# TD 1 : Quelques rappels de probabilités

**Exercice 1**

1. Donner la définition d'une variable aléatoire réelle et de sa loi de probabilité.
2. Donner la définition de la fonction de répartition d'une variable aléatoire réelle.
3. Donner la définition d'une densité.
4. Expliquer les intérêts de la fonction de répartition et de la densité.

**Exercice 2**

1. Soit $X$ une v.a.r. de fonction de répartition définie par :

$$F_X(x) = \begin{cases} 1 & \text{si} \quad x \geq 1 \\ 3/4 & \text{si} \quad 2/3 \leq x < 1 \\ 1/2 & \text{si} \quad 1/4 \leq x < 2/3 \\ 0 & \text{si} \quad x < 1/4. \end{cases}$$

   Tracer le graphe de cette fonction de répartition et calculer :

$$\mathbf{P}(X = 1/4),\ \mathbf{P}(X \geq 2/3),\ \mathbf{P}(X < 2/3),\ \mathbf{P}(X > 2/3),\ \mathbf{P}(X = 2/3),\ \mathbf{P}(X \in [1/4, 2/3]).$$

   Cette v.a.r. est-elle de loi discrète ? A densité ?

2. Mêmes questions pour une v.a.r. dont la fonction de répartition est définie par :

$$F_X(x) = \begin{cases} 1 & \text{si} \quad x \geq 2/3 \\ 3x/2 & \text{si} \quad 0 \leq x < 2/3 \\ 0 & \text{si} \quad x < 0. \end{cases}$$

3. Mêmes questions pour une v.a.r. dont la fonction de répartition est définie par :

$$F_X(x) = \begin{cases} 1 & \text{si} \quad x \geq 2/3 \\ \frac{1}{4}(x+1) & \text{si} \quad 0 \leq x < 2/3 \\ 0 & \text{si} \quad x < 0. \end{cases}$$

**Exercice 3**

1. Calculer l'espérance et la variance d'une v.a.r. de loi absolument continue de densité $f_X$ définie par $f_X(x) = \frac{1}{2}\exp(-|x|)$.
2. Même question pour une v.a.r. de densité $f_X$ définie par

$$f_X(x) = \begin{cases} -1/x^3 & \text{si} \quad x < -1, \\ 0 & \text{si} \quad -1 \leq x < 1, \\ 1/x^3 & \text{si} \quad x \geq 1. \end{cases}$$

**Exercice 4**
Soit $X$ une variable aléatoire réelle de densité $f_\theta$ définie par

$$f_\theta(x) = (1-\theta)\mathbf{1}_{[-1/2,0]}(x) + (1+\theta)\mathbf{1}_{]0,1/2]}(x),$$

où $\theta$ est un paramètre réel tel que $|\theta| \neq 1$.

1. Quelles conditions doit vérifier $\theta$ pour que $f_\theta$ soit bien une densité de probabilité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ ?

2. Calculer l'espérance de $X$.

3. Soient $X_1, \ldots, X_n$ $n$ variables aléatoires indépendantes et de même loi de densité $f_\theta$. Soient $U_n$ et $V_n$ les variables aléatoires définies par

$$U_n = \sum_{i=1}^{n} \mathbf{1}_{]-\infty,0]}(X_i) \quad \text{et} \quad V_n = \sum_{i=1}^{n} \mathbf{1}_{]0,+\infty[}(X_i).$$

   (a) Montrer que $U_n$ et $V_n$ suivent des lois binomiales dont on précisera les valeurs des paramètres.
   (b) Calculer $\mathbf{E}[U_n]$, $\mathbf{E}[V_n]$ et en déduire $\mathbf{E}[(V_n - U_n)/n]$.
   (c) Calculer $\mathbf{E}[U_n V_n]$ et $\mathbf{cov}(U_n, V_n)$.
   (d) Montrer que

$$\mathbf{V}\left[\frac{V_n - U_n}{n}\right]$$

   tend vers 0 lorsque $n$ tend vers l'infini.

**Exercice 5**
Soit $X$ la variable aléatoire réelle de densité

$$f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x} & \text{si} \quad x > 0 \quad (\lambda > 0), \\ 0 & \text{si} \quad x \leq 0. \end{cases}$$

Déterminer la loi des variables suivantes : $\lambda X$, $1/X$, $X^2$, $\sqrt{X}$.

**Exercice 6**
1. Donner la définition d'un vecteur aléatoire.

2. Donner la définition de la fonction de répartition d'un vecteur aléatoire.

3. Donner la définition de la densité d'un vecteur aléatoire.

4. Expliquer les intérêts de la fonction de répartition et de la densité.

**Exercice 7**
Soit $(X,Y)$ un couple aléatoire de densité conjointe $f$ définie par :

$$f(x,y) = cxy\ \mathbf{1}_{[0,2]\times[0,5]}(x,y).$$

1. Déterminer la valeur de $c$, puis la fonction de répartition conjointe du couple $(X,Y)$.

2. Déterminer les lois marginales de $X$ et $Y$ de deux façons différentes.

3. Calculer $\mathbf{P}(X \leq 1, Y \leq 1)$, $\mathbf{P}(X \leq Y)$.

**Exercice 8**
Soit $(X,Y)$ un couple aléatoire dont la densité conjointe est définie par :

$$f(x,y) = \begin{cases} e^{-y} & \text{si } 0 \leq x \leq y \\ 0 & \text{sinon.} \end{cases}$$

1. Déterminer la fonction de répartition conjointe du couple $(X, Y)$.

2. Déterminer les lois marginales de $X$ et $Y$.

3. Calculer $\mathbf{P}(X \leq 1, Y \leq 2)$, $\mathbf{P}(X/Y \leq 1/2)$.

**Exercice 9**
Calculer $\mathbf{P}(X^2 + Y^2 \leq 1)$ si $(X, Y)$ est un couple aléatoire de densité conjointe $f$ définie par

$$f(x, y) = \begin{cases} 2 & \text{si } 0 \leq x \leq y \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

