Université de Strasbourg Département de mathématique
L3 Mathématiques Appliquées/DUAS
S6 année 2024-2025

Travaux Dirigés de l'UE Statistique Mathématique S6

Table des matières

Rappels sur les variables aléatoires discrètes

Exercice 1

Pour chacune des lois suivantes, rappeler $\mathbb{P}(X = k)$, les valeurs possibles pour k, et calculer l'espérance et la variance :

1. Loi de Bernoulli(p) ;
2. Loi Binomiale(n, p) ;
3. Loi Géométrique(p) ;
4. Loi de Poisson(λ).

Exercice 2

Donner la définition de la fonction génératrice d'une variable X à valeurs entières positives. Cette fonction génératrice détermine complètement la loi de X. Plus précisément, montrer que, pour tout $n \geqslant 0$, on a

$$\mathbb{P}(X = n) = \frac{G_X^{(n)}(s)|_{s=0}}{n\,!}$$

où $G_X^{(n)}$ désigne la $n-$ème dérivée de G_X.

Exercice 3

Calculer la fonction génératrice pour chacune des lois :

1. Loi de Bernoulli(p) ;
2. Loi Binomiale(n, p) ;
3. Loi Géométrique(p) ;
4. Loi de Poisson(λ).

Retrouver l'espérance et la variance d'une variable aléatoire X suivant chacune de ces lois en utilisant la fonction génératrice.

1 Variables continues

Exercice 1

1. Montrer que la fonction F définie par

$$F(x) = \begin{cases} \frac{e^x}{2} & \text{pour } x < 0, \\ 1 & \text{pour } x \geqslant 0. \end{cases}$$

   est une fonction de répartition.
2. La fonction f telle que $f(x) = F'(x)$ est-elle une densité de probabilité ?

Exercice 2

Soit X une variable aléatoire réelle de densité $f(x) = x\,e^{-x^2/2}\mathbf{1}_{\{x>0\}}$.

1. Vérifier que f est une densité de probabilité.

2. Montrer que $Y = X^2$ est une variable aléatoire continue, dont on précisera la densité. Reconnaître la loi de Y .
3. Calculer l'espérance et la variance de Y.

Exercice 3

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires telle que pour chaque n, X_n a la même loi que X_1. On suppose que $\lim_{x\to+\infty} x\mathbb{P}(|X_1|>x)=0$. On pose $M_n := \max_{1\leqslant i\leqslant n}|X_i|$.

Montrer que M_n/n converge en probabilité vers 0.

Exercice 4

Soit X une variable aléatoire de fonction de répartition F continue strictement croissante. Soit $Y = F(X)$ une nouvelle variable aléatoire. Quelle est la loi de Y ?

Exercice 5

Soient X_1, X_2 deux variables indépendantes de lois exponentielles de paramètres λ_1, λ_2.

1. Calculer la loi de (m, M) où $m := \min\{X_1, X_2\},\ M := \max\{X_1, X_2\}$.
2. Calculer les lois de m et M.
3. On suppose que X_1 et X_2 représentent les durées de vie de deux ampoules distinctes numérotées 1 et 2, que l'on suppose indépendantes. Déterminer la probabilité que la première ampoule ait une durée de vie supérieure à la deuxième.
4. Supposant que $\lambda_1 = \lambda_2$, montrer que m et $M - m$ sont des variables indépendantes.

Exercice 6

Calculer les fonctions caractéristiques des lois usuelles suivantes : binomiale, Poisson, uniforme, exponentielle.

Exercice 7

Soit Y une variable aléatoire de loi exponentielle $\lambda > 0$ et ε une variable aléatoire discrète indépendante de Y et telle que $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$.

1. Montrer que la variable aléatoire $Z = \varepsilon Y$ est à densité et la calculer. Cette loi est appelée loi exponentielle symétrique.
2. Calculer la fonction caractéristique de $Z = \varepsilon Y$.

