ED Probabilité Statistique

Tran Viet Chi, chi.tran@math.univ-lille1.fr.

1 TD 1 : Probabilités discrètes

Exercice 1 (Mots de passe)
1. Combien de mots de passe de 8 symboles peut-on créer avec 66 caractères ?
2. Combien existe-t-il de nombres entiers formés de 3 chiffres distincts ?
3. Combien de plaques d'immatriculation différentes contenant 4 lettres suivies de deux chiffres peut-on faire ?

Exercice 2 (Formules sommatoires)
Montrer les relations suivantes pour $p \in \mathbb{N}$, n, N, $q \in \mathbb{N}^*$:

$$\sum_{k=0}^{n} C_n^k = 2^n, \quad \sum_{k=0}^{n} k C_n^k = n2^{n-1}, \quad \sum_{k=0}^{n} C_p^k C_q^{n-k} = C_{p+q}^n,$$

$$\sum_{k=0}^{n} \left(C_n^k\right)^2 = C_{2n}^n, \quad nC_{n-1}^{q-1} = qC_n^q, \quad C_n^p + C_n^{p+1} = C_{n+1}^{p+1}, \quad \sum_{k=n}^{N} C_k^n = C_{N+1}^{n+1}.$$

Exercice 3 (Tiroir à chaussettes)
Un tiroir contient $2n \in \mathbb{N}$ chaussettes (n paires). Yoann, qui part en voyage a décidé d'emmener $2r$ chaussettes ($r \leq n$). Au moment de faire sa valise une panne d'électricité survient. Il prend donc $2r$ chaussettes au hasard. Quelle est la probabilité qu'il n'y ait parmi ces $2r$ chaussettes aucune paire complète ? Quelle est la probabilité qu'il y ait parmi ces $2r$ chaussettes exactement k paires complètes, avec $1 \leq k \leq r$?

Exercice 4 (Tableaux noirs)
Si 10 tableaux noirs doivent être affectés à 4 écoles, de combien de manières peut-on les répartir ? Qu'en est-il si chaque école doit recevoir au moins un tableau ?

Exercice 5 (?+ ?+ ?+ ?+ ?=1000)
Combien existe-t-il de 5-uplets $(i, j, k, l, m) \in \mathbb{N}^5$ tels que $i + j + k + l + m = 1000$?

Exercice 6 (Anniversaires)
On relève les dates d'anniversaire dans un groupe de 30 personnes (jours numérotés de 1 à 365). On suppose que les jours de naissance sont équiprobables. Quelle est la probabilité que deux personnes au moins aient le même anniversaire ?

Exercice 7 (Progression arithmétique)
On munit l'ensemble des parties de

$$\Omega = \{(a, b, c) \in \mathbb{N}^3,\ 1 \leq a < b < c \leq 50 \text{ et } c - b = b - a\}$$

de l'équiprobabilité P. Quelles sont alors les probabilités des événements :

$$A = \{(a, b, c) \in \Omega, b - a = 10\}$$
$$B = \{(a, b, c) \in \Omega, b - a > 2\}$$
$$C = \{(a, b, c) \in \Omega, c \neq 50\}$$

Indication : Pour déterminer le cardinal de Ω, on pourra calculer d'abord le nombre de valeurs possibles pour a et c lorsque b est fixé.

Exercice 8 (Formule du crible et applications)

1. Soient A et B deux ensembles. Exprimer $\mathbb{P}(A \cup B)$ en fonction de $\mathbb{P}(A)$, $\mathbb{P}(B)$ et $\mathbb{P}(A \cap B)$.
2. Soient $(A_n)_{n\in[\![1,N]\!]}$ une suite d'ensembles. Montrer la formule du crible :

$$\mathbb{P}\left(\bigcup_{n\in[\![1,N]\!]} A_n\right) = \sum_{n=1}^{N}(-1)^{n-1} \sum_{1<i_1<i_2<\cdots<i_n} \mathbb{P}\left(\bigcap_{k\in[\![1,n]\!]} A_{i_k}\right).$$

3. Si de plus, $\mathbb{P}\left(\bigcap_{k\in[\![1,n]\!]} A_{i_k}\right)$ ne dépend que de n, montrer que :

$$\mathbb{P}\left(\bigcup_{n\in[\![1,N]\!]} A_n\right) = \sum_{n=1}^{N}(-1)^{n-1} C_N^n \mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right).$$

4. Un facteur distrait distribue au hasard N lettres dans les N boîtes d'un immeuble. Quelle est la probabilité pour qu'aucun habitant de l'immeuble ne reçoive sa lettre ? Quelle est sa limite lorsque $N \to +\infty$?
5. Un coursier distribue à présent r prospectus dans ces mêmes n boîtes aux lettres ($r \geq N$), en oubliant au fur et à mesure où il les a déposés. Quelle est la probabilité pour que chaque locataire reçoive au moins un prospectus.

Exercice 9 (Alcootest)
On s'intéresse à la fiabilité d'un alcootest pour automobilistes. Grâce à des études statistiques sur un grand nombre d'automobilistes, on sait que $0,5\%$ d'entre eux dépassent la dose d'alcool autorisée.
Aucun test n'est fiable à 100%. Avec celui que l'on considère, la probabilité que le test soit positif quand la dose d'alcool autorisée est dépassée, et la probabilité que le test soit négatif quand elle ne l'est pas, valent toutes deux $\rho = 0,95$.
1. Quelle est la probabilité qu'un automobiliste ayant un test positif ait réellement dépassé la dose d'alcool autorisée ?
2. Quelle devrait être la valeur de ρ pour que cette probabilité soit de 95% ?
3. Un policier affirme : *Ce test est beaucoup plus fiable le samedi soir à la sortie des boites de nuit !* Sachant que la proportion d'automobilistes ayant trop bu est alors de 30%, déterminer s'il a raison.

2 TD 2 : Intro à la théorie de l'intégrale de Lebesgue

Exercice 10 (Vrai ou faux ?)
Les affirmations suivantes sont-elles vraies ou fausses ? Justifiez brièvement ou donnez un contreexemple.
1. Soit $f : E \mapsto F$ et $\mathcal{E} \subset \mathcal{P}(F)$. La tribu engendrée par l'image réciproque de $\mathcal{E}$ coïncide avec l'image réciproque de la tribu engendrée par $\mathcal{E}$.
2. Soit X un ensemble. Alors l'ensemble de parties de X est une tribu.
3. Si A est un ensemble inclus dans un ensemble B avec B mesurable, alors A est mesurable.
4. Si f et g sont mesurables, alors $h := \sup(f, g)$ est mesurable.

