Probabilités et statistique

DUT GMP, IUT de Rennes

Jean-Marie Lion
UFR Mathématiques
Université de Rennes 1

22 juin 2015
résumé avec exercices
et sujets corrigés

La statistique est une discipline qui a pour objet l'analyse de données ainsi que la prévision. Depuis les travaux de Bayes au $XVIII^e$ siècle son développement s'appuie sur la théorie des probabilités. Celle-ci est la branche des mathématiques qui essaie d'expliquer le hasard. L'un des fondateur de cette théorie est Blaise Pascal, savant et philosophe du $XVII^e$ siècle.

Programme

Contenu
Probabilités et Statistiques

Objectifs
Maîtriser les bases des probabilités et des statistiques

Compétences visées
Étudier une variable aléatoire suivant une loi normale
Estimer une moyenne, une variance, une fréquence
Tester l'égalité de moyennes, de fréquences

1 Premiers éléments de statistique descriptive

On remarquera que la statistique emprunte le vocabulaire à la démographie. La raison est historique. La statistique est née de la démographie.

En statistique on distingue deux grands types de données, les **données quantitatives** qui représentent des grandeurs mesurables (des relevés de poids, de température, de prix) et les **données qualitatives** (des listes de couleurs de pulls vendus l'hiver 2006, de catégories professionnelles des salariés bretons, de bulletins de vote dans une urne le soir d'une élection, la liste des menus journaliers du RU, la liste des défauts de fabrication à corriger d'une voiture sortant d'une chaîne de montage). Ces suites de données s'appellent des **séries statistiques**.

Dans les deux situations (quantitatif ou qualitatif), on associe à une **population** d'**individus** une **variable** ou **caractère**. Les individus de la population étudiée seront numérotés par un entier. Ainsi x_i représentera la valeur du caractère de l'individu portant le numéro i. Si la population est composée de n individus alors $x = (x_1, ..., x_n)$ est la série statistique associée au caractère étudié. L'**effectif** est le nombre d'individus qui forment la population.

1.1 Données quantitatives

Définition 1.1 Soit $x = (x_1, ..., x_n)$ une suite de données quantitatives. L'**étendue** est la différence des valeurs extrêmes. Pour $x \in \mathbf{R}$ l'**effectif cumulé de** x est le cardinal $k(x) = \mathrm{Card}(\{i, x_i \leq x\})$ et la **fréquence cumulée** $F(x)$ est $F(x) = \frac{k(x)}{n}$. La **médiane** x_{med} divise les données en deux parties égales, celles qui lui sont plus petites et celles qui lui sont plus grandes. On a $F(x_{med}) = 1/2$. On définit aussi les trois **quartiles** $q_1, q_2 = x_{med}$ et q_3 par $F(q_i) = \frac{i}{4}$. La **moyenne** $\overline{x}$ **de** x est

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + ... + x_n).$$

On appelle **variance empirique** le nombre positif $v = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ qui est égal à $\overline{x^2} - \overline{x}^2$ où $\overline{x^2}$ désigne la moyenne de la suite des carrés : $x^2 = (x_1^2, ..., x_n^2)$ et $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n}(x_1^2 + ... + x_n^2)$.
L'**écart-type** est $\sigma = \sqrt{v}$. Le **coefficient de variation** est le rapport $\frac{\sigma}{\overline{x}}$.

Remarque 1.1 Le médiane est bien moins sensible que la moyenne aux variations des valeurs extrêmes.

Remarque 1.2 La variance, l'écart-type et le coefficient de variation mesurent comme la population se disperse par rapport à la moyenne.

Remarque 1.3 Si, au lieu de donner la valeur x_i du caractère de chaque individu i on donne pour chaque valeur c_j prise par le caractère le nombre n_j d'invidus alors on définit la **fréquence** du caractère c_j par $f_j = \frac{n_j}{n}$ et la **fréquence cumulée** F_j du caractère c_j comme la somme des fréquences des caractères inférieurs ou égaux à c_j. La moyenne $\overline{x}$ se calcule par la formule

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{d} n_j c_j = \frac{1}{n}(n_1 c_1 + ... + n_d c_d) = f_1 c_1 + ... + f_d c_d$$

où d est le nombre de valeurs différentes prises par le caractère. La variance empirique est donnée par les formules

$$v = \frac{1}{n}\sum_{j=1}^{d} n_j (c_j - \overline{x})^2 = \sum_{j=1}^{d} f_j (c_j - \overline{x})^2 = \left(\frac{1}{n}\sum_{j=1}^{d} n_j c_j^2\right) - \overline{x}^2 = \left(\sum_{j=1}^{d} f_j c_j^2\right) - \overline{x}^2.$$

En pratique on utilise les formules avec les c_i.

Remarque 1.4 On divise souvent l'étendue en intervalles appelés **classes**. On peut s'intéresser à l'effectif n_j, à l'effectif cumulé N_j, à la fréquence f_j et à la fréquence cumulée F_j d'une classe $[a_j, a_{j+1}[$ (ou $]a_j, a_{j+1}]$.) L'effectif cumulé est égal à $n_1 + ... + n_j$. La fréquence f_j est égale à $\frac{n_j}{n}$, la fréquence cumulée F_j est la somme $f_1 + ... + f_j$. On calcule la moyenne et la variance en utilisant les formules précédentes avec pour n_j l'effectif de la classe et $c_j = \frac{a_j + a_{j+1}}{2}$. Pour calculer avec précision la médiane on repère d'abord la classe médiane. C'est la classe dont la fréquence cummulée $F_{j_{med}}$ est la première à passer $\frac{1}{2}$. On calcule alors x_{med} par interpolation linéaire en posant

$$x_{med} = a_{j_{med}} + \left(\frac{a_{j_{med}+1} - a_{j_{med}}}{F_{j_{med}} - F_{j_{med}-1}}\right)\left(\frac{1}{2} - F_{j_{med}-1}\right).$$

On fait de même pour les quartiles en remplaçant $\frac{1}{2}$ par $\frac{1}{4}$ ou $\frac{3}{4}$.

Exemple 1.1 Considérons une population formée de vingt réservoirs d'essence numérotés de 1 à 20. Le caractère étudié est le volume d'essence (au litre près) contenue dans chacun des réservoirs. La série statistique est

$$\begin{array}{ccccc} x_1 = 60, & x_2 = 59, & x_3 = 57, & x_4 = 61, & x_5 = 57, \\ x_6 = 56, & x_7 = 58, & x_8 = 59, & x_9 = 60, & x_{10} = 56, \\ x_{11} = 56, & x_{12} = 59, & x_{13} = 60, & x_{14} = 56, & x_{15} = 56, \\ x_{16} = 59, & x_{17} = 56, & x_{18} = 58, & x_{19} = 53, & x_{20} = 62. \end{array}$$

Les différentes valeurs c_j prises par le caractère volume d'essence contenue dans chaque réservoir sont

$$\begin{array}{llll} c_1 = 53\ (n_1 = 1), & c_2 = 56\ (n_2 = 6), & c_3 = 57\ (n_3 = 2), & c_4 = 58\ (n_4 = 2), \\ c_5 = 59\ (n_5 = 4), & c_6 = 60\ (n_6 = 3), & c_7 = 61\ (n_7 = 1), & c_8 = 62\ (n_8 = 1). \end{array}$$

L'effectif est

$$n = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7 + n_8$$

c'est à dire

$$n = 20 = 1 + 6 + 2 + 2 + 4 + 3 + 1 + 1.$$

La moyenne du caractère peut être calculée en divisant la somme des x_i par l'effectif $n = 20$. Ça donne

$$\begin{aligned} \bar{x} &= \tfrac{1}{20}(60 + 59 + 57 + 61 + 57 + 56 + 58 + 59 + 60 + 56 + \\ &\quad 56 + 59 + 60 + 56 + 56 + 59 + 56 + 58 + 53 + 62) = 57,9. \end{aligned}$$

Elle peut être calculée à partir des c_j et des n_j. On obtient

$$\bar{x} = \frac{1}{20}(1 \times 53 + 6 \times 56 + 2 \times 57 + 2 \times 58 + 4 \times 59 + 3 \times 60 + 1 \times 61 + 1 \times 62) = 57,9.$$

Le calcul de la variance se fait aussi de plusieurs façons. On a

$$\begin{aligned} v = \tfrac{1}{20} & ((60 - 57,9)^2 + (59 - 57,9)^2 + (57 - 57,9)^2 + (61 - 57,9)^2 + (57 - 57,9)^2 + \\ & (56 - 57,9)^2 + (58 - 57,9)^2 + (59 - 57,9)^2 + (60 - 57,9)^2 + (56 - 57,9)^2 + \\ & (56 - 57,9)^2 + (59 - 57,9)^2 + (60 - 57,9)^2 + (56 - 57,9)^2 + (56 - 57,9)^2 + \\ & (59 - 57,9)^2 + (56 - 57,9)^2 + (58 - 57,9)^2 + (53 - 57,9)^2 + (62 - 57,9)^2) = 4,59. \end{aligned}$$

Mais on a aussi

$$\begin{aligned} v = \tfrac{1}{20} & (60^2 + 59^2 + 57^2 + 61^2 + 57^2 + 56^2 + 58^2 + 59^2 + 60^2 + 56^2 + \\ & 56^2 + 59^2 + 60^2 + 56^2 + 56^2 + 59^2 + 56^2 + 58^2 + 53^2 + 62^2) - 57,9^2 = 4,59. \end{aligned}$$

En utilisant les formules qui font intervenir les c_j et les n_j on obtient

$$\begin{aligned} v = \tfrac{1}{20} & (1 \times (53 - 57,9)^2 + 6 \times (56 - 57,9)^2 + 2 \times (57 - 57,9)^2 + 2 \times (58 - 57,9)^2 \\ & 4 \times (59 - 57,9)^2 + 3 \times (60 - 57,9)^2 + 1 \times (61 - 57,9)^2 + 1 \times (62 - 57,9)^2) = 4,59 \end{aligned}$$

ou encore

$$\begin{aligned} v = \tfrac{1}{20} & (1 \times 53^2 + 6 \times 56^2 + 2 \times 57^2 + 2 \times 58^2 + \\ & 4 \times 59^2 + 3 \times 60^2 + 1 \times 61^2 + 1 \times 62^2) - 57,9^2 = 4,59. \end{aligned}$$

Les fréquences f_j sont les quotients $\frac{n_j}{n.}$ On a donc

$$\begin{array}{llll} f_1 = 0,05\ , & f_2 = 0,3\ , & f_3 = 0,1\ , & f_4 = 0,1\ , \\ f_5 = 0,2\ , & f_6 = 0,15\ , & f_7 = 0,05\ , & f_8 = 0,05. \end{array}$$

