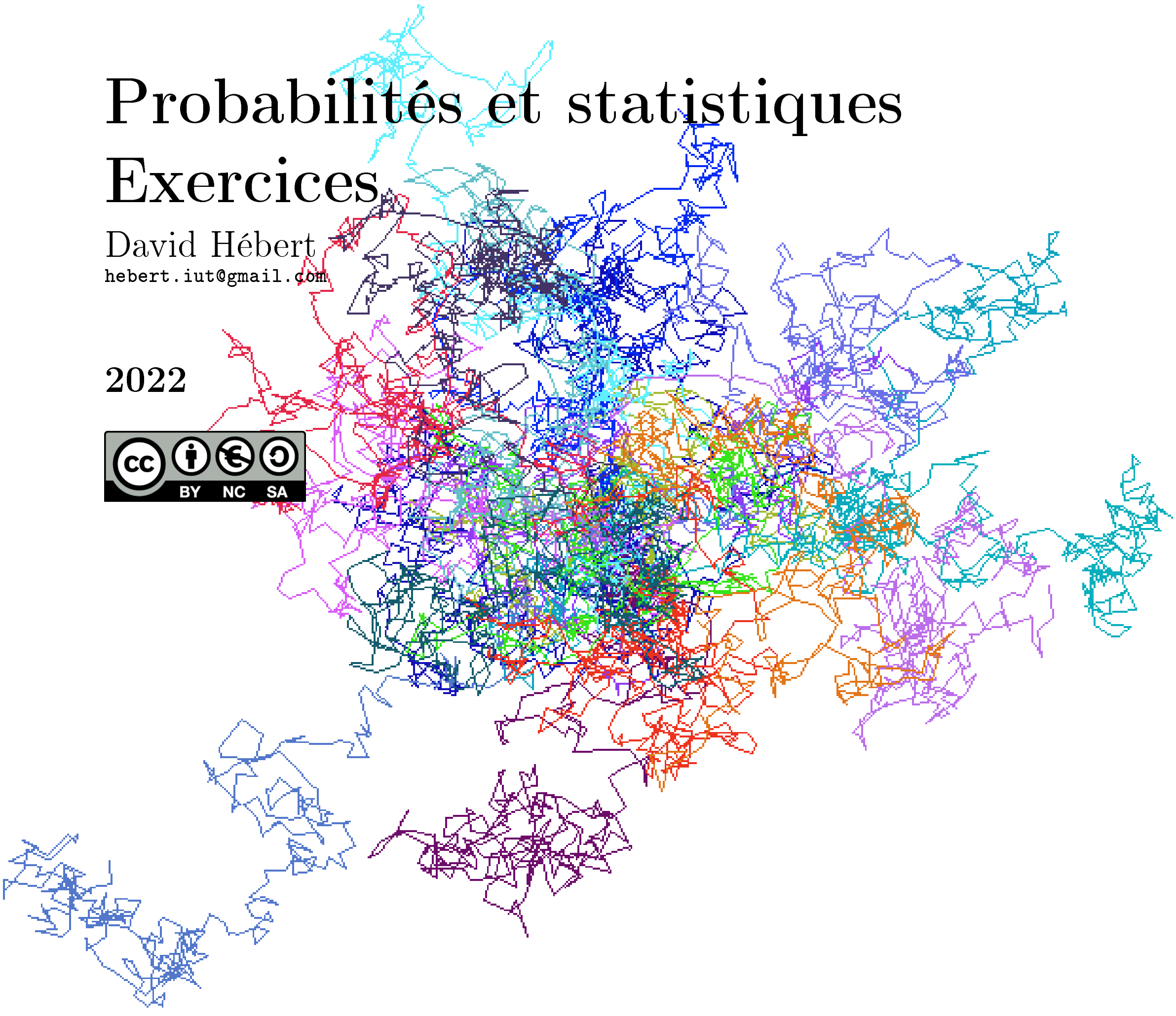
Probabilités et statistiques
Exercices

David Hébert
hebert.iut@gmail.com

2022



Combinatoire et premiers calculs de probabilité

Exercice 1

1. On dispose de 20 émeraudes ayant toutes une teinte de vert différente. Combien de bracelet différent de 9 émeraudes peut-on former ?
2. On dispose de 30 rubis ayant toutes une teinte de rouge différente. Combien de collier différent de 17 rubis peut-on former ?
3. On dispose de 53 saphirs ayant toutes une teinte de bleu différente. Combien de broche différente de 3 saphirs peut-on former ?
4. On dispose de 152 topazes ayant toutes une teinte de jaune différente. Combien de diadème différent de 98 topazes peut-on former ?

Exercice 2

1. On tire trois cartes au hasard dans un jeu de 52 cartes. Combien y a-t-il de tirage possible ?
2. Dans une urne contenant 23 cafards, on tire au hasard 7 cafards (les cafards sont tous identique et indiscernable au toucher... oui, au toucher). Combien y a-t-il de tirage possible ?
3. Dans une urne contenant 314 159 265 358 979 boules rouges indiscernables, on tire au hasard 32 384 626 boules rouges. Combien y a-t-il de tirage possible ?

Exercice 3

Lors d'un concours d'échecs, 10 joueurs ont fait une seule partie contre tous les autres (sauf eux-mêmes bien sûr). Combien de parties ont été jouées ?

Exercice 4

Certains mois ont 31 jours, d'autres en ont 30. Mais combien en contiennent 28 ?

Exercice 5

Une famille de 5 personnes va chez le photographe pour une photo où ils apparaitront tous. Pour être sûr de faire une bonne photo, le photographe veut prendre une photo en testant toutes les places possibles entre les membres de la famille.

Combien de photos ont-ils pris ?

Exercice 6

Dans chacune des expériences aléatoires suivantes décrire l'univers Ω.

1. Dans un urne on dispose de 3 boules rouges et de 3 boules bleues. On tire au hasard 2 boules (avec remise).
2. Dans un urne on dispose de 2 boules rouges et de 2 boules bleues. On tire au hasard 3 boules (avec remise).
3. Dans un urne on dispose de 2 boules rouges et de 2 boules bleues. On tire au hasard 3 boules (sans remise).

Exercice 7

Au loto, on tire 7 chiffres entre 1 et 49. Combien de combinaison sont alors possible ? Quelle est la probabilité de gagner (donner la valeur approchée) ?

Exercice 8

On lance 3 fois une pièce de monnaie équilibré.

1. Décrire l'univers Ω de cet expérience aléatoire.
2. Déterminer la probabilité des événements de Ω.

Exercice 9

On lance deux fois un dé truqué qui tombe sur un nombre paire deux fois plus souvent que sur un nombre impaire de manière équitable.

1. Décrire l'univers Ω de cet expérience aléatoire.
2. Déterminer la probabilité de chaque événement de Ω.

Exercice 10

Une urne contient trois boules blanches et deux boules noires indiscernables au toucher. On tire au hasard deux boules. On répondra à chacune des questions trois fois : une fois en supposant que le tirage est sans remise, une seconde fois en supposant que le tirage est avec remise et une troisième fois en supposant le tirage simultané.

1. Décrire l'univers Ω de cet expérience aléatoire.
2. Déterminer la probabilité de chaque événement de Ω.
3. Déterminer la probabilité de l'événement : "On tire au moins une boule blanche".

Exercice 11

Même exercice que précédemment sauf que l'on tire trois boules au lieu de deux.

Exercice 12

Une urne contient 10 jetons numérotés de 1 à 10 indiscernable au toucher. On tire au hasard 3 jetons. On répondra à chacune des questions trois fois : une fois en supposant que le tirage est sans remise, une seconde fois en supposant que le tirage est avec remise et une troisième fois en supposant le tirage simultané.

1. Quel est la probabilité qu'il y ai exactement un numéro pair ?
2. Quel est la probabilité qu'il y ai au moins un numéro pair ?
3. Quel est la probabilité qu'il y ai au plus un numéro pair ?

Exercice 13

Une urne contient des boules indiscernables : 5 jaunes, 4 noires et 7 rouges. On tire au hasard 6 boules. On répondra à chacune des questions trois fois : une fois en supposant que le tirage est sans remise, une seconde fois en supposant que le tirage est avec remise et une troisième fois en supposant le tirage simultané.

