

Examen final de Statistique — mercredi 6 mai 2026

Exercice 1 — La loi « inverse-gamma » de paramètres $\lambda, \alpha > 0$, notée $\Gamma^{-1}(\lambda, \alpha)$ dans la suite, est la loi de la variable aléatoire $1/X$ lorsque X est elle-même de loi $\Gamma(\lambda, \alpha)$. On rappelle que la densité sur \mathbb{R}_+ est donnée par $\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$.

1. Montrer que la densité sur \mathbb{R}_+ de la loi inverse-gamma est donnée par $\frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{x^{\alpha+1}} e^{-\frac{\lambda}{x}}$, et montrer que son espérance est égale à $\lambda/(\alpha - 1)$ si $\alpha > 1$.
2. Qui a la plus grande entropie : $\Gamma(1, 1)$ ou $\Gamma^{-1}(1, 2)$?

Exercice 2 — On observe deux populations X_1, \dots, X_n et Y_1, \dots, Y_m ; toutes ces variables sont indépendantes et gaussiennes de variance σ^2 inconnue, mais les n variables X_i ont pour moyenne μ tandis que les m variables Y_i ont une moyenne ν . On cherche à construire un intervalle de confiance pour la différence de moyennes $\delta = \mu - \nu$.

1. Soient \bar{X}_n, \bar{Y}_m les moyennes empiriques de chaque population. Montrer que $\hat{\delta} = \bar{X}_n - \bar{Y}_m$ est un estimateur sans biais de δ . Donner sa loi.
2. Montrer que $\hat{\sigma}^2$ est un estimateur sans biais de σ^2 , où

$$\hat{\sigma}^2 := \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{j=1}^m (Y_j - \bar{Y}_m)^2}{n + m - 2}$$

3. Montrer que, à une constante multiplicative près, $\hat{\sigma}^2$ suit une loi du chi-deux dont on précisera le degré de liberté, et que $\hat{\sigma}^2$ et $\hat{\delta}$ sont indépendants.
4. En déduire un intervalle de confiance pour δ et un test de l'hypothèse nulle $\mu = \nu$.

Exercice 3 — On observe un échantillon de variables aléatoires X_1, \dots, X_n iid selon une même loi P , à valeur dans \mathbb{N} . On cherche à tester l'hypothèse nulle $H_0 : P = \text{Poisson}(2)$ contre l'hypothèse alternative $H_1 : P = \text{Geom}(1/2)$, où on a défini la loi géométrique par $P(X = k) = (1/2)^{k+1}$ pour $k \in \{0, 1, 2, \dots\}$. Proposer un test de H_0 contre H_1 .

Exercice 4 — On observe un vecteur aléatoire $X \sim \mathcal{N}(\mu, I_n)$ dans \mathbb{R}^n .

1. Montrer que l'estimateur du maximum de vraisemblance de μ est $\hat{\mu} = X$, et que son risque quadratique $\mathbb{E}[|\hat{\mu} - \mu|^2]$ vaut n .
2. On se place maintenant dans une optique bayésienne. On choisit un a priori gaussien, à savoir $\mu \sim \mathcal{N}(m\mathbf{1}, \sigma^2 I_n)$ avec $m \in \mathbb{R}, \sigma^2 > 0$, et $\mathbf{1}$ est le vecteur de taille n dont toutes les composantes sont égales à 1. Pour commencer, on suppose m et σ^2 connus.
 - (a) Montrer que la loi a posteriori de μ est la loi gaussienne de moyenne $\frac{m\mathbf{1} + \sigma^2 X}{1 + \sigma^2}$ et de variance $\frac{\sigma^2}{1 + \sigma^2} I_n$.
 - (b) En déduire que l'estimateur MMSE est égal à $\hat{\mu}_* = (1 - \lambda)m\mathbf{1} + \lambda X$, où $\lambda = \sigma^2/(1 + \sigma^2)$.
 - (c) Calculer le risque quadratique de $\hat{\mu}_*$. Qui est meilleur : $\hat{\mu}$ ou $\hat{\mu}_*$?
3. On suppose maintenant que m et σ^2 sont inconnus.
 - (a) Montrer que la loi marginale de X est $\mathcal{N}(m\mathbf{1}, (\sigma^2 + 1)I_n)$.
 - (b) En déduire que $\hat{m} = n^{-1}(X_1 + \dots + X_n)$ est un estimateur sans biais de m .
 - (c) Quelle est la loi de $S = \sum_{i=1}^n (X_i - \hat{m})^2$?

