Université de Nice-Sophia Antipolis
L2 MPIE - Statistiques
Sylvain Rubenthaler http://math.unice.fr/~rubentha/cours.html

Exemple d'exercices pour l'examen

Documents, calculatrices et téléphones interdits. Les exercices sont indépendants.

1. (Exercice 3 de la feuille de TD 8)
2. Soient $X_1, X_2, \dots$ indépendants et de même loi de Bernoulli de paramètre p. On suppose que $p \geq p_0 = 0,4$. On veut estimer p à partir d'un échantillon $X_1, \dots, X_n$. Soit $\alpha = 0,01$.
   (a) Soit $Z \sim \mathcal{N}(0,1)$. Trouver dans la table une valeur t telle que $\mathbb{P}(|Z| \geq t) = 0,3174$.
   (b) Calculer $\mathbb{E}(X_1)$, $\mathbb{V}(X_1)$ (refaire la démonstration du cours).
   (c) Vers quoi tend $\overline{X}_n = \frac{X_1+\dots+X_n}{n}$ quand $n \to +\infty$? Est-ce une limite p.s., en probabilité ou en loi ?
   (d) Vers quoi tend $\frac{\sqrt{n}}{\sqrt{p(1-p)}}\left(\overline{X}_n - p\right)$ quand $n \to +\infty$? Est-ce une limite p.s., en probabilité ou en loi ?
   (e) Montrer que $\forall p \in [p_0; 1]$, $\frac{1-p}{p} \geq 1,5$.
   (f) Trouver n tel que $\mathbb{P}\left(\left|\left(\overline{X}_n - p\right)\right| \geq p \times \alpha\right) \leq 0,07935$.

CORRIGÉ

1. (a) Calculons

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x \times \frac{2x}{\theta^2}\mathbf{1}_{[0;\theta]}(x)dx \\ &= \int_0^\theta \frac{2x^2}{\theta^2}dx = \left[\frac{2x^3}{3\theta^2}\right]_0^\theta = \frac{2\theta}{3}\,.\end{aligned}$$

Soit $\overline{X}_n = \frac{X_1+\dots+X_n}{n}$. Prenons pour estimateur de θ la variable $Y_n = 3\overline{X}_n/2$. Nous avons :

$$\mathbb{E}(Y_n) = \frac{3}{2} \times \frac{n\mathbb{E}(X_1)}{n} = \theta\,.$$

Donc Y_n est un estimateur sans biais de θ. Par la loi des grands nombres :

$$Y_n \underset{n\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} \theta\,.$$

Donc Y_n converge en probabilité vers θ donc c'est un estimateur convergent.

(b) Calculons la vraisemblance (on écrit la densité pour des X_i de densité ayant un paramètre u à la place de θ):

$$L(X_1,\dots,X_n;u) = \prod_{i=1}^n \frac{2X_i}{u^2}\mathbf{1}_{[0;u]}(x_i) = \begin{cases} = \frac{2^n X_1\dots X_n}{u^{2n}} & \text{si } 0 \leq X_1,\dots,X_n \leq u\,, \\ = 0 & \text{sinon}\,. \end{cases}$$

Donc $\widehat{\theta}_n = max_{1\leq i\leq n}X_i$. Calculons pour $t \in \mathbb{R}$:

$$\mathbb{P}(\widehat{\theta}_n \leq t) = \begin{cases} 0 & \text{si } t \leq 0\,, \\ \mathbb{P}(X_1 \leq t,\dots,X_n \leq t) & \text{si } 0 \leq t \leq \theta\,, \\ 1 & \text{si } t \geq \theta\,. \end{cases}$$

Pour $0 \leq t \leq \theta$, par indépendance des X_i :

$$\begin{aligned}
\mathbb{P}(X_1 \leq t, \ldots, X_n \leq t) &= \prod_{i=1}^{n} \mathbb{P}(X_i \leq t) \\
&= \prod_{i=1}^{n} \left(\int_0^t \frac{2x}{\theta^2} dx \right) \\
&= \prod_{i=1}^{n} \frac{t^2}{\theta^2} = \frac{t^{2n}}{\theta^{2n}} .
\end{aligned}$$

