq E(Y) = aE(\mathbb{1}_{[X \geq a]}) = aP([X \geq a])$$

puisque $\mathbb{1}_{[X \geq a]}$ suit une loi de Bernoulli de paramètre $p = P([X \geq a])$. D'où l'inégalité de Markov. $\square$

Corollaire 2

Soit X une variable aléatoire admettant un moment d'ordre r, alors on a :

$$\forall a > 0, \quad P(|X| \geq a) \leq \frac{E(|X|^r)}{a^r}.$$

Preuve. La variable $|X|^r$ est positive et admet une espérance par hypothèse. On en déduit que pour tout $a > 0$:

$$P(|X| \geq a) = P(|X|^r \geq a^r) \overset{\text{Inég. de Markov}}{\leq} \frac{E(|X|^r)}{a^r}.$$

$\square$

Remarque. On retiendra en particulier que si X admet un moment d'ordre 2 (ou de manière équivalente une variance), alors on a :

$$\forall a > 0, \quad P(|X| \geq a) \leq \frac{E(X^2)}{a^2}.$$

Propriété 3 (*Inégalité de Bienaymé* (1796 - 1878) - *Tchebychev* (1821 - 1894))

Soit X une variable aléatoire admettant une variance. Alors on a :

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

Preuve.

□

Remarque. L'inégalité de Bienaymé-Tchebychev permet d'obtenir une majoration de $P(|X - E(X)| \geq \varepsilon)$ sans connaitre la loi de X : il suffit simplement de connaître son espérance et sa variance. Cette majoration est cependant assez grossière, souvent bien supérieure à $P(|X - E(X)| \geq \varepsilon)$.

1.2 Convergence en probabilités

Définition.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires, et X une variable aléatoire. On dit que $(X_n)_{n\in\mathbb{N}}$ *converge en probabilité vers* X, et on note $X_n \xrightarrow{P} X$, lorsqu'on a :

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

Remarques.

- On peut remplacer l'inégalité large $P(|X_n - X| \geq \varepsilon)$ par une inégalité stricte $P(|X_n - X| > \varepsilon)$ puisque :
$$0 \leq P(|X_n - X| > \varepsilon) \leq P(|X_n - X| \geq \varepsilon) \leq P(|X_n - X| > 2\varepsilon).$$
- Par passage à l'événement contraire, cette définition est équivalente à :
$$\forall \varepsilon > 0, \ \lim_{n\to+\infty} P(|X_n - X| < \varepsilon) = 1 \ \text{ ou encore } \ \forall \varepsilon > 0, \ \lim_{n\to+\infty} P(|X_n - X| \leq \varepsilon) = 1.$$
Cette définition s'interprète donc ainsi : quelque soit la « marge d'erreur » $\varepsilon > 0$ que l'on se donne, la probabilité que les valeurs prises par X_n et X soient proches de moins de ε est très grande lorsque n est grand.
- La limite pour la convergence en probabilité est unique presque surement : si $X_n \xrightarrow{P} X$ et $X_n \xrightarrow{P} X'$ alors $X = X'$ p.s. (c'est à dire $P(X = X') = 1$).

Exemple. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes et suivant toutes la même loi uniforme sur $[0, 1]$. Posons pour tout $n \in \mathbb{N}^*$,

$$Y_n = \max(X_1, \ldots, X_n).$$

Montrer que Y_n converge en probabilité vers la variable aléatoire constante $Y = 1$.

Remarques.

- Pour montrer la convergence en probabilité, il suffit de le montrer pour tout ε « petit » (dans l'exemple précédent, pour $0 < \varepsilon < 1$). En effet si on a établi la convergence pour $\varepsilon > 0$, alors on a la convergence pour tout $\varepsilon' \geq \varepsilon$ puisque :
$$0 \leq P(|X_n - X| \geq \varepsilon') \leq P(|X_n - X| \geq \varepsilon).$$
- On a ici un exemple d'une suite de variables à densité convergeant en probabilité vers une variable discrète. Le contraire est aussi possible.

Propriété 4

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue. Si $X_n \xrightarrow{P} X$ alors $f(X_n) \xrightarrow{P} f(X)$.

Remarque. Le résultat ci-dessus est encore valable lorsque les variables X_n, $n \in \mathbb{N}$, et X sont à valeurs dans un intervalle I et f est une fonction continue sur I.

1.3 Loi faible des grands nombres

Théorème 5 (*Loi faible des grands nombres*)

Hypothèses :
- $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables **mutuellement indépendantes**.
- Les variables X_n admettent la même espérance m et la même variance.

Pour tout $n \in \mathbb{N}^*$, on pose $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$. Alors $\overline{X_n}$ converge en probabilité vers la variable aléatoire constante égale à m, soit encore :
$$\overline{X_n} \xrightarrow{P} m.$$

Preuve.

□

Remarque. On peut montrer encore mieux (loi forte des grands nombres - Hors Programme) : sous les mêmes hypothèses, il existe un évènement A de probabilité 1 tel que :

$$\forall \omega \in A, \quad \overline{X_n}(\omega) \to m.$$

En appliquant la loi faible des grands nombres au cas de variables de Bernoulli, on obtient le résultat suivant.

Corollaire 6 (*Théorème d'or de Bernoulli* (1654 - 1705))

Soit (X_n) une suite de variables de Bernoulli indépendantes de même paramètre p. Alors $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_i$ converge en probabilité vers la variable constante égale à p, soit :

$$\overline{X_n} \xrightarrow{P} p.$$

Conséquence.

On considère une succession d'expériences identiques indépendantes, au cours de chacune desquelles un événement A est susceptible de se produire avec probabilité $p = P(A)$. Pour tout $n \geq 1$, on définit la variable X_n par :

$$X_n = \begin{cases} 1 & \text{si l'événement } A \text{ se réalise au cours de la } n\text{-ième expérience,} \\ 0 & \text{sinon.} \end{cases}$$

Les variables aléatoires X_n sont indépendantes et suivent toutes la loi de Bernoulli de paramètre p.