5. Si f est mesurable, alors $|f|$ est mesurable.

6. Si $|f|$ est mesurable, alors f est mesurable.

7. L'image réciproque par une fonction mesurable bijective d'une tribu est une tribu.

8. $\mathbf{1}_A$ est une fonction mesurable si et seulement si A est un ensemble mesurable.

9. Si μ est la mesure de comptage sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ et si f est une fonction mesurable de $\mathbb{N}$ dans $\mathbb{R}_+$, alors $\int f d\mu$ n'est autre que $\sum_{n\in\mathbb{N}} f(n)$.

10. On n'a l'égalité $\int_A f(x)d\mu(x) = \int(\mathbf{1}_A(x)f(x))d\mu(x)$ que si f est étagée positive.

11. Soit f Borélienne positive telle que $\forall n \in \mathbb{N}$, $f(n) = +\infty$ et soit λ la mesure de Lebesgue sur $\mathbb{R}_+$, alors $\int f d\lambda = +\infty$.

12. Pour toute fonction mesurable f de $\mathbb{R}$ dans $\mathbb{R}_+$ et pour toute mesure μ sur $\mathbb{R}$, la fonction $F(u) = \int_{-\infty}^{u} f(x)d\mu(x)$ est continue.

Exercice 11 (Familles générant la tribu Borélienne de $(0, 1)$)
Soit $\mathcal{B}(]0, 1[)$ la tribu Borélienne sur $]0, 1[$.

1. Montrer que tout ouvert de $]0, 1[$ peut s'écrire comme réunion dénombrable d'intervalles de $]0, 1[$ de la forme $]r - \delta, r + \delta[$ où r et δ sont des rationnels de $]0, 1[$.
2. Montrer que $\mathcal{B}(]0, 1[)$ est engendrée par chacune des familles suivantes :

- $\mathcal{C}_1 = \{[a, b],\ a \leq b,\ a, b \in]0, 1[\}$

- $\mathcal{C}_1 = \{[a, b],\ a \leq b,\ a, b \in]0, 1[\cap \mathbb{Q}\}$

- $\mathcal{C}_1 = \{]0, t],\ t \in]0, 1[\}$

- $\mathcal{C}_2 = \{]0, 1/2^n[,\ ]k/2^n, k/2^{n+1}[,\ n \in \mathbb{N},\ k \in \mathbb{N},\ 1 \leq k \leq 2^n - 1\}$.

3. On considère pour tout $n \in \mathbb{N}$ la tribu sur $]0, 1[$ suivante :

$$\mathcal{B}_n := \sigma\left(\left\{\left]0, \frac{1}{2^n}\right[,\ \left[\frac{k}{2^n}, \frac{k}{2^{n+1}}\right[,\ k \in \mathbb{N},\ 1 \leq k \leq 2^n - 1\right\}\right).$$

Montrer que la suite des $(\mathcal{B}_n)_{n\in\mathbb{N}}$ est croissante au sens de l'inclusion mais que $\cup_n \mathcal{B}_n$ n'est pas une tribu.

Exercice 12 (Fonction mesurable Lebesgue-intégrable mais non Riemann-intégrable)
On pose $f = \mathbf{1}_{[0,1]\cap\mathbb{Q}}$, c'est-à-dire $f(x) = 1$ si $x \in [0, 1] \cap \mathbb{Q}$ et $f(x) = 0$ sinon.

1. Montrer que f n'est pas Riemann intégrable sur $[0, 1]$.

2. On rappelle que $[0, 1] \cap \mathbb{Q}$ est dénombrable et on note $(r_n)_{n\in\mathbb{N}^*}$ une numérotation de cet ensemble. On pose $f_n = \mathbf{1}_{r_1,\cdots r_n}$ pour $n \geq 1$. Montrer que f_n est Riemann intégrable sur $[0, 1]$. Montrer que (f_n) converge simplement vers f. Que peut-on en déduire ?

3. Montrer que f est mesurable et calculer l'intégrale (de Lebesgue) de f ?

Exercice 13 (Calcul d'intégrales par rapport à des mesures autres que Lebesgue)

Calculer les intégrales suivantes :

$$\int_{\mathbb{R}} \sin(x)\mu(dx) \text{ où } \mu(dx) = \sum_{n\in\mathbb{N}} \delta_{\pi n}(dx)$$
$$\int_{\mathbb{R}} x^3\mu(dx) \text{ où } \mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\lambda(dx), \quad \lambda \text{ étant la mesure de Lebesgue sur } \mathbb{R}$$
$$\int_{\mathbb{R}} \mathbf{1}_{[1,2]}(x)e^x \sin(\pi x/2)\mu(dx) \text{ où } \mu(dx) = \delta_2(dx) + e^{-x}\lambda(dx).$$

Exercice 14 (Convergences d'intégrales)
Déterminer si elles existent les limites suivantes :

$$\lim_{n\to+\infty} \int_0^{+\infty} e^{-n\sin^2(x)} f(x)dx \text{ où } f \text{ est une fonction intégrable}$$
$$\lim_{n\to+\infty} \int_0^{n} \left(1-\frac{x}{n}\right)^n e^{x/2}dx, \quad \lim_{n\to+\infty} \int_0^{n} \ln(x)\left(1-\frac{x}{n}\right)^n dx.$$

Exercice 15 (Dérivation sous le signe somme)
Soit $f : x \mapsto \int_0^{+\infty} \frac{e^{-xt}}{1+t^2}dt$.
1. Montrer que f est définie pour tout $x \in \mathbb{R}_+$ et à valeurs réelles.
2. Montrer que f est deux fois dérivable sur $\mathbb{R}_+^*$ et qu'elle vérifie $f''(x) + f(x) = 1/x$. En déduire f.

Exercice 16 (Fonction de deux variables)
Soit $f(x,y) = 1/(1-xy)$.
1. Montrer que $\int_{[0,1]^2} f(x,y)dx\,dy = \sum_{n\geq 1} 1/n^2$.

2. En faisant le changement de variable $x = (u+v)/\sqrt{2}$, $y = (u-v)/\sqrt{2}$, montrer que $\sum_{n\geq 1} 1/n^2 = \pi^2/6$.

3 TD 3 : Variables aléatoires discrètes

Exercice 17 (Dauphins)
Une famille de dauphins est constituée de six femelles et de quatre mâles. On choisit au hasard dans cette famille un groupe de quatre dauphins. Soit X la variable aléatoire représentant le nombre de femelles du groupe.
1. Quelles sont les valeurs possibles de X ?
2. Déterminer la loi de X.
3. Calculer $\mathbb{E}(X)$.
4. Calculer $\mathrm{Var}(X)$.