La moyenne et la variance de la série statistique précédente se calculent en utilisant les formules qui reposent sur les f_j. On obtient

$$\bar{x} = 0,05\times 53 + 0,3\times 56 + 0,1\times 57 + 0,1\times 58 + 0,2\times 59 + 0,15\times 60 + 0,05\times 61 + 0,05\times 62 = 57,9$$

et

$$\begin{aligned} v = \quad & (0,05\times(53-57,9)^2 + 0,3\times(56-57,9)^2 + 0,1\times(57-57,9)^2 + 0,1\times(58-57,9)^2 \\ & 0,2\times(59-57,9)^2 + 0,15\times(60-57,9)^2 + 0,05\times(61-57,9)^2 + 0,05\times(62-57,9)^2 = \quad 4,59 \end{aligned}$$

ou encore

$$\begin{aligned} v = \quad & 0,05\times 53^2 + 0,3\times 56^2 + 0,1\times 57^2 + 0,1\times 58^2 + \\ & 0,2\times 59^2 + 0,15\times 60^2 + 0,05\times 61^2 + 0,05\times 62^2 - 57,9^2 = \quad 4,59. \end{aligned}$$

Les effectifs cumulés valent

$$\begin{array}{cccc} N_1 = 1 \ , & N_2 = 7 \ , & N_3 = 9 \ , & N_4 = 11 \ , \\ N_5 = 15 \ , & N_6 = 18 \ , & N_7 = 19 \ , & N_8 = 20 \end{array}$$

et les fréquences cumulées

$$\begin{array}{cccc} F_1 = 0,05 \ , & F_2 = 0,35 \ , & F_3 = 0,45 \ , & F_4 = 0,55 \ , \\ F_5 = 0,75 \ , & F_6 = 0,9 \ , & F_7 = 0,95 \ , & F_8 = 1. \end{array}$$

On en déduit que la classe du 1er quartile est la 2ème classe (celle de 56), la classe médiane est la 4ème classe (celle de 58) et la classe du 3ème quartile est la 5ème classe (celle de 59).

Définition 1.2 Soit $(x,y) = ((x_1,y_1),\dots,(x_n,y_n))$ une suite de données quantitatives couplées. La **covariance** est

$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = v(xy) - v(x)v(y).$$

Le **coefficient de corrélation** est le rapport $r = \dfrac{cov(x,y)}{\sigma(x)\sigma(y)} \in [-1,1]$.

Proposition 1.1 *Parmi les droites affines $Y = aX+b$, celle qui minimise le nombre*

$$\sum_{i=1}^{n}(y_i - ax_i - b)^2$$

est la droite appelée **droite de régression linéaire***. Elle passe par $(\bar{x},\bar{y})$ et sa représentation paramétrique est*

$$X \mapsto Y = [\bar{y} - \frac{cov(x,y)}{v(x)}\bar{x}] + \frac{cov(x,y)}{v(x)}X.$$

Proposition 1.2 *Le coefficient de corrélation est égal à 1 ou -1 si et seulement si les (x_i,y_i) sont sur la droite de régression linéaire. Les séries statistiques sont d'autant plus corrélées linéairement que le coefficient de corrélation est proche de 1 ou -1.*

Remarque 1.5 La méthode qui consiste à minimiser $\sum_{i=1}^{n}(y_i-ax_i-b)^2$ s'appelle la **méthode des moindres carrés**. Elle est due à Gauss et Legendre (XIX^e siècle).

1.2 Données qualitatives

Les données qualitatives peuvent parfois s'ordonner (classement) mais elles ne s'additionnent jamais. Elles nécessitent un traitement qui diffère de celui des données quantitatives.

Définition 1.3 Soit $x = (x_1, ..., x_n)$ une suite de données qualitatives (la **variable** ou le **caractère**) à valeurs dans un ensemble $\{c_1, ..., c_d\}$ dont les éléments sont appelés **modalités** Pour chaque $j \in \{1, ..., d\}$ on appelle **effectif** de la modalité c_j le nombre n_j de i tel que $x_i = c_j$ et on appelle **fréquence** de la modalité c_j le rapport (positif) $f_j = \dfrac{n_j}{n}$. On a

$$n_1 + ... + n_d = n \text{ et } f_1 + ... + f_d = 1.$$

Le **mode** est la modalité de fréquence (ou d'effectif) la plus élevée.

Exemple 1.2 Considérons les couleurs de dix feutres numérotés d'une trousse d'écolier : $x_1 = noir$, $x_2 = jaune$, $x_3 = rouge$, $x_4 = bleu$, $x_5 = vert$, $x_6 = bleu$, $x_7 = rouge$, $x_8 = rouge$, $x_9 = noir$, et $x_{10} = rouge$. Ordonnons les couleurs par effectifs croissants. On obtient $c_1 = jaune$, $n_1 = 1$, $f_1 = 0,1$, $c_2 = vert$, $n_2 = 1$, $f_2 = 0,1$, $c_3 = bleu$, $n_3 = 2$, $f_3 = 0,2$, $c_4 = noir$, $n_4 = 2$, $f_4 = 0,2$ et $c_5 = rouge$, $n_5 = 4$, $f_5 = 0,4$. Le mode est le rouge.

1.3 Représentation de séries statistiques

On représente habituellement une série statistique à l'aide d'un tableau, d'un histogramme, d'une courbe ou d'un nuage de points. Considérons par exemple les températures moyennes mensuelles (en ^{o} C) et les précipitations mensuelles (en mm) à Condate au cours des cinquante dernières années.

mois	Janv.	Fév.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
temp.	5	6	8	10	13	16	19	18	16	13	8	6
pluies	61	52	49	45	58	46	43	47	57	64	68	69

Résumons en un tableau les paramètres des deux séries statistiques.

	température	précipitations
étendue	14	26
médiane	11,5	54,5
moyenne	11,5	54,9
variance	22,8	77,4
écart-type	4,8	8,8

Représentons par des histogrammes les températures et les précipitations.

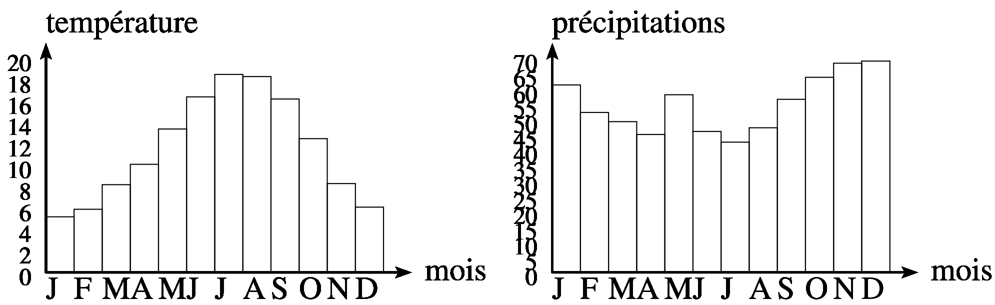


FIGURE 1 – Températures et les précipitations moyennes à Condate

Calculons la corrélation entre la température et les précipitations.

covariance	-22
coefficient de corrélation	-0,53
pente de la droite de régression	-0,97
ordonnée à l'origine de la droite de régression	66

Le coefficient de corrélation est loin de -1 et de 1. La température et les précipitations sont peu corrélées. Finissons par une représentation graphique de la distribution (température, précipitations) à l'aide d'un **nuage** de points et de de la droite de corrélation.

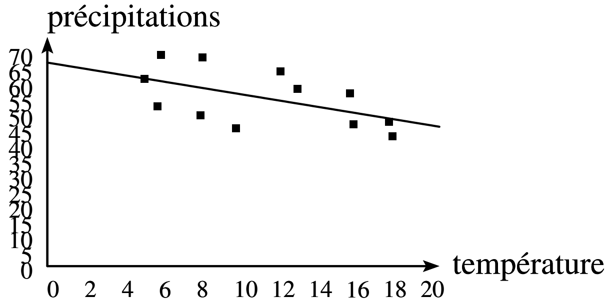


FIGURE 2 – Le graphe (température, pluie) à Condate (en ^{o}C et mm)

2 Probabilités et variables aléatoires

2.1 Probabilités

L'objet fondamental des probabilités est **l'expérience aléatoire** c'est à dire une expérience qui a plusieurs issues possibles mais qui sont incertaines. Pour tenter de comprendre une expérience aléatoire, le probabiliste va lui associer un modèle mathématique appelé **espace probabilisé**. Le choix de ce modèle est délicat et il faut toujours avoir en tête que ce n'est qu'un modèle, une vue de l'esprit, mais pas la réalité.

Avant de donner la définition mathématique des espaces probabilisés décrivons trois expériences aléatoires.

Le jeu de pile ou face repose sur le principe suivant. Lorsqu'on lance la pièce on ne sait pas si c'est pile ou face qui sera visible et chacun des deux a autant de chances d'apparaître : pile et face ont chacun une chance sur deux de tomber. On a évidemment une chance sur une de voir pile ou face.

Le lancer d'un dé repose sur un principe similaire. Les six faces du dé ont autant de chances d'apparaître, une chance sur six chacune. En lançant le dé on a bien sûr six chances sur six de sortir un chiffre entre 1 et 6. Puisque trois faces du dé portent un numéro pair et les trois autres portent un numéro impair on admet que la chance de sortir un numéro pair en lançant le dé est de trois fois une chance sur six c'est à dire une chance sur deux. De même, puisque deux faces portent un numéro supérieur ou égal à 5 et deux autres portent un numéro inférieur ou égal à 2, on conçoit que la chance de sortir un numéro supérieur ou égal à 5 est deux fois une chance sur six c'est à dire une chance sur trois, la chance de sortir un numéro inférieur ou égal à 2 est deux fois une chance sur six c'est à dire une chance sur trois et la chance de sortir un numéro supérieur ou égal à 5 ou inférieur ou égal à 2 est de quatre sur six. C'est l'addition de la chance de sortir un numéro supérieur ou égal à 5 et de la chance de sortir un numéro inférieur ou égal à 2. En revanche la chance de sortir un numéro qui soit pair ou supérieur ou égal à 5 est de quatre chances sur six (c'est à dire deux chances sur trois). Elle n'est pas égale à la chance de sortir un numéro pair additionnée de la chance de sortir un numéro supérieur ou égal à 5 car 6 est à la fois pair et supérieur ou égal à 5.