1. Donner l'univers décrivant cet expérience ?
2. Calculer les probabilités des événements suivants :
   — A : "On a obtenu au moins deux boules noires".
   — B : "On a obtenu exactement deux boules rouges et au moins une boule jaune".
3. Les événements A et B sont-ils indépendants ?

Exercice 14

Dans un lycée, on sait que 55% des élèves sont des filles. On sait également que 35% des filles et 30% des garçons mangent à la cantine. On choisit un élève au hasard. Quel est la probabilité que cet élève ne déjeune pas à la cantine ?

Exercice 15

Un appareil ménager peut présenter après sa fabrication deux défauts. On appelle A l'événement "l'appareil présente un défaut d'apparence" et F l'événement "l'appareil présente un défaut de fonctionnement".
On suppose que les événements A et F sont indépendants.
On sait que la probabilité que l'appareil présente un défaut d'apparence est égale à $0,02$ et que la probabilité que l'appareil présente au moins l'un des deux défauts est égale à $0,069$. On choisit au hasard un des appareils. Quelle est la probabilité que l'appareil présente le défaut F ?

Exercice 16

On dispose de deux urnes U_1 et U_2 et d'un dés cubique bien équilibré dont les faces sont numérotés de 1 à 6. L'urne U_1 contient trois boules rouges et une boule noire tandis que l'urne U_2 contient trois boules rouges et deux boules noires.

Une partie se déroule de la manière suivante : le joueur lance le dé; si le résultat est 1 alors il tire au hasard une boule dans l'urne 1 sinon dans l'urne 2.

1. Représenté la situation à l'aide d'un arbre pondéré.
2. Montrer que la probabilité de tirer une boule noire est de $\frac{3}{8}$.
3. Sachant que l'on a tiré une boule noire, quelle est la probabilité d'avoir obtenu 1 au lancé de dé.

Exercice 17

On lance deux dés. Un de couleur verte, l'autre de couleur rouge.

1. Donner l'univers Ω décrivant cette expérience.
2. Considérons les événements suivants :
   — A : "Le nombre obtenu avec le dé vert est pair".
   — B : "Le nombre obtenu avec le dé rouge est impair".
   — C : "Les deux nombres obtenus sont de même parité".
   Calculer la probabilité des événements suivants : A, B, C, $A \cap B$, $A \cap C$, $B \cap C$ et $A \cap B \cap C$.
3. Les événements A et B sont-ils indépendants ? De même pour B et C ? De même pour C et A ?
4. La famille (A, B, C) est-elle indépendante ?

Exercice 18

Soit n un entier supérieur ou égal à 2. On considère une famille avec n enfants et les événements :
— A : "Il y a dans cette famille au moins un garçon et au moins une fille".
— B : "Il y a dans cette famille au plus un garçon".
Discuter, suivant les valeurs de n, de la dépendance des événements A et B.

Exercice 19

Une urne contient des boules indiscernables : 4 boules rouges, 5 boules vertes et 7 boules bleues. On note respectivement R, V et B l'ensemble des boules rouges, vertes et bleues et on note $U = R \cup V \cup B$.

Partie A. On tire trois boules de l'urne sans remise et sans tenir compte de l'ordre de sortie.

1. Déterminer l'univers Ω de cette expérience ainsi que son cardinal.
2. Écrire l'événement "Toutes les boules sont de la même couleur" comme partie de Ω et calculer sa probabilité.
3. Même question avec l'événement "On obtient exactement deux boules rouges".

Partie B. On tire une boule, on note sa couleur et on la remet dans l'urne et ce 3 fois de suite.

1. Déterminer l'univers Ω de cette expérience ainsi que son cardinal.
2. Soit $\omega \in \Omega$. Est-ce que le singleton $\{\omega\}$ est un événement ?
3. Si $(E, F, G) \in \{R, V, B\}^3$, vérifier que $E \times F \times G$ est un événement.
4. Vérifier que l'assertion "Toutes les boules obtenues sont de la même couleur" définit un événement. Calculer sa probabilité.
5. Même question avec l'événement "On obtient exactement deux boules rouges".

Exercice 20

[Extrait du Bac S 2012 - Polynésie]
Une urne contient 100 petits cubes de bois dont 60 sont bleus et les autres rouges. Parmi les cubes bleus, 40% ont leur faces marqués d'une cercle, 20% marqués d'un losange et les autres marqués d'une étoiles. Parmi les cubes rouges, 20% ont leur face marqué d'un cercle, $x\%$ marqué d'un losange pour un certain $x \in [0; 80]$ et les autres sont marqués d'une étoiles.

Partie A. On tire au hasard un cube de l'urne.

1. Démontrer que la probabilité que soit tiré un cube marqué d'un losange est égale à $0.12 + 0.004x$.
2. Déterminer x pour que la probabilité de tirer un cube marqué d'un losange soit égale à la probabilité de

tirer un cube marqué d'une étoile.

3. Déterminer x pour que les événements "tirer un cube bleu" et "tirer un cube marqué d'un losange" soient indépendant.
4. Supposons que $x = 50$. Calculer la probabilité que soit tiré un cube bleu sachant qu'il est marqué d'un losange.

Partie B. On tire au hasard et simultanément 3 cubes de l'urne (on arrondira les résultats au millième).

1. Quelle est la probabilité de tirer au moins un cube rouge ?
2. Quelle est la probabilité que les cubes tirés soient de la même couleur ?
3. Quelle est la probabilité de tirer exactement un cube marqué d'un cercle ?

Variables aléatoires discrètes

Exercice 21

Une urne contient 5 boules indiscernables au toucher : 3 rouges et 2 vertes.

On tire au hasard et simultanément deux boules.

1. Calculer les probabilités des évènements $R =$ "Les deux boules sont rouges" et $V =$ "Les deux boules sont vertes".
2. On appelle X la variable aléatoire qui, à chaque tirage associe le nombre de boules vertes obtenues.
   (a) Déterminer la loi de X.
   (b) Calculer $E(X)$.

Exercice 22

Une urne contient 4 boules blanches et une boule noire indiscernable au toucher. Vous tirez deux boules de l'urne avec remise.

— Vous perdez 10€ si vous obtenez deux boules blanches.
— Vous perdez 5€ si vous obtenez exactement une boule noire.
— Vous gagnez 50€ si vous obtenez deux boules noires.

On note X la variable aléatoire du gain algébrique d'une partie.

1. Déterminer la loi de X.
2. Calculer l'espérance et l'écart-type de X.
3. Mêmes questions si l'urne contient deux boules blanches et trois boules noires.

Exercice 23

Dans une salle sans fenêtre à l'université 6 étudiants dont un tueur compulsif, 3 idiots et 2 futurs Nobel sont enfermés. Le tueur compulsif n'est pas suicidaire.

— Un idiot éteint la lumière.
— Dans le noir, le tueur compulsif tue au hasard.
— Les cris d'agonies du tué font qu'un futur Nobel rallume la lumière.
— Un idiot éteint la lumière à nouveau.
— Le tueur tue à nouveau au hasard.
— Si un futur Nobel est en vie, il allume la lumière ouvre la porte et sauve tout le monde, sinon, dans le noir, le tueur tue tout le monde.

On note X le nombre de futur Nobel tués au cours de cette "expérience aléatoire".

1. Représenter la situation sur un arbre pondéré.
2. Quelle est la probabilité que tout le monde (sauf le tueur) soit tué ? sauvé ?
3. Déterminer la loi de X.
4. Quel est l'espérance de X ?

Exercice 24

Une variable aléatoire Y suit une loi binomiale de paramètres 20 et $\frac{1}{5}$. Calculer $P(Y \geqslant 2)$.

Exercice 25

Sur des petits bout de papier, on a noté le nom des 25 étudiants d'une classe. Chacun des étudiants tire au hasard et avec remise un papier du chapeau et tarte l'étudiant dont le nom est inscrit. On note X le nombre d'étudiant qui se tarte eux-même.

1. Quelle est la loi de X ?
2. En moyenne, combien d'étudiant se tarte eux-même ?
3. Quelle est la probabilité qu'au moins un étudiant se tarte lui-même ?

Exercice 26

(Pondichéry, avril 2009)

On dispose de deux dés cubiques dont les faces sont numérotées de 1 à 6. Ces dés sont en apparence identique mais l'un est bien équilibré et l'autre truqué. Avec le dé truqué, la probabilité d'obtenir 6 lors d'un lancé est égale à $\frac{1}{3}$.

Les résultats seront donnés sous forme de fractions irréductibles.