Exercice 18 (Grossiste)
Un grossiste estime que la demande en tonnes de denrées périssables est une variable aléatoire X de loi

k	0	1	2	3	4	5
$\mathbb{P}(X=k)$	$0,05$	$0,15$	$0,2$	$0,35$	$0,15$	$0,1$

1. Calculer la demande moyenne.
2. Calculer la variance de X.
3. Calculer la probabilité que la demande soit

1. inférieure ou égale à deux tonnes.
2. supérieure ou égale à une tonne et inférieure ou égale à trois tonnes.
3. strictement supérieure à deux tonnes.

4. Le stock du grossiste est de trois tonnes. Il gagne cinq mille euros par tonne vendue et perd deux mille euros par tonnes invendue. On note Y la variable aléatoire discrète représentant son bénéfice ou sa perte.

1. Exprimer Y en fonction de X.
2. Calculer $\mathbb{E}(X)$.
3. Calculer $\mathrm{Var}(X)$.

Exercice 19 (Ajustement d'un paramètre)
Soit $a \in \mathbb{R}^+$ et X la variable aléatoire discrète à valeurs dans $\{1,2,3,4,5\}$ définie par la loi de probabilité

$$\mathbb{P}(X=k) = \frac{1}{6a}(k-a)^2.$$

1. Pour quelle(s) valeur(s) de a la loi ci-dessus est-elle bien une loi de probabilité ?

On rappelle les formules $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$ et $\displaystyle\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2$.

2. Calculer $\mathbb{E}(X)$ et $\mathrm{Var}(X)$.

Exercice 20 (Encore des urnes)
Dans une urne contenant initialement $n \in \mathbb{N}^*$ boules numérotées de 1 à n, on effectue deux tirages successifs suivant la technique suivante : si on tire au premier coup la boule numéro k, alors celle-ci est remise dans l'urne avec k boules supplémentaires portant toutes le numéro k. On effectue alors le second tirage.

On appelle X_1 la variable égale au numéro de la boule tirée au premier coup, et X_2 celle égale au numéro de la boule tirée au second coup.
1. Déterminer la loi de X_1, son espérance et sa variance.
2. Déterminer la loi de X_2 et vérifier que :

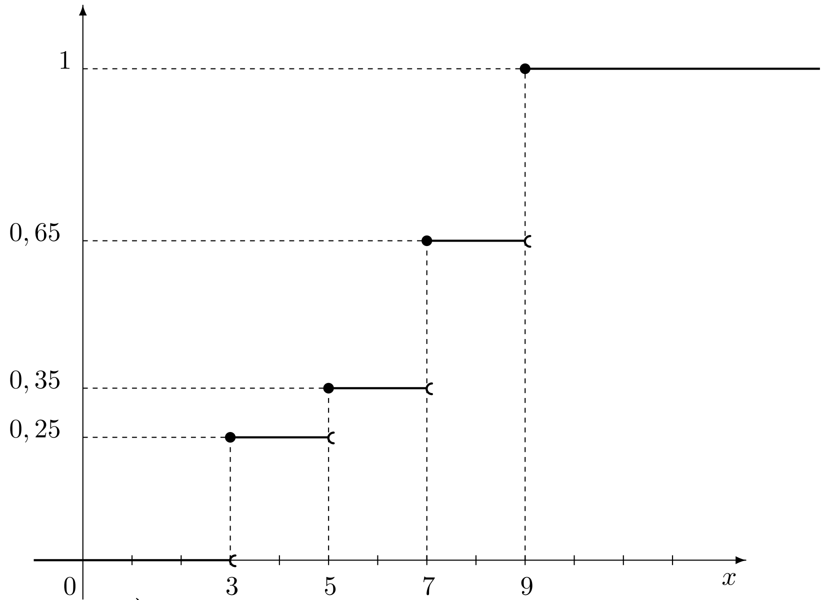
$$\sum_{k=1}^{n} \mathbb{P}(X_2=k) = 1.$$

3. Montrer que :

$$\mathbb{E}(X_2) = \frac{1-n}{2} + \frac{3n+1}{2}\sum_{k=1}^{n}\frac{1}{n+k}.$$

indication : $k/(n+k) = k - n + n^2/(n+k)$. **4.** Déterminer un équivalent simple de $E(X_2)$ lorsque $n \to +\infty$.

Exercice 21 (Lire une fonction de répartition)
Voici le graphe de la fonction de répartition F_X d'une variable aléatoire X donnée.
En déduire la loi de X, qu'on présentera sous forme de tableau.



Exercice 22 (Les cravates)
Dans cet exercice, on précisera pour chaque loi :

- *son nom si elle en a un,*
- *les valeurs que peut prendre la variable aléatoire correspondante,*
- *et la probabilité de chacune de ces valeurs.*

1. Monsieur Zzzz possède 50 cravates, dont une seule est à rayures. Tous les matins, il prend une cravate au hasard dans l'armoire, et tous les soirs il remet la cravate du jour à sa place.
On observe Monsieur Zzzz pendant 20 jours et on appelle X le nombre de fois où il porte une cravate à rayures. Quelle est la loi de la variable aléatoire X ?
Application numérique : Calculer $P(X = 1)$.

2. Monsieur Zzzz part en voyage. Il met dans sa valise 20 cravates prises au hasard dans l'armoire.
Quelle est la loi du nombre V de cravates à rayures contenues dans la valise ?
Application numérique : Calculer $P(V = 1)$.

3. Monsieur Zzzz possède aussi 10 chemises dont 3 sont bleues. Il prend 5 chemises au hasard et les mets dans sa valise.
Quelle est la loi du nombre Y de chemises bleues contenues dans la valise ?
Application numérique : Calculer $P(Y = 1)$.

Exercice 23 (Calcul d'espérance)
Soit X une variable de Poisson de paramètre $\lambda > 0$. Calculer $\mathbb{E}\left(\dfrac{1}{1+X}\right)$.

Exercice 24 (Oeufs de tortues)
Le nombre X d'œufs pondus par une tortue au cours d'une ponte suit une loi de Poisson de paramètre $\lambda > 0$. Un œuf a la probabilité p d'arriver à éclosion. Quelle est la loi du nombre Y de bébés tortues à chaque ponte ?