**Exercice 10**
Soit $X$ et $Y$ deux variables aléatoires réelles indépendantes de lois absolument continues, de densités respectives définies par

$$f_X(x) = 2\exp(-2x)\mathbf{1}_{[0,+\infty[}(x) \quad \text{et} \quad f_Y(y) = y\exp(-y)\mathbf{1}_{[0,+\infty[}(y).$$

1. Donner la densité conjointe du couple $(X, Y)$.

2. Calculer pour tout $u \in [0, +\infty[$, $P(2X + Y = u)$, $P(2X + Y \leq u)$. Que peut-on en déduire ?

**Exercice 11**
Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire dont les variables marginales sont indépendantes identiquement distribuées (échantillon), chacune de fonction de répartition notée $F$.

1. Déterminer la loi de $\max\{X_1, \ldots, X_n\}$, et celle de $\min\{X_1, \ldots, X_n\}$.

2. Dans le cas d'un couple de variables aléatoires, déterminer la loi conjointe de

$$(\max\{X_1, X_2\}, \min\{X_1, X_2\}) .$$

**Exercice 12**
Soit $(X, Y)$ un couple aléatoire dont la densité conjointe est la fonction $f$ définie par

$$f(x, y) = \frac{1}{2}(x + y)\exp(-(x + y))\mathbf{1}_{[0,+\infty[^2}(x, y).$$

1. Déterminer la loi de la variable $X + Y$.

2. Déterminer la loi conjointe du couple aléatoire $(X + Y, X - Y)$, puis celle du couple aléatoire $(X + Y, Y)$.

**Exercice 13**
Soit $(X, Y)$ un couple aléatoire dont la densité conjointe est donnée par :

$$f(x, y) = \frac{1}{2\sqrt{x}} e^{-y} \mathbf{1}_{x>0, y>0, x<y^2}.$$

1. Déterminer les densités marginales de $X$ et $Y$.

2. Les variables marginales $X$ et $Y$ sont-elles indépendantes ?

3. Les variables $X$ et $Y - \sqrt{X}$ sont-elles indépendantes ?

4. Les variables $Y$ et $X/Y^2$ sont-elles indépendantes ?

# TD 2 : Théorie de l'estimation

**Exercice 1**
On dispose de $n$ observations $x_1, \ldots, x_n$ à valeurs dans une espace $\mathcal{H}$.

1. Expliquer les différentes étapes de la modélisation statistique. On donnera notamment la définition précise d'un modèle.
2. Quelle est la distinction entre un modèle paramétrique et un modèle non-paramétrique ?
3. Qu'est-ce qu'un estimateur ?
4. Qu'est-ce que le biais d'un estimateur ? Sa variance ?

**Exercice 2**
On considère $n$ variables aléatoires $X_1, \ldots, X_n$ i.i.d. d'espérance $\mu$ et de variance $\sigma^2$ finie.

1. Proposer un estimateur (raisonnable) de $\mu$.
2. Calculer son biais et sa variance.
3. Même question pour $\sigma^2$.

**Exercice 3**
Soit $X_1, \ldots, X_n$ $n$ variables aléatoires i.i.d. de loi de Bernoulli de paramètre $p$ inconnu.

1. Proposer un estimateur (raisonnable) de $p$.
2. Calculer son biais et sa variance.

**Exercice 4**
Soit $X_1, \ldots, X_n$ $n$ variables aléatoires i.i.d. de loi de Poisson de paramètre $\lambda$ inconnu.

1. Proposer un estimateur (raisonnable) de $\lambda$.
2. Calculer son biais et sa variance.

**Exercice 5**
Soit $X_1, \ldots, X_n$ $n$ variables aléatoires i.i.d. de loi normale $\mathcal{N}(\mu, \sigma^2)$.

1. On suppose ici que $\sigma^2$ est connu.
   (a) Proposer un estimateur (raisonnable) de $\mu$.
   (b) Calculer son biais et sa variance.
2. On suppose maintenant que $\mu$ est connu.
   (a) Proposer un estimateur (raisonnable) de $\sigma^2$.
   (b) Calculer son biais.
3. On suppose enfin que $\mu$ et $\sigma^2$ sont inconnus.

(a) Proposer un estimateur raisonnable de $\mu$. Calculer son biais et sa variance.

(b) Proposer un estimateur raisonnable de $\sigma^2$.

(c) Calculer son biais. On pourra écrire

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)+(\mu-\bar{X})\right]^2.$$

**Exercice 6**

Soit $X_1,\dots,X_n$ $n$ variables aléatoires i.i.d. de loi de Uniforme sur $[0,\theta]$ avec $\theta\in\mathbb{R}^+$ inconnu. On considère les estimateurs

$$\hat{\theta}_m=2\bar{X}_n \quad \text{et} \quad \hat{\theta}_{MV}=\max(X_1,\dots,X_n).$$

1. Expliquer la logique de construction de ces estimateurs.
2. Discuter du biais de $\hat{\theta}_{MV}$.
3. Calculer le biais, la variance et le risque quadratique de $\hat{\theta}_m$.
4. Calculer la fonction de répartition $\hat{\theta}_{MV}$ et montrer que la densité de $\hat{\theta}_{MV}$ est donnée par

$$f_{\hat{\theta}_{MV}}(x)=n\frac{x^{n-1}}{\theta^n}\mathbf{1}_{]0,\theta[}(x).$$

5. Déduire de la question précédente l'espérance, la variance de $\hat{\theta}_{MV}$ et le risque quadratique de $\hat{\theta}_{MV}$.
6. Donner une constante $c$ dépendant de $n$ telle que $\tilde{\theta}=c\hat{\theta}_{MV}$ soit sans biais.
7. Calculer la variance et le risque quadratique de $\tilde{\theta}$.
8. Comparer les risques quadratiques des estimateurs $\hat{\theta}_m$, $\hat{\theta}_{MV}$ et $\tilde{\theta}$.
9. Plus généralement, on considère un estimateur $\bar{\theta}$ de la forme $\alpha\hat{\theta}_{MV}$ avec $\alpha\in\mathbb{R}$.
   (a) Calculer le risque quadratique de $\bar{\theta}$.
   (b) Calculer la valeur de $\alpha$ qui minimise ce risque quadratique.
   (c) Conclure.

