# Chapitre 7

4 théorèmes à connaître.

# Théorèmes limites

## 7.1 Introduction

Les théorèmes limites importants en actuariat peuvent être classés en deux catégories:

1. Ceux reliés à la "loi des grands nombres": ils concernent des conditions pour lesquelles la moyenne d'une suite de variables aléatoires converge vers la moyenne espérée.

2. Ceux reliés au "théorème central limite": ils concernent des conditions pour lesquelles la somme d'un grand nombre de variables aléatoires a une distribution qui est approximativement normale.

⇨ En actuariat, la loi des grands nombres nous permet d'expliquer l'impact du regroupement d'assurés au sein d'un portefeuille. Pour ce qui est du théorème central limite, on peut l'utiliser pour une première approximation de la distribution du montant total des sinistres d'un portefeuille, c'est-à-dire la distribution de la somme de tous les sinistres individuels reliés à un portefeuille.

## 7.2 Loi des grands nombres

Deux résultats sont nécessaires pour démontrer la loi des grands nombres: l'inégalité de Markov et l'inégalité de Tchebychev.

**Proposition 7.1** ***(Inégalité de Markov)*** *Soit $X$ une variable aléatoire non négative. Alors pour $\forall a > 0$, on a*

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

⇒ Quand on connaît juste $E[X]$

**Preuve.** Soit $X$ une variable aléatoire continue et non négative.

$$\begin{aligned} E[X] &= \int_0^\infty x f_X(x)\mathrm{d}x \\ &= \int_0^a x f_X(x)\mathrm{d}x + \int_a^\infty x f_X(x)\mathrm{d}x \\ &\geq \int_a^\infty x f_X(x)\mathrm{d}x \\ &\geq \int_a^\infty a f_X(x)\mathrm{d}x. \end{aligned}$$

$$\rightarrow \quad E[X] \geq \int_a^\infty a f_X(x)\mathrm{d}x = a\left(\Pr(X > a)\right)$$

$$\Pr(X > a) \leq \frac{E[X]}{a}.$$

Soit $X$ une variable aléatoire discrète et non négative.

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} x \Pr(X=x) \\ &= \sum_{x=0}^{a-1} x \Pr(X=x) + \sum_{x=a}^{\infty} x \Pr(X=x) \\ &\geq \sum_{x=a}^{\infty} x \Pr(X=x) \\ &\geq \sum_{x=a}^{\infty} a \Pr(X=x) \end{aligned}$$

$$E[X] \geq \sum_{x=a}^{\infty} a \Pr(X=x) = a\left(\Pr(X \geq a)\right)$$

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

■

**Proposition 7.2** ***(Inégalité de Tchebychev)*** *Soit $X$ une variable aléatoire de moyenne finie $\mu$ et variance finie $\sigma^2$. Alors pour toute valeur de $k > 0$, on a*

finie = qui existe

$$\Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

**Preuve.** Soit $Y = \left(\frac{X-\mu}{\sigma}\right)^2$ une variable aléatoire non négative. À l'aide de l'inégalité de Markov avec $a = k^2$, on a

$$\Pr(Y \geq k^2) \leq \frac{E[Y]}{k^2}$$

et donc

$$\Pr\left(\left(\frac{X-\mu}{\sigma}\right)^2 \geq k^2\right) \leq \frac{E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]}{k^2} = \frac{1}{k^2}$$

Chance que ça dépasse un certain nombre d'écart-type par rapport à la moyenne,

Ainsi, on a

$$\begin{aligned} \Pr\left(\left|\frac{X-\mu}{\sigma}\right| \geq k\right) &\leq \frac{1}{k^2} \\ \Rightarrow \Pr(|X-\mu| \geq k\sigma) &\leq \frac{1}{k^2}. \end{aligned}$$

Par l'interprétation

■

**Remarque 7.3** *L'inégalité de Tchebychev peut également être écrite comme suit*

$$\Pr(|X-\mu| \geq k^*) \leq \frac{\sigma^2}{k^{*2}}$$

*car en posant $k^* = k\sigma$ dans la première forme choisie pour l'inégalité, on a*

$$\begin{aligned} \Pr(|X-\mu| &\geq k\sigma) \leq \frac{1}{k^2} \\ \Pr(|X-\mu| &\geq k^*) \leq \frac{1}{\left(\frac{k^*}{\sigma}\right)^2} \\ \Rightarrow \Pr(|X-\mu| \geq k^*) &\leq \frac{\sigma^2}{(k^*)^2}. \end{aligned}$$

Par les calculs/pitonnage

*Plug N' play pour ce chapitre!



