Agrégation Interne de Mathématiques
Thierry Champion
séance du 2 décembre 2016

Théorèmes limites

Notations - Introduction

Dans tout ce problème, $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé. Toutes les variables aléatoires considérées ici sont des variables aléatoires définies sur Ω. Si Y est une telle variable aléatoire, on note respectivement $\mathbb{E}(Y)$, $\mathbb{V}(Y)$ et $\sigma(Y)$ l'espérance, la variance et l'écart-type de Y lorsque ceux-ci existent.

On note $(X_n)_{n\geq 1}$ une suite de variables aléatoires *indépendantes deux à deux* et de même loi, et on pose $S_n = X_1 + \ldots + X_n$. On rappelle les énoncés du programme :

Théorème (Loi faible des grands nombres)**.** *On suppose que X_1 admet une espérance, dans ce cas $\left(\frac{S_n}{n}\right)_n$ converge en probabilités vers $\mathbb{E}(X_1)$:*

$$\forall \varepsilon > 0, \qquad \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| \geq \varepsilon\right) = 0$$

Théorème (Loi forte des grands nombres)**.** *On suppose que X_1 admet une espérance, dans ce cas $\left(\frac{S_n}{n}\right)_n$ converge presque sûrement vers $\mathbb{E}(X_1)$:*

$$\mathbb{P}\left(\lim_{n\to+\infty} \frac{S_n}{n} = \mathbb{E}(X_1)\right) = 1$$

Théorème (Théorème Central Limite)**.** *On suppose que X_1 admet une espérance et une variance non nulle ($\mathbb{V}(X_1) \neq 0$), dans ce cas $\left(\frac{S_n - n\mathbb{E}(X_1)}{\sigma(X_1)\sqrt{n}}\right)_n$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite*

$$\forall a < b, \quad \mathbb{P}\left(\frac{S_n - n\mathbb{E}(X_1)}{\sigma(X_1)\sqrt{n}} \in]a, b[\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

Remarque : dans ce dernier énoncé, $n\mathbb{E}(X_1) = \mathbb{E}(S_n)$ et $\sigma(X_1)\sqrt{n} = \sqrt{\sigma(S_n)}$, on pourrait donc réécrire

$$\forall a < b, \quad \mathbb{P}\left(\frac{S_n - \mathbb{E}(S_n)}{\sigma(S_n)} \in]a, b[\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

On admettra que si Y est une variable aléatoire prenant des *valeurs positives* et ayant une espérance alors

Inégalité de Markov. Pour tout $\alpha > 0$ on a $\mathbb{P}(Y \geq \alpha) \leq \dfrac{\mathbb{E}(Y)}{\alpha}$.

On admettra également les deux résultats suivants

Espérance et majoration. Si $0 \leq Y \leq Z$ et Z admet une espérance alors Y admet une espérance et $\mathbb{E}(Y) \leq \mathbb{E}(Z)$.

Espérance et produit. Si Y et Z sont indépendantes et admettent une espérance alors YZ admet une espérance et $\mathbb{E}(Y\,Z) = \mathbb{E}(Y)\,\mathbb{E}(Z)$.

Les trois parties sont independantes.

1 Loi faible des grands nombres

On suppose dans cette partie que X_1 admet une espérance et une variance.

1. Déduire de l'inégalité de Markov que pour tout $n \geq 1$ on a

$$\forall \varepsilon > 0, \qquad \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}(X_1)}{n\,\varepsilon^2}$$

   En déduire la loi faible des grands nombres dans ce cas

2. **Application 1.** On suppose que X_1 suit la loi uniforme sur $[0,1]$. Soit f une fonction continue sur $[0,1]$, en appliquant la question précédente, à partir de quelle valeur de n aura-t-on :

$$\mathbb{P}\left(\left|\frac{f(X_1) + \ldots + f(X_n)}{n} - \int_0^1 f(t)dt\right| \geq 10^{-2}\right) \leq 1\%$$

   On exprimera ce résultat en fonction du maximum de $|f|$ sur $[0,1]$.

3. **Application 2.** On suppose que X_1 suit la loi de Bernoulli de paramètre p (pour un certain $p \in]0,1[$). À partir de quelle valeur de n a-t-on l'inégalité :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq 10^{-2}\right) \leq 1\%$$

4. **Grandes déviations.** Dans cette question on suppose à nouveau que X_1 suit la loi de Bernoulli de paramètre p (pour un certain $p \in]0,1[$). L'objectif est de trouver une meilleure borne que celle de la question précédente. Dans cette question, $n \geq 1$ et $\varepsilon > 0$.

   a) Soit $\lambda > 0$, montrer que

$$\mathbb{P}\left(\frac{S_n}{n} - p \geq \varepsilon\right) \leq e^{-n\lambda\varepsilon}\,\mathbb{E}\left(e^{\lambda(X_1 - p)}\right)^n$$

b) Calculer $\mathbb{E}\left(e^{\lambda(X_1-p)}\right)$ et trouver la valeur $\lambda(p,\varepsilon)$ qui minimise la fonction $g_{p,\varepsilon}$: $\lambda \mapsto e^{-\lambda\varepsilon}\,\mathbb{E}\left(e^{\lambda(X_1-p)}\right)$ sur $\mathbb{R}_+^*$.

c) On note $I(p,\varepsilon) = g_{p,\varepsilon}(\lambda(p,\varepsilon))$, justifier que $I(p,\varepsilon) < 1$.

d) Pour $p = \frac{1}{2}$, justifier que

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq 2I(p,\varepsilon)^n$$

Application : reprendre la question 1.3 avec cette nouvelle estimation.

