# 1 Rappels

## 1.1 Transformée d'une loi exponentielle

Soit $X$ une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$. Déterminer la loi de la variable aléatoire $e^{-\lambda X}$.

## 1.2 Existence de la covariance

On rappelle qu'une variable aléatoire réelle $X$ est dite de carré intégrable, et l'on note $X \in L_2(\Omega)$, si $\mathbb{E}[X^2] < \infty$.
Soit $X$ et $Y$ deux variables aléatoires de $L_2(\Omega)$. Montrer que $|\operatorname{cov}(X, Y)| < \infty$.

## 1.3 Covariance - suite

Soit $(X, Y, Z)$ un triplet aléatoire vérifiant $X + Y + Z = 1$. On suppose $\operatorname{var}(X) \leq \operatorname{var}(Y) \leq \operatorname{var}(Z) < \infty$. Montrer que

1. $\operatorname{cov}(Z, X) \leq 0$ et $\operatorname{cov}(Z, Y) \leq 0$.
2. $\operatorname{cov}(X, Y) \geq 0 \Leftrightarrow \operatorname{var}(X) + \operatorname{var}(Y) \leq \operatorname{var}(Z)$
3. $|\operatorname{cov}(X, Z)| \leq |\operatorname{cov}(Y, Z)|$.

## 1.4 Critère de convergence en probabilité

On rappelle la définition de la convergence en moyenne d'ordre $r > 0$ : Soit $(X_n), n \geq 1$ une suite de variables aléatoires. On suppose qu'il existe $r > 0$ tel que pour tout $n \geq 1$, $\mathbb{E}[|X_n|^r] < \infty$. On dit que $(X_n), n \geq 1$ converge vers 0 en moyenne d'ordre $r$ si $\mathbb{E}[|X_n|^r] \to 0$ quand $n \to \infty$. On dit que $(X_n), n \geq 1$ converge vers $X$ en moyenne d'ordre $r$ si la suite $(X_n - X), n \geq 1$ converge vers 0 en moyenne d'ordre $r$.

Montrer que la convergence en moyenne d'ordre $r > 0$ vers 0 entraîne la convergence en probabilité vers 0.

# 2 Modèles statistiques

## 2.1 Sommes de lois de Poisson

1. Montrer que la somme de deux variables indépendantes suivant une loi de Poisson de même paramètre $\lambda > 0$ suit une loi de Poisson de paramètre $2\lambda$.
2. Soit $X$ et $Y$ deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs $\lambda$ et $\mu$. Montrer que $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.

## 2.2 Assurance-vie

La force de persuasion d'un démarcheur en assurance-vie se traduit par la probabilité $p$ qu'après s'être présenté au domicile de quelqu'un, cette personne souscrive au contrat proposé. Déterminer la loi du nombre $N$ de visites nécéssaires à la souscription de $n$ contrats, $n \in \mathbb{N}^\star$.

## 2.3 Modélisation avec la loi exponentielle

On modélise la durée de vie d'une ampoule par une loi exponentielle de paramètre $\lambda > 0$. Proposer un modèle statistique associé à l'observation d'un échantillon de taille $n$.

## 2.4 Somme de lois de Poisson - suite

Le nombre de spams reçus par jour dans une boite mail est supposé suivre une loi de Poisson de paramètre $\lambda > 0$. On relève le nombre de spams lors pendant $n$ jours.

1. Proposer une modèle statistique de la situation.
2. Quelle est la loi du nombre total de spams ?

# 3 Estimateurs - Propriétés

## 3.1 Formules de Huyghens

1. Soit $X$ une variable aléatoire de $L_2(\Omega)$, montrer
$$\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$
2. Soit $X_1, \ldots, X_n$ des variables aléatoires et $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Montrer
$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

## 3.2 Consistance d'estimateurs de la variance

La variance d'échantillon
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$
est un estimateur naturel de la variance de la population, appelé variance empirique. Un autre estimateur, plus usuel, est la variable
$$\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$
Montrer la consistance de ces deux estimateurs vers la variance de population.