**Exemple 7.4** *Soit $X$ une variable aléatoire non négative représentant le montant d'un sinistre telle que $E[X] = 250$. (a) Trouver une borne supérieure à la probabilité que le montant d'un sinistre soit supérieure ou égal à 400. (b) Sachant que $E[X] = 250$ et $Var(X) = 2500$, trouver une borne inférieure à $\Pr(150 \leq X \leq 350)$.*

**Solution.** (a) Par l'inégalité de Markov, on a → car on a juste $E[X]$

$$\Pr(X \geq 400) \leq \frac{250}{400} = 0.625$$

(b) Par l'inégalité de Tchebychev, on a

$$\begin{aligned}
\Pr(150 \leq X \leq 350) &= \Pr(|X - 250| \leq 100) \\
&= 1 - \Pr(|X - 250| \geq 100) \\
&\geq 1 - \frac{Var(X)}{100^2} \\
&= 1 - \frac{2500}{10000} \\
&= 0.75
\end{aligned}$$

De façon similaire, on a

$$\Pr(|X - 250| \geq (k)(50)) \leq \frac{1}{k^2}.$$

Ainsi pour $k = 2$, on a

$$\Pr(X \geq 350, X \leq 150) \leq \frac{1}{2^2} = 0.25$$

d'où

$$\Pr(150 \leq X \leq 350) \geq 1 - 0.25 = 0.75.$$

■

**Exemple 7.5** *Soit les variables aléatoires $X$ et $Y$ telles que $X \sim U(0, 10)$ et $Y \sim Exp(0.01)$. Comparer les valeurs exactes des probabilités avec la borne obtenue à l'aide de l'inégalité de Tchebychev. (a) $\Pr(|X - 5| \geq 4)$. (b) $\Pr(|Y - 100| \geq 150)$.*

**Solution.** (a)

$$\begin{aligned}
\Pr(|X - 5| \geq 4) &= 1 - \Pr(|X - 5| < 4) \\
&= 1 - \Pr(1 < X < 9) \\
&= 1 - (F_X(9) - F_X(1)) \\
&= 1 - \left(\frac{9}{10} - \frac{1}{10}\right) \\
&= 1 - \frac{4}{5} \\
&= \frac{1}{5} = 0.2.
\end{aligned}$$

ou de façon similaire

$$\Pr(|X - 5| \geq 4) = \Pr(X < 1) + \Pr(X > 9) = 0.2.$$

Inégalité de Tchebychev: on a $X \sim U(0, 10)$, d'où $E[X] = \frac{0+10}{2} = 5, Var(X) = \frac{(10-0)^2}{12} = \frac{25}{3}$

$$\Pr(|X - 5| \geq 4) \leq \frac{Var(X)}{16} = \frac{\frac{25}{3}}{16} = \frac{25}{48} = 0.5208.$$

ou de façon similaire

$$\Pr\left(|X - 5| \geq \left(\frac{4}{5}\sqrt{3}\right)\frac{5}{\sqrt{3}}\right) \leq \frac{1}{\left(\frac{4}{5}\sqrt{3}\right)^2} = 0.5208.$$

La borne supérieure de Tchebychev est plutôt loin de la valeur exacte. (b)

$$\begin{aligned}\Pr(|Y-100| \geq 150) &= \Pr(Y < -50) + \Pr(Y > 250) \\ &= 0 + 1 - \left(1 - e^{-(0.01)(250)}\right) \\ &= 0.0821\end{aligned}$$

Inégalité de Tchebychev, on a $Y \sim Exp(0.01)$, d'où $E[Y] = \frac{1}{0.01} = 100, Var(Y) = \frac{1}{(0.01)^2} = 10000$

$$\Pr(|X-100| \geq 150) \leq \frac{10000}{(150)^2} = 0.4444$$

la borne n'est pas très précise... mais ces deux inégalités vont nous servir pour la loi des grands nombres.

ou de façon similaire

$$\Pr(|Y-100| \geq (1.5)(100)) \leq \frac{1}{(1.5)^2} = 0.4444.$$

La borne supérieure de Tchebychev est plutôt loin de la valeur exacte. ■

**Remarque 7.6** *Les inégalités de Markov et Tchebychev permettent de trouver des bornes pour des probabilités lorsque seulement la moyenne ou la moyenne et la variance de la distribution sont connues.*

**Remarque 7.7** *Étant donné que l'inégalité de Tchebychev est valide pour tout choix de distribution de $X$, la borne obtenue sur la probabilité n'est pas très "précise". Cette inégalité est toutefois un outil théorique utilisé dans plusieurs résultats importants comme la loi des grands nombres.*

IMPORTANT, sinon on ne peut pas utiliser.