2 Vecteurs aléatoires continus

Exercice 1

Soient X et Y deux variables aléatoires indépendantes de loi respective $\mathrm{Gamma}(a, \lambda)$ et $\mathrm{Gamma}(b, \lambda)$ avec $a, b, \lambda \in]0, \infty[$.

1. Calculer la loi du couple $\left(X + Y, \frac{X}{X+Y}\right)$.
2. Montrer que $X + Y$ et $\frac{X}{X+Y}$ sont indépendantes et identifier leur loi.

Exercice 2

Soient X_1, X_2 des variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$.

1. Montrer que X_1^2 suit la loi $\chi^2(1)$, i.e., une Gamma$(1/2, 1/2)$.
2. Montrer que $X_1^2 + X_2^2$ suit la loi $\chi^2(2)$, i.e., une Gamma$(1, 1/2)$.

Exercice 3

Soient X, Y deux variables aléatoires réelles indépendantes, de même loi $\mathcal{N}(0, \sigma^2)$.

1. Montrer que $X+Y$ et $X-Y$ sont indépendantes et ont la même loi, laquelle ?
2. Calculer la densité de X/Y.

Exercice 4

Soient X_1, X_2 et X_3 des variables aléatoires indépendantes de densité $f(x) = e^{-x}, x > 0$. On pose

$$Y_1 = \frac{X_1}{X_1 + X_2}, Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}, Y_3 = X_1 + X_2 + X_3.$$

Calculer la densité de (Y_1, Y_2, Y_3) et en déduire l'indépendance de ces trois nouvelles variables.

Exercice 5

Soit $f_{X|Y}(x|y) = c_1 x/y^2, 0 < x < y, 0 < y < 1$ et $f_Y(y) = c_2 y^4, 0 < y < 1$, la densité conditionnelle de X sachant $Y = y$ et la densité marginale de Y, respectivement.

1. Déterminer les constantes c_1 et c_2.
2. Déterminer la densité conjointe de X et Y.
3. Calculer $\mathbb{P}\left[\frac{1}{4} < X < \frac{1}{2} | Y = \frac{5}{8}\right]$.
4. Calculer $\mathbb{P}\left[\frac{1}{4} < X < \frac{1}{2}\right]$.

Exercice 6

Soit $X|\{\Theta = \theta\}$ de loi exponentielle(θ) et Θ de loi Gamma(α, λ).

1. Démontrer que X est de loi de Pareto avec densité

$$f_X(x) = \frac{\alpha \lambda^\alpha}{(\lambda + x)^{\alpha+1}}, x > 0.$$

2. Démontrer que $\mathbb{E}[X] = \lambda(\alpha - 1)^{-1}$ si $\alpha > 1$.

Exercice 7

Soit $f \colon \mathbb{R}^n \to \mathbb{R}$ la fonction définie par

$$f(x_1, \ldots, x_n) = \prod_{j=1}^n \mathbf{1}_{0 < x_j < 1} - \prod_{j=1}^n g(x_j),$$

où $g(t) := \mathbf{1}_{\{0 < t < 1/2\}} - \mathbf{1}_{\{1/2 < t < 1\}}$.

1. Démontrer que f est une densité.

2. Soit $(X_1, \ldots, X_n)$ un vecteur aléatoire ayant pour densité f. Déterminer la densité de $(X_1, \ldots, X_{n-1})$.
3. La collection de variables aléatoires $(X_i)_{i=1}^n$ est-elle indépendante ?

3 Théorèmes limites

Exercice 1

Une urne contient trois boules numérotées 1, 2, 3. On effectue une suite de n tirages indépendants avec remise, chaque boule ayant la même probabilité d'être tirée. On note X_i le numéro de la boule tirée lors du $i-$ème tirage.
Soit $S = X_1 + ... + X_n$. Trouver le plus petit nombre s positif tel que

$$\mathbb{P}[2n - s \leqslant S \leqslant 2n + s] \geqslant 0.9$$

1. au moyen de l'inégalité de Markov,
2. au moyen de l'inégalité de Tchébychev.