2 Loi forte des grands nombres

Dans cette partie, on démontre la loi forte des grands nombres dans le cas où on suppose que $\mathbb{E}(X_1) = 0$ et que $\mathbb{E}(X_1^4) \leq K$ pour un certain un réel K.

1. En considérant les variables aléatoires $(X_1^2 - 1)^2$ et $(X_1^2 - X_2^2)^2$ démontrer que X_1^2 admet une espérance et que $\mathbb{E}(X_1^2)^2 \leq \mathbb{E}(X_1^4)$.
2. Démontrer de même que la variable aléatoire X_1^3 admet une espérance.
3. Démontrer que $\mathbb{E}(S_n^4) = \mathbb{E}\left(\sum_{k=1}^{n} X_k^4 + 6\sum_{1\leq i<j\leq n} X_i^2 X_j^2\right)$.

   En déduire que $\mathbb{E}\left(\frac{S_n^4}{n^4}\right) \leq \frac{3K}{n^2}$.
4. A l'aide des résultats précédents et de l'inégalité de Markov, montrer que

$$\forall \varepsilon > 0, \qquad \sum \mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) < +\infty$$

5. Soit $\varepsilon > 0$, montrer que

$$\forall N \in \mathbb{N}^*, \qquad \left(\limsup_{n\to+\infty} \left|\frac{S_n}{n}\right| \geq \varepsilon\right) \subset \bigcup_{k\geq N}\left(\left|\frac{S_k}{k}\right| \geq \varepsilon\right)$$

   En déduire que $\mathbb{P}\left(\limsup_{n\to+\infty} \left|\frac{S_n}{n}\right| \geq \varepsilon\right) = 0$.
6. Conclure des questions précédentes la loi forte des grands nombres dans ce cas.

3 Théorème Central Limite

Dans cette partie on se propose de justifier une utilisation courante du Théorème Central Limite : étant donné une variable aléatoire X_1 ayant une espérance et une variance $\mathbb{V}(X_1) \neq 0$, on veut déterminer un intervalle de confiance pour l'espérance

$m = \mathbb{E}(X_1)$ sans connaitre l'écart-type de X_1. Pour cela, on propose en général l'intervalle de confiance à 95%

$$\left[\overline{x} - 1,96\,\frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}}{\sqrt{n}}\ ;\ \overline{x} + 1,96\,\frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}}{\sqrt{n}}\right]$$

ou $x_i = X_i(\omega)$ et $\overline{x} = \frac{S_n(\omega)}{n}$ sont des réalisations indépendantes de X_1. Dans cette formule, on a remplacé l'écart-type de X_1, inconnu, par l'écart-type empirique $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}$. On va démontrer que la suite $(U_n)_n$ de variables données par

$$U_n = \frac{S_n - n\mathbb{E}(X_1)}{\sqrt{\frac{n}{n-1}\sum_{i=1}^{n}(X_i - \frac{S_n}{n})^2}} = V_n\,\frac{\sigma(X_1)}{\sqrt{W_n}}$$

converge en loi vers la loi normale centrée réduite, avec $V_n = \frac{S_n - n\mathbb{E}(X_1)}{\sigma(X_1)\sqrt{n}}$ et $W_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \frac{S_n}{n})^2$.

1. Pour $n \geq 1$, démontrer que $\mathbb{E}(W_n) = V(X_1)$.
2. Pour $n \geq 1$, démontrer que $W_n = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{S_n}{n}\right)^2\right)$. En déduire que
   — pour $n \geq 1$, $\mathbb{E}(W_n) = V(X_1)$ (on dit que W_n est un estimateur sans biais),
   — $(W_n)_n$ converge presque sûrement vers $\mathbb{V}(X_1)$.
3. On démontre maintenant que si $(Q_n)_n$ est une suite de variables aléatoires qui converge en loi vers une variable aléatoire à densité Q et si $(R_n)_n$ est une suite de variables aléatoires qui converge presque sûrement vers une constante $c > 0$ alors $(R_n\,Q_n)_n$ converge en loi vers $c\,Q$, c'est-à-dire :

$$\forall t \in \mathbb{R}, \qquad \lim_{n\to+\infty} \mathbb{P}(R_n\,Q_n \leq t) = \mathbb{P}(c\,Q \leq t) \tag{R Q}$$

   a) Soit $\varepsilon > 0$, montrer que

$$\bigcap_{N\geq 1}\bigcup_{n\geq N}(|R_n - c| \geq \varepsilon) \subset \Omega \setminus (\lim_{n\to+\infty} R_n = c)$$

   b) En déduire que

$$\forall \varepsilon > 0,\ \forall \delta \in]0,1[\,,\ \exists N,\ \forall n \geq N, \qquad \mathbb{P}(|R_n - c| \leq \varepsilon) > 1 - \delta$$

   c) Démontrer (R Q).
4. Démontrer que $(U_n)_n$ converge en loi vers la loi normale centrée réduite. En conclure la validité de l'intervalle de confiance proposé.