Exercice 25 (Variable géométrique)
1. Soit X une variable géométrique de paramètre $p \in]0,1[$. Montrer que $\mathbb{P}(X = n + k | X > n) = \mathbb{P}(X = k)$.
2. Soit X une variable aléatoire sur $\mathbb{N}$ telle que $\mathbb{P}(X = 1) = p$ avec $0 < p < 1$ et pour tout k entier positif ou nul, $\mathbb{P}(X = 1 + k | X > 1) = \mathbb{P}(X = k)$. Montrer que X suit une loi géométrique de paramètre p.

Exercice 26 (Pile ou face)
Sébastien joue à pile ou face avec une pièce truquée où il a une probabilité p de faire "pile".

1. Quelle est la loi du nombre de fois où il doit jouer pour obtenir une victoire ? Quelle est l'espérance de cette loi ? Quelle en est la variance ?

2. Séb joue maintenant avec Aldéric une suite de parties indépendantes. Lors de chacune d'elles, ils ont respectivement les probabilités p et $q = 1 - p$ de gagner. Le vainqueur est celui des deux joueurs qui le premier obtient 2 victoires de plus que son adversaire. Quelle est la probabilité pour qu'Aldéric soit le vainqueur final ?

4 TD 4 : Variables aléatoires continues

Exercice 27 (Moments de la loi normale)
Calculer les moments centrés d'une loi normale d'espérance m et de variance σ^2.

Exercice 28 (Lois normales et application aux stats)
Soient $X_1, \cdots X_n$ des variables aléatoires gaussiennes indépendantes de loi $\mathcal{N}(m, \sigma^2)$. On pose :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i,\ \Sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - m)^2,\ S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

1. Quelle est la loi de $\bar{X}$?

2. Quelle est la loi de $n\Sigma^2/\sigma^2$?

3. Montrer que $\bar{X}$ et S^2 sont indépendantes.

4. Montrer que nS^2/σ^2 suit une loi du χ^2 à $n-1$ degrés de liberté.

Exercice 29 (Inverse généralisé d'une fonction de répartition et applications)
Soit F la fonction de répartition d'une loi de probabilité μ sur $\mathbb{R}$. On pose pour $t \in [0,1]$, $F^{-1}(t) = \inf\{s \in \mathbb{R},\ F(s) \geq t\}$.

1. Calculer $F(F^{-1}(t))$ et $F^{-1}(F(x))$.

2. On suppose que U est une variable aléatoire de loi uniforme sur $[0,1]$. Quelle est la loi de $F^{-1}(U)$? A quoi cette propriété peut-elle servir ?

3. Calculer F et F^{-1} pour la loi exponentielle de paramètre λ.

Exercice 30 (Liens entre fonction de répartition, espérance. Application à la médiane)
Soit X une variable aléatoire intégrable à valeurs dans $\mathbb{R}$. On notera $\mu(dx)$ la loi de X.

1. Montrer que son espérance est donnée par :

$$\mathbb{E}(X) = \int_0^{+\infty} S(x)\, ds$$

où $S(x) = 1 - F(x) = \mathbb{P}(X > x)$ est la fonction de survie de X.
2. Plus généralement, établir que pour tout réel a :

$$\mathbb{E}(|X-a|) = \int_a^{+\infty} S(x)dx + \int_{-\infty}^a F(x)dx.$$

3. On suppose que X est une variable à densité. Montrer alors que le minimum de $\mathbb{E}(|X-a|)$ est atteint pour toutes les valeurs a telles que $F(a) = 1/2$. Comment s'interprètent ces valeurs ?

Exercice 31 (Utilisation de fonctions caractéristique)
Soient U et V deux variables aléatoires réelles indépendantes qui suivent une loi normale centrée réduite.

1. Quelle est la loi de $W = U/V$?

2. En déduire la fonction caractéristique de la loi de Cauchy de paramètre $\alpha > 0$ dont la densité est $x \mapsto \alpha/(\pi(x^2+\alpha^2))$. On pourra utiliser l'identité suivante qui est vraie pour une fonction intégrable ψ et un réel $\gamma > 0$,

$$\int_{-\infty}^{+\infty} \psi\left(x - \frac{\gamma}{x}\right) dx = \int_{-\infty}^{+\infty} \psi(x)dx.$$

3. Soient $X_1, \cdots X_n$ des variables aléatoires indépendantes de loi de Cauchy de paramètre α. Quelle est la loi de $\sum_{i=1}^n X_i/n$?
4. Réciproquement, soit $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de même loi μ symétrique. On suppose que pour tout entier $n \in \mathbb{N}^*$, $\sum_{i=1}^n X_i/n$ a pour loi μ. Déterminer μ.

5 TD 5 : Vecteurs aléatoires, espérance conditionnelle

Exercice 32 (Couple de variables aléatoires)
Soient X et Y les variables aléatoires discrètes dont la loi jointe est donnée par le tableau suivant :

$X \setminus Y$	-1	0	2	5
0	0,10	0,05	0,15	0,05
1	0,15	0,20	0,25	0,05

1. Vérifier que ce tableau définit bien une loi de probabilité bivariée.
2. Quelle est la loi marginale de X ?
3. Quelle est la loi marginale de Y ?
4. Calculez $\mathbb{P}(Y \geq 0 \,|\, X = 1)$.
5. Calculez les espérances de X, Y, leurs variances, et la covariance de X et Y.

Exercice 33 (Couple de variables aléatoires (2))
Soient X et Y les variables aléatoires discrètes dont la loi jointe est donnée par le tableau suivant :

$X \setminus Y$	0	1	2
0	0	1/5	1/5
1	1/5	1/5	1/5

1. Donner la loi et la fonction de répartition de X.
2. Donner la loi de Y.
3. Les variables X et Y sont-elles indépendantes ?

Exercice 34 (Casino)
Dans un casino, James participe à une suite de parties indépendantes. On désigne par $X_1, X_2, \cdots$ ses gains (positifs ou négatifs) définis comme la différence de la somme gagnée et de la mise, à chaque partie. On note S_n la fortune après n parties : $S_n = X_1 + \cdots + X_n$. On suppose que le jeu est équitable, c'est-à-dire que $\forall i$, $\mathbb{E}(X_i) = 0$. On pose $\mathcal{F}_n = \sigma(X_1, \cdots X_n) = \sigma(S_1, \cdots S_n)$.

1. Que représente $\mathcal{F}_n$?

2. Calculer $\mathbb{E}(S_{n+1} \,|\, \mathcal{F}_n)$.

James veut adopter une stratégie de jeu : à la n-ième partie, il mise la somme H_n. Ainsi sa fortune après n parties devient $\widetilde{S}_n = \sum_{k=1}^n H_k X_k$.