(d) En déduire que si $n > 3$, alors

$$\hat{\lambda} = 1 - \frac{n-3}{S}$$

est un estimateur sans biais de λ .

4. L'estimateur de James-Stein de μ est défini par $\hat{\mu}_{\text{JS}} = (1 - \hat{\lambda})\hat{m}\mathbf{1} + \hat{\lambda}X$. Montrer qu'il est sans biais.

Exercice 5 — *Loi de Hotelling et erreur de prédiction.* On se place dans un modèle linéaire classique et gaussien avec n observations de la forme (x_i, y_i) , où $y_i = x_i^\top \beta + \varepsilon_i$. Ici, les $y_i \in \mathbb{R}$ sont des variables à expliquer, $x_i \in \mathbb{R}^{p,1}$ sont des variables explicatives, $\beta \in \mathbb{R}^{p,1}$ est le vecteur des coefficients et les $\varepsilon_i \in \mathbb{R}$ sont des erreurs iid de loi $\mathcal{N}(0, \sigma^2)$. On suppose que $p < n - 2$.

1. Donner la forme et la loi de l'estimateur des moindres carrés $\hat{\beta}$ de β .
2. Après avoir estimé $\hat{\beta}$, on reçoit une nouvelle réalisation $\xi \sim \mathcal{N}(0, I_p)$, $\xi \in \mathbb{R}^{p,1}$, et on cherche à prédire la vraie valeur de sortie du modèle $y = \xi^\top \beta + \varepsilon'$, où $\varepsilon' \sim \mathcal{N}(0, \sigma^2)$ est une nouvelle erreur, indépendante des ε_i . On pose naturellement $\hat{y} = \xi^\top \hat{\beta}$. Montrer que l'erreur de prédiction $\mathbb{E}[|y - \hat{y}|^2]$ est égale à

$$\sigma^2 \mathbb{E}[1 + \langle \xi, (X^\top X)^{-1} \xi \rangle]$$

où X est la matrice $n \times p$ dont les lignes sont les x_i^\top .

3. On va calculer la loi de $T = \langle \xi, (X^\top X)^{-1} \xi \rangle$, appelée *statistique de Hotelling*. On suppose dorénavant que les variables explicatives x_i ont elles-mêmes été tirées selon la loi $\mathcal{N}(0, I_p)$. On note $M = X^\top X$.
 - (a) Montrer que pour n'importe quelle matrice $O \in \mathbb{R}^{p,p}$ orthonormale ne dépendant pas de X , la variable T a la même loi que $\langle O\xi, (X^\top X)^{-1} O\xi \rangle$.
 - (b) En choisissant judicieusement O , en déduire que T a la même loi que $|\xi|^2 (M^{-1})_{1,1}$.
 - (c) Quelle est la loi de $|\xi|^2$?
 - (d) (*Aparté algébrique.*) Soit H une matrice par blocs, que l'on suppose symétrique et inversible, soit H^{-1} sa matrice inverse. On suppose que ces matrices peuvent s'écrire par blocs sous la forme

$$H = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}, \quad H^{-1} = \begin{pmatrix} a & b \\ b^\top & c \end{pmatrix}.$$