**Aide :** On rappelle que si $\{X_n\}$ et $\{Y_n\}$ sont deux suites de variables aléatoires qui convergent respectivement vers $X$ et $Y$ en probabilité alors $\{X_n + Y_n\}$ et $\{X_n Y_n\}$ convergent en probabilité vers $X + Y$ et $XY$ respectivement.

## 3.3 Propriétés de la moyenne d'échantillon

Soit $X_1, \ldots, X_n$ un $n$-échantillon dont la loi commune a pour espérance $\mathbb{E}[X] = \mu$ et la variance $\mathrm{var}(X) = \sigma^2$.

1. Montrer que $\bar{X}$ est un estimateur sans biais de $\mathbb{E}[X]$.
2. Montrer que la variance de la moyenne d'échantillon est $\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}$.

## 3.4 Maximum de lois uniformes

Soit $U_1, \ldots, U_n$ des variables aléatoires indépendantes de même loi uniforme sur $[0, \theta]$, $\theta$ étant un paramètre inconnu. On pose $T_n = \sup(U_1, \ldots, U_n)$. Montrer que $T_n$ est un estimateur convergeant de $\theta$.

## 3.5 Décomposition du risque quadratique

Soit $\hat{\theta}$ un estimateur d'un paramètre réel $\theta$. Montrer la formule de décomposition biais-variance du risque quadratique :
$$\mathcal{R}(\hat{\theta}, \theta) = (\mathbb{E}_\theta[\hat{\theta}] - \theta)^2 + \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])^2].$$

## 3.6 Estimateur sans biais de la variance

Soit $X_1, \ldots, X_n$ un $n$-échantillon de loi commune d'espérance $\mathbb{E}[X] = \mu$ et de variance $\mathrm{var}(X) = \sigma^2$. Soit la variance empirique de l'échantillon
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

1. Montrer
$$\mathbb{E}\left[S^2\right] = \frac{n-1}{n}\sigma^2.$$
2. En déduire un estimateur sans biais de $\sigma^2$.

### 3.7 Borne supérieure d'une loi uniforme (suite)

Soit $U_1, \ldots, U_n$ des variables i.i.d. selon une une loi uniforme sur $[0, \theta]$. On s'intéresse dans cet exercice à des estimateurs de $\theta$, la borne supérieure du support de la loi commune.

1. Rappeler la loi de $T_n = \sup(U_1, \ldots, U_n)$.
2. Calculer le biais de cet estimateur.
3. En déduire un estimateur sans biais de $\theta$.
4. Calculer la variance de ces deux estimateurs.
5. Comparer le risque quadratique des deux estimateurs.

# 4 Construction d'estimateurs

## 4.1 Méthode des moments pour une loi Gamma

On rappelle qu'une variable aléatoire $X$ à valeurs dans $\mathbb{R}^+$ suit une loi Gamma de paramètres $p > 0$ et $\lambda > 0$, et on note $X \sim \Gamma(p, \lambda)$, si elle est absolument continue avec comme densité

$$f_{p,\lambda}(x) = \frac{\lambda}{\Gamma(p)} e^{-\lambda x} (\lambda x)^{p-1} 1_{[0,+\infty[}(x)$$

où $\Gamma(.)$ est la fonction dite fonction Gamma (ou seconde fonction d'Euler) définie pour tout réel $p > 0$ par

$$\Gamma(p) = \int_0^{+\infty} e^{-x} x^{p-1} dx.$$

Notons que cette famille de lois inclut celle des lois exponentielles puisque pour $p = 1$, la loi $\Gamma(1, \lambda)$ coïncide avec la loi $\mathcal{E}(\lambda)$.
On rappelle

$$X \sim \Gamma(p, \lambda) \Rightarrow \mathbb{E}(X) = \frac{p}{\lambda} \text{ et } \mathrm{var}(X) = \frac{p}{\lambda^2}.$$

1. Quels sont, en fonction de $p$ et $\lambda$, les deux premiers moments de la loi Gamma ?
2. Exprimer les paramètres $\lambda$ et $p$ en fonction des moments $m_1$ et $m_2$.
3. En déduire l'estimateur des moments de $(m_1, m_2)$.

