STATISTIQUE MATHÉMATIQUE - TRAVAUX DIRIGÉS : CORRIGÉS.

ENSAE - Mastères spécialisés - 2025/2026

ENSAI École nationale de la statistique et de l'analyse de l'information

1 Modèles statistiques et estimateurs

1.1 Révisions sur les lois et quelques calculs classiques.

1°. Soit X une variable aléatoire de loi exponentielle de paramètre λ. Déterminer la loi de $Y = [X]$, partie entière de X.

2°. Soit $X_1, ..., X_n$, n v.a.i.i.d. de loi exponentielle de paramètre λ et $M_n = \mathrm{Min}(X_1, ..., X_n)$. Déterminer la loi de M_n.

3°. Déterminer la loi du minimum de deux lois uniformes sur $[0, 1]$, indépendantes.

4°. Soit X une variable aléatoire suivant une loi de Laplace. Déterminer la loi de $Y = |X|$ et calculer $\mathbb{P}[X \geq 2]$, $\mathbb{P}[X < 0]$, $\mathbb{P}[X = 0]$.

5°. Soit X une variable aléatoire de loi uniforme sur $[-\pi/2, \pi/2]$. Déterminer la loi de $Y = \tan X$.

Corrigé :

1°. Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$, c'est-à-dire de densité

$$f_X(x) = \lambda e^{-\lambda x}\mathbf{1}_{x\geq 0}.$$

Pour tout $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(Y = k) &= \mathbb{P}(k \leq X < k+1)\\ &= \int_k^{k+1} \lambda e^{-\lambda x}\,dx = e^{-\lambda k} - e^{-\lambda(k+1)}\\ &= q^k(1-q)\end{aligned}$$

avec $q = e^{-\lambda}$. Ainsi, Y suit une loi géométrique (décalée) de paramètre $p = 1 - e^{-\lambda}$.

2°. La fonction de répartition de M_n est, pour tout $t \geq 0$,

$$\begin{aligned}\mathbb{P}(M_n \leq t) &= 1 - \mathbb{P}(M_n > t)\\ &= 1 - \mathbb{P}(X_1 > t, \ldots, X_n > t)\\ &= 1 - e^{-n\lambda t}.\end{aligned}$$

par indépendance et identique distribution des X_i. Donc M_n suit une loi exponentielle de paramètre $n\lambda$.

3°. Pourr tout $t \in [0, 1]$,

$$\mathbb{P}(M \leq t) = 1 - \mathbb{P}(M > t) = 1 - \mathbb{P}(U_1 > t, U_2 > t) = 1 - (1-t)^2.$$

La densité de M est donc

$$f_M(t) = \frac{d}{dt}\mathbb{P}(M \leq t) = 2(1-t)\mathbf{1}_{[0,1]}(t).$$

4°. Loi de Y : pour tout $y \geq 0$,

$$\mathbb{P}(Y \leq y) = \mathbb{P}(-y \leq X \leq y) = \int_{-y}^{y} \frac{\lambda}{2} e^{-\lambda|x|}\,dx = 1 - e^{-\lambda y}.$$

Donc Y suit une loi exponentielle de paramètre λ.

Calculs des probabilités :

$$\mathbb{P}(X \geq 2) = \int_2^{+\infty} \frac{\lambda}{2} e^{-\lambda x}\,dx = \frac{1}{2}e^{-2\lambda},$$

$$\mathbb{P}(X < 0) = \int_{-\infty}^{0} \frac{\lambda}{2} e^{\lambda x}\,dx = \frac{1}{2},$$

$$\mathbb{P}(X = 0) = 0 \quad \text{(car X est à densité).}$$

5°. arctan est une bijection de $\mathbb{R}$ dans $]-\pi/2, \pi/2[$. Pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(Y \leq x) = \mathbb{P}(\tan X \leq x) = \mathbb{P}(X \leq \arctan x) = F(\arctan x).$$

avec $F(t) = \frac{1}{\pi}\mathbb{1}_{[-\pi/2,\pi/2]}(t)$. Ainsi,

$$\mathbb{P}(Y \leq x) = \frac{1}{2} + \frac{1}{\pi}\arctan x.$$

La densité de Y est donc

$$f_Y(x) = \frac{d}{dx}\mathbb{P}(Y \leq x) = \frac{1}{\pi(1+x^2)}.$$

Ainsi, Y suit une loi de Cauchy standard.