En écrivant les blocs de la première colonne de la matrice HH^{-1} , montrer la *formule de Schur*,

$$a = (A - BC^{-1}B^\top)^{-1}.$$

(e) On note $Z_1, \dots, Z_p \in \mathbb{R}^{n,1}$ les *colonnes* de X , et on note

$$Z = \begin{pmatrix} Z_1^\top \\ \vdots \\ Z_p^\top \end{pmatrix} \in \mathbb{R}^{(p-1) \times n}.$$

Donner la loi des Z_i et montrer que

$$\frac{1}{(M^{-1})_{1,1}} = |Z_1|^2 - \langle Z_1, Z^\top (ZZ^\top)^{-1} Z Z_1 \rangle.$$

En déduire que tout ceci est égal à $|(I_n - Z^\top (ZZ^\top)^{-1} Z) Z_1|^2$.

(f) Montrer que $\frac{1}{(M^{-1})_{1,1}} \sim \chi^2(n - p + 1)$, puis en déduire la loi de $\langle \xi, (X^\top X)^{-1} \xi \rangle$.

4. Conclure que l'erreur de prédiction est égale à

$$\sigma^2 \left(1 + \frac{p}{n - p - 1} \right).$$

Comment cette erreur varie-t-elle en fonction de p ?

Solution de l'exercice 1

Si f est une fonction test, alors

$$\begin{aligned}\mathbb{E}[f(1/X)] &= \int_0^\infty f(1/x) \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \int_0^\infty f(y) \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{-\alpha+1} e^{-\frac{\lambda}{y}} \frac{dy}{y^2} \\ &= \int_0^\infty f(y) \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} e^{-\frac{\lambda}{y}} dy\end{aligned}$$

et on obtient la densité demandée. Pour la moyenne, il suffit de remarquer que

$$\mathbb{E}[1/X] = \int_0^\infty x^{-1} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1)}{\lambda^{\alpha-1}} = \frac{\lambda}{\alpha-1},$$

cette formule n'étant bien entendu valable que si $\alpha - 1 > 0$, c'est-à-dire $\alpha > 1$. Donc la moyenne est donnée par

$$\mathbb{E}[1/X] = \frac{\lambda}{\alpha-1}$$

Pour l'entropie, il suffit de remarquer que les deux lois ont la même espérance, qui est précisément 1. Le théorème de Boltzmann-Gibbs dit que celle qui a la plus grande entropie sous la contrainte $\mathbb{E}[X] = 1$ est le modèle exponentiel associé à X , c'est-à-dire la loi exponentielle de paramètre 1, autrement dit $\Gamma(1, 1)$. Il n'y a pas besoin de faire les calculs.

Solution de l'exercice 2

Il est clair que $\bar{X}_n \sim N(\mu, \sigma^2/n)$ et $\bar{Y}_m \sim N(\nu, \sigma^2/m)$. Donc $\hat{\delta} = \bar{X}_n - \bar{Y}_m \sim N(\delta, \sigma^2(\frac{1}{n} + \frac{1}{m}))$. Nous avons maintes fois vu en cours que

$$\mathbb{E}[\sum (X_i - \bar{X}_n)^2] = (n-1)\sigma^2$$

et de façon analogue que

$$\mathbb{E}[\sum (Y_j - \bar{Y}_m)^2] = (m-1)\sigma^2$$

donc par linéarité de l'espérance, on voit que

$$\mathbb{E} \left[\sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{j=1}^m (Y_j - \bar{Y}_m)^2 \right] = (n+m-2)\sigma^2$$

et donc $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$ ce qui montre que $\hat{\sigma}^2$ est un estimateur sans biais de σ^2 . En fait, $\hat{\sigma}^2$ est une moyenne pondérée des variances empiriques de chaque population, que l'on notera momentanément $\hat{\sigma}_X^2$ et $\hat{\sigma}_Y^2$. En cours, nous avons vu (c'est une conséquence du théorème de Cochran) que $(n-1)\hat{\sigma}_X^2/\sigma^2 \sim \chi^2(n-1)$ et $(m-1)\hat{\sigma}_Y^2/\sigma^2 \sim \chi^2(m-1)$. Comme $\hat{\sigma}_X^2$ et $\hat{\sigma}_Y^2$ sont indépendants, on en déduit que