Considérons maintenant un mètre coloré alternativement en blanc sur des portions de neuf centimètres et en noir sur des portions de un centimètre. On sectionne ce mètre au hasard. On admet facilement qu'on a trois chances sur dix de le sectionner avant la graduation *30 cm*, qu'on a une chance sur dix de le sectionner après la graduation *90 cm*, et qu'on a une chance sur dix de le sectionner sur un portion noire. On le sectionne avant la graduation *30 cm* ou après la graduation *90 cm* avec quatre chances sur dix. Ceci est bien égal à la chance de le sectionner avant la graduation *30 cm* additionnée de la chance de le sectionner après la graduation *90 cm*. En revanche on le sectionne avant la graduation *30 cm* ou sur une portion noire avec trente sept chances sur cent. Ceci n'est pas égal à la chance de le sectionner avant la graduation *30 cm* additionnée de la chance de le sectionner sur une portion noire car on peut sectionner ce mètre sur une portion noire qui se situe avant la graduation *30 cm*.

Lancer la pièce, lancer le dé ou sectionner au hasard le mètre constituent des **expériences aléatoires**. Les ensemble {pile, face}, $\{1,2,3,4,5,6\}$ et [0 cm, 100 cm] sont les ensembles des **issues**, **épreuves** ou **événements élémentaires** associés à ces expériences aléatoires. *Pile tombe*, *4 sort*, *on sectionne à 17,3 cm* sont des **événements élémentaires** alors que *un nombre pair sort* ou *on sectionne sur une portion noire* sont appeles **événements** associés à ces expériences. À chaque événement on a associé un nombre, sa chance d'arriver, qui est compris entre 0 et 1. Ce nombre s'appelle la **probabilité** de l'événement. On a ainsi muni l'ensemble des issues d'une **probabilité**. On parle alors d'**espace probabilisé**.

Certains événements sont disjoints (on dit **incompatibles**). C'est le cas d'événements élémentaires associés à une même expérience aléatoire. C'est aussi le cas de *sortir un numéro supérieur ou égal à 5* et *sortir un numéro inférieur ou égal à 2* dans l'expérience du lancer d'un dé. C'est encore le cas de *sectionner le mètre avant la graduation 30 cm* et de *sectionner le mètre après la graduation 90 cm*. On a vu que lorsque les événements étaient disjoints leurs chances ou leurs probabilités s'additionnent.

Donnons maintenant une définition mathématique des objets associés à une expérience aléatoire.

Définition 2.1 On appelle **espace probabilisé** $(\Omega, \mathcal{F}, p)$ la donnée d'un ensemble Ω, d'une famille $\mathcal{F}$ de sous-ensembles de Ω et d'une application p de $\mathcal{F}$ dans $[0,1]$ qui vérifient les propriétes suivantes :

— $\Omega \in \mathcal{F}$ et $p(\Omega) = 1$
— Si $A \in \mathcal{F}$ alors $\Omega \setminus A \in \mathcal{F}$
— Si $A_1, ..., A_n, ... \in \mathcal{F}$ alors la réunion des A_i est dans $\mathcal{F}$
— Si $A_1, ..., A_n, ... \in \mathcal{F}$ sont deux à deux disjoints alors

$$p(A_1 \cup ... \cup A_n \cup ...) = p(A_1) + ... + p(A_n) + ...$$

L'ensemble Ω est appelé **univers** et ses éléments **issues**, **épreuves** ou **événements élémentaires**. Les éléments de $\mathcal{F}$ s'appellent les **événements** et l'application p est une **probabilité**. Deux événements A et B sont dits **indépendants** si $p(A \cap B) = p(A)p(B)$, ils sont **incompatibles** s'ils sont d'intersection vide (disjoints) et ils sont dits **équiprobables** si $p(A) = p(B)$. Un événement de probabilité 1 est dit **certain**. C'est le cas de Ω. Un événement de probabilité nulle est est dit **impossible**. C'est le cas de l'événement vide $\emptyset$. Si Ω est fini ou dénombrable on parle de **probabilité discrète** et sinon de **probabilité continue**.

Exemples 2.1 Dans l'expérience de lancer du dé, les événements *un nombre pair sort* et *un nombre impair sort* ne sont pas indépendants. En effet la probabilité de chacun d'eux est $\frac{1}{3}$ alors que la probabilité de sortir un nombre qui est pair et impair à la fois est nulle. En revanche les événements *un nombre pair sort* (de probabilité $\frac{1}{2}$) et *sortir un numéro supérieur ou égal à 5* (de probabilité $\frac{1}{3}$) sont indépendants car la probabilité de sortir un numéro pair supérieur ou égal à 5 est $\frac{1}{6}$ qui est égal à $\frac{1}{2}\frac{1}{3}$.

Dans l'expérience du mètre sectionné les événements *sectionner le mètre avant la graduation 30 cm* (de probabilité $\frac{3}{10}$) et *on sectionne sur une portion noire* (de probabilité $\frac{1}{10}$) sont indépendants car la probabilité de sectionner sur une portion noire avant 30 cm est de $\frac{3}{100}$ qui est égal à $\frac{3}{10}\frac{1}{10}$. En revanche les événements *sectionner le mètre avant la graduation 17,3 cm* (de probabilité $0,173$) et *on sectionne sur une portion noire* (de probabilité $\frac{1}{10}$) sont indépendants car la probabilité de sectionner sur une portion noire avant 17,3 cm est de $\frac{1}{100}$ qui n'est pas égal à $\frac{1}{10} \times 0,173$.

Exemple 2.2 Probabilité uniforme sur un ensemble fini Soit n un entier naturel non nul. On pose $\Omega = \{1,...,n\}$, on note $\mathcal{F}$ la famille de tous les sous-ensembles de Ω et on considère la probabilité p définie par $p(k) = \frac{1}{n}$ pour tout $k \in \Omega$. Cette probabilité est la **probabilité uniforme sur** Ω. Si A est un sous-ensemble de Ω alors

$$p(A) = \frac{\text{Cardinal de } A}{n}.$$

Cet exemple est la généralisation du jeu de pile ou face ($n = 2$) et du lancer d'un dé ($n = 6$). Cette probabilité permet de modéliser le tirage d'une carte dans un jeu de 32 cartes et affirme qu'on a une chance sur huit de tirer un as au hasard.

Exemple 2.3 Le lancer de deux pièces Modélisons l'expérience aléatoire qui consiste à lancer deux pièces indiscernables et à regarder le tirage obtenu. Il y a trois issues possibles *obtenir deux piles* (I_{pp}), *obtenir deux faces* (I_{ff}) et *obtenir un pile et un face* (I_{fp}). Il est faux de penser que les trois événements sont équiprobables. En effet si maintenant on marque l'une des pièces on constate que l'événement *obtenir un pile et un face* (I_{fp}) peut être décomposé en deux événements disjoints *on obtient pile avec la pièce marquée et face avec l'autre* et *on obtient face avec la pièce marquée et pile avec l'autre*. On conclut alors que $p(I_{pp}) = p(I_{ff}) = \frac{1}{4}$ et $p(I_{pf}) = \frac{1}{2}$. Une expérience équivalente au lancer de deux pièces dont l'une est marquée est de lancer deux fois la même pièce. Si on remplace pile et face par garçon et fille on conclut qu'une fratrie de deux enfants possède une chance sur deux d'être formée d'un garçon et d'une fille, une chance sur quatre de deux filles et une chance sur quatre de deux garçons.

Exemple 2.4 Le joueur obstiné Un joueur décide de lancer une pièce de monnaie jusqu'au premier pile. C'est une expérience aléatoire qui est modélisée par un espace probabilisé qui n'est plus fini mais dénombrable. À tout entier naturel non nul n est associé l'événement élémentaire *on fait* $n-1$ *lancers qui donnent un face puis un lancer qui donne un pile* (A_n). Puisque l'issue d'un lancer est indépendant du passé on en déduit que la probabilité de cet événement est $p(A_n) = \frac{1}{2^n}$.

Exemple 2.5 Probabilité uniforme sur un segment Ici Ω est un segment $[a,b]$. La **probabilité uniforme sur** Ω est la probabilité qui à tout segment $[\alpha,\beta]$ inclus dans $[a,b]$ associe $p([\alpha,\beta]) = \frac{\beta-\alpha}{b-a}$. On admet l'existence de cette probabilité. La preuve rigoureuse est due à Borel et Lebesgue au début du XX^e siècle. Cette probabilité permet de modéliser l'expérience du mètre sectionné.

Exemple 2.6 Le lancer d'une fléchette On lance une fléchette au hasard dans une cible de rayon 1. Si $[\alpha,\beta] \subset [0,1]$ la probabilité pour que la fléchette tombe à une distance comprise entre α et β est proportionnelle à l'aire de cette zone et vaut $\beta^2 - \alpha^2$.

Exemple 2.7 Série statistique À une série statistique $x = (x_1,...,x_n)$ est toujours associé un espace probabilisé. Si les données sont qualitatives on munit l'ensemble des modalités $\{c_1,...,c_d\}$ de la probabilité définie par $p(c_j) = f_j$ où f_j désigne la fréquence de la modalité. Si les données sont quantitatives on munit $\mathbf{R}$ de la probabilité définie par $p(]x,y]) = F(y) - F(x)$ où F est la fréquence cumulée.

Définition 2.2 Soit $(\Omega, \mathcal{F}, p)$ un espace probabilisé et $A \in \mathcal{F}$ tel que $p(A) > 0$. La **probabilité conditionnelle sachant** A est la probabilité définie par

$$p(B/A) = \frac{p(A \cap B)}{p(A)}.$$

Exemple 2.8 La probabilité de sectionner le mètre sur une portion noire (événement B) sachant qu'il est coupé avant 17,3 cm (événement A) est $p(B/A) = \frac{0{,}01}{0{,}173} \simeq 0,0578$.

Proposition 2.1 *Si* $A_1,...,A_n \in \mathcal{F}$ *alors*

$$p(A_1 \cap ... \cap A_n) = p(A_1)p(A_2/A_1)...p(A_n/A_1 \cap ... \cap A_{n-1}).$$

L'exemple du joueur obstiné pouvait être traité à l'aide de probabilités conditionnelles et de la formule ci-dessus.

2.2 Variables aléatoires

On vient de voir des espaces probabilisés associés à des expériences aléatoires. Il existe aussi parfois des nombres naturellement associés aux événements élémentaires provenant d'expériences aléatoires. On peut penser par exemple à des gains associés à certains tirages dans une loterie, au nombre de faces obtenues en lançant deux pièces, la mesure de l'endroit ou on sectionne le mètre, la distance à laquelle la fléchette se trouve du centre. Ces nombres qui sont des fonctions d'événements aléatoires s'appellent **variables aléatoires**.