**Proposition 7.8** ***(Loi des grands nombres)*** *Soit $X_1, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées telles que $E[X_i] = \mu$ et $Var[X_i] = \sigma^2 > 0$ $(i = 1, ..., n)$. Alors, pour $\forall \varepsilon > 0$, on a*

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + ... + X_n}{n} - \mu\right| \geq \varepsilon\right) \to 0.$$

**Preuve.** Posons $S_n = X_1 + ... + X_n$. On a donc

$$\begin{aligned}\Pr\left(\left|\frac{X_1 + ... + X_n}{n} - \mu\right| \geq \varepsilon\right) &= \Pr\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \\ &= \Pr\left(\left|\frac{S_n - n\mu}{n}\right| \geq \varepsilon\right) \\ &= \Pr(|S_n - n\mu| \geq n\varepsilon).\end{aligned}$$

Selon l'inégalité de Tchebychev, on a

car les v.a. sont indép. et identiquement distribuées.

$$\begin{aligned}\Pr(|X-\mu| \geq k^*) &\leq \frac{\sigma^2}{k^{*2}} \\ \Pr(|S_n - n\mu| \geq n\varepsilon) &\leq \frac{Var(S_n)}{n^2\varepsilon^2} = \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.\end{aligned}$$

De façon similaire, on a

$$\Pr\left(|S_n - E[S_n]| \geq k\sqrt{Var[S_n]}\right) \leq \frac{1}{k^2}$$

avec

$$\begin{aligned}E[S_n] &= E[X_1 + ... + X_n] = n\mu \\ Var(S_n) &= Var(X_1 + ... + X_n) = n\sigma^2.\end{aligned}$$

On a donc

$$\begin{aligned} k\sqrt{Var\left[S_n\right]} &= n\varepsilon \\ k\sqrt{n\sigma^2} &= n\varepsilon \\ k &= \frac{n\varepsilon}{\sqrt{n\sigma^2}} \\ &= \frac{\sqrt{n}\varepsilon}{\sigma} \end{aligned}$$

Ainsi,

$$\Pr\left(\left|S_n - E\left[S_n\right]\right| \geq k\sqrt{Var\left[S_n\right]}\right) \leq \frac{1}{k^2}$$

peut s'écrire

$$\Pr\left(\left|S_n - E\left[S_n\right]\right| \geq \frac{\sqrt{n}\varepsilon}{\sigma}\sqrt{n\sigma^2}\right) \leq \frac{1}{\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)^2} = \frac{\sigma^2}{n\varepsilon^2}$$

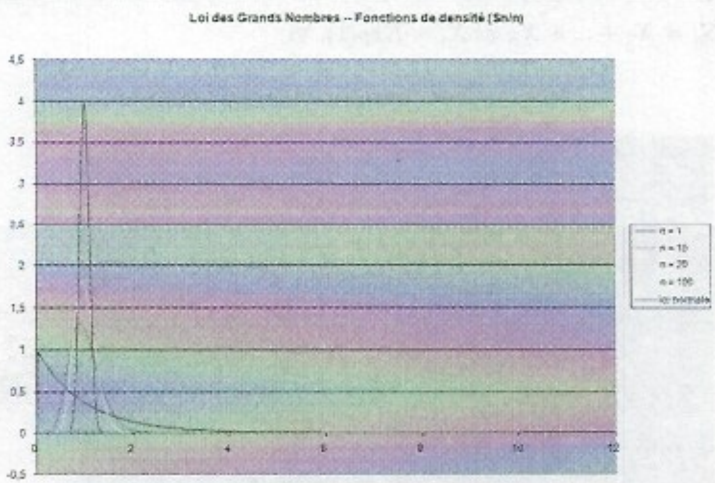
Donc pour tout $\varepsilon > 0$ fixé, on a

$$\lim_{n\to\infty} \Pr\left(\left|S_n - n\mu\right| \geq n\varepsilon\right) \leq \left(\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2}\right) \to 0.$$

■

**Remarque 7.9** *La loi des grands nombres stipule que la probabilité que la moyenne échantillonnale diffère de la vraie moyenne de plus d'une très petite quantité tend vers 0 lorsque le nombre d'observation dans l'échantillon augmente. La moyenne échantillonale varie donc de très peu de la vraie moyenne lorsque est très grand.*

**Exemple 7.10** *Soit la variable aléatoire $S_n = X_1 + ... + X_n$ où $X_i \sim Exp(1), \forall i$. Comparaison des graphiques des fonctions de densité de la variable aléatoire $\frac{S_n}{n}$ pour $n = 1, 10, 20, 100$.*




→ plus on prend un grand nombre de valeur, plus on tend vers une valeur.