$$\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{j=1}^m (Y_j - \bar{Y}_m)^2}{\sigma^2} \sim \chi^2(n+m-2),$$

et donc que

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n+m-2} \chi^2(n+m-2).$$

Par ailleurs, le théorème de Cochran dit aussi que $\hat{\sigma}_X^2$ et \bar{X}_n sont indépendants, et que $\hat{\sigma}_Y^2$ et \bar{Y}_m sont indépendants. Comme les deux populations sont elles-mêmes indépendantes, on en déduit que $\hat{\sigma}^2$ et $\hat{\delta}$

peuvent tous deux s'écrire comme fonctions de variables aléatoires qui sont indépendantes, et donc qu'ils sont indépendants.

Finalement, on a

$$\frac{\hat{\delta} - \delta}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0, 1)$$

et

$$\frac{\hat{\sigma}}{\sigma} \sim \sqrt{\frac{\chi^2(n+m-2)}{n+m-2}}$$

ce qui démontre que

$$T = \frac{\hat{\delta} - \delta}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\frac{\hat{\delta} - \delta}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\frac{\hat{\sigma}}{\sigma}} \sim \mathcal{T}(n+m-2).$$

En notant t_α le quantile symétrique d'ordre $1 - \alpha$ de la loi de Student à $n + m - 2$ degrés de liberté, on a donc $\mathbb{P}(|T| \leq t_\alpha) = 1 - \alpha$. Ainsi, en pivotant, on voit que l'intervalle de confiance pour δ est donné par

$$\left[\hat{\delta} - t_\alpha \hat{\sigma} \sqrt{\frac{1}{n} + \frac{1}{m}}, \hat{\delta} + t_\alpha \hat{\sigma} \sqrt{\frac{1}{n} + \frac{1}{m}} \right].$$

Pour le test, on rejettera l'hypothèse nulle si 0 n'est pas dans cet intervalle.

Solution de l'exercice 3

Trivial.

Solution de l'exercice 4

La vraisemblance du modèle étant proportionnelle à $\ell(\mu) = \exp(-|X - \mu|^2/2)$, on voit immédiatement que maximiser la vraisemblance revient à trouver le minimum de $\mu \mapsto |X - \mu|^2$, lequel est évidemment atteint en X . Son risque quadratique vaut

$$\mathbb{E}[|X - \mu|^2] = n \times 1 = n.$$

La loi a posteriori de μ sachant X est proportionnelle à

$$e^{-|\mu - m\mathbf{1}|^2/2\sigma^2} \times e^{-|X - \mu|^2/2}.$$

Le double de l'opposé du terme dans l'exponentielle vaut

$$\frac{|\mu|^2}{\sigma^2} + \frac{|m\mathbf{1}|^2}{\sigma^2} - \frac{2\langle \mu, m\mathbf{1} \rangle}{\sigma^2} + |X|^2 + |\mu|^2 - 2\langle \mu, X \rangle.$$

En regroupant les termes en μ , on voit que ceci est égal à

$$|\mu|^2 \left(\frac{1}{\sigma^2} + 1 \right) - 2\langle \mu, \frac{m\mathbf{1}}{\sigma^2} + X \rangle + \heartsuit$$

où \heartsuit ne dépend pas de μ . En factorisant, on voit donc que ce terme est égal à

$$\left(1 + \frac{1}{\sigma^2} \right) \left| \mu - \frac{m\mathbf{1} + \sigma^2 X}{1 + \sigma^2} \right|^2 + \heartsuit\heartsuit$$

où encore \heartsuit ne dépend pas de μ . On en déduit que la loi a posteriori de μ sachant l'observation X est une loi gaussienne

$$\mathcal{N}\left(\frac{m\mathbf{1} + \sigma^2 X}{1 + \sigma^2}, \frac{\sigma^2}{1 + \sigma^2}\right).$$