Le théorème d'or de Bernoulli affirme que la fréquence $\overline{X_n}$ de réalisation de l'événement A au cours des n premières expériences converge en probabilité vers $p = P(A)$. Ainsi,

la fréquence statistique de réalisation d'un événement tend vers la probabilité de cet événement.

Ce résultat est fondamental : c'est lui qui justifie que le formalisme que nous employons correspond bien à l'intuition que l'on se fait d'une probabilité.

Remarque. On peut donner un ordre de grandeur de la vitesse de convergence de $\overline{X_n}$ vers p.

Cette inégalité permet de donner le nombre de répétitions de l'expérience à faire pour approcher p avec une marge d'erreur fixée.

Exemple. On dispose d'une pièce truquée et on souhaite évaluer la probabilité p qu'elle tombe sur pile. Si on note X_i la variable de Bernoulli qui vaut 1 si la pièce tombe sur pile au i-ème lancé, on sait d'après le théorème d'Or de Bernoulli que la fréquence $\overline{X_n} = \dfrac{X_1 + \cdots + X_n}{n}$ d'apparition du pile au cours de n lancers tend en probabilité p. On n'est cependant pas à l'abri d'une série de lancers où on aurait pas « eu de chance », c'est-à-dire telle que $\overline{X_n}$ prenne une valeur éloignée de la valeur de p. On va donc chercher le nombre n de lancers à effectuer pour approcher p à 0.1 près en quantifiant ce risque, c'est-à-dire en étant sûr de notre résultat avec une probabilité de 99%.

Exemple. On modélise cette situation sur `Scilab`.

```
p = rand()
function F=frequence_pile(n)
      F=mean(grand(1,n,"bin",1,p))
endfunction
```

En exécutant la fonction `frequence_pile` pour $n = 2500$, ce qui correspond à 2500 lancés de la pièce truquée, on obtient (par exemple) :

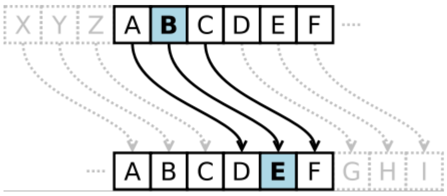
```
--> frequence_pile(2500)
ans =
      0.7656
```

Ainsi on peut affirmer que $0.6656 \leq p \leq 0.8656$ avec une probabilité de 99%. En appelant la variable p sur `Scilab`, on constate qu'en fait $p = 0.7560439$.

Application. Décryptage par analyse fréquentielle d'un chiffrement par substitution.
Le chiffrement par substitution consiste à fixer une permutation de l'alphabet, par exemple :

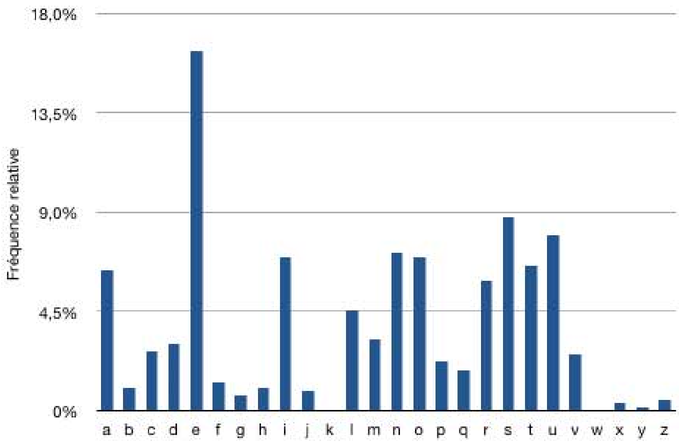
Texte non crypté	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Texte crypté	SMKRATNGQJUDZLPVYOCWIBXFEH

et à substituer dans un message chacune des lettres par la lettre correspondante. Ces techniques de chiffrement sont utilisées depuis bien longtemps (l'empereur César lui-même s'en servait).



Un exemple de chiffrement par substitution : le chiffrement par décalage.

Le défaut de ce genre de méthodes est qu'elles ne résistent pas au décryptage par une analyse fréquentielle. Son principe repose sur le fait que dans chaque langue, les lettres apparaissent avec une certaine fréquence.



Fréquences théoriques d'apparition des lettres en français.

D'après le théorème d'Or de Bernoulli, la fréquence d'apparition de chaque lettre dans le message crypté devrait[1] correspondre approximativement à sa fréquence théorique, ce qui permet de retrouver le message d'origine. Une condition est cependant nécessaire : que le message soit suffisamment long pour refléter la répartition théorique des fréquences des lettres.

1.4 Pour prouver une convergence en probabilité

La convergence en probabilité est souvent difficile à montrer. Si l'énoncé ne nous y aide pas, on pourra penser aux points suivants.

Méthode.

Pour montrer la convergence en probabilité d'une suite (X_n) de variables aléatoires, on pourra procéder comme suit :

- *commencer par s'assurer que ce n'est pas directement une conséquence de la loi faible des grands nombres ;*
- *essayer d'utiliser l'inégalité de Bienaymé-Tchebychev : si $E(X_n)$ existe et est constante égale à m, et si $V(X_n)$ existe et tend vers 0, alors on montre que $X_n \xrightarrow{P} m$;*
- *essayer d'utiliser l'inégalité de Markov : si $E(X_n)$ existe et converge vers une constante $m \in \mathbb{R}$, alors on peut essayer de montrer que $X_n \xrightarrow{P} m$;*
- *si aucune de ces méthodes ne fonctionne, il faudra calculer directement $P(|X_n - X| \geq \varepsilon)$, puis passer à la limite.*

[1] Là aussi, on peut « jouer de malchance » : imaginez que le message crypté soit le roman *La disparition* de Georges Perec, long de 300 pages et ne comportant pas une seule fois la lettre e.