Exercice 2

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de loi exponentielle de paramètre λ_n. Étudier la convergence en loi dans le cas où $\lim_{n\to\infty} \lambda_n = \lambda \in [0, \infty[$.

Exercice 3

Soit $(U_i)_{i\geqslant 1}$ une suite de variables aléatoires indépendantes de loi uniforme sur l'intervalle $[0, \theta]$, où $\theta > 0$. On pose pour $n \geqslant 1, X_n = \max_{1\leqslant i\leqslant n} U_i$.

1. Montrer que $(X_n)_{n\geqslant 1}$ converge en probabilité et déterminer sa limite. On pourra calculer $\mathbb{P}(|X_n - \theta| > \varepsilon)$ pour $\varepsilon > 0$.
2. Etudier la convergence en loi de la suite $(n(\theta - X_n))_{n\geqslant 1}$.

Exercice 4 Calcul de limites d'intégrales

1. En considérant une suite de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$, calculer à l'aide de la loi (faible) des grands nombres

$$\lim_{n\to\infty} \int_{[0,1]^n} f\left(\frac{x_1 + ... + x_n}{n}\right) dx_1...dx_n$$

   où f est une application continue bornée de $\mathbb{R}$ dans $\mathbb{R}$.

2. Soit $(Y_k, k \geqslant 1)$ une suite de variables aléatoires indépendantes de loi de Poisson de paramètre $\alpha > 0$. Déterminer la loi de $Z_n = \sum_{k=1}^n Y_k$.
3. Montrer que la suite des moyennes empiriques $(\overline{Y}_n = Z_n/n, n \geqslant 1)$ converge vers une limite que l'on déterminera. Calculer, en s'inspirant de la question 1,

$$\lim_{n\to\infty} \sum_{k\geqslant 0} e^{-\alpha n} \frac{(\alpha n)^k}{k!} f\left(\frac{k}{n}\right)$$

   où $\alpha > 0$ et f est une application continue bornée de $\mathbb{R}$ dans $\mathbb{R}$.

Exercice 5

Soit $(X_n, n \geqslant 1)$ une suite de variables aléatoires indépendantes de loi de Poisson de paramètre 1. On note $S_n = \sum_{k=1}^{n} X_k$.

1. Montrer que $\frac{S_n - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.
2. Déterminer la loi de S_n, puis montrer que $\lim_{n\to\infty} e^{-n}\left(\sum_{k=0}^{n} n^k/k!\right) = \frac{1}{2}$.

4 Exhausitivité, famille libre, statistique complète

Exercice 1

Soient $X_1, \cdots, X_n$ des variables aléatoires indépendantes et de même loi, de densité

$$f(x) = \frac{1}{\log\alpha}\frac{1}{x}\mathbf{1}_{x\in(\theta,\alpha\theta)}$$

où $\alpha > 1$ est un paramètre connu et $\theta > 0$ est inconnu. Soit $X_{1,n} < ... < X_{n,n}$ l'échantillon ordonné.

1. Montrez que $\phi(X_1, \cdots, X_n) = (X_{1,n}, X_{n,n})$ est une statistique exhaustive minimale.
2. Montrez que $\psi(X_1, \cdots, X_n) = \left(\frac{X_{n,n}}{X_{1,n}}, X_{n,n}X_{1,n}\right)$ est aussi exhaustive minimale.
3. Montrez que $\frac{X_{n,n}}{X_{1,n}}$ est une statistique libre.
4. Est-ce que $\psi(X_1, \cdots, X_n)$ est une statistique complète ?

Exercice 2

Soient $X_1, \cdots, X_n$ des variables aléatoires indépendantes et de même loi, uniforme sur l'intervalle $(\theta, \theta+1)$ où $\theta \in \mathbb{R}$. Soit $X_{1,n} < ... < X_{n,n}$ l'échantillon ordonné.

1. Montrer que la statistique $(X_{1,n}, X_{n,n})$ est une statistique exhaustive minimale.
2. En déterminant explicitement la loi de l'étendue $R = X_{n,n} - X_{1,n}$, montrez que R est une statistique libre.
3. Établir le même résultat que en 2, mais en utilisant le fait que le modèle est un modèle à paramètre de position.