L'estimateur MMSE est donc donné par

$$\hat{\mu}_* = \mathbb{E}[\mu|X] = \frac{m\mathbf{1} + \sigma^2 X}{1 + \sigma^2} = m\lambda\mathbf{1} + X(1 - \lambda)$$

où $\lambda = 1/(1 + \sigma^2)$. Son risque quadratique vaut

$$\mathbb{E}[|\hat{\mu}_* - \mu|^2] = \mathbb{E}\left[\left|\frac{m + \sigma^2 X}{1 + \sigma^2} - \mu\right|^2\right] = \mathbb{E}[|\lambda(\mu - m) + (1 - \lambda)(X - \mu)|^2]$$

c'est-à-dire, par indépendance, $\lambda^2 \times n\sigma^2 + (1 - \lambda)^2 \times n \times 1$. En factorisant et en réduisant sous le même dénominateur, on obtient

$$n \frac{\sigma^2}{\sigma^2 + 1} < n.$$

C'est mieux que l'estimateur du maximum de vraisemblance! Évidemment, à ce stade, il n'y a rien d'étonnant : nous avons utilisé un a priori assez fort, qui impliquait une information supplémentaire ignorée par l'estimateur du maximum de vraisemblance, à savoir que les données sont toutes proches de la moyenne m .

Maintenant, calculons la marginale de X . Il est clair que $X = \mu + \epsilon$ où $\epsilon \sim \mathcal{N}(0, I_n)$, et $\mu = m\mathbf{1} + \xi$ où $\xi \sim \mathcal{N}(0, \sigma^2 I_n)$. Donc $X \sim \mathcal{N}(m\mathbf{1}, (\sigma^2 + 1)I_n)$. Il est donc évident que les coordonnées X_i de X sont indépendantes et centrées en m , donc leur moyenne empirique est évidemment un estimateur sans biais de m , tandis que par le théorème de Cochran, $S/(1 + \sigma^2)$ suit une loi $\chi_2(n - 1) = \Gamma((n - 1)/2, 1/2)$. Le calcul de $\mathbb{E}[1/S]$ se fait facilement, soit en invoquant l'exercice 1, soit en refaisant le calcul :

$$\begin{aligned} \mathbb{E}\left[\frac{1 + \sigma^2}{S}\right] &= \frac{1}{\sqrt{2}^{n-1} \Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty e^{-s/2} s^{\frac{n-1}{2}-1} \frac{1}{s} ds \\ &= \frac{1}{\sqrt{2}^{n-1} \Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty e^{-s/2} s^{\frac{n-3}{2}-1} ds \\ &= \frac{\sqrt{2}^{n-3} \Gamma\left(\frac{n-3}{2}\right)}{\sqrt{2}^{n-1} \Gamma\left(\frac{n-1}{2}\right)} \\ &= \frac{1}{2} \times \frac{2}{n-3} \\ &= \frac{1}{n-3}. \end{aligned}$$

On en déduit que

$$\mathbb{E}\left[\frac{n-3}{S}\right] = \frac{1}{1 + \sigma^2}$$

et donc que $1 - (n - 3)/S$ est bien un estimateur sans biais de λ .

Solution de l'exercice 5

$\hat{\beta} = (X^\top X)^{-1} X^\top Y \sim \mathcal{N}(\beta, \sigma^2 (X^\top X)^{-1})$. Donc en particulier $y - \hat{y} = \xi^\top \beta + \epsilon' - \xi^\top \hat{\beta} = \xi^\top (\beta - \hat{\beta}) + \epsilon'$ suit une loi $\mathcal{N}(0, \sigma^2 \xi^\top (X^\top X)^{-1} \xi + \sigma^2)$. Donc l'erreur de généralisation est bien égale à $\mathbb{E}[\sigma^2(1 + \xi^\top (X^\top X)^{-1} \xi)]$.