Définition 2.3 Soit $(\Omega, \mathcal{F}, p)$ un espace probabilisé. Une variable **aléatoire discrète** X associée à cet espace probabilisé est une application de Ω dans $\mathbf{R}$ qui prend un nombre de valeurs fini ou dénombrable.

Remarque 2.1 On peut faire la somme ou le produit de deux variables aléatoires dès qu'elles sont définies sur le même espace probabilisé.

Considérons une variable aléatoire discrète X sur un espace probabilisé $(\Omega, \mathcal{F}, p)$. Puisque l'ensemble des valeurs de X est discret, ces éléments peuvent être numérotes $x_1, \ldots, x_i, \ldots$ $(1 \leq i \leq n)$ ou $i \in \mathbf{N}$ suivant que X prend un nombre fini ou dénombrable de valeurs. On définit sur l'ensemble des valeurs une probabilité p_X appelée **loi de la variable aléatoire** en posant pour chaque i, $p_X(x_i) = p_i = p(X^{-1}(x_i))$.

Définition 2.4 L'**espérance** de X est le nombre

$$E(X) = \sum_i p_i x_i = \overline{X}$$

s'il existe (ce qui n'est pas toujours le cas).

Remarque 2.2 L'espérance est la version probabiliste de la moyenne ou du barycentre.

Définition 2.5 La **variance** de X est le nombre positif

$$V(X) = \sum_i p_i (x_i - \overline{X})^2$$

s'il existe (ce qui n'est pas toujours le cas). L'**écart-type** $\sigma(X)$ est la racine carrée de la variance de X : $\sigma(X) = \sqrt{V(X)}$.

Proposition 2.2

$$V(X) = E(X^2) - (E(X))^2$$

Exemple 2.9 On associe au jeu de pile ou face la variable aléatoire X qui vaut 1 quand on tombe sur pile et 0 quand on tombe sur face. On a $p_1 = p_0 = \frac{1}{2}$, $E(X) = \frac{1}{2}$ et $V(X) = \frac{1}{4}$.

Définition 2.6 Soit $(\Omega, \mathcal{F}, p)$ un espace probabilisé. Une variable **aléatoire continue** X associée à cet espace probabilisé est une application de Ω dans $\mathbf{R}$ qui prend plus qu'un nombre dénombrable de valeurs et qui est telle que si $x \in \mathbf{R}$ alors $X^{-1}((-\infty, x])$ est un événement. On appelle **fonction de répartition de la variable aléatoire** X la fonction définie par $F(x) = p(X^{-1}((-\infty, x]))$. On définit sur $\mathbf{R}$ une probabilité appelée **loi de la variable aléatoire** en posant $p_X(]\alpha, \beta]) = F(\beta) - F(\alpha)$.

Considérons une variable aléatoire continue X sur un espace probabilisé $(\Omega, \mathcal{F}, p)$.

Proposition 2.3 *La fonction de répartition F associée à X est croissante, à valeurs dans $[0, 1]$ et vérifie*

$$\lim_{x \to -\infty} F(x) = 0, \lim_{x \to +\infty} F(x) = 1.$$

On suppose dorénavant que la fonction de répartition F de la variable aléatoire X est dérivable (sauf peut-être en un nombre fini de points) de dérivée $f = F'$.

Définition 2.7 La dérivée f de F s'appelle **densité** de X ou de p_X.

Définition 2.8 L'**espérance** de X est le nombre

$$E(X) = \int_{\infty}^{+\infty} xF'(x)dx = \int_{\infty}^{+\infty} xf(x)dx$$

s'il existe (ce qui n'est pas toujours le cas).

Remarque 2.3 Comme dans le cas discret l'espérance est la version probabiliste de la moyenne ou du barycentre.

Définition 2.9 La **variance** de X est le nombre positif

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= \int_{\infty}^{+\infty} x^2 f(x)dx - (\int_{\infty}^{+\infty} xf(x)dx)^2 \\ &= \int_{\infty}^{+\infty} (x - E(X))^2 f(x)dx \end{aligned}$$

s'il existe (ce qui n'est pas toujours le cas). L'**écart-type** $\sigma(X)$ est la racine carrée de la variance de X : $\sigma(X) = \sqrt{V(X)}$.

Remarque 2.4 (définition et propriétés) Une variable aléatoire X suit la **loi de Poisson** $\mathcal{P}(\lambda)$ **de paramètre** $\lambda > 0$ si pour tout $k \in \mathbf{N}$ $p(X = k) = p_k = \dfrac{\lambda^k}{k}\exp(-\lambda)$. Alors $E(X) = V(X) = \lambda$.

2.3 Inégalité de Bienaymé-Tchebychev

Comme pour les séries statistiques quantitatives, la variance d'une variable aléatoire mesure la concentration de la variable autour de sa valeur moyenne. C'est ce qu'exprime l'inégalité de Bienaymé-Tchebychev.

Proposition 2.4 *(***Inégalité de Bienaymé-Tchebychev** *(XIX^e siècle)) Si X est une variable aléatoire d'espérance de variance finies alors*

$$p(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2} \quad \textit{pour tout} \quad \varepsilon > 0.$$

2.4 Indépendance

Soient X et Y deux variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{F}, p)$.

Proposition 2.5 *Dans les cas discrets et continus on a*

$$E(\lambda X) = \lambda E(X), E(X+Y) = E(X) + E(Y), V(\lambda X + c) = \lambda^2 V(X).$$

Définition 2.10 La **covariance** de deux lois X et Y est le nombre

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$

Proposition 2.6 *$Cov(X,Y)$ existe dès que $V(X)$ et $V(Y)$ existent.*

Définition 2.11 Les variables aléatoires X et Y sont dites **indépendantes** si :

— pour tous les i, j les événements $X^{-1}(x_i)$ et $Y^{-1}(y_j)$ sont indépendants dans le cas discret,
— pour tous les $x, y \in \mathbf{R}$ les événements $X^{-1}((-\infty, x])$ et $Y^{-1}((-\infty, y])$ sont indépendants dans le cas continu.

Proposition 2.7 *Si X et Y sont deux variables aléatoires indépendantes alors*

$$\begin{array}{rcl} Cov(X,Y) & = & 0 \\ E(XY) & = & E(X)E(Y) \\ V(X+Y) & = & V(X)+V(Y) \end{array}$$

dès que $E(X)$, $E(Y)$, $V(X)$ et $V(Y)$ existent.

Remarque 2.5 (proposition) Si X_1 et X_2 indépendantes suivent des lois de Poisson $\mathcal{P}(\lambda_1)$ et $\mathcal{P}(\lambda_2)$ alors $X = X_1 + X_2$ suit la loi de Poisson de paramètre $\mathcal{P}(\lambda_1 + \lambda_2)$.

En appliquant l'inégalité de Bienaymé-Tchebychev à une suite de variables aléatoires indépendantes et de même loi on obtient la loi faible des grand nombres.

Théorème 2.1 *(***Loi faible des grands nombres***) Soient $(X_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires indépendantes et de même loi, de moyenne m et de variance v. Alors la suite de variables aléatoire $(\Sigma_n)_{n\in\mathbf{N}}$ définies par*

$$\Sigma_n = \frac{X_1 + \ldots + X_n}{n}$$

vérifie

$$p(|\Sigma_n - m| \geq \varepsilon) \leq \frac{v}{n\varepsilon^2}.$$

Exemple 2.10 On lance n fois une pièce et on compte le nombre de piles obtenus. Ceci se modélise on considérant la somme S_n de n variables aléatoires X_i qui prennent les valeurs 0 (pour face) et 1 (pour pile) indépendantes et de même loi d'espérance $\frac{1}{2}$ et de variance $\frac{1}{4}$. La variable aléatoire S_n compte le nombre de piles. On a donc en posant $S_n = n\Sigma_n$,

$$p(|S_n - \frac{n}{2}| \geq n\varepsilon) \leq \frac{1}{4n\varepsilon^2}.$$

On applique ce résultat pour différentes valeurs de n et ε.
— $n = 10$ et $\varepsilon = \frac{1}{10}$: ça donne $p(|S_{10} - 5| \geq 1) \leq 2,5$ ce qui ne sert à rien car $2,5 \geq 1$.
— $n = 100$ et $\varepsilon = \frac{1}{10}$: ça donne $p(|S_{10} - 50| \geq 10) \leq 0,25 < 1$.
— $n = 100$ et $\varepsilon = \frac{1}{20}$: ça donne $p(|S_{10} - 50| \geq 5) \leq 1$ ce qui ne sert à rien.
— $n = 1000$ et $\varepsilon = \frac{1}{10}$: ça donne $p(|S_{10} - 500| \geq 100) \leq 0,025 < 1$.
— $n = 1000$ et $\varepsilon = \frac{1}{20}$: ça donne $p(|S_{10} - 500| \geq 50) \leq 0,1 < 1$.

2.5 Exercice

Comme exercice on peut calculer les espérances et les variances des différentes lois associées aux exemples.

3 La loi binomiale

Définition 3.1 On appelle **épreuve de Bernoulli** ($XVIII^e$) de paramètre $p \in [0,1]$ une expérience aléatoire qui a deux issues, l'une appelée **succès** de probabilité p et l'autre appelée **échec** de probabilité $1-p$. On lui associe la variable aléatoire Y qui affecte la valeur 1 au succès et la valeur 0 à l'échec. On dit que Y suit une **loi de Bernoulli** $\mathcal{B}(p)$ **de paramètre** p.

Exemples 3.1 Le jeu de pile ou face est une épreuve de Bernoulli de paramètre $\frac{1}{2}$, le lancer d'un dé avec le 1 comme seule issue gagnante est une épreuve de Bernoulli de paramètre $\frac{1}{6}$ et sectionner le mètre avec comme seule issue gagnante le sectionner avant 17,3 cm est une épreuve de Bernoulli de paramètre $0,173$.

Proposition 3.1 *Si Y est une loi de Bernoulli de paramètre p alors $E(Y) = p$ et $V(Y) = p(1-p)$.*

Si on s'intéresse au nombre de succès après avoir répéter de façon indépendante une épreuve de Bernoulli on considère la variable aléatoire X qui est égale à $X = Y_1 + ... + Y_n$ où les Y_i sont des variables aléatoires indépendantes qui suivent une loi de Bernoulli de paramètre p. La loi X compte le nombre de succès de n épreuves de Bernoulli indépendantes et de même paramètre p.