1.2 Moyenne et variance de la moyenne et de la variance empirique d'un échantillon.

On considère $X_1, ..., X_n$, n v.a.i.i.d. et l'on note $\overline{X}$ leur moyenne empirique et S^2 leur variance empirique :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ et } S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1}$$

1°. Déterminer la moyenne et la variance de $\overline{X}$.

2°. Déterminer la moyenne et la variance de S^2.

1.3 Calculs sur un échantillon uniforme

Soit X une variable aléatoire de loi uniforme sur $[0, 2a]$ et $X_1 ..., X_n$ un n échantillon de X. Soit $\overline{X}$ la moyenne empirique des X_i et M le maximum.

1°. Calculer $\mathbb{E}[X]$ et $\mathbb{V}(X)$.

2°. Calculer $\mathbb{E}[\overline{X}]$ et $\mathbb{V}(\overline{X})$.

3°. Calculer $\mathbb{E}[M]$ et $\mathbb{V}(M)$.

4°. Comparer les résultats des questions 2° et 3°.

Corrigé :

1°. Calcul de $\mathbb{E}[X]$ et $\mathbb{V}(X)$:

$$\mathbb{E}[X] = \frac{0+2a}{2} = a,$$

$$\mathbb{V}(X) = \frac{(2a)^2}{12} = \frac{a^2}{3}.$$

2°. Calcul de $\mathbb{E}[\overline{X}]$ et $\mathbb{V}(\overline{X})$:

$$\mathbb{E}[\overline{X}] = \mathbb{E}[X] = a,$$

$$\mathbb{V}(\overline{X}) = \frac{\mathbb{V}(X)}{n} = \frac{a^2}{3n}.$$

3°. Calcul de $\mathbb{E}[M]$ et $\mathbb{V}(M)$: La fonction de répartition de M est $F_M(t) = \left(\frac{t}{2a}\right)^n$ pour $t \in [0,2a]$, donc la densité de M est

$$f_M(t) = \frac{n}{2a}\left(\frac{t}{2a}\right)^{n-1}.$$

On en déduit

$$\mathbb{E}[M] = \int_0^{2a} t \cdot \frac{n}{2a}\left(\frac{t}{2a}\right)^{n-1} dt = \frac{2an}{n+1},$$

$$\mathbb{E}[M^2] = \int_0^{2a} t^2 \cdot \frac{n}{2a}\left(\frac{t}{2a}\right)^{n-1} dt = \frac{4a^2 n}{n+2},$$

$$\mathbb{V}(M) = \mathbb{E}[M^2] - \mathbb{E}[M]^2 = \frac{4a^2 n}{(n+2)(n+1)^2}.$$

4°. Comparaison des résultats des questions 2° et 3° : La moyenne empirique $\overline{X}$ est un estimateur sans biais de a et sa variance décroît en $1/n$, tandis que le maximum M est biaisé (mais asymptotiquement sans biais) mais sa variance décroît plus vite en $1/n^2$.

1.4 Moyenne et variance empirique d'un échantillon gaussien.

On considère le modèle d'échantillonnage gaussien $X_1, \dots, X_n$ de n v.a.i.i.d. de loi $\mathcal{N}(m, \sigma^2)$ et l'on note $\overline{X}$ la moyenne empirique et S^2 la variance empirique :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ et } S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (2)$$

1°. Parmi les variables aléatoires suivantes, lesquelles sont des statistiques?

$\overline{X}$? S^2? nS^2/σ^2? $T = \sqrt{n-1}(\overline{X} - m)/S$?

2°. Déterminer le modèle image par la moyenne empirique.

3°. Déterminer le modèle image par la statistique $(\overline{X}, S^2)$ en opérant de la façon suivante : écrivez S^2 comme fonction de $Y_i = X_i - \overline{X}$ puis en déduire que S^2 est indépendante de $\overline{X}$; déterminer la loi de nS^2/σ^2. Enfin, écrivez le modèle image sous la forme d'un triplet.

4°. Déterminer la loi de T.

1.5 Statistiques d'ordre.

On considère le modèle statistique d'échantillonnage où $\mathcal{P}$ est l'ensemble des lois de probabilités définies sur $(\mathbb{R}, B(\mathbb{R}))$. Soit F la fonction de répartition d'une loi $P \in \mathcal{P}$.