5 Estimation ponctuelle, méthode des moments, méthode du maximum de vraisemblance

Exercice 1

On considère un échantillon de taille n de loi de Poisson $\mathcal{P}(\lambda)$.

1. Calculer l'espérance et la variance d'une loi de Poisson.
2. Calculer l'estimateur des moments de λ.
3. Calculer l'estimateur du maximum de vraisemblance de λ.

Exercice 2

Dans un laboratoire d'astrophysique, on s'intéresse à un certain type de particules cosmiques. Un détecteur a relevé les durées d'attente en heures entre les réceptions successives des premières particules captées. On admet que ces durées sont des réalisations de variables aléatoires indépendantes et de loi uniforme $\mathcal{U}(a, b)$.

1. Estimer les paramètres a et b de cette loi par la méthode des moments.
2. Même question pour le maximum de vraisemblance.

Exercice 3

Pour chacune des deux lois suivantes, on se donne des variables aléatoires indépendantes $X_1, \ldots, X_n$. Déterminer quels sont les estimateurs des moments et du maximum de vraisemblance.

1. $f(x) = \theta x^{\theta-1}\mathbf{1}_{x\in[0,1]}$
2. $f(x) = e^{-(x-\theta)}\mathbf{1}_{x\in[\theta,+\infty)}$.

Exercice 4

On considère un échantillon de variables aléatoires indépendantes $X_1, \ldots, X_n$ dont la densité est donnée par

$$f(x) = \frac{1}{\theta_2}e^{-(x-\theta_1)/\theta_2}\mathbf{1}_{x\in[\theta_1,+\infty)},$$

où $\theta_1 \in \mathbb{R}, \theta_2 \in \mathbb{R}_+^\star$.
Déterminer les estimateurs du maximum de vraisemblance de θ_1 et θ_2.

Exercice 5

La loi de Pareto a été utilisée initialement comme modèle de la distribution des revenus. Elle est définie par la fonction de répartition

$$F(x) = (1 - (\theta_1/x)^{\theta_2})\mathbf{1}_{x\in[\theta_1,+\infty)}.$$

Déterminer les estimateurs du maximum de vraisemblance de θ_1, θ_2.

Exercice 6

On considère le modèle de la loi uniforme $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{U}_{[0,\theta]} : \theta > 0\})$ et $X_1, ..., X_n$ des variables aléatoires indépendantes dans ce modèle, l'échantillon ordonné sera noté $X_{1,n} < ... < X_{n,n}$.
1. Donner une statistique exhaustive pour θ.
2. Calculer la densité de la loi $X_{1,n}$ conditionnelle à $\{X_{n,n} = x_n\}$. En déduire l'expression de $\mathbb{E}(X_{1,n}|X_{n,n} = x_n)$ puis de $\mathbb{E}(X_{1,n}|X_{n,n})$.
3. Déterminer alors un estimateur $\widetilde{\theta}$ de θ, estimateur amélioré par rapport à $\widehat{\theta} = X_{1,n} + X_{n,n}$ par le théorème de Rao-Blackwell suivant :
Théorème. *Soit* $(E, \mathcal{E}, \mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\})$ *un modèle paramétrique et* $\underline{X} = (X_1, ..., X_n)$ *un échantillon dans ce modèle. Soit* $T(\underline{X})$ *un estimateur de* $g(\theta)$ *de carré intégrable.*
Si le modèle possède une statistique exhaustive $S(\underline{X})$ *pour le paramètre* θ*, alors l'estimateur*