3. Quelle hypothèse raisonnable de mesurabilité doit-on faire pour modéliser correctement le problème ?

4. Avec cette hypothèse, calculer $\mathbb{E}(\widetilde{S}_{n+1} \,|\, \mathcal{F}_n)$. Commenter.

Exercice 35 (Calcul de normes L^p)
Soit $p \in [1, +\infty[$, soient X et Y deux variables aléatoires réelles indépendantes de L^p avec $\mathbb{E}(Y) = 0$. Calculer $\mathbb{E}(X + Y \,|\, X)$ et en déduire que $\|X\|_p \leq \|X + Y\|_p$.

Exercice 36 (Calculs)
Soient X et Y deux variables aléatoires réelles. On suppose que la loi de X a pour densité

$$f(x) = \frac{2}{\ln(2)^2} \mathbf{1}_{[0,1]}(x) \frac{\ln(1+x)}{1+x}$$

et que la loi conditionnelle de Y sachant $X = x$ a pour densité :

$$\mathbf{1}_{[0,x]}(y) \frac{1}{(1+y)\ln(1+x)}.$$

1. X et Y sont-elles indépendantes ?

2. Quelle est la loi de Y ?

3. Quelle est la loi de X sachant $Y = y$?

4. Déterminer $\mathbb{E}(Y \,|\, X)$.

6 TD 6 : Théorèmes limites

Exercice 37 (Estimateurs d'une proportion et d'une variance)
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et de même loi de Bernoulli $\mathcal{B}(1,p)$ avec $0<p<1$. Pour tout $n\in\mathbb{N}^*$, on pose :

$$S_n=\sum_{k=1}^{n}X_k \quad \text{et } \bar{X}_n=\frac{S_n}{n}.$$

1. Montrer que $\forall n\in\mathbb{N}^*$, $\mathbb{E}\left(\bar{X}_n\right)=p$ et que $\lim_{n\to+\infty}\bar{X}_n=p$ p.s.
2. Quelle est la variance σ^2 d'une variable aléatoire suivant la loi $\mathcal{B}(1,p)$?
3. On se propose d'approcher σ^2 par la suite $(U_n)_{n\in\mathbb{N}^*}$ où $U_n=\bar{X}_n(1-\bar{X}_n)$.

1. Calculez $\mathbb{E}\left(U_n\right)$ et $\lim_{n\to+\infty}U_n$ (précisez le sens de cette limite).
2. Proposez une autre approximation de σ^2, V_n qui vérifierait $\mathbb{E}(V_n)=\sigma^2$.

Exercice 38 (Intervalles de confiance)
Dans une population de $N=30\,000\,000$ individus, la proportion d'individus présentant de plus de 1 m 77 est $p=0,4$. On prélève un échantillon de taille $n=1\,600$ et on note X_n le nombre d'individus de l'échantillon de plus de 1 m 77.
1. Quelle est la limite $\lim_{n\to+\infty}X_n/n$?
2. Minorez la probabilité des événements :

$$\{0.30\le X_n/n\le 0.50\},\quad \{0.35\le X_n/n\le 0.45\},\quad \{0.38\le X_n/n\le 0.42\}$$

3. Calculez la longueur minimale L de la fourchette telle que :

$$\mathbb{P}\left(0,40-\frac{L}{2}\le\frac{X_n}{n}\le 0.40+\frac{L}{2}\right)\ge 0.95$$

Exercice 39 (Convergence d'une suite de v.a. Gaussiennes)
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles de lois respectives $\mathcal{N}(m_n,\sigma_n^2)$. Montrer qu'il y a équivalence entre :

(i) X_n converge en loi vers une loi qui n'est pas une loi constante,

(ii) les suites $(m_n)_{n\in\mathbb{N}}$ et $(\sigma_2^1)_{n\in\mathbb{N}}$ convergent et la limite de $(\sigma_2^1)_{n\in\mathbb{N}}$ n'est pas 0.s

Exercice 40 (Formule de Stirling)
Soit $(X_i)_{i\in\mathbb{N}}$ des variables aléatoires réelles indépendantes identiquement distribuées de loi de Poisson de paramètre 1. On note $S_n=X_1+\cdots+X_n$.

1. Quelle est la loi de S_n ?
2. En notant $[.]_-$ la partie négative, étudier la convergence en loi de :

$$\left[\frac{S_n-n}{\sqrt{n}}\right]_-$$

3. Montrer que

$$\mathbb{E}\left[\frac{S_n-n}{\sqrt{n}}\right]_-\to\frac{1}{\sqrt{2\pi}}$$

4. En déduire la formule de Stirling :

$$n!\equiv\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

7 TP 1 : Débuter avec R

7.1 Séquences, Vecteurs, Tableaux (arrays)

7.1.1 Déclaration et définition

1. Saisir le vecteur a en tapant `a <- c(10, 5, 3, 6, 21)`. Taper a. Puis $a[2]$ et $a[1,3]$.

2. Saisir le vecteur b en tapant cette fois `b<- array(data=c(15, 3, 12, 2, 1),dim=c(1,5))`. Demander b.

3. Pour voir la différence entre une liste et un vecteur, demander `nrow(a)`, `ncol(a)`, `dim(a)` et de même pour b.

4. Générer un vecteur c de dimension 5 dont toutes les composantes sont 1 en utilisant la commande `array`.

5. Générer un vecteur d en tapant `d<-seq(from=1, to=10, by=2)`.

6. Si on veut générer des séquences d'entiers consécutifs, on peut utiliser la commande `1:5`. Créer un vecteur e de dimension 5 et de composantes la séquence des 5 premiers entiers.

7. Générer une matrice diagonale de dimension 3 dont les éléments de la diagonale sont 1, 5 et 9 à l'aide de la commande `diag`.

7.1.2 Manipulations de base

1. Taper `2*a+b+1`. Qu'obtient-on ?

2. Taper `e[3]`. Qu'obtient-on ?

3. Taper `cos(a)`, `exp(a)`. Qu'obtient-on ?

4. Taper `a*e` puis `a%*%e`. Comparer.

5. Taper maintenant `b%*%e`. Que se passe-t-il ? Corriger par `f<- t(b)%*%e`.

6. Demander `dim(f)`. Taper `f[2,3]`, `f[,3]`, `f[2:5,]` puis `f[2:3,4]`. Qu'obtient-on ?

7. Taper `cbind(b,e)` puis `rbind(b,e)`. Pourquoi faut-il éviter d'utiliser `cbind` ou `rbind` avec des objets du type de a ? (Essayer `cbind(a,b)` et `rbind(a,b)`).