1. On lance le dé bien équilibré trois fois de suite et on désigne par X la variable aléatoire donnant le nombre de 6 obtenus.
   (a) Quelle loi de probabilité suit la variable aléatoire X.
   (b) Quelle est son espérance ?
   (c) Calculer $P(X = 2)$.
2. On choisi au hasard l'un des deux dés, les choix étant équiprobables. On lance le dé choisi trois fois de suite.
   On considère les événements :
   — D : "Le dé choisi est est le dé bien équilibré" ;
   — A : "Obtenir exactement deux 6".
   (a) Calculer la probabilité des événements
   — E_1 : "Choisir le dé bien équilibré et obtenir exactement deux 6",
   — E_2 : "Choisir le dé truqué et obtenir exactement deux 6"
   (on pourra construire un arbre pondéré).
   (b) En déduire que $P(A) = \frac{7}{48}$.
   (c) Ayant choisi au hasard l'un des deux dés et l'ayant lancé trois fois de suite, on a obtenu exactement deux 6. Quelle est la probabilité d'avoir choisi le dé truqué ?
   (d) On choisit au hasard et de manière équiprobable l'un des deux dés et on le lance $n > 1$ fois de suite. On note B_n l'évènement "Obtenir au moins un 6 parmi ces n lancers". Déterminer, en fonction de n, $P(B_n)$.

Exercice 27

Pour embaucher ses cadres, une entreprise fait appel à un cabinet de recrutement. La procédure tenue est la suivante : le cabinet effectue une première sélection sur dossier dont 40% sont retenues. Les candidats sélectionnés passent ensuite un entretient à la suite duquel 70% sont retenues. Ces derniers sont alors convoqués à un ultime entretient avec le DRH de l'entreprise qui ne gardera que 25% des candidats.

Partie A. On choisit au hasard le dossier d'un candidat.

1. Représenter la situation à l'aide d'un graphe pondéré.
2. Calculer la probabilité que le candidat soit retenu à l'issue du premier entretient.
3. Montrer que la probabilité que le candidat ne soit pas recruté est de 93%.

Partie B. On dispose de 5 dossiers de candidatures étudiés indépendants les uns des autres. On note X la variable aléatoire donnant le nombre de personne recrutées parmi ces 5 candidats.

1. Quel est la loi de X ?
2. Quel est la probabilité que exactement 2 candidats soient recrutés ; on arrondira le résultat au millième.

Partie C. Quel est le nombre minimum de dossiers que le cabinet doit traiter pour que la probabilité d'embaucher au moins un candidat soit supérieur à 99,9% ?

Exercice 28

Soit k un entier supérieur ou égal à 2. Une urne contient k boules noires et 3 boules blanches indiscernables au toucher. Une partie consiste à prélever au hasard successivement et avec remise deux boules dans cette urne. On établis la règle de jeu suivante :

— un joueur perd 9€ si les deux boules tirées sont de couleurs blanches ;
— un joueur perd 1€ si les deux boules tirées sont de couleurs noires ;
— un joueur gagne 5€ si les deux boules tirées sont de couleurs différentes. On dit dans ce cas qu'il gagne la partie.

Partie A. On suppose ici que $k = 7$.

1. Déterminer la probabilité de gagner une partie.
2. Soit n un entier supérieur ou égal à 3. Un joueur joue n parties identiques et indépendantes. On note X la variable aléatoire qui comptabilise le nombre de partie gagnées par le joueur et p_n la probabilité que le joueur gagne au moins une fois au cours des n parties.
   (a) Quel est la loi de X ?
   (b) Exprimer p_n en fonction de n et calculer p_{10} (arrondi au millième).
   (c) Déterminer le nombre minimale de parties que le joueur doit jouer afin que la probabilité de gagner au moins une fois soit supérieure à 99%.

Partie B. On note Y_k la variable aléatoire égale au gain algébrique du joueur.

1. Justifier que $P(Y_k = 5) = \dfrac{6k}{(k+3)^2}$.
2. Donner la loi de probabilité de Y_k.
3. Déterminer les valeurs de k pour lesquelles ce jeu est favorable au joueur (c'est à dire que lorsque son gain moyen est positif).

Exercice 29

Partie A. Un groupe de 50 coureurs, portant des dossards numérotés de 1 à 50, participent à une course cycliste qui comprend 10 étapes et au cours de laquelle aucun abandon n'est constaté. A la fin de chaque étape, un groupe de 5 coureurs est choisi au hasard pour subir un contrôle antidopage. Ces désignations étant indépendantes.

1. A l'issue de chaque étape, combien peut-on former de groupes différents de 5 coureurs ?
2. A l'issu d'une étape, on choisi au hasard un coureur au hasard parmi les 50 cyclistes. Montrer que la probabilité que ce participant ai subit un contrôle antidopage est 0.1.
3. On note X la variable aléatoire qui comptabilise le nombre de contrôles subis par un coureur sur l'ensemble des 10 étapes.
   (a) Quel est la loi de X ?
   (b) Calculer les probabilités suivantes en arrondissant au dix-millième.
      i. Le gagnant de la course a été contrôlé exactement 5 fois.
      ii. Le second de la course n'a pas été contrôlé
      iii. Le troisième de la course à été contrôlé au moins une fois.

Partie B. On note T l'événement "Sur un test antidopage, le contrôle est positif" dont les statistiques donne $P(T) = 5\%$. Un contrôle antidopage n'est pas fiable à 100% :
— si un coureur est dopé, le test est positif dans 97% des cas,
— si un coureur n'est pas dopé, le test est positif dans 1% des cas.

1. Calculer la probabilité qu'un coureur pris au hasard soit dopé.
2. Sachant qu'un coureur est contrôlé positif, quel est la probabilité qu'il ne soit pas dopé ?

Exercice 30

Dans une association sportive, un quart des femmes et un tiers des hommes adhèrent à la section tennis. On sait également que 30% des membres de cette association adhèrent à la section tennis.

Partie A. On choisit au hasard un membre de cette association.

1. Montrer que la probabilité que le membre choisi soit une femme est de $\dfrac{2}{5}$.
2. On choisi un membre parmi les adhérents de la section tennis. Quel est la probabilité qu'il s'agisse d'une

femme ?

Partie B. Pour financer une sortie, les membres de l'association organisent une loterie.

1. Chaque semaine, un membre de l'association est choisi au hasard et de manière indépendante, pour tenir la loterie.
   (a) Déterminer la probabilité pour qu'en 4 semaines consécutives, il y ai exactement deux fois un membre qui adhère à la section tennis parmi les membre choisi pour tenir la loterie.
   (b) Pour tout $n \in \mathbb{N}_{>0}$, on note p_n la probabilité pour qu'en n semaines consécutives, il y ai au moins un membre qui adhère à la section tennis parmi les membres choisi pour tenir la loterie. Calculer p_n.
   (c) Déterminer le plus petit entier n tel que $p_n \geqslant 99\%$.
2. Pour la loterie, on utilise une urne contenant 100 jetons ; 10 jetons exactement sont gagnants et rapporte 20€ chacun, les autres ne rapportent rien. Pour jouer à cette loterie, un joueur doit payer 5€ puis tire au hasard et de façon simultanée deux jetons dans l'urne. On note X la variable aléatoire associant le gain algébrique réalisé par un joueur lors d'une partie de cette loterie.
   (a) Déterminer la loi de X.
   (b) Calculer et interpréter le gain moyen de à ce jeu.

Probabilités continues

Exercice 31

Révision de calcul intégrale : calculer les intégrales suivantes.