**Exercice 7**

Soit $X_1,\dots,X_n$ $n$ variables aléatoires de loi exponentielle de paramètre $\lambda>0$. On cherche à estimer $\mu=1/\lambda$. Comparer les erreurs quadratiques des estimateurs

$$\hat{\mu}_1=\bar{X}_n \quad \text{et} \quad \hat{\mu}_2=\frac{n}{n+1}\bar{X}_n=\frac{1}{n+1}\sum_{i=1}^{n}X_i.$$

**Exercice 8**

1. Expliquer la méthode du maximum de vraisemblance.
2. Qu'est-ce qu'un estimateur VUMSB ?
3. Comment peut-on montrer qu'un estimateur est VUMSB ?

**Exercice 9**

1. Déterminer les estimateurs du maximum de vraisemblance pour les modèles de Bernoulli $\mathcal{B}(p)$, de Poisson $\mathcal{P}(\lambda)$, exponentiel $\mathcal{E}(\lambda)$ et uniforme $\mathcal{U}_{[0,\theta]}$.
2. Déterminer l'information de Fisher de ces modèles.

3. Les estimateurs du maximum de vraisemblance sont-ils VUMSB ?

**Exercice 10**
Soit $X_1, \dots, X_n$ $n$ variables aléatoires i.i.d. de loi normale $\mathcal{N}(\mu, \sigma^2)$. Les 4 premiers moments de cette loi sont donnés par :

$$\mathbf{E}[X] = \mu, \quad \mathbf{E}[X^2] = \sigma^2 + \mu^2, \quad \mathbf{E}[X^3] = 3\mu\sigma^2 + \mu^3 \quad \text{et} \quad \mathbf{E}[X^4] = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4.$$

1. On suppose ici que $\sigma^2$ est connu.
   (a) Calculer l'estimateur du maximum de vraisemblance de $\mu$.
   (b) Calculer l'information de Fisher du $n$-échantillon.
   (c) Conclure.
2. On suppose maintenant que $\mu$ est connu.
   (a) Calculer l'estimateur du maximum de vraisemblance de $\sigma^2$.
   (b) Calculer l'information de Fisher du $n$-échantillon.
   (c) Conclure.

**Exercice 11**
Soit $X_1, \dots, X_n$ $n$ variables aléatoires réelles indépendantes qui suivent une loi géométrique de paramètre $p \in ]0, 1[$. On rappelle que la loi géométrique de paramètre $p$ est à support dans $\mathbb{N}^\star$ et vérifie

$$\mathbf{P}(X_1 = x) = (1-p)^{x-1}p, \quad x \in \mathbb{N}^\star.$$

On a de plus

$$\mathbf{E}[X_1] = \frac{1}{p} \quad \text{et} \quad \mathbf{V}[X_1] = \frac{1-p}{p^2}.$$

On cherche à estimer le paramètre $\lambda = 1/p$.

1. Exprimer la vraisemblance du modèle en fonction de $\lambda$ et en déduire l'estimateur du maximum de vraisemblance $\hat{\lambda}$.
2. Calculer la borne de Cramer-Rao du modèle considéré.
3. L'estimateur $\hat{\lambda}$ est-il VUMSB ? Justifier.

**Exercice 12**
On considère $n$ variables aléatoires $X_1, \dots, X_n$ i.i.d. de loi $\mathbf{P}_\theta$ définie par

$$\mathbf{P}_\theta(\{0\}) = \frac{1}{4}, \quad \mathbf{P}_\theta(\{1\}) = \frac{3\theta}{4} \quad \text{et} \quad \mathbf{P}_\theta(\{2\}) = \frac{3(1-\theta)}{4}$$

avec $\theta \in ]0, 1[$.

1. Calculer l'espérance et la variance de $X_1$.
2. En déduire l'estimateur des moments de $\theta$.
3. Calculer le biais et la variance de cet estimateur.
4. Déterminer l'information de Fisher $I_n(\theta)$ du $n$ échantillon.
5. L'estimateur des moments est-il VUMSB ?

**Exercice 13**
**Rappels :**

- La densité de la loi Gamma de paramètres $p \in \mathbb{N}^\star$ et $\theta > 0$, notée $\Gamma(p,\theta)$, est donnée par

$$f(x) = \frac{\theta}{(p-1)!} \exp(-\theta x)(\theta x)^{p-1} \mathbf{1}_{]0,+\infty[}(x).$$

- Si $X_1, \ldots, X_n$ sont $n$ variables aléatoires indépendantes de loi $\Gamma(p,\theta)$, alors $X_1 + \ldots + X_n$ suit une loi $\Gamma(np,\theta)$.
- Pour tout $\theta > 0$ et pour tout entier $p \geq 1$ on a

$$\int_0^{+\infty} \exp(-\theta x) x^{p-1}\,\mathrm{d}x = \frac{(p-1)!}{\theta^p},$$

avec la convention $0! = 1$.