$$\widetilde{T}(\underline{X}) = \mathbb{E}(T(\underline{X})|S(\underline{X}))$$

de $g(\theta)$ a un risque quadratique inférieur à $T(\underline{X})$ pour tout $\theta \in \Theta$.
Si $T(\underline{X})$ est un estimateur sans biais de $g(\theta)$ alors $\widetilde{T}(\underline{X})$ est également sans biais pour $g(\theta)$ et l'inégalité sur les risques quadratiques se traduit sur les variances.
4. La statistique $X_{n,n}$ est-elle complète ? Pour cela on rappelle que
Statistique complète *Une statistique exhaustive T d'un modèle statistique paramétrique avec T à valeur dans $\mathbb{R}^d$ est dite complète si pour toute fonction borélienne $f : \mathbb{R}^d \to \mathbb{R}$ telle que $f(T)$ soit intégrable, on ait :*

$$\forall \theta \in \Theta, \qquad \mathbb{E}(f(T)) = 0 \Longrightarrow f = 0 \qquad \mathbb{P}_\theta - ps.$$

Conclure en utilisant le résultat suivant :
Théorème de Lehmann-Scheffé *Soit $(E, \mathcal{E}, \mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\})$ un modèle paramétrique et $\underline{X} = (X_1, ..., X_n)$ un échantillon dans ce modèle. Soit $T(\underline{X})$ un estimateur sans biais de $g(\theta)$ de carré intégrable et $S(\underline{X})$ une statistique exhaustive et complète de θ. Alors l'estimateur amélioré de Rao-Blackwell $\widetilde{T}(\underline{X}) = \mathbb{E}(T(\underline{X})|S(\underline{X}))$ est optimal dans la classe des estimateurs sans biais de $g(\theta)$.*

6 Biais et erreur quadratique moyenne

Exercice 1

On considère un échantillon de loi exponentielle $\mathcal{E}(\lambda)$.
1. Donner les estimateurs par la méthode des moments et du maximum de vraisemblance.
2. Ces estimateurs sont-ils sans biais ?

Exercice 2

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi normale $\mathcal{N}(0, \theta)$ où $\theta > 0$. Vérifiez que $\frac{1}{n}\sum_{i=1}^n X_i^2$ est un estimateur sans biais de θ.

Exercice 3

Soient X_1, X_2, X_3 des variables aléatoires indépendantes de loi uniforme $\mathcal{U}(0, \theta)$. On note $X_{1,n} < X_{2,n} < X_{3,n}$ l'échantillon ordonné.
1. Montrer que $4X_{1,n}$, $2X_{2,n}$ et $4X_{3,n}/3$ sont des estimateurs sans biais de θ.
2. Calculer leur variance.

Exercice 4

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi uniforme $\mathcal{U}(\theta, \theta + 1)$. On souhaite estimer θ par $X_{1,n} = \min_{1 \leqslant k \leqslant n} X_k$.
1. Calculer le biais de $X_{1,n}$.
2. Calculer la variance de $X_{1,n}$.
3. En déduire que $X_{1,n}$ est un estimateur convergent.

Exercice 5

Soient $\theta > 0$ et $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi uniforme $\mathcal{U}(0, \theta)$. On pose $\tilde{\theta}_n = 2\bar{X}_n$ et $\hat{\theta}_n = \max_{1 \leqslant k \leqslant n} X_k$.

1. Montrer que $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance et $\tilde{\theta}_n$ l'estimateur des moments de θ.
2. Regarder la convergence de $\hat{\theta}_n$ vers θ (en probabilité, presque surement et en erreur quadratique moyenne).
3. Regarder la convergence de $\tilde{\theta}_n$ vers θ (en probabilité, presque surement et en erreur quadratique moyenne).
4. Comparer les deux estimateurs en terme de risque quadratique.

Exercice 6

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi gamma $\mathcal{G}(m, 1/\theta)$ où m est un entier connu. La densité est donc

$$f_X(y) = \frac{1}{(m-1)!\theta^m} y^{m-1} e^{-y/\theta} \mathbf{1}_{y \in [0,\infty)}.$$

1. Montrer que $\hat{\theta} = 1/(nm) \sum_{i=1}^n X_i$ est un estimateur sans biais de θ.
2. Calculer sa variance. En déduire que c'est un estimateur convergent.