8. Taper `c(a,b)`. Qu'obtient-on ?

7.2 Listes

1. Taper

`Lst <- list(name="Fred", wife="Mary", no.children=3,child.ages=c(9,7,4))`.

Demander `Lst`.

2. Demander `Lst[[1]]`. Retrouver le résultat en tapant `Lst$name`

3. Demander `Lst[[4]]`. Qu'obtient-on. Demander l'âge du troisième enfant.

7.3 Exercices

Exercice 41 (Résolution d'un système linéaire et inversion d'une matrice carrée)
Soient :

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 3 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

1. Saisir la matrice A (par colonnes) et le vecteur b. La matrice A est-elle inversible ? Calculer son déterminant grace à la commande `det(A)`.

2. Résoudre $Ax = b$ avec la commande `solve(A,b)`.

3. Pour `eigen(A, symmetric=FALSE, only.values = FALSE)`. Le résultat de cette commande est une liste. En extraire les valeurs propres, dans une matrice que l'on appellera D et les vecteurs propres dans une matrice de passage que l'on appellera P.

4. calculer P^{-1} à l'aide de la commande `solve(P)` et retrouver A à partir de P, D et P^{-1}.

Exercice 42 (Visualisation des différentes convergences en probabilité)
Simuler $N = 10000$ variables aléatoires i.i.d. de loi $\mathcal{N}(m, \sigma^2)$ avec $m = 2$ et $\sigma^2 = 1$.
1. Calculer et ranger dans dans un vecteur les moyennes successives

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

pour n variant de 10 à N. Puis tracer le graphe de $\bar{X}_n$ en fonction de n. Commentaire ?

2. Calculer et ranger dans un vecteur les quantités

$$\zeta_n = \sqrt{n} \frac{\bar{X}_n - m}{\sigma}$$

pour n variant de 10 à N. Tracer le graphe de ζ_n en fonction de n. Commentaire ?
3. Calculer (et garder en mémoire) $M = 1000$ valeurs de ζ_N que l'on notera $(\zeta_N^j)_{j \in [\![1,M]\!]}$. Tracer l'histogramme de ces valeurs et superposer lui la densité de la loi normale $\mathcal{N}(0,1)$. Tracer le graphique quantile-quantile d'adéquation avec la loi normale centrée réduite. Commentaire ?

8 TD 7 : Maximum de vraisemblance

Exercice 43 (Information de Fisher)
Calculer l'information de Fisher dans les modèles statistiques suivants :
1. Loi de Poisson de paramètre λ,
2. Loi normale $\mathcal{N}(m, \sigma^2)$

Exercice 44 (EMV et lois asymptotiques)
Calculer l'estimateur du maximum de vraisemblance lorsque l'échantillon iid $(X_1, \cdots X_n)$ suit :

1. une loi de Bernoulli $\mathcal{B}(1,p)$, p étant le paramètre. Donner la loi limite de $\sqrt{n}(\widehat{p}_{\text{emv}} - p)$.

2. une loi normale $\mathcal{N}(m, \sigma^2)$, (m, σ^2) étant le paramètre. Donner la loi limite de

$$\sqrt{n}\begin{pmatrix} \widehat{m} - m \\ \widehat{\sigma^2} - \sigma^2 \end{pmatrix}$$

3. une loi uniforme $\mathcal{U}([a,b])$, (a,b) étant le paramètre. Donner la loi limite du vecteur

$$n\begin{pmatrix} \widehat{a} - a \\ b - \widehat{b} \end{pmatrix}$$

Exercice 45 (Exo sur les durées de vie)
On dispose de n observations $y_1, \cdots y_n$ sur les durées de vie de certains composants industriels. On suppose que les variables aléatoires $Y_1, \cdots Y_n$ associées sont iid de loi exponentielle $\mathcal{E}(1/\theta)$ où $\theta > 0$.

1. Soit F la fonction de répartition de Y_1. On cherche à estimer la fonction de survie $S(t) = 1 - F(t)$. Calculer S.

2. Calculer l'EMV de θ et en déduire un estimateur convergent $\widehat{S}(t)$ de $S(t)$ pour $t \in \mathbb{R}_+$ fixé. Pourquoi $\widehat{S}(t)$ est-il biaisé pour n fixé fini ?

3. Calculer la loi limite de $\sqrt{n}(\widehat{S}(t) - S(t))$.

Soit $t \in \mathbb{R}_+$ fixé et T la variable aléatoire définie par $T = \mathbf{1}_{Y_1 > t}$. On note $\Sigma_n = Y_1 + \cdots Y_n$.

4. Déterminer la loi de Y_1 conditionnellement à Σ_n.

5. Calculer $T^* = \mathbb{E}(T \mid \Sigma_n)$. Comment s'appelle cet estimateur ?

6. Montrer que T^* est l'estimateur sans biais de $S(t)$ optimal parmi les estimateurs sans biais.

7. Peut-on dire si T^* est efficace à distance finie.

9 TP 2 : Maximum de vraisemblance

Exercice 46 (Maximum de vraisemblance pour la loi Gamma)
Les lois Gamma $\Gamma(k,\theta)$ sont des lois de probabilité définies à partir :

1. d'un paramètre d'échelle $\theta > 0$
2. d'un paramètre de forme $k > 0$

par des densités de probabilité de la forme :

$$f(x\,;\,\theta,k) = \frac{x^{k-1}\exp\left(-\frac{x}{\theta}\right)}{\Gamma(k)\theta^k}\,\mathbf{1}_{]0,+\infty[}(x), \quad \text{où} \quad \Gamma(k) = \int_0^{+\infty} t^{k-1}e^{-t}dt. \tag{9.1}$$

Les lois Gamma sont utilisées pour modéliser des variables aléatoires positives, souvent des durées. Par exemple, pour des événements se réalisant suivant un processus de Poisson de paramètre θ, l'événement N (pour N fixé) suit une loi $\Gamma(N,\theta)$.

Partie A (théorique)

1. Vérifier que lorsque $k = 1$, on retrouve une loi exponentielle.
2. Soient $(X_i)_{i\in[\![1,n]\!]}$ n variables aléatoires indépendantes de lois respectives $\Gamma(\alpha_i,\,\theta)$, $i \in [\![1,n]\!]$. Montrer que $\sum_{i=1}^n X_i$ suit une loi $\Gamma\left(\sum_{i=1}^n \alpha_i, \theta\right)$ (Propriété d'infinie divisibilité).