Définition 3.2 On appelle **loi binomiale de paramètres** (n,p) la loi $\mathfrak{B}(n,p)$ suivie par la somme X de n variables aléatoires indépendantes $Y_1,...,Y_n$ qui suivent une loi de Bernoulli de paramètre p.

Proposition 3.2 *Si $0 \leq k \leq n$ sont deux entiers on pose $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Alors*

- $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ *(Formule du binôme de Newton (XVII^e siècle)*
- *Le nombre de sous-ensembles à k éléments dans un ensemble à n éléments est $\binom{n}{k}$.*

Proposition 3.3 *Si X suit une loi binomiale de paramètres (n,p) alors*

$$p_k = p(X^{-1}(k)) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Exemple 3.2 Si on lance n fois une pièce de monaie, la variable aléatoire *nombre de piles obtenus* suit une loi binomiale $\mathfrak{B}(n, \frac{1}{2})$:

$$p_k = p(\text{on obtient k piles en n lancers}) = \binom{n}{k} \frac{1}{2^n}.$$

Exemple 3.3 Dans une usine de voitures on fabrique 700 voitures par jour. La probabilité pour qu'une voiture ait besoin d'une retouche de finition est de $0,01$ et ne dépend pas des autres voitures. Le nombre de voitures produites par jour et ayant besoin d'une retouche de finition suit donc une loi binomiale de paramètre $(700,\ 0,01)$. La probabilité pour que 10 voitures aient besoin d'une retouche dans la journée est

$$p_{10} = p(X^{-1}(10)) = \binom{700}{10} 0,01^{10} 0,99^{690},$$

c'est à dire $p_{10} = 0,0710$. Un calcul donne $p_4 = 0,0909$, $p_5 = 0,1278$, $p_6 = 0,1495$, $p_7 = 0,1498$, $p_8 = 0,1310$ et $p_8 = 0,1018$. On en déduit que la probabilité pour que des retouches soient nécessaires pour quatre à dix voitures est de $p = p_4 + p_5 + p_6 + p_7 + p_8 + p_9 + p_{10}$ c'est à dire $p = 0,82$. Il faut prévoir l'effectif de l'atelier des retouches en conséquence.

Proposition 3.4 *Si la variable aléatoire X suit une loi binomiale de paramètres (n,p) alors*

$$E(X) = np, \ \ V(X) = np(1-p) \ \ et \ \ \sigma(X) = \sqrt{np(1-p)}.$$

Exemple 3.4 Dans l'usine considérée précédemment il y a en moyenne 7 voitures par jour qui ont besoin d'une retouche ($n = 700$ et $p = 0,001$) avec une variance de 6,93. Un calcul utilisant l'inégalité de Bienaymé-Tchebychev assure que la probabilté que des retouches soient nécessaires pour au plus 15 voitures est d'au moins 0,91.

Remarque 3.1 Si n est grand et p petit, la loi de Poisson $\mathcal{P}(np)$ *approxime* $\mathfrak{B}(n,p)$.

4 La loi normale

4.1 Définition et premières propriétés

Proposition 4.1 *Soient $m \in \mathbf{R}$ et $\sigma > 0$. La fonction définie par*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

est continue, strictement positive, strictement croissante sur $(-\infty, m]$, strictement décroissante sur $[m, +\infty)$, elle vérifie $f(m+t) = f(m-t)$ et

$$\begin{aligned}
\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx &= 1, \\
\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx &= m, \\
\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx &= \sigma^2.
\end{aligned}$$

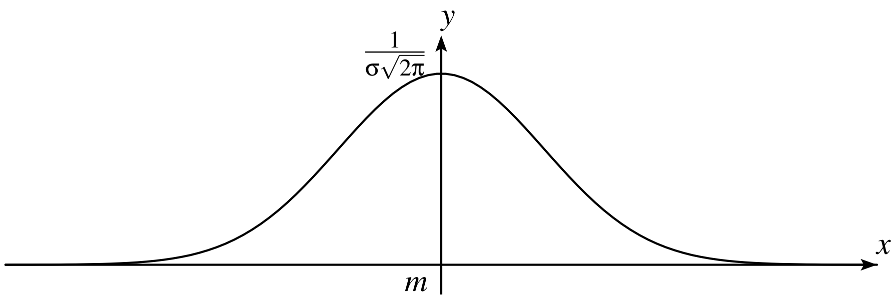


FIGURE 3 – La cloche de Gauss

Définition 4.1 La variable aléatoire X suit une **loi normale** $\mathcal{N}(m, \sigma^2)$ (ou de Gauss et Laplace (XIX^e siècle)) de moyenne m et d'écart-type $\sigma \geq 0$ si sa densité est la fonction

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}.$$

Si $m = 0$ la loi et dite **centrée**, si $\sigma = 1$ elle est dite **réduite**.

Proposition 4.2 *Si la variable aléatoire X suit une* **loi normale** $\mathcal{N}(m, \sigma^2)$ *alors la variable aléatoire* $Y = \dfrac{X - m}{\sigma}$ *suit la loi normale centrée réduite* $\mathcal{N}(0, 1)$.

Remarque 4.1 La loi normale est assez simple puisqu'elle est caractérisée totalement par deux paramètres, son espérance (sa moyenne) et sa variance (ou l'écart-type). Plus σ est petit plus la probabilité se concentre vers la moyenne.

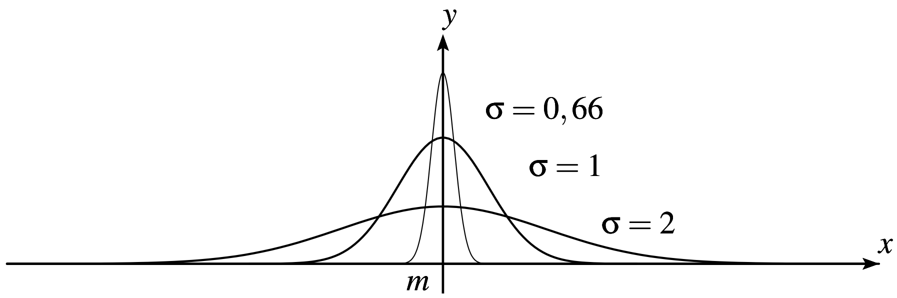


FIGURE 4 – variation de la concentration en fonction de σ

Pour pouvoir faire du calcul de probabilités en relation avec la loi normale on utilise une table (papier, électronique) qui donne avec une bonne précision la fonction de répartition

$$\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$$

de la loi normale centrée réduite et les propriétés de symétrie et d'homogénéité des lois normales.

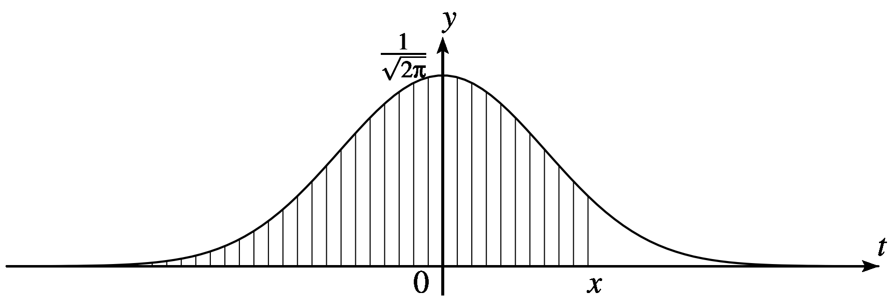


FIGURE 5 – L'aire hachurée est égale à $\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$

Proposition 4.3 *Si* $x \in \mathbf{R}$, *on a* $\Pi(x) = 1 - \Pi(-x)$, *i.e.* $\Pi(x) - \Pi(-x) = 2\Pi(x) - 1$ *et* $\Pi(0) = \frac{1}{2}$.

Proposition 4.4 *Si la variable aléatoire X suit une* **loi normale** $\mathcal{N}(m, \sigma^2)$ *alors la variable aléatoire* $Y = \dfrac{X - m}{\sigma}$ *suit la loi normale centrée réduite* $\mathcal{N}(0,1)$. *Ainsi si* $a < b$ *on a*

$$p(X^{-1}((-\infty, a])) = \Pi(\frac{a-m}{\sigma}) \;\; et \;\; p(X^{-1}([a,b])) = \Pi(\frac{b-m}{\sigma}) - \Pi(\frac{a-m}{\sigma})$$

et si $\alpha < \beta$ *on a*

$$p(X^{-1}((-\infty, m + \alpha\sigma])) = \Pi(\alpha) \;\; et \;\; p(X^{-1}([m + \alpha\sigma, m + \beta\sigma])) = \Pi(\beta) - \Pi(\alpha).$$

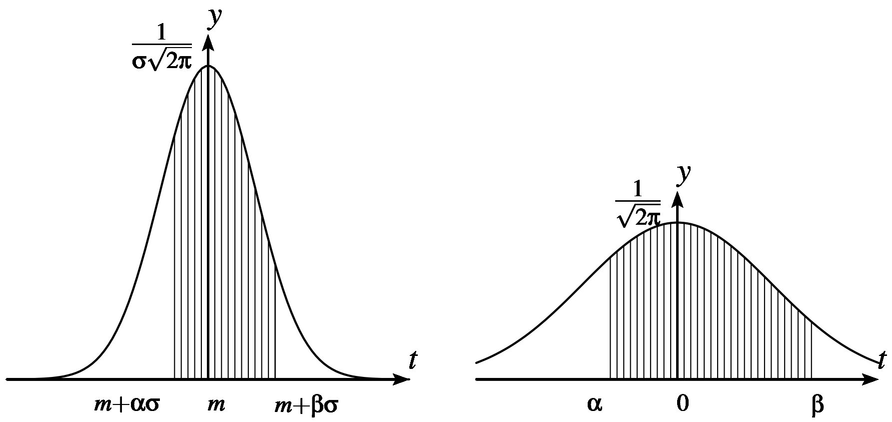


FIGURE 6 – Les aires hachurées sont égales et représentent la probabilité $p(X^{-1}([m+\alpha\sigma, m+\beta\sigma])$ (ici $\alpha < 0 < \beta$))

Corollaire 4.1 *Si la variable aléatoire X suit une* **loi normale** $\mathcal{N}(m, \sigma^2)$ *et si $0 < a$ on a*

$$p(X^{-1}([m-a, m+a])) = 2\Pi(\frac{a}{\sigma}) - 1.$$

Exemple 4.1 Si une population se distribue selon une variable aléatoire qui suit une loi normale $\mathcal{N}(m, \sigma^2)$ alors on déduit de la lecture d'une table de la loi normale centrée réduite que

- 38% de la population est dans l'intervalle $[m-0,5\sigma, m+0,5\sigma]$
- 68% de la population est dans l'intervalle $[m-\sigma, m+\sigma]$
- 86% de la population est dans l'intervalle $[m-1,5\sigma, m+1,5\sigma]$
- 95% de la population est dans l'intervalle $[m-2\sigma, m+2\sigma]$
- 99% de la population est dans l'intervalle $[m-2,5\sigma, m+2,5\sigma]$
- 99,7% de la population est dans l'intervalle $[m-3\sigma, m+3\sigma]$

Signalons le bon comportement par addition des variables aléatoires qui suivent des lois normales.