Comme les lignes de X sont iid invariantes par rotations, X a la même loi que XO^\top , et donc $(X^\top X)^{-1}$ a la même loi que $((XO^\top)^\top(XO^\top))^{-1} = (OX^\top XO^\top)^{-1} = O^\top(X^\top X)^{-1}O$. Donc T a la même loi que $\langle O\xi, (O^\top X^\top XO)^{-1}O\xi \rangle$. On commence par échantillonner ξ , puis on choisit pour O la matrice orthogonale qui envoie ξ vers $|\xi|e_1$ afin d'obtenir $\langle |\xi|e_1, (X^\top X)^{-1}|\xi|e_1 \rangle = |\xi|^2(M^{-1})_{1,1}$. Clairement la loi de $|\xi|^2$ est $\chi^2(p)$.

Pour la formule de Schur, il suffit d'écrire deux blocs de $HH^{-1} = I$:

$$I = Aa + Bb^\top \quad 0 = B^\top a + Cb^\top$$

d'où on déduit que $b^\top = -C^{-1}B^\top a$ puis $I = Aa - BC^{-1}B^\top a = (A - BC^{-1}B^\top)a$. Donc finalement on retrouve la formule de Schur : $a = (A - BC^{-1}B^\top)^{-1}$. Appliqué à la matrice $X^\top X$, qui peut s'écrire par blocs sous la forme

$$\begin{pmatrix} |Z_1|^2 & R \\ R^\top & ZZ^\top \end{pmatrix}$$

avec $R_i = \langle Z_i, Z_1 \rangle$ et donc $R^\top = ZZ_1$ (vecteur colonne), on obtient effectivement

$$((M^{-1})_{1,1})^{-1} = |Z_1|^2 - Z_1^\top Z^\top (ZZ^\top)^{-1} ZZ_1.$$

Ceci n'est rien d'autre que $|(I - P)Z_1|^2$ où P est la projection orthogonale sur le sous-espace engendré par les colonnes de Z^\top , qui sont précisément Z_2, \dots, Z_p . Donc $I - P$ est un projecteur orthogonal sur un espace de dimension $n - (p - 1)$. Attention, il faut bien s'assurer que les Z_i sont linéairement indépendants. C'est le cas car la densité des Z_i est ac par rapport à Lebesgue sur \mathbb{R}^{np} .

Le théorème de Cochran suffit à conclure pour dire que $|(I - P)Z_1|^2 \sim \chi^2(n - (p - 1))$. Finalement, nous avons démontré que

$$\langle \xi, (X^\top X)^{-1} \xi \rangle \sim \frac{\chi^2(p)}{\chi^2(n - p + 1)}.$$

C'est presque une loi de Fisher : plus précisément,

$$\langle \xi, (X^\top X)^{-1} \xi \rangle \times \frac{n - p + 1}{p} \sim \mathcal{F}(p, n - p + 1).$$

Pour calculer l'espérance de cette loi, on utilise l'indépendance : en effet, $M_{1,1}^{-1}$ ne dépend que de X et pas de ξ . Donc,

$$\mathbb{E}[\langle \xi, (X^\top X)^{-1} \xi \rangle] = \mathbb{E}[\chi^2(p)] \times \mathbb{E}\left[\frac{1}{\chi^2(n - p + 1)}\right].$$

L'espérance de la loi du $\chi^2(k)$ est bien sûr k ; pour l'espérance de l'inverse, il suffit de remarquer que la loi du $\chi^2(k)$ est une loi $\Gamma(k/2, 1/2)$ et donc l'exercice 1 montre que l'espérance de l'inverse est égale à $(1/2)/(k/2 - 1) = 1/(k - 2)$. On en conclut que

$$\mathbb{E}[\langle \xi, (X^\top X)^{-1} \xi \rangle] = \frac{p}{n - p - 1}.$$

Ce qui est frappant, c'est que la fonction $p \mapsto p/(n - 1 - p)$ a pour dérivée $((n - 1 - p) + p)/(n - 1 - p)^2 > 0$: autrement dit, l'erreur de généralisation est croissante.