Map 311 - Introduction aux probabilités et à la simulation aléatoire.

PC 8 - Année 2016

Vincent Bansaye - Lucas Gerin

Méthodes d'estimation

EXERCICE 1 -(Somme normalisée)
Soit $(X_i)_{i\geq 1}$ une suite de v.a. i.i.d. d'espérance 0 et de carré intégrable. Quelle est la limite en loi de

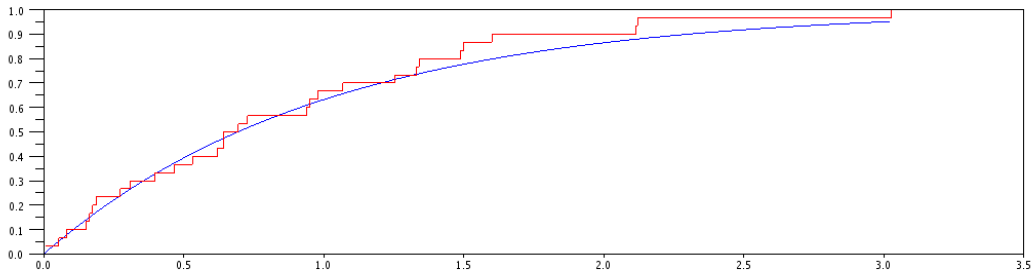
$$\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n X_i^2}} \quad ?$$

EXERCICE 2 -(Fonction de répartition empirique). Soit $(X_k)_{k\geq 1}$ une suite de v.a. i.i.d. de fonction de répartition F, que l'on suppose continue sur $\mathbb{R}$. La fonction F est inconnue, on cherche à l'estimer à partir d'un échantillon.

Pour $n \geq 1$, on définit la *fonction de répartition empirique* F_n associée à $X_1, \dots, X_n$:

$$F_n(t) = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{X_k \leq t}.$$

La fonction F_n est donc aléatoire, on va montrer qu'elle donne une bonne estimation de F.



Les fonctions de répartition empirique et théorique pour un échantillon de 30 *v.a. exponentielles.*

1. Démontrer que F_n converge simplement vers F : pour tout $t \in \mathbb{R}$, $F_n(t) \to F(t)$ p.s.

On souhaite démontrer que la convergence est en fait uniforme. Soit $\varepsilon > 0$, et soit une subdivision[1] $-\infty = a_0 < a_1 < \dots < a_{K-1} < a_K = +\infty$ tels que, pour tout i, $F(a_{i+1}) - F(a_i) \leq \varepsilon$.

2. Démontrer que si $x \in [a_i, a_{i+1}]$,

$$F_n(a_i) - F(a_i) - \varepsilon \leq F_n(x) - F(x) \leq F_n(a_{i+1}) - F(a_{i+1}) + \varepsilon,$$

   et conclure :

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \to 0 \text{ p.s.}$$

[1] Au fait, pourquoi est-ce que cette subdivision existe ?

EXERCICE 3 -(Maximum de vraisemblance)

1. Pour $\lambda > 0$, soit $X_1, \ldots, X_n$ un échantillon de Poisson(λ). Déterminer la fonction de vraisemblance $L(x_1, \ldots, x_n; \lambda)$, puis l'estimateur du maximum de vraisemblance (EMV) $\hat{\lambda}_n$ de λ. Est-ce que l'EMV est sans biais ? Est-ce qu'il converge vers λ ?
2. Pour $\theta > -1$, soit $X_1, \ldots, X_n$ un échantillon de densité $x \mapsto (1+\theta)x^\theta$ sur $[0, 1]$. Déterminer la fonction de vraisemblance $L(x_1, \ldots, x_n; \theta)$, puis l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ. Est-ce que l'EMV est sans biais ? Est-ce qu'il converge vers θ ?

EXERCICE 4 -(Estimation optimale de σ^2)
On considère un échantillon de n variables aléatoires indépendantes $X_1, \ldots, X_n$ de loi $\mathcal{N}(0, \sigma^2)$, avec σ^2 inconnue. On cherche à l'estimer par un estimateur de la forme

$$S_c = \frac{1}{c} \sum_{k=1}^{n} X_k^2,$$

où c est un réel positif. Pour évaluer la qualité de l'estimateur S_c, on introduit deux quantités :

- Le *biais* de S_c est le réel $b(c) = \mathbb{E}[S_c] - \sigma^2$.
- Le *risque quadratique* est la quantité $\mathbf{R}(c) = \mathbb{E}\left[(S_c - \sigma^2)^2\right]$.

1. Comment choisir c pour que S_c soit sans biais ?
2. Vérifier que le risque quadratique peut s'écrire sous la forme suivante

$$\mathbf{R}(c) = \mathrm{Var}(S_c) + b(c)^2.$$

3. On rappelle que $\mathbb{E}[X_1^4] = 3\sigma^4$. Comment choisir c pour que S_c soit de risque quadratique minimal ?

EXERCICE 5 -(Processus autorégressif : estimation)

Soient a un réel et $(X_k)_{k\geq 0}$ le processus défini par $X_0 = 0$ et

$$X_{k+1} = aX_k + \varepsilon_{k+1}$$

où $(\varepsilon_k)_{k\geq 1}$ est une suite de $\mathcal{N}(0, 1)$ i.i.d. (en particulier ε_{k+1} est indépendante de X_k). On cherche à estimer le paramètre a à partir d'un échantillon $(X_1, \ldots, X_n)$.

1. Écrire la densité de $(X_1, X_2, X_3, \ldots, X_n)$. En déduire la fonction de vraisemblance $L(x_1, \ldots, x_n; a)$,
2. Calculer l'estimateur du maximum de vraisemblance $\hat{a}_n$ de a.