Proposition 4.5 *Si X_1 et X_2 sont des variables indépendantes qui suivent des lois normales $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ alors la somme $X = X_1 + X_2$ est une variable aléatoire qui suit la loi normale $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Définition 4.2 Soient $(X_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires et X une variable aléatoire. On dit que la suite $(X_i)_{i\in\mathbf{N}}$ converge en loi vers la variable aléatoire X si pour tous les $a < b$ on a

$$\lim_{i\to+\infty} p(X_i^{-1}([a,b])) = p(X^{-1}([a,b])).$$

4.2 Le théorème limite central

Un intérêt de la loi normale est qu'elle explique le comportement de phénomènes répétés en précisant la loi faible des grands nombres.

Théorème 4.1 *(***Théorème limite central***) Soient $(X_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires indépendantes et de même loi, de moyenne m et décart-type σ. Alors la suite de variables aléatoire $(\overline{S}_n)_{n\in\mathbf{N}}$ définies par*

$$\overline{S}_n = \frac{X_1 + \ldots + X_n - nm}{\sqrt{n}\sigma}$$

converge en loi vers une variable aléatoire qui suit la loi normale centrée réduite.

Ce théorème fondamental a une longue histoire qui commence avec De Moivre ($XVIII^e$ siècle). Elle se poursuit avec Gauss, Laplace (XIX^e siècle), Lévy et Lindeberg (XX^e siècle) et elle n'est pas encore achevée

Remarque 4.2 Dans la pratique si les X_i sont des Bernoulli de paramètre p (c'est à dire $X_1 + ... + X_n$ suit une binomiale de paramètres (n, p) alors on approxime cette dernière par la loi normale $\mathcal{N}(np, np(1-p))$ dès que np et $n(1-p)$ dépassent 20. Dans le cas d'une variable continue on fait l'approximation de $\overline{S}_n$ par la loi normale centrée réduite dès que n dépasse 30.

Exemple 4.2 Une personne qui remplit au hasard un questionnaire de cent questions est en train de réaliser cent épreuves de Bernoulli indépendantes et de paramètre $\frac{1}{2}$. Le nombre de bonnes réponses dans le questionnaire est une variable aléatoire X qui suit une loi binomiale $\mathfrak{B}(100, \frac{1}{2})$ donc d'espérance 50, de variance 25 et d'écart-type 5. La probabilité d'obtenir entre 45 et 55 bonnes réponses est 72%. Si on fait le calcul en approximant loi la binomiale $\mathfrak{B}(100, \frac{1}{2})$ (qui est la loi de la somme de 100 variables aléatoires suivant une loi de Bernoulli de paramètre $\frac{1}{2}$ donc de moyenne $\frac{1}{2}$ et d'écart-type $\frac{1}{2}$) par une loi normale comme dans le théorème limite central on considère alors la variable aléatoire $Y = \frac{X-50}{5}$ qui est supposée suivre une loi voisine de la loi normale centrée réduite. On a donc

$$p(X^{-1}([45, 55]) = p(Y^{-1}([-1, 1]) \simeq \Pi(1) - \Pi(-1) = 68\%.$$

L'expérience du questionnaire est équivalente à cent lancers d'une pièce. On a déjà vu qu'appliquer la loi faible des grands nombres dans ce cadre ne permet pas de conclure ($m = \frac{1}{2}, v = \frac{1}{4}, n = 100, \varepsilon = \frac{1}{20}$).

5 Échantillonnage et estimation

5.1 Échantillonnage

On considère une population $\mathcal{P}$ de N individus. On s'intéresse à deux caractères des individus de cette population, un caractère quantitatif *continu* et un caractère qualitatif bimodal (il prend deux valeurs P et F). On va étudier ces caractères sur une sous-population appelée **échantillon**.

5.1.1 Le caractère quantitatif

On note μ la moyenne de ce caractère, v sa variance empirique et σ son écart-type. Ce caractère quantitatif est modélisé par une variable aléatoire X dont la loi est définie par

$$p(X^{-1}([a, b]) = \frac{\mathrm{Card}(\{i \in \mathcal{P}, x_i \in [a, b]\})}{N}$$

dont l'espérance est μ et la variance v. Prélevons un **échantillon** de n individus au hasard. On suppose que l'échantillon est grand mais que la taille de la population est encore plus grande. Pour chacun des n individus de l'échantillon on note x_i la valeur du caractère considéré. Ce sont des variables aléatoires de même espérance μ et de même variance v que X.

La moyenne du caractère sur l'échantillon Vu les hypothèses sur les tailles de l'échantillon et de la population, selon le théorème limite central on peut supposer que la moyenne des x_i

$$\overline{x} = \frac{x_1 + ... + x_n}{n}$$

est une variable aléatoire qui suit la loi normale $\mathcal{N}(\mu, \frac{v}{n})$.

5.1.2 Le caractère qualitatif

On suppose que pour k individus de la population le caractère vaut P et donc pour $N-k$ le caractère vaut F. On pose $p = \frac{k}{N}$. Le caractère d'un individu pris au hasard vaut P avec la probabilité p et F avec la probabilité $1-p$: il suit une loi de Bernoulli de paramètre p.

On prélève un échantillon de n individus au hasard. Si on suppose que la population est beaucoup plus grande que l'échantillon on peut supposer que les caractères des individus de cet échantillon sont indépendants et suivent la même loi de Bernoulli de paramètre p. Par conséquent le nombre d'individus dans cet échantillon dont le caractère vaut P est une variable aléatoire X qui suit une loi binomiale $\mathfrak{B}(n,p)$.

La fréquence d'apparition du caractère P dans l'échantillon La fréquence des individus de l'échantillon dont le caractère vaut P est la variable aléatoire $F = \frac{X}{n}$. L'espérance de F est $E(F) = p$ et sa variance est $V(F) = \frac{p(1-p)}{n}$. D'après le théorème limite central, si n est grand la loi de F est approximée par la loi normale $\mathcal{N}(p, \frac{p(1-p)}{n})$.

5.2 Estimation

L'objet de l'**estimation** est de déduire des propriétés d'un échantillon $\mathcal{E}$ les propritétés de la population $\mathcal{P}$ dont il est issu. Les propriétés recherchées sont par exemple la moyenne μ d'un caractère quantitatif, sa variance V ou la fréquence p d'apparition du caractère P dans la population totale. Les données connues sont l'effectif n de l'échantillon, la moyenne m et la variance empirique v du caractère quantitatif pour l'échantillon et la fréquence f à la quelle le caractère le caracère modal vaut P dans l'échantillon.

La théorie de l'estimation a été introduite par Fisher au début du XX^e siècle.

5.2.1 L'estimation de la variance

Si la taille de la population est grande par rapport à la taille de l'échantillon on peut supposer que les caractères $x_1, \ldots, x_n$ des individus qui forment un échantillon pris au hasard sont des variables aléatoires indépendantes qui suivent une même loi d'espérance la moyenne μ de la population totale et de variance la variance V de la population totale. On en déduit que la moyenne $\overline{x} = \frac{x_1+\ldots+x_n}{n}$ est une variable aléatoire d'espérance μ. Ainsi la moyenne m d'un échantillon est une estimation *en moyenne* de la moyenne de la population. C'est plus compliqué pour la variance. On montre que la variance de l'échantillon est une variable aléatoire dont l'espérance n'est pas V mais $\frac{n-1}{n}V$. C'est pourquoi $\frac{n}{n-1}v$ est une estimation *en moyenne* de la variance de la population.

5.2.2 L'estimation de la moyenne

Dans la situation précédente, pour faire l'estimation de la moyenne μ avec un **risque** contrôlé on considère (en utilisant le théorème central limite) que la moyenne $\overline{x}$ d'un échantillon au hasard est une variable aléatoire qui suit la loi normale $\mathcal{N}(\mu, \frac{V}{n})$ si la variance V de la population est connue ou $\mathcal{N}(\mu, \frac{v}{n-1})$ si on ne connaît que la variance de l'échantillon. Par conséquent, si $\lambda > 0$, la probabilité

$$p(\overline{x} \in [\mu-\lambda, \mu+\lambda])$$

est égale à $2\Pi(\lambda\sqrt{\frac{n}{V}}) - 1$. Or les événements $\overline{x} \in [\mu-\lambda, \mu+\lambda]$, $\mu \in [\overline{x}-\lambda, \overline{x}+\lambda]$ et $|\mu-\overline{x}| \leq \lambda$ sont identiques. Par conséquent, puisque l'échantillon dont on dispose est pris au hasard la probabilité $p(\mu \in [m-\lambda, m+\lambda])$ est égale à $2\Pi(\lambda\sqrt{\frac{n}{V}}) - 1$. L'estimation de la moyenne μ avec le risque α revient à rechercher un intervalle $I = [m-\lambda, m+\lambda]$ dit **intervalle de confiance** tel que la probabilté $p(\mu \in I) = p(|\mu-m| \leq \lambda)$ soit égal **au seuil de confiance** $c = 1-\alpha$:

$$p(\mu \in I) = p(|\mu-m| \leq \lambda) = c = 1-\alpha.$$

Le nombre λ solution du problème vérifie

$$\Pi(\lambda\sqrt{\frac{n}{V}}) = 1 - \frac{\alpha}{2}.$$

Il est nécessaire de supposer que n est grand.

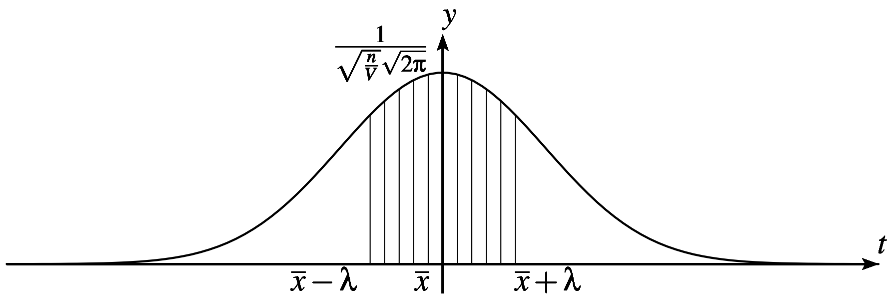


FIGURE 7 – L'aire hachurée représente la probabilité $p(\mu \in [\bar{x}-\lambda, \bar{x}+\lambda]) = 1-\alpha)$

La notion d'intervalle de confiance est due à Neyman (XX^e siècle).