2 Convergence en loi

2.1 Définition

Définition.

Soit (X_n) une suite de variables aléatoires et X une variable aléatoire toutes définies sur $(\Omega, \mathscr{A}, P)$. Notons F_{X_n} et F_X leurs fonctions de répartition respectives.

On dit que $(X_n)_{n\in\mathbb{N}}$ *converge en loi vers* X, et on note $X_n \xrightarrow{\mathscr{L}} X$, si pour tout $x \in \mathbb{R}$ **en lequel** F_X **est continue**, on a :

$$\lim_{n\to+\infty} F_{X_n}(x) = F_X(x).$$

Rappel. Si X est une variable à densité, F_X est continue sur $\mathbb{R}$, et on étudiera donc la limite de $F_{X_n}(x)$ pour tout $x \in \mathbb{R}$.

Remarques.

- Si $X_n \xrightarrow{\mathscr{L}} X$, cela signifie concrètement que pour n grand, la loi de X_n « ressemble beaucoup » à la loi de X.
- La convergence en loi ne dépend que des lois des variables aléatoires. En particulier, si $X_n \xrightarrow{\mathscr{L}} X$ et si Y suit la même loi que X, alors $X_n \xrightarrow{\mathscr{L}} Y$. Il n'y a donc pas unicité de la limite pour la convergence en loi.

Méthode.

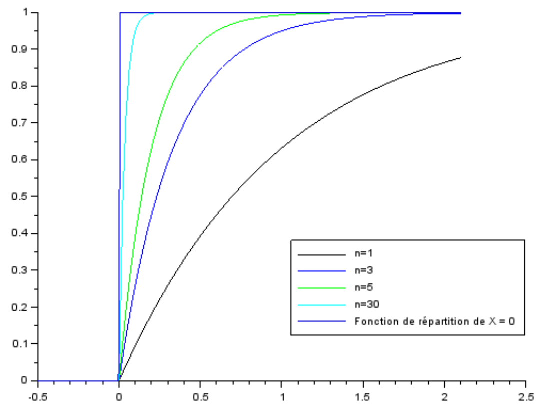
Pour étudier la convergence en loi d'une suite (X_n) de variables aléatoires, il faut :

- *déterminer les fonctions de répartitions F_{X_n} des X_n ;*
- *fixer $x \in \mathbb{R}$, et étudier la convergence de la suite de réels $(F_{X_n}(x))_{n\in\mathbb{N}}$. On note $F(x)$ sa limite.*
- *Vérifier si F est bien la fonction de répartition d'une variable aléatoire, c'est-à-dire :*
  - $\lim_{x\to-\infty} F(x) = 0$ *;*
  - $\lim_{x\to+\infty} F(x) = 1$ *;*
  - *F est croissante ;*
  - *F est continue à droite en tout point.*

  La valeur aux points de discontinuité de F pourra pour cela être rediscutée.

Si ces conditions sont bien vérifiées, alors X_n converge en loi vers toute variable aléatoire de fonction de répartition F.

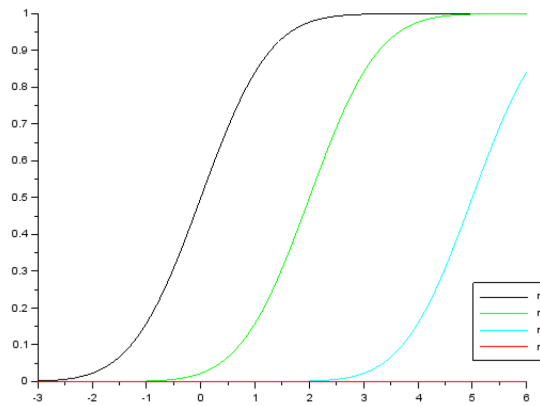
Exercice. Soit (X_n) une suite de variables aléatoires telles que $X_n \hookrightarrow \mathscr{E}(n)$. Étudier la convergence en loi de (X_n).



Courbes représentatives de F_{X_n} et de F_X

Remarque. L'exemple précédent montre qu'une suite de variables aléatoires à densité peut converger en loi vers une variable aléatoire discrète (l'inverse étant aussi possible).

Exercice. Soit (X_n) une suite de variables aléatoires telles que $X_n \hookrightarrow \mathscr{N}(n, 1)$. Étudier la convergence en loi de (X_n).



Courbes représentatives des F_{X_n}.

Exercice. Pour tout $n \in \mathbb{N}^*$, soit X_n une variable aléatoire de fonction de répartition F_{X_n} définie par :

$$\forall x \in \mathbb{R}, \quad F_{X_n}(x) = \frac{1}{(1 + e^{-x-\ln(n)})^n}.$$

Étudier la convergence en loi de la suite (X_n).

Propriété 7

Si $X_n \xrightarrow{\mathscr{L}} X$, alors pour tous $a < b$ points de continuité de F_X, on a :

$$\lim_{n\to+\infty} P(a < X_n \leq b) = P(a < X \leq b).$$

Preuve. On a :

$$P(a < X_n \leq b) = F_{X_n}(b) - F_{X_n}(a) \underset{n\to+\infty}{\longrightarrow} P(a < X \leq b).$$

□

Propriété 8 (Hors programme)

Soit (X_n) une suite de variables aléatoires, et X une variable aléatoires définies sur $(\Omega, \mathscr{A}, P)$.
Si $X_n \xrightarrow{P} X$, alors $X_n \xrightarrow{\mathscr{L}} X$. La réciproque est fausse.

Remarque. Même si ce résultat n'est pas au programme, il est intéressant de l'avoir en tête lorsqu'on cherche à montrer qu'une suite (X_n) converge en probabilité vers une variable X inconnue : pour « deviner » la loi de X, et donc avoir une idée de qui est X, on commence par déterminer la convergence en loi de (X_n), ce qui est plus simple.