Partie B - Estimateur du maximum de vraisemblance

1. Dessiner la densité de la loi Gamma pour $k = 3$ et $\theta = 0.5$.

2. On se donne n variables aléatoires $(X_i)_{i\in[\![1,n]\!]}$ indépendantes identiquement distribuées de loi $\Gamma(k,\theta)$. Ecrire la log-vraisemblance renormalisée des observations $\ell(x_1,\cdots x_n\,;\,k,\theta)/n$ et programmer-la en tant que fonction **R** (On utilisera la commande `gamma(k)` pour calculer $\Gamma(k)$).

3. Simuler un tel échantillon avec $n = 1000$, $k = 3$ et $\theta = 0.5$. Nous nous proposons dans cet exercice de déterminer les estimateurs du maximum de vraisemblance $\widehat{k}$ et $\widehat{\theta}$.

4. Représenter par un graphique en 2D la fonction $(k,\theta) \in \mathbb{R}_+^2 \mapsto \ell(x_1,\cdots x_n\,;\,k,\theta)/n \in \mathbb{R}$. On calculera pour cela la valeur de la fonction sur une grille de $]0,5]^2$ de pas $0,05$, puis on utilisera les commandes `persp`, `image` et `contour`. En particulier, pour `contour`, jouer avec les options `levels nlevels` et modifier le cadrage avec les options `xlim=c(\cdots,\cdots) ylim=c(\cdots,\cdots)`. Déterminer graphiquement l'estimateur du maximum de vraisemblance (EMV).

5. Déterminer par le calcul l'EMV $\widehat{\theta}_n$ du paramètre θ.

6. Peut-on obtenir une expression explicite pour l'EMV $\widehat{k}_n$ du paramètre k ? Montrer qu'il vérifie l'équation suivante où ψ est la fonction digamma définie par $\psi(k) = \Gamma'(k)/\Gamma(k)$) :

$$\ln(k) - \psi(k) = \ln\left(\frac{1}{n}\sum_{i=1}^n x_i\right) - \frac{1}{n}\sum_{i=1}^n \ln(x_i). \tag{9.2}$$

7. Etudier la convexité de la fonction $k \mapsto \ln(k) - \psi(k)$.

8. Nous souhaitons résoudre numériquement (9.2) en utilisant la méthode de Newton-Raphson.
8.1. Rappeler le principe de cette méthode.
8.2. Déterminer l'expression pour passer de l'approximation $k^{(m)}$ à l'approximation $k^{(m+1)}$ de la solution de (9.2) (on fera intervenir la fonction trigamma ψ', dérivée de la fonction ψ).
8.3. Pour déterminer la condition initiale, nous allons utiliser l'approximation suivante, qui fait l'objet de la partie **D** (estimation par la méthode des moments) :

$$\widetilde{\theta}_n = \frac{s_n^2}{\bar{x}_n}, \quad \widetilde{k}_n = \frac{\bar{x}_n^2}{s_n^2}, \tag{9.3}$$

où $\bar{x}_n$ et s_n^2 sont respectivement la moyenne et la variance empirique (commandes `mean` et `var`).
8.4. Rassembler ces résultats dans une fonction EMV1 d'argument $(x_1,\cdots,x_n)$ et retournant :
– l'EMV $(\widehat{k}_n,\widehat{\theta}_n)$,

– le nombre d'itérations faites dans l'algorithme de Newton-Raphson pour obtenir le résultat ainsi,
– la valeur de $\ln(k) - \psi(k) - \ln(\bar{x}_n) + \overline{\ln(x)}$.

On utilisera les commandes `digamma(k)` ou `psigamma(k,1)` pour calculer $\psi(k)$ et les commandes `trigamma(k)` ou `psigamma(k,2)` pour calculer $\psi'(k)$.

9. Nous souhaitons comparer les résultats précédents à ceux retournés par les méthode `nlm` et `mle` pré-implémentée dans **R**.
9.1. Montrer que $\widehat{k}_n$ minimise la fonction suivante :

$$\widetilde{\ell}(x_1, \cdots x_n\,;\, k) = -k\left(\sum_{i=1}^{n}\frac{\ln(x_i)}{n} - 1 - \ln(\bar{x}) + \ln(k)\right) + \ln(\Gamma(k)) \tag{9.4}$$

9.2. Tracer la courbe de la fonction $k \mapsto \widetilde{\ell}(x_1, \cdots x_n\,;\, k)$. Déterminer graphiquement son minimum.
9.3. Calculer $\widehat{k}_n^{(2)}$ en minimisant la fonction $k \mapsto \widetilde{\ell}(x_1, \cdots x_n\,;\, k)$ à l'aide de la commande `nlm`. On prendra les mêmes conditions initiales qu'en (9.3).
9.4. Calculer $\widehat{k}_n^{(3)}$ en utilisant la commande `mle`, après avoir préalablement appelé la librarie `stats4` (Taper `library(stats4)`).
9.5. Rassembler les résultats précédents dans une fonction EMV2 qui prend en argument $(x_1, \cdots, x_n)$ et retourne les EMV $(\widehat{k}_n^{(2)}, \widehat{k}_n^{(3)})$.
9.5. Pour comparer les trois estimateurs de k :
– Générer $N = 1000$ échantillons de $n = 100$ observations indépendantes et de loi $\Gamma(k = 3, \theta = 0.5)$.
– Pour chacun des échantillons généré, calculer $\widehat{k}_n$, $\widehat{k}_n^{(2)}$ et $\widehat{k}_n^{(3)}$ en utilisant les fonctions obtenues aux questions **8** et **9**. Conserver les valeurs trouvées dans un tableau.
– A la fin de la manœuvre, nous avons trois séries de $N = 1000$ réalisations (couplées) des estimateurs $\widehat{k}_n$, $\widehat{k}_n^{(2)}$ et $\widehat{k}_n^{(3)}$. Utiliser-les pour comparer ces deux distributions (moyennes, QQ-plots ...)

Partie C Consistence et vitesse de convergence

1. (théorique) L'EMV dans cet exercice est-il convergent ? asymptotiquement normal ?