Donc la densité de $\widehat{\theta}_n$ est la fonction f suivante (on dérive $t \mapsto \mathbb{P}(\widehat{\theta}_n \leq t)$ là où c'est possible)

$$f(t) = \begin{cases} 0 & \text{si } t < 0 \,, \\ \frac{2nt^{2n-1}}{\theta^{2n}} & \text{si } 0 < t < \theta \,, \\ 0 & \text{si } \theta < t \,. \end{cases}$$

(Remarque : on ne calcule pas la densité aux points $0, \theta$ mais ce n'est pas grave car la valeur en ces points n'a pas d'importance pour le calcul d'intégrale qui suit.) Calculons :

$$\begin{aligned}
\mathbb{E}(\widehat{\theta}_n) &= \int_{-\infty}^{+\infty} t f(t) dt \\
&= \int_0^{\theta} t \times \frac{2nt^{2n-1}}{\theta^{2n}} dt \\
&= \left[\frac{2nt^{2n+1}}{(2n+1)\theta^{2n}} \right]_0^{\theta} = \frac{2n\theta}{(2n+1)} .
\end{aligned}$$

Donc $\widehat{\theta}_n$ est un estimateur avec biais de θ.
Nous avons toujours $\widehat{\theta}_n \leq \theta$ et $\forall \epsilon > 0$,

$$\mathbb{P}(\widehat{\theta}_n \leq \theta - \epsilon) = \frac{(\theta - \epsilon)^{2n}}{\theta^{2n}} = \left(\frac{\theta - \epsilon}{\theta} \right)^{2n} \underset{n \to +\infty}{\longrightarrow} 0 .$$

Donc $\widehat{\theta}_n \underset{n \to +\infty}{\overset{\text{proba.}}{\longrightarrow}} \theta$. Donc $\widehat{\theta}_n$ est un estimateur convergent.

2. (a) Nous avons $\mathbb{P}(|Z| \geq t) = 2\mathbb{P}(Z \geq t) = 2(1 - \mathbb{P}(Z \leq t))$. On veut donc $\mathbb{P}(Z \leq t) = 1 - 0,3174/2 = 1 - 0,1587 = 0,8413$. On prend donc $t = 1$.

   (b) $\mathbb{E}(X_1) = p \times 1 + (1-p) \times 0 = 0$, $\mathbb{E}(X_1^2) = p \times 1^2 + (1-p) \times 0^2 = p$, $\mathbb{V}(X_1) = p - p^2 = p(1-p)$.

   (c) Par la loi des grands nombres : $\overline{X}_n \underset{n \to +\infty}{\overset{\text{p.s.}}{\longrightarrow}} \mathbb{E}(X_1) = p$.

   (d) Par le théorème central-limite : $\frac{\sqrt{n}}{\sqrt{p(1-p)}}(\overline{X}_n - p) \underset{n \to +\infty}{\overset{\text{loi}}{\longrightarrow}} Z$ avec $Z \sim \mathcal{N}(0,1)$.

   (e) On étudie rapidement $f : p \mapsto \frac{1-p}{p}$. Nous avons $f'(p) = -\frac{1}{p^2} + 1 \leq 0$ pour $p \in [0;1]$. Donc, $p \in [p_0; 1] \Rightarrow \frac{1-p}{p} \geq \frac{1-p_0}{p_0} = 1,5$.

   (f) Nous remarquons tout d'abord que

$$\mathbb{P}\left(\sqrt{\frac{n}{p(1-p)}} \left| (\overline{X}_n - p) \right| \geq p \times \alpha \sqrt{\frac{n}{p(1-p)}} \right) = \mathbb{P}(\left| (\overline{X}_n - p) \right| \geq p) .$$

   Soit $Z \sim \mathcal{N}(0,1)$. Par le théorème central-limite, on fait l'approximation

$$\mathbb{P}\left(\sqrt{\frac{n}{p(1-p)}} \left| (\overline{X}_n - p) \right| \geq p \times \alpha \sqrt{\frac{n}{p(1-p)}} \right) = \mathbb{P}\left(|Z| \geq p \times \alpha \sqrt{\frac{n}{p(1-p)}} \right) .$$

On veut donc n tel que $p \times \alpha\sqrt{\frac{n}{p(1-p)}} \geq 1$. Donc, on veut

$$n \quad \geq \quad \frac{(1-p)}{p\alpha^2}\,.$$

Il suffit donc de prendre

$$n \geq \frac{(1-p_0)}{p_0\alpha^2} = 1,5 \times 10^4\,.$$