- Si $X$ est une variable aléatoire réelle qui admet une densité $f$ par rapport à la mesure de Lebesgue $\lambda$, l'espérance de $X$ est donnée par

$$\mathbf{E}[X] = \int_{\mathbb{R}} x f(x)\,\mathrm{d}\lambda(x).$$

Si de plus $\varphi : \mathbb{R} \to \mathbb{R}$ est une fonction réelle telle que $\mathbf{E}[\varphi(X)]$ existe, on a

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x)\,\mathrm{d}\lambda(x).$$

On considère $X_1, \ldots, X_n$ un échantillon composé de $n$ variables aléatoires indépendantes et identiquement distribuées selon une loi $\Gamma(2,\theta)$. On cherche à estimer $\theta > 0$.

1. Montrer que $\mathbf{E}[X_1] = \frac{2}{\theta}$ et calculer $\mathbf{V}[X_1]$.
2. Calculer la borne de Cramer Rao pour le modèle considéré.
3. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ du paramètre $\theta$ et montrer que $\hat{\theta}_n$ peut se mettre sous la forme
$$\hat{\theta}_n = \frac{2n}{Y_n}$$
où $Y_n$ est une variable aléatoire de loi $\Gamma(2n,\theta)$.
4. L'estimateur $\hat{\theta}_n$ est-il consistant ? Justifier.
5. Calculer le biais de $\hat{\theta}_n$ et en déduire un estimateur $\tilde{\theta}_n$ sans biais du paramètre $\theta$.
6. Calculer $\mathbf{V}(\hat{\theta}_n)$ et en déduire $\mathbf{V}(\tilde{\theta}_n)$.
7. L'estimateur $\tilde{\theta}_n$ est-il VUMSB (de variance uniformément minimum parmi les estimateurs sans biais) ? Si non, est-il asymptotiquement VUMSB ?

## TD 3 : Théorèmes de convergence

**Exercice 1**
Donner les définitions des différents mode de convergence pour des suites de variables aléatoires.

**Exercice 2**
Soit $X_1, \ldots, X_n$ $n$ v.a.r indépendantes d'espérance $\mu$ et de variance $\sigma^2$.

1. A l'aide de l'inégalité de Bienaymé Tchebychev, montrer que $\bar{X}_n$ converge en probabilité vers une quantité à déterminer.
2. Quel autre résultat aurait-on utiliser ?

**Exercice 3**
Spot $(X_n)_n$ une suite de variables aléatoires indépendantes définies par

$$\mathbf{P}(X_n = \sqrt{n}) = \frac{1}{n} \quad \text{et} \quad \mathbf{P}(X_n = 0) = 1 - \frac{1}{n}.$$

1. Est-ce que $X_n$ converge en probabilité vers 0 ?
2. Est-ce que $X_n$ converge en moyenne quadratique vers 0 ?

**Exercice 4**
Spot $(X_n)_n$ une suite de variables aléatoires indépendantes définies par

$$\mathbf{P}(X_n = n^{0.4}) = \frac{1}{n} \quad \text{et} \quad \mathbf{P}(X_n = 0) = 1 - \frac{1}{n}.$$

1. Calculer $\mathbf{E}[|X_n|^p]$.
2. Conclure en terme de convergence en moyenne d'ordre $p$.

**Exercice 5**
Soit $(X_n)_n$ une suite de v.a.r indépendantes de loi uniforme $\mathcal{U}_{[-1/n,1/n]}$.

1. Est-ce que $X_n$ converge en probabilité vers 0 ?
2. Est-ce que $X_n$ converge en loi vers 0 ?

**Exercice 6**

1. Rappeler la définition de la fonction caractéristique.
2. Calculer la fonction caractéristique de la loi de Bernoulli $\mathcal{B}(p)$ et de la loi binomiale $\mathcal{B}(n,p)$.
3. Soit $(X_n)_n$ une suite de v.a.r indépendantes de loi binomiale $\mathcal{B}(n, p_n)$ avec $\lim_{n\to\infty} np_n = \lambda$.
   (a) Montrer que $X_n \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda)$. On admettre que pour la loi de Poisson $\mathcal{P}(\lambda)$, la fonction caractéristique est donnée par $\varphi(t) = e^{\lambda(e^{it}-1)}$.
   (b) Déduire du résultat précédent une approximation de $\mathbf{P}(X_{100} = 35)$ avec $X_n$ qui suit une loi binomial $\mathcal{B}(n, 40/n)$.

**Exercice 7**

1. Ecrire le TCL pour une suite de v.a.r. de Bernoulli $\mathcal{B}(p)$.
2. Ecrire le TCL pour une suite de v.a.r. de loi de Poisson $\mathcal{P}(1)$.
3. Déduire de la question précédente une approximation de la loi de Poisson $\mathcal{P}(n)$.
4. On s'intéresse à la probabilité $\alpha_n = \mathbf{P}(n-1 \leq Y \leq n+1)$ où $Y$ suit une loi de Poisson $\mathcal{P}(n)$. On note $\beta_n$ la valeur approchée de cette probabilité en utilisant l'approximation de la loi précédente. Compléter le tableau suivant :

| $n$ | 5 | 10 | 50 | 100 |
|---|---|---|---|---|
| $\alpha_n$ | | | | |
| $\beta_n$ | | | | |

**Exercice 8**

Soit $(X_n)_n$ une suite de v.a.r indépendantes admettant un moment d'ordre 2. On pose

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \leq x} \quad \text{et} \quad F(x) = \mathbf{P}(X \leq x).$$

1. Montrer que $\forall x \in \mathbb{R},\ F_n(x) \overset{p.s.}{\to} F(x)$.
2. Ecrire un TCL pour $(F_n(x))_n$.

**Exercice 9**

Soit $(X_n)_n$ une suite de v.a.r. indépendantes de loi $\mathcal{E}(\lambda)$. Proposer une suite $(\varphi_n)_n$ ainsi que variable aléatoire $Y_n$ telles que

$$\varphi_n(Y_n - \lambda) \overset{\mathcal{L}}{\to} \mathcal{N}(0,1).$$

**Exercice 10**

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. On pose $S_n = \sum_{i=1}^n X_i$. Montrer que si $S_n/\sqrt{n} \overset{\mathcal{L}}{\to} Y$, alors $S_n/n \overset{\mathbf{P}}{\to} 0$, c'est-à-dire la suite $(X_n)$ vérifie la loi faible des grands nombres.

