Fiche 1 - Statistiques - M1 ingé math

Emeline Schmisser, emeline.schmisser@math.univ-lille1.fr, bureau 314 (bâtiment M3).

Exercice 1 : Moyenne et variance empirique
On considère un échantillon $X_1, \dots, X_n$ de variables indépendantes identiquement distribuées d'espérance m et de variance σ^2.

1. Rappeler la définition de la variance de X_1. Montrer que :
$$\mathrm{Var}\left(X_1\right) = \mathbb{E}\left(X_1^2\right) - \left(\mathbb{E}\left(X_1\right)\right)^2$$
On appelle moyenne et variance empirique les estimateurs suivants :
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \qquad S_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$$
2. Montrer que
$$S_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2$$
3. Est-ce que la moyenne empirique est un estimateur sans biais ? Qu'en est-il pour la variance empirique ? Est-ce que ces estimateurs sont asymptotiquement sans biais ?
4. Proposer un estimateur sans biais S_n^{*2} de σ^2.
5. Peut-on avoir $\mathbb{E}\left(S_n^*\right) = \sigma$? (Ne pas calculer $\mathbb{E}\left(S_n^*\right)$).
6. Les estimateurs $\bar{X}_n$, S_n^2 et S_n^{*2} sont-ils convergents ?

Exercice 2 : Risque quadratique
Soit T_n un estimateur du paramètre θ. Le risque quadratique de T est la quantité
$$R(T_n) = \mathbb{E}\left((T_n - \theta)^2\right)$$

1. Montrer que $R(T_n) = \mathrm{Var}\left(T_n\right) + \left(\mathbb{E}\left(T_n - \theta\right)\right)^2$
2. Montrer que si $R(T_n)$ tend vers 0 quand n tend vers l'infini, alors T_n est un estimateur convergent de θ.
3. Montrer que si T_n est un estimateur sans biais ou asymptotiquement sans biais de θ dont la variance tend vers 0, alors T_n est un estimateur convergent de θ.

Exercice 3 : Exponentielle translatée
Soit $X_1, \dots, X_n$ un échantillon i.i.d. d'une v.a. X de densité $f(x) = \exp\{-(x-\theta)\}\mathbf{1}_{x \geq \theta}$ avec $\theta \in \mathbb{R}$. Il s'agit de la loi exponentielle translatée.

1. Montrer que $Y_n = \min\{X_1, \dots, X_n\}$ converge vers θ en probabilité.
2. Calculer la fonction de répartition de Y_n. En déduire la densité de Y_n.
3. Calculer le biais de Y_n et en déduire un estimateur sans biais de θ.

Exercice 4 : Loi uniforme
Soient $(X_1, \dots, X_n)$ des variables aléatoires i.i.d de loi uniforme sur $[0, a]$.

1. Proposer au moins deux estimateurs différents de a.

2. Calculer leurs biais. En déduire deux estimateurs non biaisés de a.
3. Calculer la variance des estimateurs. Est-ce que ces estimateurs sont convergents ?
4. Quel est le meilleur estimateur au sens du risque quadratique ?

Exercice 5 : Loi de Poisson
Soient $X_1, \dots X_n$ des variables aléatoires de loi de Poisson $\mathcal{P}(\lambda)$.

1. Calculer $\mathbb{E}(X_i)$.
2. En déduire un estimateur de λ.
3. Est-ce que cet estimateur est sans biais ? convergent ?

Exercice 6 : Loi Gamma
Soit $(X_1, \dots, X_n)$ un échantillon i.i.d d'une variable aléatoire X de loi $\Gamma(\alpha, \beta)$. La densité de X est

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

et nous avons $\mathbb{E}(X_i) = \alpha\beta$ et $\mathrm{Var}(X_i) = \alpha\beta^2$.

1. Proposer deux estimateurs $\hat{\alpha}$ et $\hat{\beta}$ de α et β.
2. Ces estimateurs sont-ils convergents ?

Exercice 7 : Mesures dominantes
On considère n variables i.i.d. $X_1, \dots, X_n$.

1. Si la loi des X_i est continue, proposer une mesure dominante pour le modèle.
2. Si les X_i suivent une loi binomiale de paramètre p, proposer une mesure dominante pour le modèle.
3. Si les X_i suivent une loi de Poisson de paramètre λ, proposer une mesure dominante pour le modèle.
4. On suppose que X_i a une probabilité α de valoir a et une probabilité $1 - \alpha$ de suivre une loi normale $\mathcal{N}(0, 1)$. Proposer une mesure dominante pour le modèle.

Exercice 8 : Système industriel
Un système industriel fonctionne en utilisant deux machines de types différents. Les durées de vie X_1 et X_2 de ces deux machines sont deux lois exponentielles de paramètres respectifs λ_1 et λ_2.

1. Calculer la fonction de survie d'une v.a. exponentielle X de paramètre $\lambda > 0$. On rappelle que la fonction de survie $S_X(x)$ vérifie
$$S_X(x) = \mathbb{P}(X > x) = 1 - F_X(x)$$
2. Calculer la probabilité pour que le système ne tombe pas en panne avant la date t. En déduire la fonction de survie du système. Quelle loi suit le système ?
3. Calculer la probabilité pour que la panne soit due à une défaillance de la première machine.