Exercice 7

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi normale $\mathcal{N}(\mu, \sigma^2)$. On considère les estimateurs de σ^2 de la forme

$$T(a) = a \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

1. Calculer le biais de $T(a)$ et trouver la valeur de a pour laquelle $T(a)$ est sans biais.
2. Trouver la valeur de a qui minimise le risque quadratique de tels estimateurs.

7 Efficacité

Exercice 1

Un système électronique est constitué de 4 composants identiques : un composant principal et trois composants d'appoint. Quand le premier est défaillant, le second le remplace instantanément. Quand le second est défaillant, le troisième le remplace instantanément. Quand le troisième est défaillant, le quatrième le remplace instantanément. La panne du quatrième composant entraîne la panne du système. Les durées de vies des composants sont indépendantes et identiquement distribuées selon une loi exponentielle $\mathcal{E}(1/\theta)$. On considère $X_1, \ldots, X_n$ des variables aléatoires indépendantes représentants les durées de vie de n systèmes.

1. Quelle est la loi de X_i ?
2. Trouver l'estimateur de maximum de vraisemblance $\widehat{\theta}$. Donner son espérance et sa variance.
3. Vérifier que le modèle est régulier.

4. Est-ce que $\widehat{\theta}$ est efficace ?
5. Quelle est la loi asymptotique de $\widehat{\theta}$?

Exercice 2

L'entreprise Miaou distribue un aliment pour chat dans un contenant métallique dont le poids après remplissage, représenté par la variable X, est calibré à une valeur nominale. On souhaite vérifier la calibration de la chaine de production. Des études antérieures sur cette chaine avaient montré que le poids était distribué selon une loi normale. On considère donc $X_1, \ldots, X_n$ des variables aléatoires indépendantes gaussiennes $\mathcal{N}(\mu, \sigma^2)$ avec σ^2 connue.

1. Vérifier que le modèle est régulier.
2. Montrer que $\widehat{\mu} = \overline{X}_n$ est un estimateur efficace de μ.

Exercice 3

On considère $X_1, \ldots, X_n$ des variables aléatoires indépendantes de densité

$$f(x) = 3\theta^3 (x + \theta)^{-4} \mathbf{1}_{\{x>0\}}.$$

1. Montrer que $2\overline{X}_n$ est un estimateur sans biais de θ.
2. Est-il convergent ?
3. Est-il efficace ?

Exercice 4

Afin de dimensionner la capacité des serveurs, un ingénieur réalise une étude sur la durée de connexion des clients au serveur. On admet que cette durée est distribuée selon une loi exponentielle $\mathcal{E}(\lambda)$. Donc, on considère $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi exponentielle $\mathcal{E}(\lambda)$.

1. Trouver l'estimateur du maximum de vraisemblance de λ. Est-il biaisé ? Si oui, en déduire un estimateur sans biais, noté $\widehat{\lambda}$.
2. Calculer l'efficacité de $\widehat{\lambda}$.

Exercice 5

On s'intéresse à la loi de Laplace (ou loi double exponentielle) introduite par Laplace en 1774 comme la première loi des erreurs. Il introduira la seconde loi des erreurs en 1778 avec la loi normale. On considère $X_1, \ldots, X_n$ des variables aléatoires indépendantes de densité

$$f(x) = \frac{1}{2\sigma} e^{-|x|/\sigma}.$$

1. Le modèle est-il régulier ? Si oui, déterminer l'information de Fisher (à une observation) du modèle paramétrique.
2. Déterminer l'estimateur $\widehat{\sigma}$ de σ par la méthode de maximum de vraisemblance. Est-il biaisé ?
3. Établir son comportement asymptotique.

Exercice 6

Soient (X_1, X_2, X_3) des variables aléatoires indépendantes et de densité

$$f(x) = \theta^{-1} e^{-x/\theta}, x \geqslant 0.$$

1. Calculer la MSE de $\widehat{\theta}_1 = \overline{X}_3$.
2. Calculer la MSE de $\widehat{\theta}_2 = X_{2,3}$, où $X_{1,3} < X_{2,3} < X_{3,3}$.
3. Comparer.