**Exercice 11**

Soit $X_n$ la distance euclidienne entre deux points choisis indépendamment et uniformément dans le cube unité $[0,1]^n$ de dimension $n$. Montrer que $X_n/\sqrt{n} \to 1/\sqrt{6}$. En quel sens cette convergence a-t-elle lieu ?

**Exercice 12**

Soit $X_1, X_2, \ldots$ une suite de v.a.r. indépendantes, de même loi d'espérance $m$ et de variance $\sigma^2 < +\infty$. Montrer que

$$\frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} X_i X_j \overset{\mathbf{P}}{\to} m^2 \text{ lorsque } n \to +\infty.$$

**Exercice 13**

Soit $X_1, X_2, \ldots$ une suite de v.a.r. indépendantes, de même loi de Cauchy.

1. Montrer que la moyenne $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ suit également la loi de Cauchy (utiliser les fonctions caractéristiques). Pourquoi cela ne contredit-il pas la loi des grands nombres ?
2. Soit $M_n = \max\{X_1, \ldots, X_n\}$. Montrer que $\pi M_n/n$ converge en loi vers une loi limite que l'on précisera.

**Exercice 14**

Soit $X$ une v.a.r. suivant la loi de Poisson de paramètre $\lambda$.

1. Montrer que
$$\frac{X - \mathbf{E}[X]}{\sqrt{\mathbf{V}(X)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ lorsque } \lambda \to +\infty.$$
2. Comparer avec le Théorème Central Limite lorsque $\lambda \in \mathbb{N}^\star$.
3. En déduire un calcul approché de $\mathbf{P}(X \leq 30)$ lorsque $\lambda = 25$. Comparer avec la valeur exacte.
4. Mêmes questions si $X$ suit maintenant une loi gamma $\gamma(\lambda, 1)$.

**Exercice 15**

Soit $X_n$ et $Y_m$ deux v.a.r. indépendantes suivant des lois de Poisson de paramètres respectifs $n$ et $m$. Montrer que
$$\frac{(X_n - n) + (Y_m - m)}{\sqrt{X_n + Y_m}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ lorsque } m, n \to +\infty.$$

**Exercice 16**

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes, de même loi d'espérance $m$ et de variance $\sigma^2 < +\infty$. Pour tout entier $n \geq 2$, on définit les v.a.r.
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad S_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

1. Calculer $\mathbf{E}[S_n]$.
2. Montrer que $S_n \xrightarrow{p.s.} \sigma^2$, lorsque $n \to +\infty$.
3. En déduire que
$$T_n = \sqrt{n}\frac{(\bar{X}_n - m)}{\sqrt{S_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ lorsque } n \to +\infty.$$

**Exercice 17**

Soit $\theta > 0$ un nombre réel fixé et $Z = (X, Y)'$ un vecteur aléatoire dont la densité $f$ est définie sur $\mathbb{R}^2$ par
$$f(x, y) = C \exp(-\theta y)\mathbf{1}_{0<x<y}.$$

1. Montrer que $C = \theta^2$.
2. Déterminer les lois marginales de $Z$ ainsi que $\mathbf{E}[X]$ et $\mathbf{V}[X]$. Les variables aléatoires $X$ et $Y$ sont-elles indépendantes ?
3. Calculer la densité conditionnelle de $Y|X = x$.
4. En déduire $\mathbf{E}[Y|X]$ puis $\mathbf{E}[Y]$.
5. Calculer $\mathbf{P}(Y > 2X)$.
6. On pose $U = X$ et $V = Y - X$.
   (a) Déterminer la loi du couple $(U, V)'$.
   (b) Les variables aléatoires $U$ et $V$ sont-elles indépendantes ?
7. A l'aide de la question précédente calculer $\mathbf{cov}[X, Y]$. En déduire que $\mathbf{V}[Y] = \frac{2}{\theta^2}$.
8. Soit $(X_i, Y_i)_{i\geq 1}$ une suite de couples aléatoires indépendants et de même loi de densité $f$. Pour $n \geq 1$, on pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

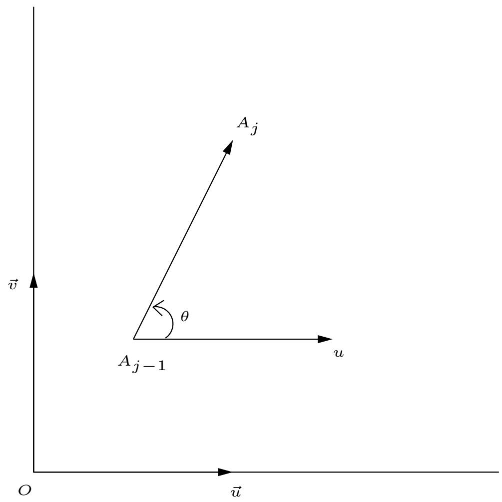


Figure 1: Exemple de saut.

(a) Montrer que, pour une suite $(a_n)_{n\geq 1}$ à préciser, $a_n(\bar{Y}_n - \bar{X}_n - 1/\theta)$ converge en loi vers une loi à déterminer.

(b) Montrer que, pour une suite $(b_n)_{n\geq 1}$ à préciser, $\frac{b_n}{\bar{X}_n}(\bar{Y}_n - \bar{X}_n - 1/\theta)$ converge en loi vers une loi à déterminer.