2.2 Cas des variables aléatoires discrètes

Propriété 9

On suppose que les variables aléatoires $(X_n)_{n\in\mathbb{N}}$ et X sont **toutes à valeurs dans** $\mathbb{Z}$.
Alors $X_n \xrightarrow{\mathscr{L}} X$ si et seulement si on a :

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

Preuve. On fait la preuve dans le cas où $X_n(\Omega), X(\Omega) \subset \mathbb{N}$ pour tout $n \in \mathbb{N}$.

□

Méthode.

Pour étudier la convergence en loi d'une suite (X_n) de variables aléatoires discrètes ***toutes à valeurs dans*** $\mathbb{Z}$*, on procèdera comme suit :*

(1) déterminer pour tout $n \in \mathbb{N}$ et $k \in \mathbb{Z}$, $p_{n,k} = P(X_n = k)$;

(2) pour tout $k \in \mathbb{Z}$ ***fixé****, étudier la convergence* ***en*** *n de la suite de réels $(p_{n,k})$. On note p_k sa limite si elle existe ;*

(3) vérifier si $(p_k)_{k\in\mathbb{Z}}$ définie une loi de probabilité, c'est à dire :

- *$\forall k \in \mathbb{Z}$, $p_k \geq 0$;*
- *la famille $(p_k)_{k\in\mathbb{Z}}$ est sommable[a] et sa somme vaut 1.*

Si ces conditions sont bien vérifiées, alors (X_n) converge en loi vers toute variable aléatoire discrète X de loi $P(X = k) = p_k$ pour tout $k \in \mathbb{Z}$.

[a] Dans le cas de variables à valeurs dans $\mathbb{N}$, cela correspond à montrer la convergence (absolue) de la série $\sum_{k\geq 0} p_k$ et que sa somme vaut 1.

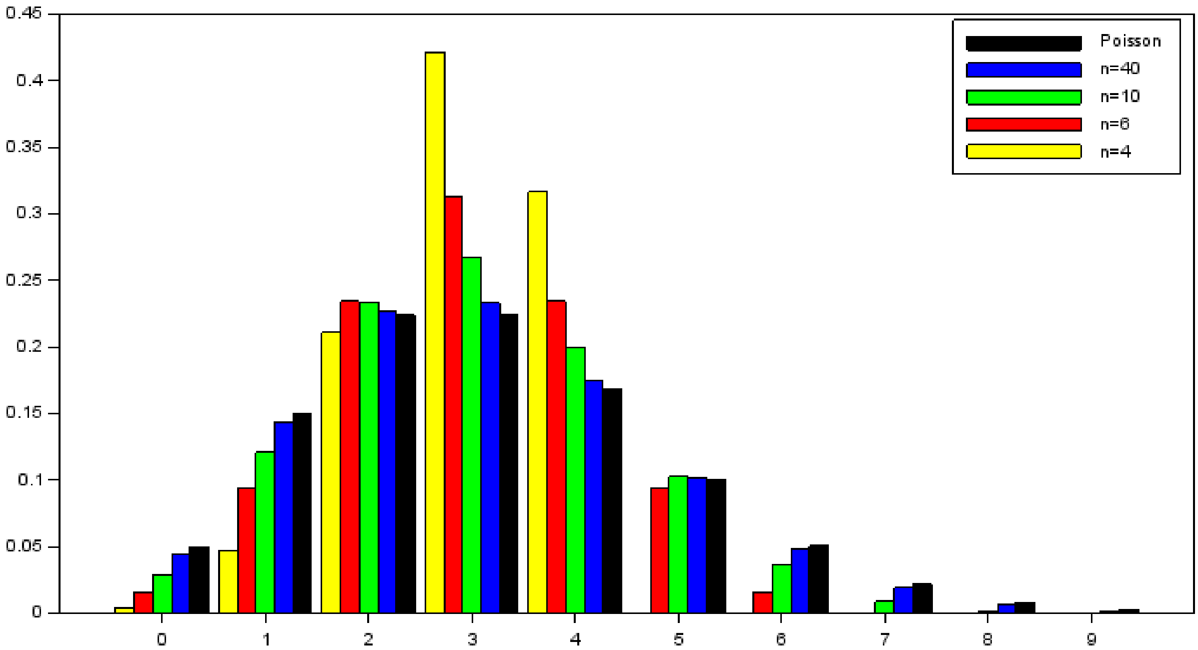
Propriété 10 (Convergence en loi de lois binomiales vers la loi de Poisson)

Soit $\lambda > 0$. On suppose que pour tout $n \in \mathbb{N}^*$, $X_n \hookrightarrow \mathscr{B}\left(n, \dfrac{\lambda}{n}\right)$.

Alors $X_n \xrightarrow{\mathscr{L}} X$ où X est une variable aléatoire suivant une loi de Poisson $\mathscr{P}(\lambda)$.

Preuve.

$\square$



Convergence en loi de $\mathscr{B}(n, 3/n)$ *vers* $\mathscr{P}(3)$.

2.3 Convergence en loi et opérations

Propriété 11

Soit f une fonction continue sur $\mathbb{R}$ à valeurs réelles. Si $X_n \xrightarrow{\mathscr{L}} X$ alors $f(X_n) \xrightarrow{\mathscr{L}} f(X)$.

Remarque. Lorsque la variable limite X est constante égale à c, le résultat ci-dessus est encore valable si la fonction f est seulement continue au point c.

Si $X_n \xrightarrow{\mathscr{L}} X$ et si $Y_n \xrightarrow{\mathscr{L}} Y$, on n'a pas nécessairement $X_n + Y_n \xrightarrow{\mathscr{L}} X + Y$. Cependant on peut montrer le résultat suivant :

Propriété 12 (*Théorème de Slutsky* (1880 - 1948))

On suppose que $X_n \xrightarrow{\mathscr{L}} X$ et $Y_n \xrightarrow{P} c$, où c est une constante. Alors on a :

- $X_n + Y_n \xrightarrow{\mathscr{L}} X + c$;
- $X_n Y_n \xrightarrow{\mathscr{L}} cX$.