5.2.3 L'estimation de la fréquence

Pour faire l'estimation de la fréquence p avec un **risque** contrôlé on considère (en utilisant le théorème central limite) que la fréquence F d'un échantillon pris au hasard est une variable aléatoire qui suit la loi normale $\mathcal{N}(p, \frac{f(1-f)}{n-1})$. Ici on a estimé la variance de la population à partir de la variance de l'échantillon connu. Par conséquent, si $\lambda > 0$, la probabilité

$$p(F \in [p-\lambda, p+\lambda])$$

est égale à $2\Pi(\lambda\sqrt{\frac{n-1}{f(1-f)}}) - 1$. Or les événements $F \in [p-\lambda, p+\lambda]$, $p \in [F-\lambda, F+\lambda]$ et $|p-F| \leq \lambda$ sont identiques. Par conséquent, puisque l'échantillon dont on dispose est pris au hasard la probabilité $p(p \in [f-\lambda, f+\lambda])$ est égale à $2\Pi(\lambda\sqrt{\frac{n-1}{f(1-f)}}) - 1$. L'estimation de la fréquence p avec le risque α revient à rechercher un intervalle $I = [f-\lambda, f+\lambda]$ dit **intervalle de confiance** tel que la probabilté $p(p \in I) = p(|p-f| \leq \lambda)$ soit égal **au seuil de confiance** $c = 1-\alpha$:

$$p(p \in I) = p(|p-f| \leq \lambda) = c = 1-\alpha.$$

Le nombre λ solution du problème vérifie

$$\Pi(\lambda\sqrt{\frac{n-1}{f(1-f)}}) = 1 - \frac{\alpha}{2}.$$

Il est nécessaire de supposer que n est grand.

5.3 Exemple

Un constructeur automobile généraliste souhaite déterminer le nombre parmi ses cinq millions de clients de ceux qui lisent les publicités qu'il leur adresse personnellement. Il va utiliser un sondage sur 200 clients. La première étape de ce sondage consiste à former un échantillon représentatif. C'est l'échantillonnage. Choisissons comme caractère représentatif des clients le prix d'achat du dernier véhicule acheté par le client. On suppose que ce prix d'achat suit une loi normale de moyenne 15000 euros et d'écart-type 5000 euros. On considérera l'échantillon **pris au hasard** comme acceptable si la moyenne et l'écart-type du prix d'achat calculés sur l'échantillon sont proches de 15000 et 5000. Il s'avère que 20% des sondés disent lire les publicités du constructeur. Recherchons avec 5% de risque un intervalle de confiance pour le taux de lecture de la publicité. On considère que la fréquence de lecture F d'un échantillon au hasard est une variable aléatoire qui suit la loi normale $\mathcal{N}(p, \frac{0{,}20 \cdot 0{,}80}{199})$ où p est la fréquence pour la population. On recherche donc λ tel que $p(|p - 0,20| \leq \lambda) = 0,95$ c'est à dire

$$\Pi(\frac{\lambda}{\sqrt{\frac{0{,}20 \cdot 0{,}80}{199}}}) = 0,975$$

ou encore $\Pi(\frac{\lambda}{0{,}0284}) = 0,975$. On trouve $\lambda = 5,5\%$. On peut donc estimer avec un risque de 5% que le taux de lecture des publicités est compris entre 14,5% et 25,5%.

6 Test de validité d'hypothèse

La notion de test d'hypothèse est due à Pearson et Neyman (XX^e siècle).

6.1 Appartenance d'un échantillon à une population

On souhaite déterminer si un échantillon de taille n peut appartenir à une population connue caractérisée par un seul caractère quantitatif (ou au moins en être raisonablement "proche"). Plus précisément on connaît la moyenne m d'un caractère quantitatif sur l'échantillon, la moyenne μ de ce caractère pour la population totale et la variance V de ce caractère pour la population. On souhaite savoir avec un risque contrôlé α (*risque de première espèce*) si la moyenne m est assez grande pour que l'échantillon puisse provenir de la population ou au moins en être "raisonnablement proche".

Avec les hypothèses sur les tailles de n et N (*grande et très grande*) on peut admettre que la moyenne $\bar{x}$ du caractère sur un échantillon extrait au hasard suit la loi normale $\mathcal{N}(\mu, \frac{V}{n})$.

On recherche $\lambda \in \mathbf{R}$ tel que la probabilité $p(\bar{x} \geq \mu - \lambda)$ soit égale au seuil de confiance $c = 1 - \alpha$:

$$p(\bar{x} \geq \mu - \lambda) = 1 - \alpha = c.$$

Le nombre λ solution du problème vérifie

$$\Pi(\lambda\sqrt{\frac{n}{V}}) = 1 - \alpha.$$

Si $m \geq \mu - \lambda$ alors avec une probabilité de $1 - \alpha$ l'échantillon peut appartenir à la population ou au moins en être "raisonnablement proche" pour qu'on ne puisse pas le distinguer. Sinon, on rejette l'hypothèse d'appartenance à la population.

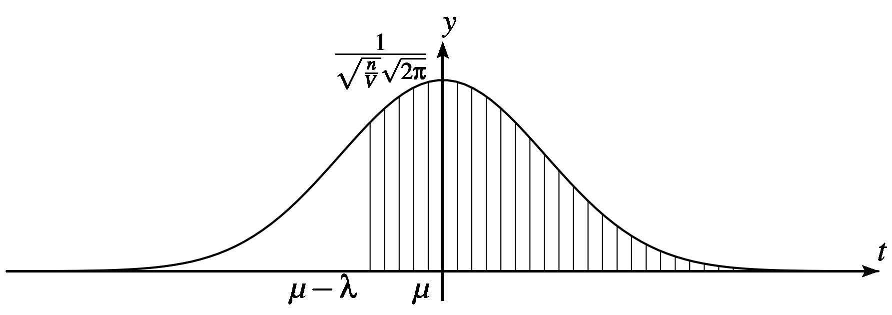


FIGURE 8 – L'aire hachurée représente la probabilité $p(\bar{x} \geq \mu - \lambda) = 1 - \alpha$

Ce test présente le paradoxe suivant. Plus le risque α est petit, plus λ est grand et plus on a de chance d'accepter l'hypothèse à tord. En revanche on réduit le risque de rejeter l'hypothèse à tord (*risque de première espèce*).

On peut bien sûr résoudre le même type de question avec un caractère qualitatif. On peut mesurer le risque de supposer qu'un échantillon appartient à la population considérée ou au moins en est "raisonnablement proche".

Exemple 6.1 Un sous-traitant d'un grand groupe industriel s'est engagé à fournir de façon bi-quotidienne aux dix usines du groupe des vitrages dont au plus 1% présentent un défaut d'aspect (6000 livraisons par an). Une livraison ne respectant cette contrainte est dite défectueuse. Le contrat stipule que si plus de 10% des livraisons sont défectueuses le contrat peut être rompu. Le contrôle-qualité sonde en un an et sur l'ensemble du groupe 300 livraisons au hasard et 15% des livraisons testées sont défectueuses. On veut savoir avec un risque de 5% si le fournisseur respecte ses engagements. Si on fait confiance au fournisseur, on peut admettre que la la fréquence F de livraisons satisfaisantes dans un échantillon au hasard de 300 livraisons suit la loi normale $\mathcal{N}(0,9\ ,\frac{0,15\cdot 0,85}{299})$ (l'écart-type est estimé à l'aide de l'échantillon). On recherche $\lambda \in \mathbf{R}$ tel que la probabilité $p(F \geq 0,9-\lambda)$ soit égale au seuil de confiance $c = 0,95$:

$$p(F \geq 0,9-\lambda) = 0,95.$$

Le nombre λ solution du problème vérifie

$$\Pi(\frac{\lambda}{\sqrt{\frac{0,15\cdot 0,85}{299}}}) = 0,95$$

c'est à dire $\lambda = 3,4\%$. Puisque 85%$<$ 90%-3,4% on ne doit pas considérer que le fournisseur respecte ses engagements.

6.2 Égalité des moyennes, des fréquences

6.2.1 Égalité des moyennes

On souhaite déterminer si deux échantillons de tailles n_1 et n_2 sont issus de deux populations pour lesquelles un caractère quantitatif donné à même moyenne (ou au moins possède des moyennes raisonnablement proches) connaissant les moyennes m_1 et m_2 du caractère quantitatif sur les échantillons, ainsi que les variances v_1 et v_2 sur ces échantillons. On va tester **l'égalité des moyennes**.

Avec les hypothèses sur les tailles de n_1, n_2 des échantillons et N_1 et N_2 des populations (*grandes et très grandes*) on peut admettre que la moyenne $\bar{x}_i$ du caractère sur un échantillon de taille n_i extrait au hasard de la population i suit la loi normale $\mathcal{N}(M_i, \frac{v_i}{n_i-1})$ ou M_i est la moyenne du caractère sur la population i. Ici on a estimé la variance de la population i à partir de la variance de l'échantillon connu. On peut aussi supposer les variables aléatoires $\bar{x}_i$ indépendantes. Sous l'hypothèse d'égalité des moyennes ($M_1 = M_2$), la variable aléatoire $\bar{x}_1 - \bar{x}_2$ suit la loi normale $\mathcal{N}(0, \frac{v_1}{n_1-1} + \frac{v_2}{n_2-1})$.

On recherche $\lambda \in \mathbf{R}$ tel que la probabilité $p(|\bar{x}_1 - \bar{x}_2| \leq \lambda)$ soit égale au seuil de confiance $c = 1-\alpha$:

$$p(|\bar{x}_1 - \bar{x}_2| \leq \lambda) = 1-\alpha = c.$$

Le nombre λ solution du problème vérifie

$$\Pi(\frac{\lambda}{\sqrt{\frac{v_1}{n_1-1} + \frac{v_2}{n_2-1}}}) = 1 - \frac{\alpha}{2}.$$

Si $|m_1 - m_2| \leq \lambda$ alors avec une probabilité de $1-\alpha$ les deux échantillons proviennent de la même population. Sinon, on rejette l'hypothèse d'égalié des moyennes.