2. Ecrire une fonction qui prend un entier n en argument et :
– génère un échantillon *iid* de taille n dans la loi $\Gamma(k = 3, \theta = 0.5)$,
– calcule $\widehat{k}_n - 3$ à l'aide de la fonction EMV1 obtenue à la partie **B.8**,

(c'est une fonction aléatoire, puisque $\widehat{k}_n$ est une variable aléatoire qui dépend de l'échantillon généré)

3. Tracer le graphe de $n \mapsto \widehat{k}_n - 3$ pour $n \in [\![10, 3\,000]\!]$, n variant de 10 en 10.

4. On cherche à montrer que $\sqrt{n}(\widehat{k} - k)$ est normalement asymptotique.
– Générer $N = 1000$ échantillons de $n = 100$ observations indépendantes et de loi $\Gamma(k = 3, \theta = 0.5)$.
– Pour chacun de ces échantillons, calculer $\sqrt{n}(\widehat{k} - k)$. On obtient $N = 1000$ réalisations de cette variable aléatoire.
– Faire un QQ-plot pour la comparer avec une loi normale.

5. Par quelle quantité doit-on renormaliser $\sqrt{n}(\widehat{k} - k)$ pour obtenir une loi asymptotique $\mathcal{N}(0,1)$? Reprendre la question 4 avec une version renormalisée de $\sqrt{n}(\widehat{k} - k)$ et comparer avec la loi $\mathcal{N}(0,1)$.

10 TP 3 : Régression linéaire

Exercice 47 (Ozone de l'air)
La table `ozone.dta` contient les variables suivantes, pour une série de journées (qui sont ici nos individus) :
- l'identifiant de la journée,
- le maximum d'ozone (variable `max03`)
- l'heure à laquelle le maximum d'ozone a été obtenu (`heure`),
- les températures à 6h, 9h, 12h, 15h, 18h (resp. `T6` à `T18`)
- la nébulosité à 6h, 9h, 12h, 15h, 18h (resp. `Ne6` à `Ne18`)
- la projection du vent sur l'axe est-ouest à 12h (`Vx`),
- le maximum d'ozone de la veille (`max03v`).

Le but est de modéliser la valeur des pics d'ozone en fonction de grandeurs physiques facilement mesurables (température, heure, nébulosité, vent) afin d'avoir des approximations de la qualité de l'air faciles et rapides à obtenir.

Rque : Ce jeu de données ne correspond pas à la même période que celui utilisé au TD2.

Partie A Explication du pic d'ozone par la température à midi

Dans cette première partie, nous souhaitons étudier les liens entre la valeur du pic d'ozone `max03` et la température à midi `T12`.

1. Importer les données avec la commande :
`donnees<-read.table(chemin,header = TRUE)`
où *chemin* est le chemin d'accès du fichier `ozone.dta`, par exemple `H:/TISD/ozone.dta`.

2. Analyser les variables `max03` et `T12` indépendamment (moyenne, écart-type, boxplot, histogramme avec densité superposée ...). Reconnaît-on l'allure de lois usuelles ?

3. Dessiner `max03` en fonction de "T12" avec la commande `plot`. Qu'en pensez-vous ?

4. Effectuer la régression de `max03` en fonction de `T12` avec la commande
`resmco<-lm(donnees$max03 ~ donnees$T12)`.

5. Extraire les coefficients de la régression à l'aide de la commande `coef`. Vérifier que l'on retrouve les mêmes valeurs avec les formules du cours.

6. Extraire de `resmco` la droite de régression en utilisant la commande `fitted` et la superposer au nuage de points obtenu à la question **3**. Recommencer en utilisant la commande `abline`, et recommencer en utilisant la série `T12` et les coefficients de la régression.

7. Demander une analyse de la régression avec la commande `summary.lm(resmco)`. Commenter.

8. Faire une analyse de la variance avec la commande `anova.lm(resmco)`.

9. Dessiner l'intervalle de confiance de la droite de régression.

10. Extraire les résidus de la régression avec la commande `residuals`. Vérifier que l'on obtient la même chose "à la main" en utilisant les séries `max03`, `T12` et les coefficients de la régression.

11. Tracer la densité estimée des résidus, leur évolution en fonction du temps, puis dessiner les résidus en fonction de `T12`. Enfin, calculer la moyenne des résidus et la covariance entre ces résidus et `T12`.

12. En supposant les résidus Gaussiens, tester la significativité du coefficient de régression correspondant à la variable `T12`.

Partie B Explication du pic d'ozone par une régression linéaire multiple

1. Dessiner `max03` en fonction des différentes variables. Quelles sont celles qui sont *a priori* intéressantes ?

2. Effectuer la régression de `max03` en fonction de toutes les variables et utiliser la commande `summary.lm` pour obtenir les détails de la régression.

3. Effectuer "à la main" une procédure "backward" pour sélectionner les variables : on estime le modèle, on retire la variable la moins significative et on recommence jusqu'à ce que toutes les variables soient significatives.

Exercice 48 (Simulations)
1. Simuler deux vecteurs de 100 variables $\mathcal{U}[0,1]$ indépendantes, x_1 et x_2. Définir $\beta_1 = 0,5$, $\beta_2 = -4$ et $\beta_3 = 3,8$.

2. Créer une fonction qui :
- simule un vecteur ε de longueur 100 suivant une loi normale de moyenne 0 et de variance $\sigma^2 = 0,2$.
- calculer ensuite le vecteur $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 + \varepsilon$.
- retourne les estimations MCO de β_1, β_2, β_3 et σ^2.

3. Appeler 1000 fois la fonction précédente et dessiner une approximation de la distribution des estimateurs β_1, β_2, β_3 et σ^2.

Exercice 49 (Test du χ^2)
Sur un échantillon de 20 sauteurs à ski, 20 slalomeurs et 20 descendeurs, on note que respectivement 0, 5 et 10 athlètes présentent un surpoids.

1. Etablir le tableau de contingence du croisement des variables "discipline" et "présence d'un surpoids".

2. Etablir les distributions marginales.

3. Etablir les distributions conditionnelles des disciplines sachant que le sportif est/n'est pas en surpoids. Commentaires ?

4. Faire un test d'indépendance du χ^2. Conclure.

Table des matières

1 TD 1 : Probabilités discrètes **1**

2 TD 2 : Intro à la théorie de l'intégrale de Lebesgue **3**

3 TD 3 : Variables aléatoires discrètes **5**

4 TD 4 : Variables aléatoires continues **8**

5 TD 5 : Vecteurs aléatoires, espérance conditionnelle **9**

6 TD 6 : Théorèmes limites **10**

7 TP 1 : Débuter avec R **11**

8 TD 7 : Maximum de vraisemblance **13**

9 TP 2 : Maximum de vraisemblance **14**

10 TP 3 : Régression linéaire **16**