1. $\displaystyle\int_{-1}^{1} \frac{1}{2}x^3 + 6x \ dx$
2. $\displaystyle\int_{0}^{-3} (1-x)^{\frac{1}{3}} \ dx$
3. $\displaystyle\int_{3}^{+\infty} \frac{1}{x^5} \ dx$
4. $\displaystyle\int_{-1}^{1} xe^x \ dx$
5. $\displaystyle\int_{-\infty}^{0} \mathbb{1}_{[-1;1]}(x) + \mathbb{1}_{\{x\geqslant -1\}}(x) dx$
6. $\displaystyle\int_{3}^{5} x\mathbb{1}_{\{x<4\}}(x) \ dx$
7. $\displaystyle\int_{\mathbb{R}} e^x \mathbb{1}_{\{x\leqslant 0\}}(x) + e^{-x}\mathbb{1}_{\{x>0\}}(x) \ dx$
8. $\displaystyle\int_{\mathbb{R}} \lambda e^{-\lambda x} \mathbb{1}_{\{x\geqslant 0\}}(x) \ dx$
9. $\displaystyle\int_{\mathbb{R}} x^2 \mathbb{1}_{[0;1]}(x) + x^3 \mathbb{1}_{[-1;1]}(x) \ dx$
10. $\displaystyle\int_{\mathbb{R}} \left(\sum_{k=0}^{n} x^{2k+1} \mathbb{1}_{[-k-1;k+1]}(x) \right) \ dx$

Exercice 32

On considère la fonction

$$\begin{array}{rcl} p : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & k(1-x)^{\frac{1}{3}} \mathbb{1}_{[0,1]}(x) \end{array}$$

1. Déterminer k pour que p soit une densité de probabilité.
2. Calculer sa fonction de répartition et la tracer dans un repère orthonormé.
3. Soit X une variable aléatoire ayant une densité de probabilité p.
   (a) Calculer $P(0.5 < X \leqslant 1)$.
   (b) Calculer l'espérance de X.
   (c) Calculer sa variance.

Exercice 33

Même exercice que précédemment avec la fonction

$$\begin{array}{rcl} p : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & k\left(xe^x \mathbb{1}_{[-1;0]}(x) + x^2 e^{-2x} \mathbb{1}_{[0;1]}(x)\right) \end{array}$$

Exercice 34

On considère la fonction

$$\begin{array}{rcl} p : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & e^{-|x|}\mathbb{1}_{[-\ln(2);\ln(2)]}(x) \end{array}$$

1. Vérifier que p est une densité de probabilité.
2. On pose $Y = |X|$.
   (a) Déterminer la fonction de répartition de Y.
   (b) En déduire la fonction de densité de probabilité de Y (on admettra qu'elle existe).

Exercice 35

Soit X une variable aléatoire ayant la fonction de répartition suivante

$$\begin{array}{rcl} F : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & (1 - e^{-\frac{x^2}{2}})\mathbb{1}_{\{x>0\}}(x) \end{array}$$

1. Vérifier que F est une fonction de répartition.
2. Donner la fonction de densité de F.
3. Calculer l'espérance de X.

Exercice 36

Soit $\lambda > 0$. On dira qu'une variable aléatoire X suit la loi exponentielle si sa fonction de densité de probabilité est $x \mapsto \lambda e^{-\lambda x}\mathbb{1}_{\{x>0\}}(x)$.

1. Vérifier qu'il s'agit bien d'une fonction de densité de probabilité.
2. Déterminer en fonction de λ la fonction de répartition F_X.
3. Déterminer en fonction de λ l'espérance et la variance de X.
4. Soit $Y = X^2$.
   (a) Calculer l'espérance de Y.
   (b) Déterminer sa fonction de répartition F_Y.
   (c) En déduire la fonction de densité de probabilité de Y (on admettra qu'elle existe).

Exercice 37

Une enquête a été effectuée auprès de familles de 4 personnes afin de connaître leur achat de lait en 1 mois. Sur l'ensemble des personnes interrogées, la consommation a une distribution de type normale avec une moyenne de 20 litres et un écart-type de 6 litres. En vue d'une campagne publicitaire, on souhaite connaître le pourcentage des faibles consommateurs (moins de 10 litres par mois) et celui des grands consommateurs (plus de 30 litres par mois).

Exercice 38

Une étude effectuée auprès de jeunes enfants montre que les premiers mots apparaissent, en moyenne, à $11,5$ mois avec un écart-type de $3,2$ mois. La distribution des âges étant normale, évaluer la proportion d'enfants ayant acquis leurs premiers mots

1. Avant 10 mois.
2. Après 18 mois.
3. Entre 8 et 12 mois.

Exercice 39

On mesure la taille en centimètres de 2500 hommes ; la distribution obtenue suit une loi normale de moyenne égale à $169\,\mathrm{cm}$ et d'écart-type égal a $5,6\,\mathrm{cm}$.

1. Quel est le pourcentage d'hommes dont la taille est inférieure à $155\,\mathrm{cm}$?
2. Quel est le pourcentage d'hommes dont la taille est comprise entre $155\,\mathrm{cm}$ et $175\,\mathrm{cm}$?
3. Quel est l'intervalle, centré sur la valeur moyenne de la taille, qui contient 60% de la population en question ?

Exercice 40

On suppose que les étudiants d'un cours de probabilité ont des notes normalement distribuées autour d'une moyenne de $9,7$, avec un écart-type de 4.

1. Trouver la probabilité pour qu'un étudiant ait une note supérieure à 13.
2. Trouver la probabilité pour qu'un échantillon aléatoire de 10 étudiants ait une note moyenne supérieure à 13.

Exercice 41

Calculer l'espérance mathématique et la variance d'une variable aléatoire normale X sachant que $P(X \geqslant -2) \simeq 0,2$ et $P(X < -5) \simeq 0,7$.

Exercice 42

Calculer l'espérance mathématique et la variance d'une variable aléatoire normale X sachant que $P(X \leqslant 2) = 0,5793$ et $P(X > 5) = 0,2119$.

Exercice 43

La durée de vie d'un certain type d'appareil est modélisée par une variable aléatoire suivant une loi normale de moyenne et d'écart-type inconnus. Les spécifications impliquent que 80% de la production des appareils ait une durée de vie entre 120 et 200 jours et que 5% de la production ait une durée de vie inférieure à 120 jours.

1. Déterminer la moyenne et la variance.
2. Quelle est la probabilité d'avoir un appareil avec une durée de vie comprise entre 200 et 230 jours ?

Statistiques unidimensionnelle

Exercice 44

On considère une population de 300 étudiants d'un IUT d'informatique. Les caractères suivants sont-ils qualitatifs ordinal, qualitatif nominal ou quantitatif.

Caractère	Quantitatif	Qualitatif ordinal	Qualitatif nominal
Notes à un examen			
Mention à un semestre			
Couleur des yeux			
Numéro de téléphone			
Temps passer sur le web (en heures)			
Sexe			
Taille			
Département de naissance			
Age			
Groupe sanguin			
Niveau dans un jeu			
Poids			
Nombre d'heure de TV par semaine			
Prénoms			

Exercice 45

Une étude sur le chiffre d'affaire de plusieurs PME a permis d'obtenir les résultats suivants (exprimé en milliers d'euro) :

Minimum	3500
Moyenne	4900
Écart-type	650
Mode	4550

Écart interquartile	1100
Médiane	4600
Premier quartile	4100
Étendue	5000

1. Classer ces paramètres en deux catégories : caractéristique de position ou de dispersion.
2. Quel est le chiffre d'affaires le plus grand parmi ces PME ?
3. Déterminer le troisième quartile.

Exercice 46

Durant l'été 2013 certains opérateurs de téléphonie mobile ont publié leur résultat du premier semestre ; on s'intéresse en particulier au nombre d'abonné. Une partie de ces informations est résumée dans le tableau suivant, exprimé en million d'abonnés :

	Orange	SFR	Bouygue	Free
Juin 2012	26.475		11.251	3.6
Juin 2013			11.286	6.795

1. SFR a déclaré que fin juin 2013 son parc d'abonné mobile totalisait 17.372 million de client en croissance de 5.8% par rapport à fin juin 2012. Compléter alors la colonne correspondant dans le tableau.
2. Calculer, pour Bouygue et Free, le pourcentage de croissance entre juin 2012 et juin 2013.
3. En ne considérant que Bouygue et SFR, quelle est la croissance moyenne ? Même question pour Bouygue, SFR et Free.
4. Supposons qu'orange ai un pourcentage de croissance moyen entre bouygue et SFR ; quel sera alors le nombre de client au premier semestre 2013.

Exercice 47

On s'intéresse à étudier la duré (en années) entre la date de mariage d'un couple et la date de naissance de leur premier enfant. Pour cela on sonde 500 familles à un enfant né après mariage. Une partie des données est synthétisée dans le tableau suivant :

Année	Nombre de famille	Effectif cumulé croissant	Effectif cumulé décroissant	Fréquence	Fréquence cumulé croissant	Fréquence cumulé décroissant	
1	91	91	500	18.2%	18.2%	100.0%	
2	72	163	409	14.4%	32.6%	81.8%	
3	60	223	337	12.0%	44.6%	67.4%	
4	52	275	277	10.4%	55.0%	55.7%	
5	45	320	225	9.0%	64.0%	45.0%	
6	40	360	180	8.0%	72.0%	36.0%	
7	37	397	140	7.4%	79.4%	28.0%	
8	32						
9	26						
10	25						
11	20						

1. Compléter le tableau (hormis la dernière colonne).
2. Calculer la médiane et les premier et dernier quartile.
3. Calculer la moyenne.
4. Déterminer variance et écart-type.
5. Représenter les données sous la forme d'un diagramme circulaire.