## 4.2 Cas de discussion autour de l'estimateur des moments

Au cours de la seconde guerre mondiale, les alliés utilisaient la théorie de l'échantillonnage pour mesurer la production allemande de fusées. Les Allemands donnaient à leurs fusées des numéros de série allant de 1 à $N$. Le problème pour les alliés était d'estimer la taille de l'arsenal allemand i.e. la valeur de $N$. Pour estimer $N$, on considèra que les fusées allemandes prises constituaient un échantillon aléatoire de numéros de série. Pour illustrer, supposons que 10 fusées prises à l'ennemi aient les numéros de séries suivants :

$$77, 30, 05, 39, 28, 10, 27, 12, 73, 49$$

1. Pour une population de numéro de série allant de 1 à $N$, quelle est la moyenne $\mu$ exprimée en fonction de $N$ ?
2. En déduire une expression de $N$ en fonction de $\mu$.
3. Par la méthode des moments, quel est l'estimateur de $\mu$ ? En déduire un estimateur de $N$.
4. Quelle est l'estimation de $N$ sur l'échantillon ?

## 4.3 EMV pour le paramètre d'une loi de Poisson

Déterminer l'estimateur du maximum de vraisemblance pour le paramètre d'une loi de Poisson.

## 4.4 EMV pour le paramètre d'une loi exponentielle

Déterminer l'estimateur du maximum de vraisemblance pour le paramètre $\lambda$ d'une loi exponentielle.

## 4.5 Estimateurs de la borne supérieure d'une loi uniforme

1. Déterminer l'estimateur des moments pour la borne supérieure $\theta$ d'une loi uniforme sur $[0, \theta]$.
2. Déterminer l'estimateur obtenu par maximum de vraisemblance.

### 4.6 Estimateur de l'aire d'un carré

On dispose d'une baguette de longueur $\mu$. A l'aide de cette baguette, on dessine un carré dont le côté a pour longueur la longueur de la baguette : l'aire du carré est donc $\mu^2$. On ne connait pas la longueur $\mu$ de la baguette alors on décide de la mesurer $n$ fois. Soient $X_1, \ldots, X_n$ ces mesures.

1. Poser le problème en termes statistiques et montrer que $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur sans biais de $\mu$.
2. Montrer que $\bar{X}^2$ est en revanche un estimateur biaisé de l'aire $\mu^2$. Calculer son biais.
3. On considère la variable aléatoire $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Déterminer $k$ tel que $\bar{X}^2 - kS_n^2$ soit un estimateur sans biais de $\mu^2$.

# 5 Estimation par intervalles

## 5.1 Intervalles de confiance pour une moyenne, pour une variance

On a mesuré la quantité d'alcool totale (mesurée en g/l) contenue dans 9 cidres doux du marché. On suppose que la quantité d'alcool des cidres suit une loi normale de moyenne $\mu$ et de variance $\sigma^2$. On a relevé les valeurs suivantes :

$$5.42 \quad 5.55 \quad 5.61 \quad 5.93 \quad 6.15 \quad 6.20 \quad 6.79 \quad 7.07 \quad 7.37$$

1. Déterminer l'intervalle de confiance de $\mu$ au niveau 95%
   (a) si l'on suppose que $\sigma = 0.6$ g/l ;
   (b) si $\sigma$ est inconnu.
2. Déterminer l'intervalle de confiance de la variance $\sigma^2$ au niveau 95%
   (a) si l'on suppose $\mu = 6.2$ g/l connu ;
   (b) si $\mu$ est inconnu.

## 5.2 Intervalle de confiance sur une proportion

Un sondage electoral effectué sur 1000 personnes donne 52% d'intentions de vote à l'un des candidats.

1. Donner un intervalle de confiance de niveau 95% sur le score de ce candidat.
2. Combien faudrait-il interroger de personnes pour avoir un intervalle deux fois moins large pour une même confiance ?

## 5.3 Intervalle de confiance sur une proportion

Un sondage electoral effectué sur 1000 personnes donne 52% d'intentions de vote à l'un des candidats.