1°. Quel est le modèle image par la statistique d'ordre $X_{(1)} = \min_{i=1\dots n} X_i$? Par $X_{(n)} = \max_{i=1\dots n} X_i$?

2°. Soit F_n la fonction de répartition empirique du n-échantillon. Déterminer le modèle image par la statistique $nF_n(x)$.

3°. Déterminer le modèle image par la statistique d'ordre $X_{(i)}$, pour $i = 1, \dots, n$.

1.6 Exemples de modèles exponentiels.

Pour les différents modèles proposés, on note X la variable des observations. Exhibez pour chacun une mesure dominante, spécifiez si les modèles sont exponentiels et dans l'affirmative, déterminez une statistique canonique, un paramètre canonique, puis calculer l'espérance et la variance de la statistique canonique.

1°. $(X_1, \dots, X_n)$ est un n-échantillon gaussien.

2°. X est issu d'un modèle hypergéométrique.

3°. X est issu d'un modèle binomial négatif.

4°. $X = \epsilon + (1-\epsilon)Y$ où $\epsilon \sim b(\alpha)$ suit une loi de Bernoulli de paramètre α et $Y \sim \mathcal{N}(0,1)$ est indépendante de ϵ.

1.7 Statistiques de rang.

Soit $X = (X_1, \dots, X_n)$ un n-échantillon issu d'un modèle $\mathcal{P}$ de loi sans atome sur $\mathbb{R}$. Le rang de X_i est

$$R_i(X) = \mathrm{card}\{j : X_j \le X_i\} \qquad (3)$$

On note $R(X) = (R_1(X), \dots, R_n(X))$.

1°. Faites le lien entre le rang et les statistiques d'ordre.

2°. Montrer que pour tout $i = 1, \dots, n$, on a

$$R_i(X) = \sum_{j=1}^{n} \mathbb{1}_{[X_i - X_j \ge 0]} \qquad (4)$$

3°. Montrer que pour tout $x \in \mathbb{R}^n$, $R(x)$ est une permutation de $(1, 2, \dots, n)$ $\mathbb{P}$-p.s.

4°. Quel est le modèle image par R?

1.8 Loi uniforme : différents estimateurs

Soit $(X_1, \dots, X_n)$ un n-échantillon de loi uniforme sur $[0, \theta]$.

1°. Déterminez un estimateur θ_1 de θ par la méthode des moments. Étudiez son biais, son erreur quadratique, sa convergence.

2°. Déterminez un estimateur θ_2 de θ par la méthode du maximum de vraisemblance. Étudiez son biais, son erreur quadratique, sa convergence. En déduire un estimateur θ_3 de θ, sans biais.

3°. On suppose que $n = 2m+1$. Déterminez un estimateur plug-in θ_4 de θ. Étudiez son biais, son erreur quadratique, sa convergence.

4°. Quel est le meilleur des estimateurs précédents ?

5°. On muni le modèle d'une loi *a priori* de densité :

$$\pi(\theta) = \frac{1}{\theta^2}\mathbb{1}_{]1,\infty[}(\theta) \tag{5}$$

Déterminez la loi *a posteriori* du modèle et proposer un estimateur θ_5 de θ.

1.9 Loi uniforme : encore d'autres estimateurs.

On considère un échantillon $(X_1,\dots,X_n)$ de loi uniforme sur l'intervalle $[\theta,\theta+1]$ où θ est inconnu. On pose :

$$\widehat{\theta}_1 = \overline{X}_n - \frac{1}{2} \tag{6}$$

$$\widehat{\theta}_2 = \min(X_1,\dots,X_n) \tag{7}$$

$$\widehat{\theta}_3 = \max(X_1,\dots,X_n) - 1 \tag{8}$$

1°. Démontrer que $\widehat{\theta}_1$ est l'estimateur des moments de θ. En déduire qu'il est sans biais ; déterminer son erreur quadratique et démontrer qu'il converge au sens des moindres carrés. Démontrer qu'il est fortement consistant, puis qu'il est asymptotiquement normal (A.N.) en précisant la loi limite et la vitesse de convergence.

2°. Démontrer que $\widehat{\theta}_2$ est un estimateur plug-in de θ.

3°. Déterminer

$$\mathbb{P}\left[n(\widehat{\theta}_2-\theta)\le x\right] \tag{9}$$

pour x variant dans $\mathbb{R}$ et en déduire la loi limite de $n(\widehat{\theta}_2-\theta)$ lorsque n tend vers l'infini. Préciser la vitesse de convergence.