Exercice 48

Calculer la médiane, la moyenne et l'écart-type des notes obtenus à un devoir par une classe de 32 élèves.

Notes	5	6	7	8	9	10	11	12	13	14	15
Effectifs	1	3	1	4	3	9	5	1	2	1	2

Représenter ces données par un diagramme en bâtons.

Exercice 49

On a effectué une étude statistique sur 4 nombres entiers dont les résultats sont donnés dans le tableau suivant. Retrouver les 4 entiers.

Minimum	2
Moyenne	5
Variance	5
Étendue	8

Statistique unidimensionnelle a caractère continue

Exercice 50

On étudie les revenus mensuel en euros d'un ensemble de familles d'un quartier de Gennevilliers.

Revenus	[700; 900[	[900; 1100[	[1100; 1300[	[1300; 1400[	[1400; 1500[	[1500; 1600[
Effectifs	13	219	20	46	50	82

1. Quel est le nombre de famille dont les revenus sont compris entre 700€ et 900€ ?
2. Quel est la proportion de famille dont les revenus sont compris entre 900€ et 1500€ ?
3. Quel est la moyenne des revenus ?
4. Quel est l'écart-type des revenus ?
5. Que mesure la moyenne et la variance ?
6. Quel est la classe médiane ? Déterminer la médiane par interpolation linéaire.
7. Faire l'histogramme de cette statistique et placer sur cet histogramme la médiane en abscisse. Quelle(s) observation(s) pouvez-vous faire ?

Exercice 51

La direction générale de l'agriculture et de la foret nous donne la répartition par tranches d'âges des chefs d'exploitation agricole d'une région.

Tranches d'âges	**Nombre d'exploitation**
Moins de 25 ans	580
De 25 à 29 ans	2162
De 30 à 39 ans	8063
De 40 à 49 ans	9569
De 50 à 59 ans	10660
Plus de 60 ans	15913

1. Quel est la population étudiée ? Quel est le caractère ?
2. Complétez un tableau statistique de cette série avec fréquence et fréquence cumulé croissante et décroissante (on prendra 20 ans pour âge minimal et 70 pour âge maximal).
3. Quel est la proportion de chefs d'exploitations qui ont au moins 40 ans ?
4. Quel est la proportion de chefs d'exploitations qui ont entre 25 et 60 ans ?
5. Représenter l'histogramme de cette série.
6. Par le calcul, déterminez la classe médiane puis la médiane par interpolation linéaire.

Exercice 52

Le tableau suivant donne la répartition des entreprises du secteur de l'automobile en fonction de leur chiffre d'affaire (exprimé en millions d'euros).

Chiffre d'affaire	Moins de 0.25	[0.25; 0.5[	[0.5; 1[	[1; 2.5[	[2.5; 5[	[5; 10[
Effectifs	137	106	112	154	100	33

1. Tracer l'histogramme de cette série statistique.
2. Déterminer le chiffre d'affaire médian.
3. Calculer la moyenne et l'écart-type.

Exercice 53

Ce tableau donne la distribution selon l'âge de la population de l'île de la Réunion et de la métropole.

Age	La réunion	Métropole
$[0;15[$	29%	20%
$[15;30[$	30%	22%
$[30;50[$	25%	28%
$[50;70[$	12%	20%
$[70;90[$	4%	10%

Tracer les histogrammes correspondants à ces deux séries. Donner la valeur médiane et celle de la moyenne correspondant à ces deux séries.

Exercice 54

La distribution des demandeur d'emplois selon le sexe et la classe d'âge dans une localité est la suivante :

Age	Hommes	Femmes
$[16;26[$	280	160
$[26;40[$	310	360
$[40;50[$	240	120
$[50;60[$	420	530
$[60;65[$	70	50

1. Tracer les deux courbes de fréquences cumulées croissantes.
2. Déterminer les quartiles de la variable X associant à chaque demandeur d'emplois féminin son âge.
3. Même question avec la variable Y associant à chaque demandeur d'emplois masculin son âge.

Exercice 55

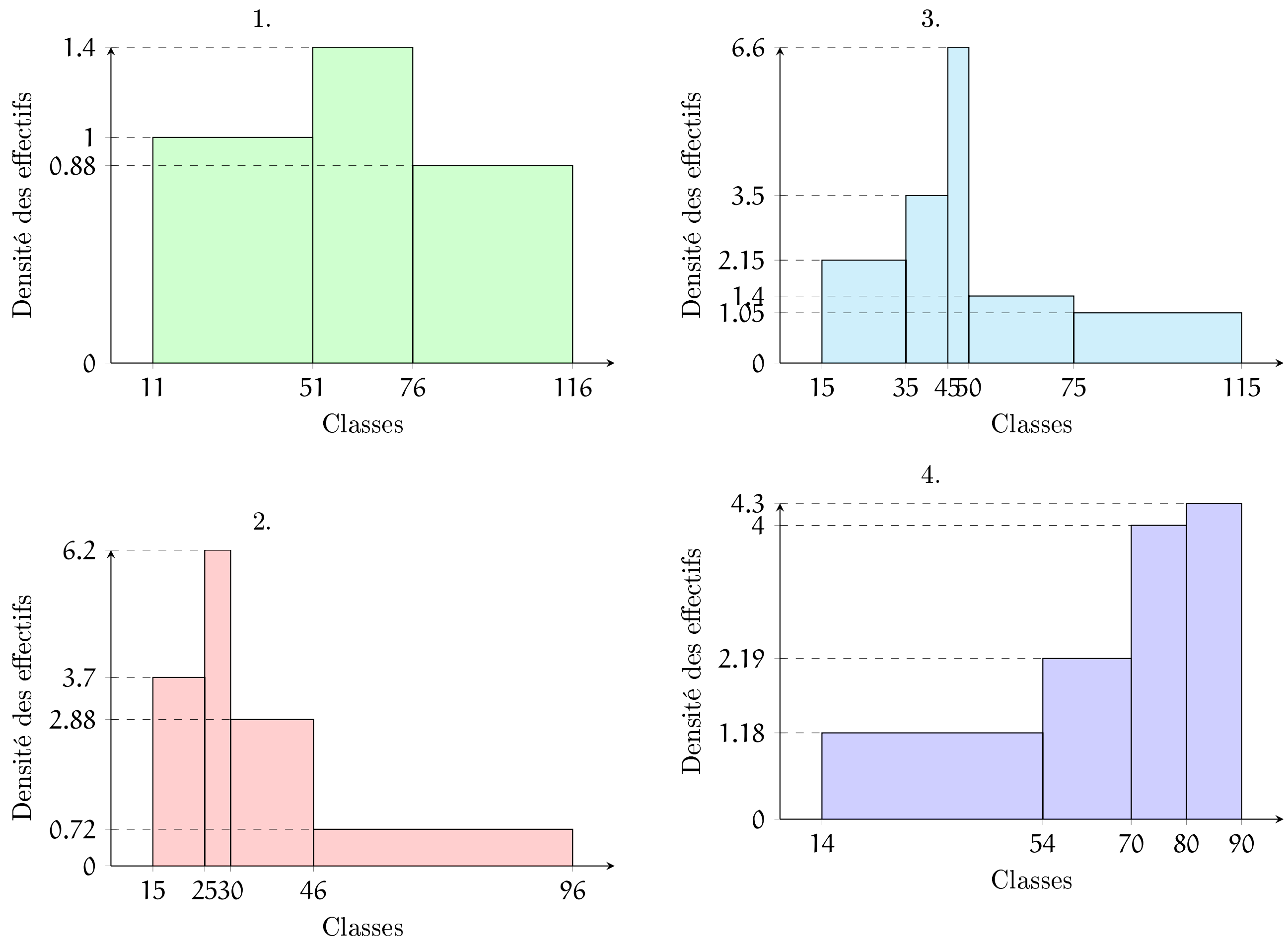
Le tableau ci-dessous représente la distribution des charges maximales (en tonnes) supportés par les câbles fabriqués par une certaine usine.