Remarque. En réalité, Y_n converge en probabilité vers une constante c si et seulement si Y_n converge en loi vers c, mais ce résultat est hors programme.

2.4 Théorème central limite

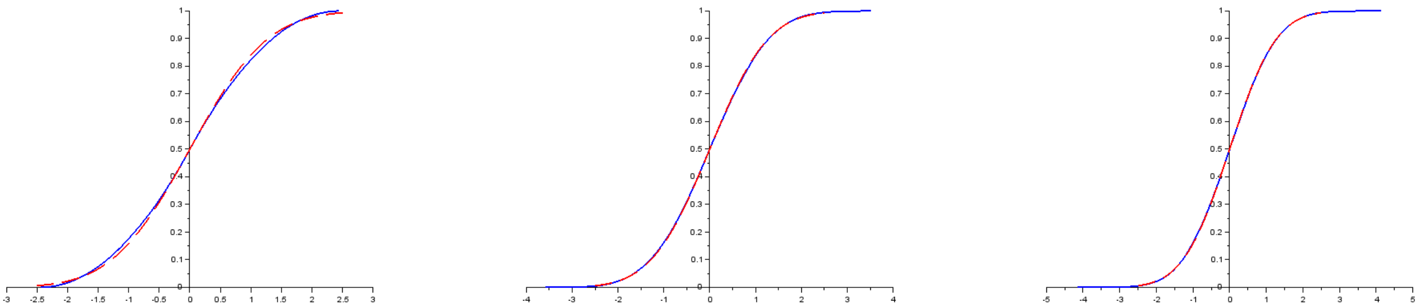
Vocabulaire. Une suite $(X_n)_{n \in \mathbb{N}}$ de variables est dite *indépendante identiquement distribuée* (en abrégé i.i.d.) lorsque les variables X_n sont mutuellement indépendantes et toutes de même loi.

Soit (X_n) une suite de variables i.i.d. d'espérance m et de variance $\sigma^2 \neq 0$. Posons $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$.

On a $E(\overline{X_n}) \overset{\text{lin. de E}}{=} \frac{1}{n}\sum_{k=1}^{n} E(X_k) = m$ et $V(\overline{X_n}) \overset{\text{ind. des } X_k}{=} \frac{1}{n^2}\sum_{k=1}^{n} V(X_k) = \frac{n\sigma^2}{n^2}$, de sorte que la variable

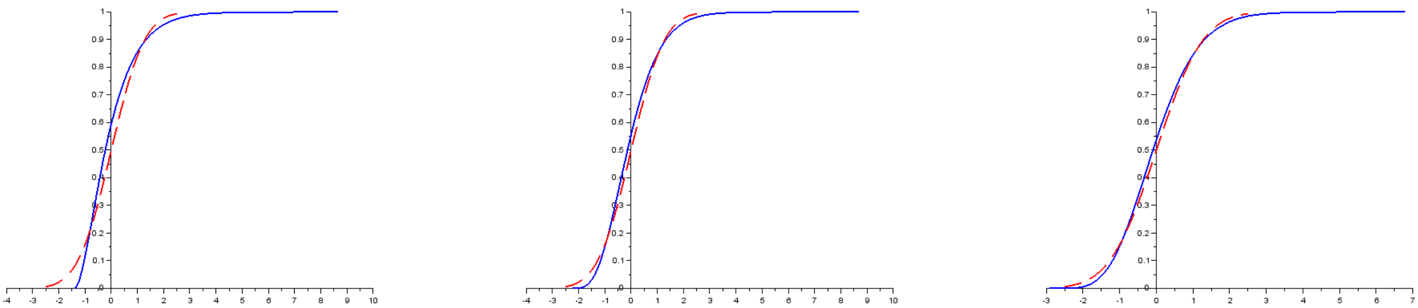
$$X_n^* = \frac{\overline{X_n} - E(\overline{X_n})}{\sqrt{V(\overline{X_n})}} = \frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\left(\frac{\overline{X_n} - m}{\sigma}\right)$$

est centrée réduite. Représentons la fonction de répartition de $\overline{X_n}^*$ pour différentes lois et différentes valeurs de $n \in \mathbb{N}^*$ en bleu, ainsi que la fonction de répartition de la loi $\mathscr{N}(0,1)$ en pointillés rouges.



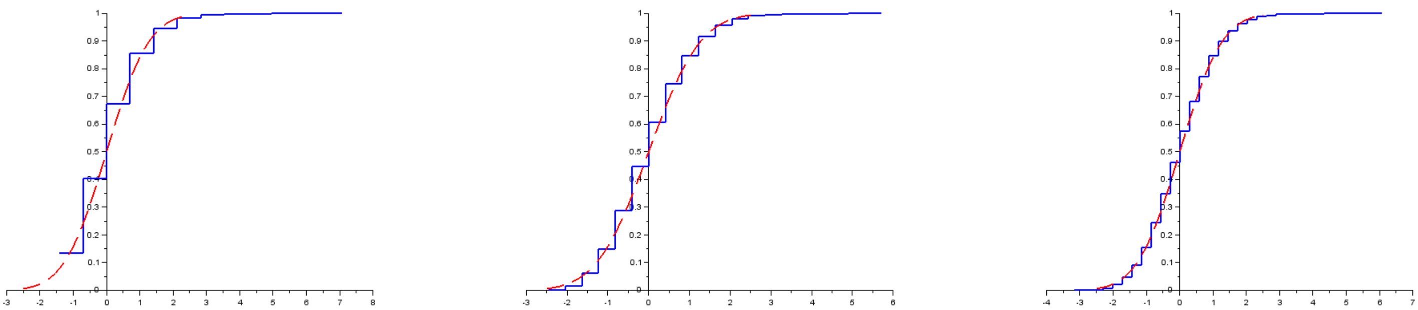
$n = 2$ $n = 6$ $n = 12$

Loi uniforme $\mathscr{U}([0,1])$.