## 7.3 Théorème central limite

**Proposition 7.11** *Soit $X_1, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées telles que $E\left[X_i\right] = \mu$ et $Var\left[X_i\right] = \sigma^2 > 0$ $(i = 1, ..., n)$. Soit la variable aléatoire $S_n = X_1 + ... + X_n$. Alors,*

$$\lim_{n\to\infty} \Pr\left(\frac{S_n - E\left[S_n\right]}{\sqrt{Var\left(S_n\right)}} \leq z\right) = \Phi\left(z\right).$$

Examen : on utilise le TCL seulement lorsque c'est demandé.



**Remarque 7.12** *Le théorème central limite stipule que la somme d'un grand nombre de variables aléatoires indépendantes et identiquement distribuées a approximativement une distribution normale.*

**Remarque 7.13** *Le théorème central limite est généralement utilisé lorsque la distribution de $X_1 + ... + X_n$ tend à être symétrique, c'est-à-dire lorsque le coefficient d'asymétrie est très petit. Sinon, l'approximation effectuée en utilisant la loi normale standard sera inadéquate.*

**Exemple 7.14** *Soit $X_i$ une variable aléatoire représentant le montant total des sinistres sur une année pour l'assuré i. Soit $S_n$ une variable aléatoire représentant le montant total des sinistres pour un portefeuille de n assurés. Si l'on suppose $X_i \sim Exp(0.01)$ pour $i = 1, ..., 50$, utiliser le théorème central limite pour approximer la probabilité que le montant total des sinistres pour l'ensemble d'un portefeuille composé de 50 assurés soit supérieur à 6000.*

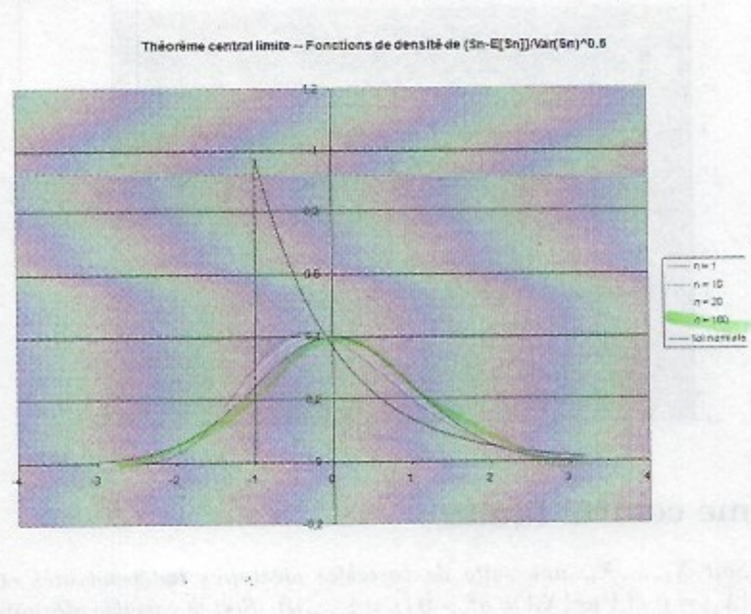
**Preuve.** $E[X_i] = \frac{1}{0.01} = 100$ et $Var(X_i) = \frac{1}{(0.01)^2} = 10000$ (et donc $\sigma_{X_i} = 100$). De plus,

$$\begin{aligned}
\Pr(S_{50} > 6000) &= 1 - \Pr\left(\frac{S_{50} - E[S_{50}]}{\sqrt{Var(S_{50})}} \leq \frac{6000 - E[S_{50}]}{\sqrt{Var(S_{50})}}\right) \\
&= 1 - \Pr\left(\frac{S_{50} - (50)(100)}{\sqrt{(50)(10000)}} \leq \frac{6000 - (50)(100)}{\sqrt{(50)(10000)}}\right) \\
&\approx 1 - \Pr(Z \leq 1.41) \\
&\approx 1 - 0.9207 \\
&= 0.0793
\end{aligned}$$

c'est seulement ici qu'il y a un approx. N(0,1) ATTENTION aux signes

La valeur exacte de $\Pr(S_{50} > 6000)$ est 0.0844. ■

**Exemple 7.15** *Comparaison des graphiques des fonctions de densité de la variables aléatoire $\frac{S_n - E[S_n]}{\sqrt{Var[S_n]}}$, $n = 1, 10, 20, 100$, où $S_n = X_1 + ... + X_n$ et $X_i \sim Exp(1)$, $\forall i$.*




Plus on augmente le nbre de donnée plus ça tend vers une loi normale.