Charge maximale	**Nombre de câbles**
$[9.3;9.7[$	2
$[9.7;10.2[$	5
$[10.2;10.7[$	12
$[10.7;11.2[$	17
$[11.2;11.7[$	14
$[11.7;12.2[$	6
$[12.2;13.2[$	4

1. Calculer moyenne et écart-type.
2. Déterminer la médiane par interpolation linéaire.
3. Calculer tous les déciles (par interpolation linéaire).

Exercice 56

Extraire des diagrammes suivant donner les informations suivantes : effectif total, liste des classes modales, effectifs de chacune des classe modales, classe médiane, classe du premier et du troisième quartile, médiane, écart-type (ou variance), étendue :



Estimateurs

Exercice 57

Les données se trouvant à la fin de l'exercice sont des tirages aléatoires de valeurs suivant une loi uniforme $\mathcal{U}_{[0;\vartheta]}$ pour un certain paramètre réel $\vartheta > 0$ que l'on souhaite estimer.

Les questions marquées d'un astérisque sont facultatives et on peut admettre ce qu'elles demandent.

Estimation ponctuelle. Soit $n \in \mathbb{N}_{>0}$ et $(X_1, \ldots, X_n)$ un échantillon de $\mathcal{U}_{[0;\vartheta]}$

0. * Montrer que pour tout réel $\alpha \neq -1$, $\mathbb{E}\left(X_1^\alpha\right) = \dfrac{\vartheta^\alpha}{\alpha+1}$.

1. On considère $T_n^{(1)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} 2X_i$.

   (a) Montrer que $T_n^{(1)}$ est un estimateur convergent et sans biais de ϑ.

   (b) Calculer $\mathbb{V}\left(T_n^{(1)}\right)$ la variance de $T_n^{(1)}$.

   (c) En déduire $\mathsf{EQ}(T_n^{(1)}, \vartheta)$ l'erreur quadratique de l'estimateur.

2. On considère $T_n^{(2)} = e\displaystyle\prod_{i=1}^{n} X_i^{\frac{1}{n}}$

   (a) Montrons que $T_n^{(2)}$ est un estimateur convergent de ϑ.

      i. * Montrer que $\mathbb{E}\left(\ln(X_1)\right) = \ln(\vartheta) - 1$.

      ii. En déduire que $L_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \ln(X_i)$ est un estimateur convergent de $\ln(\vartheta) - 1$.

      iii. Conclure.

   (b) Déterminons $\mathsf{B}(T_n^{(2)}, \vartheta)$ le biais de l'estimateur.

i. Montrer que $\mathbb{E}\left(T_n^{(2)}\right) = e\left(\frac{n}{n+1}\right)^n \vartheta$. En déduire $B(T_n^{(2)}, \vartheta)$ le biais de $T_n^{(2)}$.

ii. * Montrer $\left(\frac{n}{n+\alpha}\right)^n \sim e^{-\alpha+\frac{\alpha^2}{2n}}$

iii. En déduire que $T_n^{(2)}$ est asymptotiquement sans biais.

iv. * Montrer que $B(T_n^{(2)}, \vartheta) \sim \frac{\vartheta}{2n}$.

(c) Calculer $\mathbb{V}\left(T_n^{(2)}\right)$ la variance de $T_n^{(2)}$.

(d) * En déduire que $EQ(T_n^{(2)}, \vartheta)$ l'erreur quadratique de l'estimateur, est équivalent en l'infini à $\frac{\vartheta^2}{n}$.

3. **Application :** Avec le premier échantillon de données simulées, estimer la valeur de ϑ via $T_n^{(1)}$ et $T_n^{(2)}$ et donner l'erreur quadratique de ces estimations. Faire la même chose avec le second échantillon. Lequel de ces deux estimateurs est le meilleur ?

4. On considère $T_n^{(3)} = max(X_i | i \in [\![1, n]\!])$.

(a) * Montrer que pour tout $n \in \mathbb{N}_{>0}$ et tout $k \in \mathbb{N}$, $\int_{[0;\vartheta]^n} max(x_1, \ldots, x_n)^k \, dx_1 \cdots dx_n = \frac{n}{n+k}\vartheta^{n+k}$.

(b) * Montrer que la fonction de répartition de $T_n^{(3)}$ est

$$\forall t \in \mathbb{R},\ F_{T_n^{(3)}}(t) = \frac{t^n}{\vartheta^n}\mathbb{1}_{[0,\vartheta]}(t) + \mathbb{1}_{]\vartheta;+\infty[}(t)$$

(c) Montrer que $T_n^{(3)}$ est un estimateur convergent de ϑ asymptotiquement sans biais.

(d) Déterminer $EQ(T_n^{(3)}, \vartheta)$ et montrer que qu'il est équivalent en l'infini à $\frac{2\vartheta}{n^2}$.

5. On considère $T_n^{(4)} = \frac{n+1}{n} max(X_i | i \in [\![1, n]\!])$

(a) Montrer que T_n^4 est un estimateur convergent et sans biais de ϑ.

(b) Déterminer $\mathbb{V}\left(T_n^{(4)}\right)$ et $EQ(T_n^4, \vartheta)$.

6. **Application :** Avec le premier échantillon de données simulées, estimer la valeur de ϑ via $T_n^{(3)}$ et $T_n^{(4)}$ et donner l'erreur quadratique de ces estimations. Faire la même chose avec le second échantillon. Classer les 4 estimateurs de ϑ du moins bon au meilleur.

Intervalle de confiance. On note $T_n = T_n^{(4)}$ le meilleur estimateur étudié dans la première partie de l'exercice.

1. Montrer que $Q_{T_n}(t)$ la fonction quantile de T_n est définie pour tout $t \in [0, 1]$ par

$$Q_{T_n}(t) = \vartheta t^{\frac{1}{n}}$$

2. On admettra que l'intervalle de dispersion unilatéral supérieur de niveau $1 - \alpha$ est optimal.

(a) **(Application).** Déterminer en fonction de ϑ les intervalles de dispersions optimaux de niveau 95%, de niveau 99% et de niveau 99,9% pour les deux jeux de données de l'exercice.

(b) Montrer que $\left[T_n; T_n\alpha^{\frac{1}{n}}\right]$ est un intervalle de confiance de niveau $1 - \alpha$.

(c) **(Application).** Déterminer les intervalles de confiance de niveau 95%, 99% et 99,9% des deux jeux de données de l'exercice.

Seul au monde. Voici des données correspondant au tirage aléatoire de valeur suivant une certaine loi de probabilité inconnue.

→ Données :

−1.237	−1.444	+0.804	+0.843	+0.189	−2.705	−2.224	−1.904	−1.299	−1.426
−2.307	+0.526	+0.009	+0.537	+0.051	−2.679	−0.043	−1.356	−1.398	+0.038
+0.459	+0.449	−0.319	−2.020	+0.698	+0.464	−0.542	−1.867	−2.327	−0.469
−1.799	−0.443	−2.682	−1.736	−1.040	+0.092	−2.439	+0.315	+0.139	−0.419
−0.004	−1.862	−1.886	+0.024	−1.350	−1.276	−2.271	−1.889	−0.689	−0.964

→ Taille de l'échantillon : 50

$\rightarrow\ \overline{x} = \dfrac{1}{50}\displaystyle\sum_{i=1}^{50} x_i = -0.89356$

$\rightarrow\ \overline{v} = \dfrac{1}{50}\displaystyle\sum_{i=1}^{50} (x_i - \overline{x})^2 = 1.1569419264$

Premier modèle : loi normale. On suppose que les valeurs tirées suivent une loi normale $\mathcal{N}(\mu, \sigma)$.

Déterminer, en vous servant du cours, les intervalles de confiances symétrique de niveau 95%, 99% et 99,9% de μ et σ.

Second modèle : loi uniforme. On suppose que les valeurs suivent une loi uniforme $\mathcal{U}_{[\vartheta;1]}$ pour un certain paramètre réel $\vartheta < 1$.

Donner au moins deux estimateurs convergent de ϑ dont l'un sera sans biais et l'autre asymptotiquement sans biais. Vous comparerez la qualité de ces estimateurs et déterminerez pour le meilleur les intervalles de confiances de niveau 95%, 99% et 99,9%.