$n = 2$ $n = 6$ $n = 12$

Loi exponentielle $\mathscr{E}(1)$.



$n = 2$ $n = 6$ $n = 12$

Loi de Poisson $\mathscr{P}(1)$.

On constate graphiquement la convergence de la fonction de répartition de $\overline{X_n}^*$ vers celle de la loi $\mathscr{N}(0,1)$, et donc la convergence en loi de $\overline{X_n}^*$ vers une variable suivant une loi $\mathscr{N}(0,1)$, et ceci indépendamment de la loi suivie. On peut en fait démontrer le résultat suivant.

Théorème 13 (*Théorème central limite*)

Hypothèses :
- Les variables aléatoires $(X_n)_{n\in\mathbb{N}^*}$ sont i.i.d.
- Elles admettent une espérance m et une variance σ^2 non nulle.

Pour tout $n \in \mathbb{N}^*$, on note $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$, et $\overline{X_n}^*$ la variable centrée réduite associée :

$$\overline{X_n}^* = \frac{\overline{X_n} - E(\overline{X_n})}{\sqrt{V(\overline{X_n})}} = \sqrt{n}\left(\frac{\overline{X_n} - m}{\sigma}\right).$$

Alors $\overline{X_n}^*$ converge en loi vers une variable suivant une loi normale centrée réduite $\mathcal{N}(0,1)$.

Remarques.

- Pour tout (a,b) tel que $-\infty \leq a \leq b \leq +\infty$, on a ainsi :
$$\lim_{n\to+\infty} P(a < \overline{X_n}^* \leq b) = P(a < X \leq b) = \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t = \Phi(b) - \Phi(a)$$
où $\Phi(b) = 1$ si $b = +\infty$ et $\Phi(a) = 0$ si $a = -\infty$.
- Soit $S_n = \displaystyle\sum_{k=1}^{n} X_i$, S_n^* la variable centrée réduite associée. Alors $E(S_n) \overset{\text{lin. de } E}{=} \displaystyle\sum_{k=1}^{n} E(X_k) = nm$ et $V(S_n) \overset{\text{ind. des } X_k}{=} \displaystyle\sum_{k=1}^{n} V(X_k) = n\sigma^2$ et on a :
$$S_n^* = \frac{S_n - nm}{\sqrt{n}\sigma} = \frac{n\overline{X_n} - nm}{\sqrt{n}\sigma} = \overline{X_n}^*.$$
Par le théorème limite central, S_n^* converge aussi en loi vers une variable suivant une loi $\mathcal{N}(0,1)$.

Le saviez vous ?

Abraham de Moivre, huguenot réfugié en Angleterre à la suite de la révocation de l'Édit de Nantes, énonça le premier en 1733 le théorème central limite dans le cas particulier de variables suivant la loi de Bernoulli. Pierre-Simon de Laplace fournit la première démonstration de ce résultat en 1809, et ce, dans un cas plus général. Cependant sa formulation et sa démonstration dans un cadre rigoureux sont l'oeuvre de Pafnouti Tchebychev en 1867. Ses élèves Andreï Markov et Alexandre Liapounov affineront l'étude de leur maître et en proposeront une généralisation dans le cas où les variables ne sont pas toutes de même loi, à la condition qu'aucune des variables n'exerce une influence significativement plus importante que les autres.

Le nom du théorème fait référence à un document scientifique écrit par George Pólya en 1920, intitulé *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem*, signifiant *Sur le théorème [ayant rapport à la notion de] limite central du calcul probabiliste et le problème des moments*. Celui-ci considérait en effet ce théorème comme un élément fondamental de la théorie des probabilités.

Remarque. C'est ce résultat qui nous a permis en informatique de simuler une loi normale centrée réduite à partir d'une loi uniforme $\mathscr{U}([0,1])$. Plus particulièrement, nous avions utilisé $\overline{X_{12}}^* = \left(\displaystyle\sum_{k=1}^{12} X_k\right) - 6$ avec $X_k \hookrightarrow \mathscr{U}([0,1])$ pour simuler une loi $\mathscr{N}(0,1)$. On constate graphiquement que cette approximation semble pertinente, puisque les graphes des fonctions de répartition de $\overline{X_{12}}^*$ et de la loi $\mathscr{N}(0,1)$ semblent confondus.

Interprétation du Théorème central limite.
D'après la loi faible des grands nombres, on a $X_n \xrightarrow{P} m$. Le théorème limite central précise les fluctuations de $\overline{X_n}$ autour de m, puisque lorsque n est assez grand, on a :

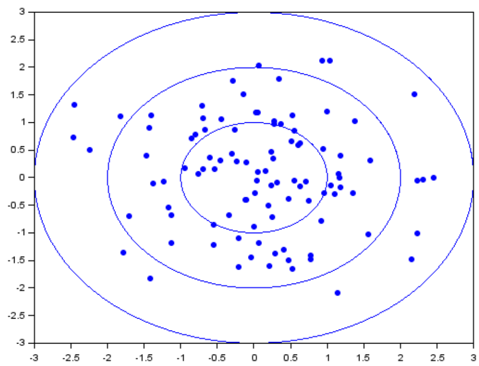
$$\overline{X_n} - m \approx X \quad \text{avec} \quad X \hookrightarrow \mathscr{N}(0, \frac{\sigma^2}{n}).$$