**Exercice 18**

Dans un plan $(O, \vec{u}, \vec{v})$, une puce part de l'origine $O = A_0$ et saute à chaque instant. De l'instant 0 à l'instant $n$, ses points de chute successifs sont notés $A_1, \ldots, A_n$. Chaque saut est supposé de même longueur égale à 1 : la longueur des segments $[A_{j-1}A_j]$ vaut donc 1 pour $j = 1, \ldots, n$. On appelle direction du saut $j$ l'angle formé par les vecteurs $\vec{u}$ et $\overrightarrow{A_{j-1}A_j}$, cet angle prend ses valeurs sur $[0, 2\pi[$ (voir Figure 1). La direction de chaque saut est supposée suivre une loi uniforme sur $[0, 2\pi[$.

1. Soit $U$ une variable aléatoire de loi uniforme sur $[0, 2\pi[$. On pose $V = (\cos(U), \sin(U))'$. Déterminer l'espérance et la matrice de variance covariance de $V$.

2. On note $S_n = \overrightarrow{OA_n}$.

   (a) Montrer que $S_n = V_1 + \ldots + V_n$ où $V_1, \ldots, V_n$ sont des vecteurs aléatoires indépendants de même loi que $V$.

   (b) Montrer que $\frac{S_n}{\sqrt{n}}$ converge en loi vers une variable aléatoire dont on précisera la loi.

3. Soit $(X, Y)'$ un vecteur gaussien centré de matrice de variance covariance identité. Soit $(R, \Theta)'$ le vecteur aléatoire correspondant aux coordonnées polaires de $(X, Y)'$.

   (a) Calculer la densité jointe de $(R, \Theta)'$.

   (b) Les variables aléatoires $R$ et $\Theta$ sont-elles indépendantes ?

   (c) Calculer la fonction de répartition de $R^2$ et en déduire que $R^2$ suit une loi exponentielle dont on précisera le paramètre.

4. On désigne par $(X_n, Y_n)'$ les coordonnées de $\overrightarrow{OA_n}$.

   (a) Montrer que
$$\frac{2}{n}(X_n^2 + Y_n^2)$$
   converge en loi vers une variable aléatoire dont on précisera la loi.

   (b) En déduire, pour $n$ suffisamment grand, une valeur approchée de $\mathbf{P}(X_n^2 + Y_n^2 \leq n \ln(2))$.

# TD 4 : Intervalles de confiance et estimation multivariée

**Exercice 1**

1. Donner la définition d'un intervalle de confiance.
2. Donner la définition d'un intervalle de confiance asymptotique.

**Exercice 2**

Soit $X_1, \ldots, X_n$ $n$ variables aléatoires de loi $\mathcal{N}(\mu, 1)$. Donner un intervalle de confiance de niveau $1-\alpha$ pour $\mu$.

**Exercice 3**

Soit $X_1, \ldots, X_n$ $n$ variables aléatoires de loi exponentielle $\mathcal{E}(\lambda)$ avec $\lambda > 0$. A l'aide des rappels de l'exercice 13 du TD2, montrer que

1. $X_1$ suit une loi $\Gamma(1, \lambda)$.
2. $S_n = X_1 + \ldots + X_n$ suit une loi $\Gamma(n, \lambda)$.
3. Calculer la densité de $\lambda S_n$ (on pourra passer par la fonction de répartition).
4. En déduire un intervalle de confiance de niveau $1-\alpha$ pour $\lambda$.

**Exercice 4**

Soit $X_1, \ldots, X_n$ $n$ variables aléatoires de loi Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$.

1. Ecrire le théorème centrale limite pour ce cas.
2. En déduire que

$$\left[\bar{X}_n + \frac{q^2_{1-\alpha/2}}{n} - \frac{1}{2}\sqrt{4\frac{\bar{X}_n q^2_{1-\alpha/2}}{n} + \frac{q^4_{1-\alpha/2}}{n^2}}, \bar{X}_n + \frac{q^2_{1-\alpha/2}}{n} + \frac{1}{2}\sqrt{4\frac{\bar{X}_n q^2_{1-\alpha/2}}{n} + \frac{q^4_{1-\alpha/2}}{n^2}}\right].$$

   est un intervalle de confiance asymptotique de niveau $1-\alpha$ pour $\lambda$. $q_{1-\alpha/2}$ désigne le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$.
3. A l'aide du théorème de Slutsky appliqué au TCL de la question 1, calculer un autre intervalle de confiance asymptotique.

**Exercice 5**

Soit $X_1, \ldots, X_n$ $n$ variables aléatoires de loi de Bernoulli $\mathcal{B}(p)$ où $p \in [0,1]$.

1. Ecrire le théorème centrale limite pour ce cas.
2. En déduire un intervalle de confiance de niveau $1-\alpha$ pour $p$.

**Exercice 6**

Soit $X_1, \ldots, X_n$ $n$ variables aléatoires de loi $\mathcal{N}(\mu, \sigma^2)$.

1. Rappeler le théorème de Cochran.

2. En déduire un intervalle de confiance de niveau $1-\alpha$ pour $\mu$ et $\sigma^2$.

**Exercice 7**
Soit $X_1, \ldots, X_{n_1}$ $n_1$ variables aléatoires indépendantes de loi $\mathcal{N}(\mu_1, \sigma_1^2)$ et $Y_1, \ldots, Y_{n_2}$ $n_2$ variables aléatoires indépendantes de loi $\mathcal{N}(\mu_2, \sigma_2^2)$. On suppose de plus que les deux échantillons sont indépendants.