Echantillon 1 .

$\rightarrow$ Données :

0.178	0.573	1.725	2.930	1.197	1.251	3.000	0.262	2.608	0.926	0.968	1.587	2.383	1.851	0.417
3.051	1.656	0.639	0.995	2.764	1.975	2.812	1.602	2.118	1.132	3.016	2.659	0.549	0.446	0.636

$\rightarrow$ Taille de l'échantillon : 30

$\rightarrow$ $\frac{1}{30}\sum_{i=1}^{30} 2x_i = 3.193733333333333$

$\rightarrow$ $e\prod_{i=1}^{30} x_i^{\frac{1}{30}} = 3.4197348436952573$

$\rightarrow$ $max(x_i | i \in [\![1, 30]\!]) = 3.051$

Echantillon 2 .

$\rightarrow$ Données :

0.596	2.089	0.117	0.783	2.814	3.085	0.556	1.938	2.079	2.071	2.552	1.640	0.656	1.096	0.267
1.165	0.680	0.775	0.091	0.887	0.210	0.832	2.444	2.325	1.471	0.605	0.009	2.388	0.912	2.942
2.851	0.170	1.025	2.403	0.755	2.402	0.330	3.026	2.254	1.845	0.147	0.326	2.040	3.016	2.982
1.967	0.813	0.603	2.425	0.782	1.961	0.993	0.802	1.504	1.906	0.275	2.167	3.065	0.131	0.135
1.261	0.830	2.127	2.763	2.753	0.178	0.218	0.293	2.254	1.770	1.445	1.540	1.554	2.865	0.119
0.848	1.914	0.964	0.958	0.241	0.659	1.882	2.681	2.645	2.334	1.399	0.024	2.236	0.912	1.870
2.541	1.623	2.403	0.912	1.849	1.608	0.144	1.181	1.667	0.658	1.159	1.277	1.824	0.462	0.172
2.098	2.435	3.003	0.049	0.362	2.565	1.052	0.153	2.668	3.098	1.819	2.515	1.385	2.475	2.495
2.425	1.413	0.917	2.630	1.050	2.262	0.168	1.093	0.587	0.461	2.425	3.072	2.575	2.595	1.036
0.478	2.979	0.876	0.344	1.383	0.141	3.023	1.399	1.407	0.894	2.579	2.414	2.870	0.014	2.827
1.258	2.776	3.047	0.504	0.650	2.485	1.677	1.059	2.263	1.698	1.273	0.625	1.954	2.019	2.187
0.154	2.541	2.821	0.385	0.322	2.764	1.628	2.893	0.598	2.587	2.162	2.553	2.205	2.489	0.573
0.849	1.842	2.993	0.371	0.143	0.378	3.031	1.975	1.815	1.874	0.068	2.845	0.731	1.928	0.962
1.907	1.962	0.535	0.478	3.079	1.024	0.185	0.389	1.093	0.133	2.646	0.555	1.752	2.907	0.649
2.049	1.324	2.462	1.946	2.670	0.368	2.801	1.928	2.087	0.811	0.832	1.871	0.254	0.387	0.623
0.072	2.303	0.762	1.046	1.696	2.638	1.901	0.106	2.429	2.439	0.011	1.470	0.081	1.049	1.212
1.643	0.670	2.779	1.957	2.687	3.122	1.699	1.272	2.468	0.208	1.452	0.708	0.477	1.920	2.821
0.892	2.615	2.268	0.163	2.436	0.790	0.073	0.808	1.483	0.097	1.834	1.836	1.259	2.246	0.641
2.303	2.846	2.589	2.856	0.011	2.240	0.887	1.572	2.254	0.561	2.467	0.881	1.167	3.130	0.243
0.323	1.010	1.584	2.904	2.996	2.497	2.010	0.679	2.281	2.289	2.189	0.548	1.396	1.846	2.798

$\rightarrow$ Taille de l'échantillon : 300

$\rightarrow$ $\frac{1}{300}\sum_{i=1}^{300} 2x_i = 3.0550666666666664$

$\rightarrow$ $e\prod_{i=1}^{300} x_i^{\frac{1}{300}} = 2.8652098608233163$

$\rightarrow$ $max(x_i | i \in [\![1, 300]\!]) = 3.13$

Statistiques bidimensionnelle

Exercice 58

On considère un modèle linéaire $Y_i = ax_i + b + \varepsilon_i$. On sait que la droite de régression linéaire $y = \hat{a}x + \hat{b}$ passe par le point de coordonnée $(2; 2.5)$. On sait de plus que $\overline{y} = 5$ et $\overline{x} = 3$. Déterminer $\hat{a}$ et $\hat{b}$.

Exercice 59

Dans le tableau ci dessous on a représenter

— Les notes en probabilités en X.
— Les notes en statistiques en Y.
Pour un caractère statistique Z on note $\overline{Z}$ la moyenne.
Dans la dernière ligne, on a réalisé la somme des colonnes.

#	X	$X-\overline{X}$	$(X-\overline{X})^2$	Y	$Y-\overline{Y}$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
1	11.1	−0.04	0.002	15	1.12	1.254	−0.045
2	7	−4.14	17.14	13.2	−0.68	0.462	2.815
3	11.5	0.36	0.13	16.1	2.22	4.928	0.799
4	15.2	4.06	16.484	17.5	3.62	13.104	14.697
5	3.7	−7.44	55.354	2.5	−11.38	129.504	84.667
6	19.5	8.36	69.89	19	5.12	26.214	42.803
7	18.9	7.76	60.218	19.7	5.82	33.872	45.163
8	11.3	0.16	0.026	17.4	3.52	12.39	0.563
9	7.4	−3.74	13.988	1.7	−12.18	148.352	45.553
10	8.3	−2.84	8.066	15.3	1.42	2.016	−4.033
11	5.7	−5.44	29.594	7	−6.88	47.334	37.427
12	1.2	−9.94	98.804	5.2	−8.68	75.342	86.279
13	6.3	−4.84	23.426	9.1	−4.78	22.848	23.135
14	17.4	6.26	39.188	17	3.12	9.734	19.531
15	12.3	1.16	1.346	13.1	−0.78	0.608	−0.905
16	15.7	4.56	20.794	19.9	6.02	36.24	27.451
17	2.7	−8.44	71.234	10.2	−3.68	13.542	31.059
18	19.4	8.26	68.228	19.6	5.72	32.718	47.247
19	14.4	3.26	10.628	17.7	3.82	14.592	12.453
20	2.3	−8.84	78.146	8.3	−5.58	31.136	49.327
21	13.7	2.56	6.554	17.1	3.22	10.368	8.243
22	16.1	4.96	24.602	17.5	3.62	13.104	17.955
23	7.2	−3.94	15.524	12	−1.88	3.534	7.407
24	17.5	6.36	40.45	19.9	6.02	36.24	38.287
25	12.7	1.56	2.434	16	2.12	4.494	3.307
Σ	278.5	0	772.25	347	0	723.93	641.185

1. Réaliser le nuage de points de ces données.
2. Déterminer la moyenne des X.
3. Déterminer l'écart-type des Y.
4. Déterminer la droite de régression linéaire.
5. La régression linéaire est-elle justifiée ?

Exercice 60

Dans le tableau ci dessous on a représenter

- L'angle, en degrés, suivant l'axe des abscisses d'une roue de voiture en X.
- L'angle, en degrés, suivant l'axe des ordonnée d'une roue de voiture en Y.

Pour un caractère statistique Z on note $\overline{Z}$ la moyenne.
Dans la dernière ligne, on a réalisé la somme des colonnes.