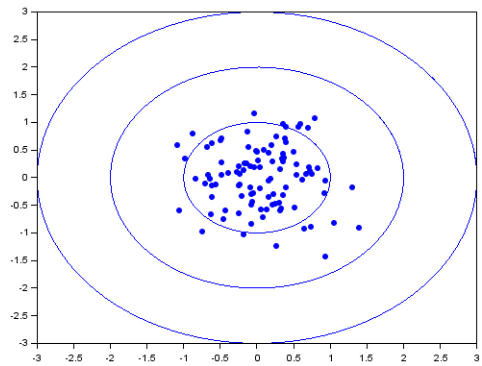
En particulier, on a :

$$X_1 + \cdots + X_n = n\overline{X_n} + nm \approx S \quad \text{avec} \quad S \hookrightarrow \mathscr{N}(nm, n\sigma^2) \quad \text{pour } n \text{ grand.}$$

Ainsi si on somme des variables aléatoires i.i.d. un grand nombre de fois, la variable aléatoire obtenue suit « approximativement » une loi normale. C'est ce résultat[2] qui justifie l'omniprésence de cette loi dans la nature, de nombreux phénomènes étant dus à l'addition d'un grand nombre de petites perturbations aléatoires.

Exemple. Prenons le cas d'un tireur à la carabine. Si (X_k, Y_k) sont les coordonnées du k-ième tir sur la cible, alors X_k et Y_k sont des sommes de petites variables indépendantes (sommes de perturbations telles que erreur de visée, tremblement, défaut de concentration, recul, perturbation par un élément extérieur comme un contre-jour ou du vent, ...). Le résultat précédent nous invite à modéliser le point d'impact (X_k, Y_k) à l'aide d'un couple de variables normales centrées.





100 *réalisations de* (X_k, Y_k) *centrées avec* $\sigma^2 = 1$. 100 *réalisations de* (X_k, Y_k) *centrées avec* $\sigma^2 = 0.5$.

L'indépendance des tirs s'exprime par l'indépendance mutuelle des couples (X_k, Y_k), et la variance de la loi normale mesure l'adresse du tireur.

3 Approximations

Objectif. Rappelons que si $X_n \xrightarrow{\mathscr{L}} X$, alors X_n suit approximativement la loi de X lorsque n est « grand ». En pratique pour simplifier un problème ou un calcul, il peut être avantageux de considérer la loi de X_n plutôt que celle de X. On donne dans la suite des conditions empiriques sur les paramètres des lois, notamment sur n, pour que cette approximation soit satisfaisante. Ces conditions ne sont pas à apprendre par cœur, le programme officiel précisant :

Les conditions sous lesquelles l'approximation est valable doivent systématiquement être reprécisées.

[2] ou plus précisément la *loi des erreurs*, qui généralise ce résultat à une somme de petites variables aléatoires dont aucune n'est prépondérante.

3.1 Approximation de la loi binomiale

Rappelons que si pour tout $n \in \mathbb{N}^*$, $X_n \hookrightarrow \mathscr{B}(n, \frac{\lambda}{n})$, alors $X_n \xrightarrow{\mathscr{L}} X$ où $X \hookrightarrow \mathscr{P}(\lambda)$.

Propriété 14

Lorsque $n \geq 30$, $p \leq 0.1$ et $np \leq 15$, on approche la loi $\mathscr{B}(n, p)$ par la loi $\mathscr{P}(np)$.

Remarque. Retenir pour cette approximation qu'il faut « n grand » et « p petit ». Remarquons pour mémoire que ces deux lois ont même espérance.

Exemple. Prenons l'exemple d'une barre d'uranium. Elle contient une très grande quantité n d'atomes radioactifs. Chacun à une probabilité p très faible de se désintégrer durant un laps de temps T. On observe de plus qu'en moyenne il y a λ désintégrations durant ce laps de temps T :

$$\lambda = n \times p.$$

En supposant l'indépendance des désintégrations atomiques, il serait rigoureux de modéliser le nombre X de désintégration par une loi binomiale $\mathscr{B}(n, p)$. En pratique, les calculs numériques avec cette loi sont rendu impossible du fait que n est très grand et inconnu, p très petit. On approxime donc la loi de X par une loi de Poisson de paramètre λ, approximation satisfaisante d'après la propriété précédente.

Soit (X_n) une suite de variables aléatoires i.i.d. suivant toutes une loi de Bernoulli $\mathscr{B}(p)$ avec $0 < p < 1$. Posons pour tout $n \in \mathbb{N}^*$, $S_n = \sum_{i=1}^{n} X_i$. Alors $S_n \hookrightarrow \mathscr{B}(n, p)$ et :

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$$

converge en loi vers (une variable aléatoire suivant une loi) $\mathscr{N}(0, 1)$ d'après le Théorème limite central. Ainsi pour n suffisamment grand,