Exercice 7

On considère $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi Bernoulli $\mathcal{B}(p)$. Montrer que $\overline{X}_n$ est un estimateur efficace.

8 Vecteurs gaussiens et espérance conditionnelle

Exercice 1

1. Soient X et ε deux variables aléatoires réelles indépendantes, X de loi $\mathcal{N}(0,1)$ et $\mathbb{P}(\varepsilon = +1) = \mathbb{P}(\varepsilon = -1) = 1/2$. On pose $Y = \epsilon X$.
   (a) Calculer la fonction de répartition de Y. En déduire sa loi
   (b) Les variables aléatoires réelles X et Y sont-elles non corrélées ?
   (c) Calculer $\mathbb{P}(X + Y = 0)$. Le vecteur (X, Y) est-il gaussien ? Les variables aléatoires réelles X et Y sont-elles indépendantes ?
2. Soient X, Y, Z trois variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$. On pose $U = X - Y$ et $V = X + Y - 2Z$.
   (a) Quelles sont les lois de U et de V ?
   (b) Ces variables aléatoires sont-elles indépendantes ?

Exercice 2

Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$. On pose

$$Y_a = X\mathbf{1}_{\{|X|>a\}} - X\mathbf{1}_{\{|X|\leqslant a\}}, \quad \forall a > 0.$$

1. Trouver la loi de Y.
2. Le vecteur (X, Y_a) est-il gaussien ?
3. X et Y_a sont-elles indépendantes ?
4. Montrer que l'on peut choisir a tel que X et Y_a soient non corrélées.

Exercice 3

Soient $X_1, \cdots, X_n$ des variables aléatoires indépendantes et de loi $\mathcal{N}(m, \sigma^2)$. On pose $Y_i = X_i - X_1, 2 \leqslant i \leqslant n$, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
1. Quelle est la loi de $(Y_2, \ldots, Y_n)^\intercal$?
2. Montrer que $(Y_2, \ldots, Y_n)^\intercal$ est indépendant de $\overline{X}_n$. En déduire que $\overline{X}_n$ et S^2 sont indépendants.

Exercice 4

Soit $X = (X_1, X_2, X_3, X_4)'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 2 \end{pmatrix}$$

1. Quelles sont les lois de X_3 et de $(X_1, X_2, X_4)'$?
2. Calculer $\mathbb{E}[X_1|X_2]$.
3. Calculer $\mathbb{E}[X_4|X_2]$.
4. En déduire deux variables indépendantes de X_2, fonctions respectivement de X_1, X_2 et de X_2, X_4.

Exercice 5

Soit

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \hookrightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Gamma = \begin{pmatrix} 4 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}\right).$$

1. Quelles sont les lois de X_1, de X_2 et de X_3 ?
2. Certaines composantes de X sont elles indépendantes ? Si oui lesquelles ?
   Certaines composantes de X sont elles orthogonales ? Si oui lesquelles ?
3. Quelle est la loi de $(X_1, X_2)'$?
4. Sans calcul, déterminer $\mathbb{E}[X_2|X_3]$ et $\mathbb{E}[(X_2 - \mathbb{E}[X_2|X_3])^2]$.
5. Calculer $\mathbb{E}[X_1|X_3]$ et $\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_3])^2]$.
6. Calculer $\mathbb{E}[X_1|(X_2\, X_3)]$ et $\mathbb{E}[(X_1 - \mathbb{E}[X_1|(X_2\, X_3)])^2]$.
7. Quelle est la loi de X_1 sachant $(X_2 = x_2, X_3 = x_3)$.
8. On pose $S = \mathbb{E}[X_1^2|X_2, X_3]$ et $\widehat{S}$ le résultat de la régression linéaire de X_1^2 sur $(1, X_2, X_3)$. Justifier que $\mathbb{E}[(X_1^2 - S)^2] \leqslant \mathbb{E}[(X_1^2 - \widehat{S})^2]$.