#	X	$X-\overline{X}$	$(X-\overline{X})^2$	Y	$Y-\overline{Y}$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
1	46	41.2	1697.44	−33	−17.76	315.42	−731.71
2	43	38.2	1459.24	−17	−1.76	3.1	−67.23
3	−38	−42.8	1831.84	5	20.24	409.66	−866.27
4	40	35.2	1239.04	−13	2.24	5.02	78.85
5	−6	−10.8	116.64	−0	15.24	232.26	−164.59
6	−45	−49.8	2480.04	−14	1.24	1.54	−61.75
7	39	34.2	1169.64	−40	−24.76	613.06	−846.79
8	−12	−16.8	282.24	−1	14.24	202.78	−239.23
9	−22	−26.8	718.24	2	17.24	297.22	−462.03
10	−10	−14.8	219.04	−19	−3.76	14.14	55.65
11	36	31.2	973.44	−8	7.24	52.42	225.89
12	49	44.2	1953.64	−24	−8.76	76.74	−387.19
13	−14	−18.8	353.44	−25	−9.76	95.26	183.49
14	−6	−10.8	116.64	−18	−2.76	7.62	29.81
15	49	44.2	1953.64	−13	2.24	5.02	99.01
16	−26	−30.8	948.64	−12	3.24	10.5	−99.79
17	−36	−40.8	1664.64	−42	−26.76	716.1	1091.81
18	−26	−30.8	948.64	−20	−4.76	22.66	146.61
19	25	20.2	408.04	−7	8.24	67.9	166.45
20	−12	−16.8	282.24	9	24.24	587.58	−407.23
21	10	5.2	27.04	−20	−4.76	22.66	−24.75
22	13	8.2	67.24	4	19.24	370.18	157.77
23	30	25.2	635.04	−35	−19.76	390.46	−497.95
24	29	24.2	585.64	−21	−5.76	33.18	−139.39
25	−36	−40.8	1664.64	−19	−3.76	14.14	153.41
Σ	120	0	23796	−381	0	4566.62	−2607.15

1. Réaliser le nuage de points de ces données.
2. Déterminer la moyenne des X.
3. Déterminer l'écart-type des Y.
4. Déterminer la droite de régression linéaire.
5. La régression linéaire est-elle justifiée ?

Exercice 61

Le tableau suivant représente l'évolution du revenu disponible brut et la consommation des ménages exprimés en euros.

Année	Revenu	Consomation
1992	8000	7389.99
1993	9000	8169.65
1994	9500	8831.71
1995	9500	8652.84
1996	9800	8788.08
1997	11000	9616.21
1998	12000	10593.45
1999	13000	11186.11
2000	15000	12758.09
2001	16000	13869.62

On cherche à expliquer la consommation des ménages par le revenu. On modélise les données, comme en cours par

$$Y_i = a x_i + b + \varepsilon_i$$

où Y_i représente la consommation des ménage l'année $1991 + i$, x_i le revenu et ε_i les termes d'erreurs.

En économétrie a est appelé la **propension marginale à consommer** et b est appelé la **consommation autonome**.

1. Tracer le nuage de point de ces donnés. Commenter.
2. Estimer (ponctuellement) la consommation autonome et la propension marginale à consommer.
3. En déduire les valeurs estimé $\hat{y}_i$.
4. Déterminer les résidus $\hat{\varepsilon}_i$ et vérifier qu'ils sont de moyenne nulle.
5. Estimer la variance de l'erreur. En déduire une estimation ponctuel de l'écart-type de l'erreur.
6. Construire l'intervalle de confiance de niveau 95% de la propension marginale à consommer.
7. En 2002 et 2003 on prévoit respectivement 16800 et 17000 pour la valeur du revenu. Déterminer les valeurs prévues de la consommation pour ces deux années.

Exercice 62

Dans le tableau ci dessous on a représenté des données statistiques X et Y.

#	X	Y
1	9	3.4
2	14	4.3
3	39	4.3
4	82	3.9
5	30	2.8
6	1	1.6
7	18	4.4
8	15	3.2
9	14	3.9
10	26	5.2

1. Réaliser le nuage de points de ces données.
2. Estimer la modélisation linéaire $Y = aX + b$. Ce modèle vous semble-t-il viable ?
3. La modélisation est en fait de la forme $Y = a \ln X + b$. Estimer les paramètre a et b, les résidus et une estimation de l'erreur. Cette régression non linéaire est-elle justifiée ?
4. Si une onzième donnée $X = 50$ quelle est une bonne prévision de Y ? Donner l'intervalle de confiance de niveau 99% de cette valeur.

Exercice 63

Dans le tableau ci dessous on a représenté des données statistiques X et Y.

#	X	Y
1	2.3	−39.52
2	3.49	−38.31
3	5	10
4	−4.88	−16.07
5	1	−14.42
6	2	−41.35
7	1.33	−11.87
8	5	15.36
9	1.89	−57.67
10	−1.48	−31.92

1. Réaliser le nuage de points de ces données.
2. Estimer la modélisation linéaire $Y = aX + b$. Ce modèle vous semble-t-il viable ?
3. La modélisation est en fait de la forme $Y = a\exp X + b$. Estimer les paramètre a et b, les résidus et une estimation de l'erreur. Cette régression non linéaire est-elle justifiée ?
4. Si une onzième donnée $X = 0$ quelle est une bonne prévision de Y ? Donner l'intervalle de confiance de niveau 99% de cette valeur.

Statistiques multidimensionnelle

Exercice 64

On considère une modélisation linéaire simple $Y_i = z_i a + b + \varepsilon_i$ où $\varepsilon_i \sim \mathcal{N}(0, \sigma)$ pour $n > 2$ observations des caractères z et y.

On choisit d'utiliser les outils de modélisation linéaire multiple et de modéliser le problème à l'aide du langage matricielle $Y = xm + \varepsilon$.

1. Donner les dimensions des matrices Y, x, m et ε.
2. Montrer que :
   (a) ${}^t xx = \begin{pmatrix} n\overline{z^2} & n\overline{z} \\ n\overline{z} & n \end{pmatrix}$
   (b) ${}^t xy = \begin{pmatrix} n\overline{zy} \\ n\overline{y} \end{pmatrix}$
3. On peut montrer que l'inverse de la matrice $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ est la matrice $\dfrac{1}{\det(A)} \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}$.
   (a) Montrer que $\det({}^t xx) = n^2\sigma_z^2$.
   (b) Montrer que $({}^t xx)^{-1} = \dfrac{1}{n\sigma_z^2} \begin{pmatrix} 1 & -\overline{z} \\ -\overline{z} & \overline{z^2} \end{pmatrix}$.
   (c) En déduire la valeur de $\hat{m}$.
   (d) Comparez ces résultats avec les estimations de la modélisation linéaire simple.
4. (a) Donnez une estimation de la matrice de variance-covariance de l'estimateur de m.
   (b) En déduire les valeurs de $\hat{\sigma_a}^2$ et $\hat{\sigma_b}^2$. Comparez vos résultats avec les variances de la modélisation linéaire simple.
   (c) Donnez une estimation de $\mathrm{Cov}(A_n, B_n)$ où A_n est l'estimateur de a et B_n l'estimateur de b.

Exercice 65

Le prof, ce petit malin, a demandé à python deux nombres de manière équiprobable, appelé x et y, parmi -1, 0 et 1 puis a appliquer une petite formules $z = a + bx + cy + \varepsilon$ où a, b et c sont des constantes et ε est la partie entière d'un tirage aléatoire d'une loi normale centré de variance σ. Dans ce modèle il y a donc 4 inconnues : a, b, c et σ. Le prof, ce petit malin, a réalisé cette expérience aléatoire cinq fois. Voici ce qu'il a obtenue.

x	y	z
0	0	2
0	-1	1
0	0	0
0	0	4
1	1	1

Le but du problème est d'estimer les 4 inconnues. On utilise le langage matricielle et on modélise le problème sous la forme $Z = Am + \varepsilon$.

1. Donner la matrice A ainsi que ${}^tAA$.
2. La matrice $B = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 10 & -5 \\ 1 & -5 & 4 \end{pmatrix}$ n'est pas la matrice inverse de ${}^tAA$. Cependant on peut démontrer qu'il existe un réel λ tel que $({}^tAA)^{-1} = \lambda B$.
   (a) Déterminer la valeur de λ.
   (b) En déduire $\hat{m}$ l'estimation ponctuelle de m.
3. Quelle opération matricielle permet d'obtenir les valeurs de $\hat{z}_i$? Réalisez cette opération.
4. Donner une estimation de σ.
5. On note $\vec{X}$ le vecteur aléatoire en dimension 3 de coordonnée A_n, B_n et C_n les estimateurs des paramètres a, b et c.
   (a) Donnez $\mathbb{V}(\vec{X})$ la matrice de variance-covariance de $\vec{X}$.
   (b) En déduire $\mathbb{V}(B_n)$ et $\mathrm{Cov}(A_n, C_n)$.
6. Question de mémoire :
   (a) Déterminer le R^2 de ce modèle.
   (b) Donnez l'intervalle de confiance symétrique de niveau 95% de B_n ainsi que l'estimation de cet intervalle.