1. Donner un intervalle de confiance de niveau 95% sur le score de ce candidat.
2. Combien faudrait-il interroger de personnes pour avoir un intervalle deux fois moins large pour une même confiance ?

## 5.4 Comparaison de moyennes

On considère une variable quantitative $Y$ dont on cherche à comparer les valeurs dans deux groupes $A$ et $B$. Pour cela $n$ mesures indépendantes de $Y$ ont été faites dans $A$ et dans $B$. On suppose la normalité des données et l'égalité des variances dans chacun des deux groupes. Ainsi, pour $i = 1..n$,

$$Y_i^A \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_A, \sigma^2) \qquad Y_i^B \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_B, \sigma^2)$$

L'objet de l'exercice est de déterminer un intervalle de confiance pour $\mu_A - \mu_B$.

1. Estimateur de la variance commune
   (a) Sous les hypothèses de l'exercice, rappelez les lois de
   $$(n-1)\frac{\hat{\sigma}_A^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{Y_i^A - \bar{Y}_A}{\sigma}\right)^2 \text{ et } (n-1)\frac{\hat{\sigma}_B^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{Y_i^B - \bar{Y}_B}{\sigma}\right)^2$$
   (b) On admet que la somme de deux lois du $\chi^2$ indépendantes est une loi du $\chi^2$ dont le degré de liberté est la somme des degrés de liberté initiaux. Quelle est la loi de
   $$(n-1)\frac{\hat{\sigma}_A^2 + \hat{\sigma}_B^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{Y_i^A - \bar{Y}_A}{\sigma}\right)^2 + \sum_{i=1}^{n} \left(\frac{Y_i^B - \bar{Y}_B}{\sigma}\right)^2 \ ?$$
2. Estimateur de $\mu_A - \mu_B$

(a) Sous les hypothèses de l'exercice, quelle est la loi de $\bar{Y}_A - \bar{Y}_B$ ?

(b) Quelle est alors la loi de

$$\frac{(\bar{Y}_A - \bar{Y}_B) - (\mu_A - \mu_B)}{\sigma\sqrt{\frac{2}{n}}} \ ?$$

(c) On choisit de prendre

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \left\{ \sum_{i=1}^{n} (Y_i^A - \bar{Y}_A)^2 + \sum_{i=1}^{n} (Y_i^B - \bar{Y}_B)^2 \right\}$$

Quelle est alors la loi de

$$T = \frac{(\bar{Y}_A - \bar{Y}_B) - (\mu_A - \mu_B)}{\hat{\sigma}\sqrt{\frac{2}{n}}} \ ?$$

3. Application numérique : intervalle de confiance sur $\mu_A - \mu_B$

La teneur en viande maigre (TVM) des carcasses de porcs joue un rôle déterminant dans l'évaluation du prix d'une carcasse. Il est donc important de connaître les facteurs ayant une influence sur la variabilité de cette quantité.
On veut comparer la population des mâles et des femelles, c'est-à-dire comparer les paramètres qui les caractérisent. Pour cela, on tire un échantillon de $n = 15$ carcasses mâles et de $n = 15$ carcasses femelles. Le tableau suivant représente les teneurs en viande maigre de carcasses de porcs en fonction du sexe :

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TVM | 56.6 | 54.8 | 59 | 60.4 | 61.8 | 62.6 | 65.0 | 58.1 | 61.4 | 60.8 | 59.2 | 58.1 | 57.5 | 55.2 | 54.6 |
| Sexe | M | M | M | M | M | F | F | F | M | M | M | F | F | M | M |
| TVM | 61.6 | 56.9 | 61.3 | 67.2 | 53.9 | 54.1 | 62.0 | 63.5 | 58.1 | 56.0 | 51.5 | 63.8 | 58.1 | 58.2 | 61.3 |
| Sexe | F | F | F | F | F | M | M | F | F | M | M | F | F | M | F |

(a) Calculer la moyenne de TVM dans chaque groupe.

(b) Calculer la variance empirique dans chaque groupe.

(c) On suppose que les conditions de l'exercice sont respectées. Donner un intervalle de confiance pour $\mu_M - \mu_F$ au niveau 95%.