9 Régression linéaire

Exercice 1

Considérons le modèle linéaire simple $Y_i = \beta_1 x_i + \beta_0 + \mathcal{E}_i$, où $\mathcal{E}_i$ sont des variables aléatoires centrées indépendantes et de variance commune σ^2. $\widehat{\beta_0}$ et $\widehat{\beta_1}$ désigneront les estimateurs par moindres carrés de β_0 et β_1.

1. Calculer $\mathrm{Var}(\widehat{\beta_1})$.
2. Calculer $\mathrm{Cov}(\widehat{\beta_1}, \overline{Y}_n)$.
3. Calculer $\mathrm{Var}(\widehat{\beta_0})$.
4. Calculer $\mathrm{Cov}(\widehat{\beta_1}, \widehat{\beta_0})$.

Exercice 2

On considère le modèle linéaire gaussien $Y_i = \beta x_i + \mathcal{E}_i$, où $\mathcal{E}_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.

1. Déterminer l'estimateur des moindres carrés $\widehat{\beta}$ de β.
2. Montrer que $\widehat{\beta}$ est un estimateur sans biais et convergent en moyenne quadratique vers β.
3. Donner la loi de $\widehat{\beta}$.
4. Déterminer un estimateur sans biais $\widehat{\sigma}^2$ de σ^2 à l'aide de la variance résiduelle.

Exercice 3

On souhaite modéliser la quantité d'ozone dans l'atmosphère par un modèle linéaire dépendant de deux variables explicatives : la température à midi et la nébulosité à midi. Notons Y_i la $i$ème variable aléatoire représentant la quantité d'ozone et $x_{T,i}$ (resp. $x_{N,i}$) la température (resp. la nébulosité). On considère le modèle de régression linéaire multiple

$$Y_i = \theta_1 + \theta_2 x_{T,i} + \theta_3 x_{N,i} + \mathcal{E}_i, \text{ pour } i = 1, \dots, n$$

où $(\theta_1, \theta_2, \theta_3)' = \theta$ est un vecteur de paramètres à estimer et $\mathcal{E}_i$ est une variable aléatoire représentant l'erreur. On suppose $(\mathcal{E}_i)$ sont iid d'espérance nulle et de variance σ^2.

1. Donner la distance des moindres carrés à minimiser.
2. En déduire les estimateurs des moindres carrés $\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3$ de $\theta_1, \theta_2, \theta_3$.
3. On définit X la matrice et le vecteur Y suivant

$$X = \begin{pmatrix} 1 & x_{T,1} & x_{N,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{T,n} & x_{N,n} \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

   Montrer dans le cas simplifié $\overline{x}_T = 0$ et $\overline{x}_N = 0$, que $\widehat{\theta} = (X^T X)^{-1} X^T Y$ où ^{T} désigne la transposée. Ainsi le modèle peut se réécrire $Y = X\theta + \mathcal{E}$. En déduire que $\widehat{\theta}$ est un estimateur sans biais de θ.
4. Donner un estimateur $\widehat{\sigma}^2$ de σ^2 la variance des erreurs en se basant sur la variance des résidus. Est-il biaisé ? On pourra utiliser le résultat (admis) suivant :
   Si A est une matrice symétrique déterministe d'ordre n et Z un vecteur aléatoire de matrice de variance-covariance $\sigma^2 I_n$, alors

$$\mathbb{E}(Z^T A Z) = \sigma^2 \operatorname{Tr}(A) + \mathbb{E}(Z^T) A \mathbb{E}(Z).$$

5. Supposons de plus que les erreurs sont de loi normale, i.e. $\mathcal{E}_i \sim \mathcal{N}(0, \sigma^2)$.
   (a) Quelle est la loi de $\widehat{\theta}_i$ pour $i = 1, 2, 3$?
   (b) En admettant que $(n-3)\widehat{\sigma}^2/\sigma^2 \sim \chi^2_{n-3}$ et que $\widehat{\sigma}^2$ est indépendant de $\widehat{\theta}_i$, donner un intervalle de confiance exact pour θ_i.