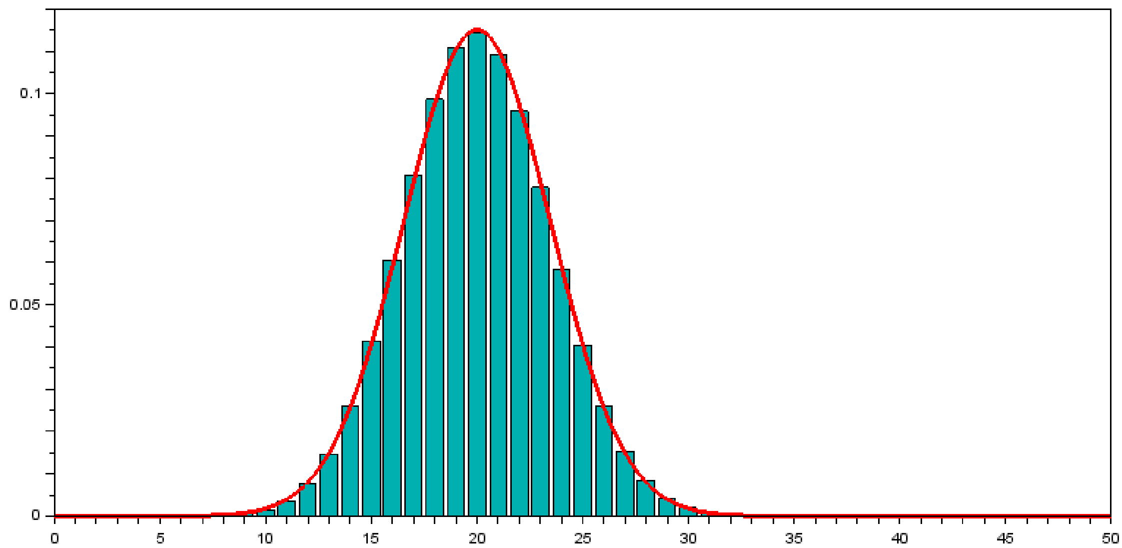
$$S_n = \sqrt{np(1-p)} S_n^* + np$$

suit approximativement une loi $\mathscr{N}(np, np(1-p))$.

Propriété 15

Lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, on approche la loi $\mathscr{B}(n, p)$ par la loi $\mathscr{N}(np, np(1-p))$.

Remarque. Retenir qu'il faut avoir « n grand » et « p ni trop petit, ni trop grand ». Remarquons pour mémoire que ces deux lois ont même espérance et même variance.



Approximation de la loi $\mathscr{B}(50, 0.4)$ *par la loi* $\mathscr{N}(20, 12)$.

Correction de continuité. On approche dans le résultat précédent une variable X suivant une loi discrète $\mathscr{B}(n, p)$ par une variable Y suivant une loi continue $\mathscr{N}(np, np(1-p))$. Mais cette approximation pose une question : comment calculer les $P(X = k)$? On ne peut pas écrire :

$$P(X = k) \approx P(Y = k)$$

puisque $P(Y = k) = 0$. La règle communément acceptée est d'utiliser que :

$$P(X = k) = P(k - 0.5 < X \leq k + 0.5)$$

et donc d'approcher $P(X = k)$ par $P(k - 0.5 \leq Y \leq k + 0.5)$. De plus, afin de garantir que la somme des probabilités reste égale à 1, on remplacera $P(X = 0)$ par $P(Y \leq 0.5)$ et $P(X = n)$ par $P(n - 0.5 \leq Y)$. On parle de *correction de continuité*.

Exercice. Soit $X \hookrightarrow \mathscr{B}(100, 0.4)$. Calculer approximativement $P(X = 35)$.

Un calcul direct donne $P(X = 35) = 0.0491$.

3.2 Approximation de la loi de Poisson

Soit (X_n) une suite de variables aléatoires i.i.d. suivant toutes une loi de Poisson $\mathscr{P}(\lambda)$. Posons pour tout $n \in \mathbb{N}^*$, $S_n = \sum_{i=1}^{n} X_i$. Alors $S_n \hookrightarrow \mathscr{P}(n\lambda)$ et :

$$S_n^* = \frac{S_n - n\lambda}{\sqrt{n\lambda}}$$

converge en loi vers $\mathscr{N}(0,1)$. Pour n « suffisamment grand »,

$$S_n = \sqrt{n\lambda} S_n^* + n\lambda$$

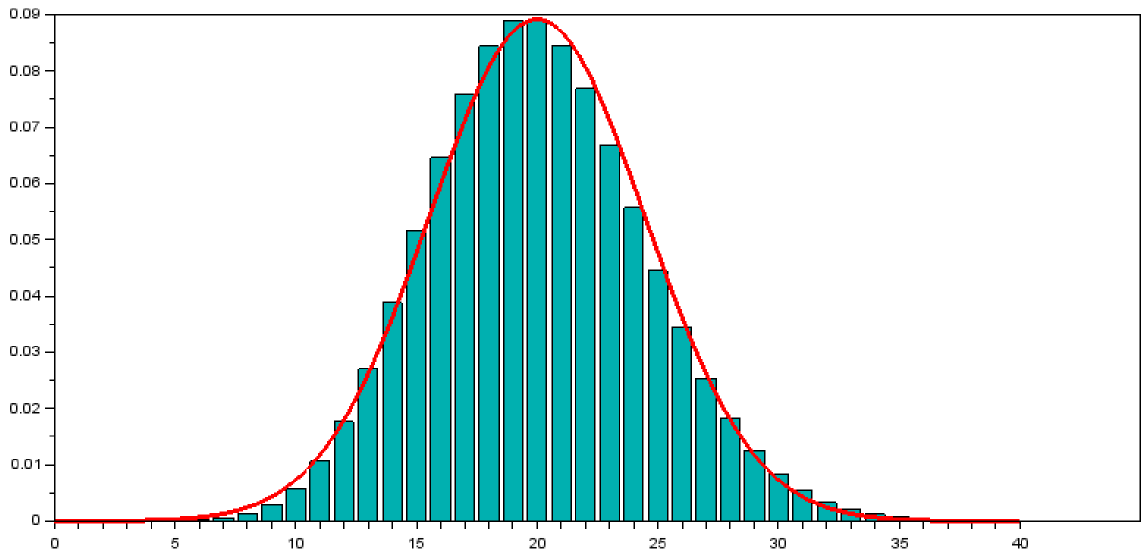
suit donc approximativement une loi $\mathscr{N}(n\lambda, n\lambda)$.

Propriété 16

Pour $\lambda \geq 15$, on approche la loi $\mathscr{P}(\lambda)$ par la loi $\mathscr{N}(\lambda, \lambda)$.

Remarques.

- Retenir qu'il faut avoir « λ assez grand ». Remarquons pour mémoire que ces deux lois ont même espérance et même variance.
- On approche ici aussi une variable aléatoire discrète par une variable aléatoire continue. On utilisera donc une correction de continuité.



Approximation de la loi $\mathscr{P}(20)$ par la loi $\mathscr{N}(20, 20)$.