4°. Déterminer la fonction de répartition de $\widehat{\theta}_2$ et sa fonction de densité. En déduire son espérance et sa variance. $\widehat{\theta}_2$ converge-t-il au sens des moindres carrés ?

5°. Démontrer que $\widehat{\theta}_3$ est un estimateur plug-in de θ, déterminer

$$\mathbb{P}\left[n(\widehat{\theta}_3-\theta)\le x\right] \tag{10}$$

pour x variant dans $\mathbb{R}$ et en déduire la loi limite de $n(\widehat{\theta}_3-\theta)$ lorsque n tend vers l'infini. Préciser la vitesse de convergence.

6°. Démontrer que la vraisemblance de l'échantillon est maximale sur l'intervalle $[x_{(n)}-1; x_{(1)}]$. En déduire un estimateur du maximum de vraisemblance. Est-il unique ?

1.10 Loi de Poisson : comparaison de deux estimateurs.

Soit X une va suivant une loi de Poisson de paramètre θ. Soit $(X_1,\dots,X_n)$ un n-échantillon de X.

1°. Déterminer deux estimateurs de θ à partir de la moyenne et de la variance de l'échantillon.

2°. Comparer ces estimateurs.

1.11 Loi de Pareto : estimation du paramètre de position par trois estimateurs.

Soit $\theta > 0$ et soit X une variable aléatoire de densité

$$f(t) = \frac{3t^2}{\theta^3}\mathbb{1}_{[0,\theta]}(t)$$

Soit $(X_1,\dots,X_n)$ un n-échantillon de X. On pose :

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k \text{ , } T_n = \max(X_1,\dots,X_n)$$

$$\text{et } Z_n = (X_1\dots X_n)^{1/n}$$

1°. Déterminer la fonction de répartition de X, puis calculer l'espérance et la variance de S_n.

2°. Déterminer un estimateur sans biais $\widehat{S_n}$ de θ de la forme αS_n. Calculer sa variance et montrer qu'il converge en moyenne quadratique vers θ.

3°. Donner une densité de T_n et démonter qu'il converge en moyenne quadratique vers θ.

4°. Calculer $\mathbb{E}\left[X^{1/n}\right]$, $\mathbb{E}[Z_n]$ et déterminer a tel que $\lim_{n\to+\infty}\mathbb{E}[\widehat{Z_n}] = \theta$ où $\widehat{Z_n} = aZ_n$. Calculer la variance et en déduire la convergence en moyenne quadratique.

5°. Lequel de ces trois estimateurs est-il préférable de choisir ?

1.12 Loi de Pareto : différents estimateurs.

Soit $n \ge 3$ un entier et α,β deux réels strictement positifs. Soit X la variable aléatoire réelle de densité

$$f(t) = \frac{\alpha\beta^\alpha}{t^{\alpha+1}}\mathbb{1}_{[t\ge\beta]} \tag{11}$$

On dit que X suit une loi de Paréto notée $\mathscr{P}(\alpha,\beta)$.

1°. Déterminer la fonction de répartition F de X, son espérance et sa variance.

2°. Déterminer la loi de la variable $Y = \ln(X/\beta)$.

3°. On suppose $\alpha > 2$ connu et on veut estimater β. On pose $Z_n = \min(X_1,\dots,X_n)$. Déterminer la loi de Z_n, son espérance, sa variance, puis déterminer un réel c_n tel que $\widehat{Z_n} = c_n Z_n$ soit un estimateur sans biais de β.

4°. Montrer que $\widehat{Z_n}$ converge en moyenne quadratique vers β.

On suppose maintenant β connu et on souhaite estimer α. on pose

$$W_n = \frac{n}{\sum_{k=1}^{n}\ln(X_k/\beta)} \tag{12}$$

5°. Déterminer la densité de $\sum_{k=1}^{n}\ln(X_k/\beta)$, calculer $\mathbb{E}[W_n]$ et en déduire que $\widehat{W}_n = \frac{n-1}{n}W_n$ est un estimateur sans biais de α.

6°. Calculer sa variance.

7°. On suppose dans cette question que $1<\alpha<2$. On effectue 402 observations de X et on trouve que $\widehat{W}_{402} = 1,4$. Donner un intervalle de confiance pour α au risque $0,05$.