Fiche 1 - Statistique - M1 Ingé Math

Tran Viet Chi, `chi.tran@math.univ-lille1.fr`, bureau 316 (bâtiment M3).

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

1. Rappeler la définition de la variance de X_1. Montrer que

$$\mathrm{Var}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2.$$

On appelle moyenne et variance empiriques les estimateurs suivants :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i, \qquad S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

2. La moyenne empirique est-elle un estimateur sans biais ? Qu'en est-il pour la variance empirique ?

3. Proposer un estimateur sans biais S_n^{*2} de la variance.

4. A-t-on aussi $\mathbb{E}(S_n^*) = \sigma$? (ne pas faire le calcul explicite).

5. $\bar{X}_n$, S_n^2 et S_n^{*2} sont-ils convergents ?

6. En utilisant le théorème central limite, construire un intervalle de confiance d'ordre $1 - \alpha$ (avec $\alpha \in]0, 1[$) pour m.

Exercice 2 (Mesure dominante)
Soit un échantillon $X_1, \dots, X_n$ de v.a. i.i.d. telles que X_i a une probabilité $\alpha \in]0, 1[$ de valoir a et une probabilité $1 - \alpha$ de suivre une loi normale $\mathcal{N}(0, 1)$:

$$X_i = a\ \mathbf{1}_{Z_i=1} + Y_i\ \mathbf{1}_{Z_i=0},$$

où les Y_i sont des v.a. i.i.d. de loi $\mathcal{N}(0, 1)$ et où Z_i sont des v.a. i.i.d. de loi $\mathcal{B}(1, \alpha)$.

1. Montrer que $(\delta_a + \lambda)^{\otimes n}$ où λ est la mesure de Lebesgue sur $\mathbb{R}$ est une mesure dominante pour le modèle considéré.

2. Montrer que

$$\frac{dP^{X_1,\dots,X_n}}{d(\delta_a + \lambda)^{\otimes n}}(x_1, \dots, x_n)) = \prod_{i=1}^n \Big(\alpha \mathbf{1}_a(x_i) + (1 - \alpha)(1 - \mathbf{1}_a(x_i))\phi(x_i)\Big)$$

où ϕ est la densité de la loi normale $\mathcal{N}(0, 1)$.

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

1. Prouver que $Y_n = \min\{X_1, \dots, X_n\}$ converge en probabilité vers θ.

2. Calculer l'espérance de Y_n et en déduire un estimateur sans biais de θ.

Exercice 4 (Risque quadratique)
On considère un modèle statistique $(E, \mathcal{E}, P_\theta, \theta \in \Theta \subset \mathbb{R}^d)$. La perte quadratique d'un estimateur T de θ est la variable aléatoire $W(T, \theta) = \|T - \theta\|^2$. Le risque quadratique de T est

$$R(T, \theta) = \mathbb{E}_\theta(W(T, \theta)),$$

où $\mathbb{E}_\theta$ signifie qu'on prend l'espérance par rapport à θ.
1. Montrer que pour tout $\theta \in \Theta$,

$$R(T, \theta) = \mathbb{E}\big(\|T - \mathbb{E}(T)\|^2\big) + \|\mathbb{E}(T) - \theta\|^2.$$

2. Montrer que si T est un estimateur sans biais de θ dont la variance tend vers zéro, alors il est convergeant.

Exercice 5 (Système industriel)
Un système fonctionne en utilisant deux machines de types différents. Les durées de vie X_1 et X_2 des deux machines sont deux v.a. indépendantes qui suivent des lois exponentielles de paramètres λ_1 et λ_2.

1. Calculer la fonction de survie d'une v.a. X exponentielle de paramètre $\lambda > 0$.

2. Calculer la probabilité pour que le système ne tombe pas en panne avant la date t. En déduire la loi de la durée de vie Z du système.

3. Calculer la probabilité pour que la panne soit due à une défaillance de la machine 1.

On dispose maintenant de n systèmes identiques et fonctionnant indépendamment les uns des autres, dont on observe les durées de vie $Z_1, \ldots, Z_n$.

4. Ecrire le modèle statistique correspondant et la vraisemblance des observations. (λ_1, λ_2) sont-il identifiables ?

5. Si on observe à la fois les durées de vie des systèmes et la cause de la défaillance (machine 1 ou 2), écrire la vraisemblance des observations. A-t-on alors suffisamment d'information pour estimer λ_1 et λ_2 ? On introduira : $S_i = \mathbf{1}_{X_2^i > X_1^i}$.

On considère maintenant un seul système utilisant une machine de type 1 et une machine de type 2. On dispose d'un stock de n_1 machines de type 1, de durées de vie $X_1^1, \ldots, X_1^{n_1}$, et d'un stock de n_2 machines de type 2, de durées de vie $X_2^1, \ldots, X_2^{n_2}$. Quand une machine en panne, on la remplace par une machine du même type, tant que le stock de machines de ce type n'est pas épuisé. Quand cela arrive, le système lui-même est en panne. On note toujours Z la durée de vie du système. Le cas $n_1 = n_2 = 0$ correspond donc à la première question (pas de stock).

6. Montrer que la loi de la somme de n variables indépendantes qui suivent une loi exponentielle de même paramètre λ est une loi gamma $\Gamma(n, \lambda)$ de densité :

$$f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbf{1}_{x>0}.$$

7. Ecrire Z en fonction des X_j^i et en déduire $\mathbb{P}(Z \geq t)$ en fonction de certaines lois gamma, dont on précisera les paramètres.

On note alors N le nombre de machines (tout type confondus) sorties du stock quand le système tombe en panne, et Z_0 la durée écoulée avant la première panne d'une machine. On note Z_i la durée écoulée entre la i-ème panne et la $(i+1)$-ème panne. La durée de vie totale du système est donc :

$$Z = \sum_{i=0}^{N} Z_i$$

et la $N + 1$-ème panne est donc la panne fatale au système.

8. Montrer que les variables Z_i sont i.i.d. et donner leur loi.

9. Préciser l'ensemble des valeurs possibles pour la variable N et en donner la loi.

10. On admet que N et les Z_i sont indépendantes. Calculer $\mathbb{E}(Z|N)$ en fonction de N, λ_1, λ_2. Donner l'expression de $\mathbb{E}(Z)$ en fonction de $\mathbb{E}(N)$, λ_1 et λ_2.