1. On suppose dans cette question que $\sigma_1 = \sigma_2 = \sigma$ et on pose
$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$
avec
$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \quad \text{et} \quad S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2.$$
   (a) Expliquer la logique de construction de $S^2$.
   (b) Quelle est la loi de $\varphi(n_1, n_2)S^2$ où $\varphi$ est une fonction à préciser ?
   (c) Quelle est la loi de $(\bar{X}_n - \bar{Y}_n)$ ?
   (d) En déduire un intervalle de confiance de niveau $1-\alpha$ pour $\mu_1 - \mu_2$.
2. On suppose maintenant que $\sigma_1 \neq \sigma_2$. A l'aide du théorème de Slutsky, construire un intervalle de confiance asymptotique de niveau $1-\alpha$ pour $\mu_1 - \mu_2$.
3. Quelle est le loi de
$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \ ?$$
4. En déduire un intervalle de confiance de niveau $1-\alpha$ pour $\sigma_1^2/\sigma_2^2$.

**Exercice 8**
Soit $X_1, \ldots, X_{n_1}$ $n_1$ variables aléatoires indépendantes de loi $\mathcal{B}(p_1)$ et $Y_1, \ldots, Y_{n_2}$ $n_2$ variables aléatoires indépendantes de loi $\mathcal{B}(p_2)$. On suppose de plus que les deux échantillons sont indépendants. En vous basant sur l'exercice précédent, donner un intervalle de confiance asymptotique de niveau $1-\alpha$ pour $p_1 - p_2$.

**Exercice 9**
On veut estimer le rendement d'un engrais pour la culture du blé. Sur 12 parcelles expérimentales, on a trouvé les rendements suivants en tonnes par hectare :

$$7.7 \quad 8.4 \quad 7.8 \quad 8.2 \quad 7.9 \quad 8.5 \quad 8.4 \quad 8.2 \quad 7.6 \quad 7.8 \quad 8.4 \quad 8.3.$$

Donner un intervalle de confiance de niveau 95% pour le rendement moyen de l'engrais (on supposera que le rendement à l'hectare est une variable gaussienne).

**Exercice 10**
Un stock comporte 10 000 pièces. Pour évaluer le nombre de pièces défectueuses dans le stock, on en tire 400 au hasard et on constate que 45 sont défectueuses.

1. Donner un intervalle de confiance de niveau 99% pour la proportion de pièces défectueuses.
2. Donner un intervalle de confiance de niveau 99% pour le nombre total de pièces défectueuses.

**Exercice 11**
On veut évaluer la différence des proportions de pièces défectueuses dans 2 procédés de fabrication différents. Pour cela on tire au hasard 1000 pièces réalisées selon le premier procédé, on en trouve 86 défectueuses. On opère de même pour 800 pièces réalisées selon le second procédé et on en trouve 92 défectueuses. Donner un intervalle de confiance de niveau 95% sur la différence des proportions de pièces défectueuses dans les deux procédés.

**Exercice 12**
Dans une ville on donne la répartition du nombre de jours sans accident, avec un accident... parmi 50 jours d'observation au cours d'une même année:

| Nb d'accidents | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Nb jours | 21 | 18 | 7 | 3 | 1 |

1. Proposer une famille de loi de probabilité permettant de modéliser ce phénomène.
2. Donner un intervalle de confiance de niveau 95% pour le nombre moyen d'accidents par jour.

**Exercice 13**
Soit $X_1, \ldots, X_n$ $n$ variables aléatoires indépendantes de loi $\mathcal{N}(\mu, \sigma^2)$. On veut estimer $\theta = (\mu, \sigma^2)$. On propose d'étudier l'estimateur $\hat{\theta} = (\hat{\mu}, S^2)$ défini par

$$\hat{\mu} = \bar{X}_n \quad \text{et} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

1. Calculer le biais de $\hat{\theta}$.
2. $\hat{\theta}$ est-il consistant ?
3. Calculer la matrice de variance covariance de $\hat{\theta}$.
4. Calculer la matrice d'information de Fisher du modèle considéré. En déduire la borne de Cramer Rao.
5. L'estimateur $\hat{\theta}$ est-il VUMSB ?

# TD 5 : Approche paramétrique vs non paramétrique

**Exercice 1**

Soit $X_1, \ldots, X_n$ n variables aléatoires réelles i.i.d de densité $f$ inconnu. On fixe un point $x$ dans $\mathbb{R}$ et on cherche à estimer $f$ au point $x$. On suppose que $f$ est bornée par $B$ et on se donne un noyau $K : \mathbb{R} \to \mathbb{R}^+$ tel que

$$\int K(u)\,\mathrm{d}u = 1, \quad \int uK(u)\,\mathrm{d}u = 0, \quad \int u^2K(u)\,\mathrm{d}u = \mathcal{C} < +\infty \quad \text{et} \quad \int K^2(u)\,\mathrm{d}u = \mathcal{K} < +\infty. \tag{1}$$

1. On note $\hat{f}$ l'estimateur de noyau $K$ et de fenêtre $h > 0$ de $f$. Rappeler sa définition et sa logique de construction.

2. Donner des exemples de noyau vérifiant (1).

3. On étudie d'abord la variance de $\hat{f}$.

   (a) Montrer que
   $$\mathbf{V}\left[K\left(\frac{X-x}{h}\right)\right] \leq hB\mathcal{K}.$$

   (b) En déduire que
   $$\mathbf{V}[\hat{f}(x)] \leq \frac{B\mathcal{K}}{nh}.$$

4. On s'intéresse maintenant au biais. On suppose que $f$ est dérivable et que sa dérivée vérifie :
   $$|f'(x) - f'(y)| \leq L|x-y|, \quad \forall x, y \in \mathbb{R}.$$

   (a) Montrer que le biais de $\hat{f}(x)$ vérifie
   $$b(\hat{f}(x)) = \int K(u)f(x+uh)\,\mathrm{d}u - f(x).$$

   (b) Montrer qu'il existe $\tau \in ]0,1[$ tel que
   $$f(x+uh) = f(x) + (uh)f'(x+\tau uh).$$

   (c) En déduire que le biais de $\hat{f}(x)$ vérifie
   $$b(\hat{f}(x)) = \int K(u)(uh)[f'(x+\tau uh) - f'(x)]\,\mathrm{d}u.$$

   (d) En déduire
   $$|b(\hat{f}(x))| \leq Ch^2$$
   où $C$ est une constante à préciser.

5. Proposer une valeur de $h$ qui minimise le risque quadratique de $\hat{f}(x)$ et calculer le risque quadratique associé à cette valeur de $h$.

**Exercice 2**
On considère le modèle de régression linéaire

$$Y_i = \beta x_i + \varepsilon_i$$

où $x_i = i/n$ et les $\varepsilon_i$ sont i.i.d tels que $\mathbf{E}[\varepsilon_i] = 0$ et $\mathbf{V}(\varepsilon_i) = \sigma^2$.

1. Calculer l'estimateur des moindres carrés $\hat{\beta}$ de $\beta$.
2. Calculer le biais et la variance de $\hat{\beta}$.
3. En déduire que le risque quadratique de $\hat{\beta}$ vérifie

$$\mathbf{E}[(\hat{\beta} - \beta)^2] = \mathrm{O}\left(\frac{1}{n}\right).$$

**Exercice 3**
On considère le modèle de régression linéaire

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

où $x_i \in \mathbb{R}$ et les $\varepsilon_i$ sont i.i.d tels que $\mathbf{E}[\varepsilon_i] = 0$ et $\mathbf{V}(\varepsilon_i) = \sigma^2$.

1. Calculer les estimateurs des moindres carrés $\hat{\beta}_0$ et $\hat{\beta}_1$ de $\beta_0$ et $\beta_1$.
2. (a) Montrer que

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

   (b) En déduire que $\hat{\beta}_0$ et $\hat{\beta}_1$ sont sans biais.
3. (a) Exprimer la variance de $\hat{\beta}_1$ en fonction de $\sigma^2$ et des $x_i$.
   (b) Montrer que $\mathbf{cov}(\bar{Y}, \hat{\beta}_1) = 0$.
   (c) En déduire la variance de $\hat{\beta}_0$.
4. Soit $\tilde{\beta}_1$ un estimateur sans biais de $\beta_1$ de la forme $\tilde{\beta}_1 = \sum_{i=1}^n \lambda_i Y_i$.
   (a) Montrer que $\sum_{i=1}^n \lambda_i = 0$ et $\sum_{i=1}^n \lambda_i = 1$.
   (b) Montrer que $\mathbf{cov}(\tilde{\beta}_1 - \hat{\beta}_1, \hat{\beta}_1) = 0$.
   (c) En déduire que $\mathbf{V}(\tilde{\beta}_1) \geq \mathbf{V}(\hat{\beta}_1)$.
   (d) Interpréter.

**Exercice 4**
On considère le modèle de régression

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

où $x_1, \ldots, x_n \in \mathbb{R}^d$ sont déterministes et $\varepsilon_1, \ldots, \varepsilon_n$ sont des variables i.i.d. d'espérance nulle et de variance $\sigma^2 < +\infty$. On désigne par $\| \,.\, \|$ la norme Euclidienne dans $\mathbb{R}^d$. On définit l'estimateur localement constant de $m$ en $x \in \mathbb{R}^d$ par :

$$\hat{m}(x) = \underset{a \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - a)^2 K\left(\frac{\|x_i - x\|}{h}\right)$$

où $h > 0$ et pour $u \in \mathbb{R}, K(u) = \mathbf{1}_{[0,1]}(u)$. On suppose que $\sum_{i=1}^n K\left(\frac{\|x_i - x\|}{h}\right) > 0$.

1. Donner la forme explicite de $\hat{m}(x)$.

2. Montrer que
$$\mathbf{V}[\hat{m}(x)] = \frac{\sigma^2}{\sum_{i=1}^n K\left(\frac{\|x_i - x\|}{h}\right)}$$
et
$$\mathbf{E}[\hat{m}(x)] - m(x) = \frac{\sum_{i=1}^n (m(x_i) - m(x)) K\left(\frac{\|x_i - x\|}{h}\right)}{\sum_{i=1}^n K\left(\frac{\|x_i - x\|}{h}\right)}.$$

3. On suppose maintenant que $m$ est Lipschitzienne de constante $L$, c'est-à-dire que $\forall (x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$
$$|m(x_1) - m(x_2)| \leq L \|x_1 - x_2\|.$$
Montrer que
$$|\text{biais}[\hat{m}(x)]| \leq Lh.$$

4. On suppose de plus qu'il existe une constante $C_1$ telle que
$$C_1 \leq \frac{\sum_{i=1}^n \mathbf{1}_{B_h}(x_i - x)}{n \text{Vol}(B_h)},$$
où $B_h = \{u \in \mathbb{R}^d : \|u\| \leq h\}$ est la boule de rayon $h$ dans $\mathbb{R}^d$ et $\text{Vol}(A)$ désigne le volume d'un ensemble $A \subset \mathbb{R}^d$. Montrer que
$$\mathbf{V}[\hat{m}(x)] \leq \frac{C_2 \sigma^2}{n h^d},$$
où $C_2$ est une constante dépendant de $C_1$ et $d$ à préciser.

5. Déduire des questions précédentes un majorant de l'erreur quadratique moyenne de $\hat{m}(x)$.

6. Calculer $h_{opt}$ la valeur de $h$ qui minimise ce majorant. Que vaut ce majorant lorsque $h = h_{opt}$. Comment varie cette vitesse lorsque $d$ augmente. Interpréter.