Ce test présente le paradoxe suivant. Plus le risque α est petit, plus λ est grand et plus on a de chance d'accepter l'hypothèse à tord. En revanche on réduit le risque de rejeter l'hypothèse à tord (*risque de première espèce*).

Exemple 6.2 Avant de décider de la commercialisation d'un carburant moins poluant un groupe pétrolier souhaite s'assurer qu'il ne réduit pas la puissance des moteurs. Son bureau d'étude mesure la puissance d'un échantillon de mille voitures représentatives du parc automobile européen alimentées avec ce nouveau carburant. Il fait le même test de puissance sur un second échantillon de 700 voitures alimentées avec le carburant classique. Pour les deux échantillons, la répartition des puissances suit une loi normale. La moyenne est 78 CV et l'écart-type 20 CV pour le premier échantillon alors que pour le second la moyenne est de 79 CV et l'écart-type 21 CV. Sous l'hypothèse d'égalité des moyennes (l'essence nouvelle n'affecte pas la puissance), on peut donc admettre que la moyenne des puissances $\bar{x}_1$ d'un échantillon de voitures alimentées avec le nouveau carburant suit la loi normale $\mathcal{N}(P,400)$ et la moyenne des puissances $\bar{x}_2$ d'un échantillon de voitures alimentées avec le carburant classique suit la loi normale $\mathcal{N}(P,441)$. De plus, la variable aléatoire $\bar{x}_1-\bar{x}_2$ suit la loi normale $\mathcal{N}(0,\frac{400}{999}+\frac{441}{699})$. On recherche $\lambda \in \mathbf{R}$ tel que la probabilité $p(|\bar{x}_1-\bar{x}_2| \leq \lambda)$ soit égale au seuil de confiance $c=0,95$:

$$p(|\bar{x}_1-\bar{x}_2| \leq \lambda) = 0,95.$$

Le nombre λ solution du problème vérifie

$$\Pi\left(\frac{\lambda}{\sqrt{\frac{400}{999}+\frac{441}{699}}}\right) = 0,975.$$

On trouve $\lambda = 1,99$. Puisque la différence 79 CV$-$78 CV est inférieure à $\lambda = 1,99$ on peut admettre avec un risque de 5% que le nouveau carburant n'affecte pas la puissance des moteurs.

6.2.2 Égalité des fréquences

On souhaite déterminer si deux échantillons de tailles n_1 et n_2 sont issus de populations dans lesquelles un caractère qualitatif à la même fréquence (ou au moins possède des fréquences raisonnablement proches), connaissant les fréquences f_1 et f_2 d'un caractère qualitatif sur les échantillons. On va faire le test de **l'égalité des fréquences**.

Avec les hypothèses sur les tailles de n_1, n_2 des échantillons et N_1 et N_2 des populations (*grandes et très grandes*) on peut admettre que la fréquence F_i du caractère sur un échantillon de taille n_i extrait au hasard de la population i suit la loi normale $\mathcal{N}(p_i,\frac{f_i(1-f_i)}{n_i-1})$ ou p_i est la fréquence du caractère sur la population i. Ici on a estimé la variance de la population i à partir de la variance de l'échantillon connu. Sous l'hypothèse d'égalité des fréquences ($p_1=p_2$) et en supposant les F_i indépendantes, la variable aléatoire F_1-F_2 suit la loi normale $\mathcal{N}(0,\frac{f_1(1-f_1)}{n_1-1}+\frac{f_2(1-f_2)}{n_2-1})$.

On recherche $\lambda \in \mathbf{R}$ tel que la probabilité $p(|F_1-F_2| \leq \lambda)$ soit égale au seuil de confiance $c = 1-\alpha$:

$$p(|F_1-F_2| \leq \lambda) = 1-\alpha = c.$$

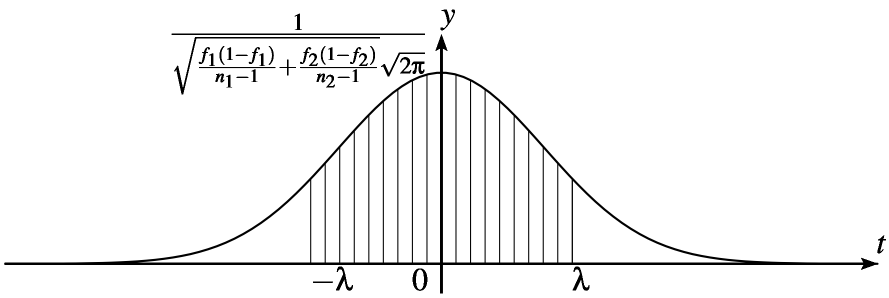


FIGURE 9 – L'aire hachurée représente la probabilité $p(|F_1 - F_2| \leq \lambda) = 1 - \alpha$

Le nombre λ solution du problème vérifie

$$\Pi\left(\frac{\lambda}{\sqrt{\frac{f_1(1-f_1)}{n_1-1} + \frac{f_2(1-f_2)}{n_2-1}}}\right) = 1 - \frac{\alpha}{2}.$$

Si $|f_1 - f_2| \leq \lambda$ alors avec une probabilité de $1-\alpha$ les deux échantillons proviennent de la même population. Sinon, on rejette l'hypothèse d'égalité des fréquences.

Comme pour le test de l'égalité des moyennes, ce test présente le paradoxe suivant. Plus le risque α est petit, plus λ est grand et plus on a de chance d'accepter l'hypothèse à tord.

Exemple 6.3 Deux ateliers fabriquant les mêmes pièces utilisent des méthodes de fabrication différentes, une classique et une autre, innovante et moins bruyante. On souhaite s'assurer que la nouvelle méthode n'engendre pas une perte de qualité. On prélève deux échantillons dans chacun des ateliers. L'échantillon *classique* est composé de 300 pièces dont 30 défectueuses (fréquence 0,10). L'échantillon innovant est composé de 200 pièces dont 24 défectueuses (fréquence 0,12). On peut admettre que la fréquence F_1 de pièces défectueuses dans un échantillon prélevé au hasard dans l'atelier classique est une variable aléatoire qui suit la loi normale $\mathcal{N}(p_1, \frac{0,1\cdot 0,9}{299})$ et la fréquence F_2 de pièces défectueuses dans un échantillon prélevé au hasard dans l'atelier innovant est une variable aléatoire qui suit la loi normale $\mathcal{N}(p_2, \frac{0,12\cdot 0,88}{199})$. Sous l'hypothèse d'égalité des fréquences ($p_1 = p_2$), la variable aléatoire $F_1 - F_2$ suit la loi normale $\mathcal{N}(0, \frac{0,1\cdot 0,9}{299} + \frac{0,12\cdot 0,88}{199})$. On recherche $\lambda \in \mathbf{R}$ tel que la probabilité $p(|F_1 - F_2| \leq \lambda)$ soit égale au seuil de confiance $c = 0,95$:

$$p(|F_1 - F_2| \leq \lambda) = 0,95.$$

Le nombre λ solution du problème vérifie

$$\Pi\left(\frac{\lambda}{\sqrt{\frac{0,1\cdot 0,9}{299} + \frac{0,12\cdot 0,88}{199}}}\right) = 0,975.$$

On trouve $\lambda = 0,066$. La différence $0,12 - 0,10 = 0,02$ est inférieure à $\lambda = 0,066$. On peut admettre avec un risque de 5% que le gain sonore n'affecte pas la qualité de fabrication.

7 Droite de Henry, capabilité, carte de contrôle

Dans cette partie nous présentons trois outils de statistique d'utilisation très courante.

7.1 Droite de Henry

7.1.1 Fonction de répartition d'une loi normale et droite de Henry

On considère une série statistique quantitative $x = (x_1, ..., x_n)$. Cette série représente par exemple les n valeurs que prend un caractère quantitatif pour chaque individu d'un échantillon. On note $\bar{x}$ la moyenne et V la variance. La droite d'équation

$$t = \frac{1}{\sqrt{V}}(x - \bar{x})$$

s'appelle **droite de Henry** associée à la série x.

Pour chaque i on note F_i la fréquence cumulée $F(x_i)$ et on note t_i le nombre tel que $\Pi(t_i) = F_{(x_i)}$. La loi de répartition de la série x est d'autant plus proche d'une loi normale si les points (x_i, t_i) sont sur la droite de Henry. La série statistique x se répartit suivant une loi normale si les points (x_i, t_i) sont sur la droite de Henry.

Si l'échantillon provient d'une population pour laquelle le caractère quantitatif est une variable aléatoire qui suit une loi normale alors, d'après le théorème limite central, les points (x_i, t_i) sont d'autant plus proches de la droite de Henry que l'échantillon est grand.

En conclusion la droite de Henry est un outil de détermination de la normalité.

7.1.2 Le papier gausso-arithmétique

Dans la pratique pour *voir* si la série statistique $x = (x_1, ..., x_n)$ suit une loi normale ou suit une loi qui est proche d'une loi normale on a recours au **papier gausso-arithmétique**. C'est un papier munit d'un système de coordonnées tel que si la série suit une loi normale alors les points (x_i, F_i) sont alignés. Les verticales sont équiréparties. En revanche les hauteurs des lignes horizontales sont déterminées à l'aide d'une table de la loi normale. En conclusion, le papier gausso-arithmétique est une déclinaison graphique de la table de la loi normale.

7.2 Capabilité opérationnelle

Dans cette partie on ne considère que des caractères quantitatifs.

Lorsqu'on produit plusieurs exemplaires d'un même objet on souhaite que les caractéristiques des exemplaires soient conforment aux prescriptions fixées par le concepteur ou l'utilisateur.

Considérons un objet simplissime caractérisé par une seule cote. Les prescriptions vont donner la **cote nominale** (c'est la cote idéale) N, la **cote maximale admissible** T_S et la **cote minimale admissible** T_I. La **cote réelle** x_i d'un exemplaire produit devra être comprise entre T_I et T_S pour que l'exemplaire soit admissible.

Lorsqu'on produit un grand nombre d'exemplaires de cet objet, le théorème limite central permet de supposer que les cotes x_i des exemplaires produits est une variable aléatoire qui suit une loi normale $\mathcal{N}(\mu, V)$ de moyenne μ et d'écart-type $\sigma = \sqrt{V}$. L'écart-type est ici estimé à partir d'un échantillon assez gros (une centaine d'exemplaires).

Un des objectifs de la production est de maximiser la proportion d'exemplaires admissibles : on cheche à tendre vers le 0 *défaut*. Pour mesurer la réalisation de cet objectif on considère deux indices appelés **indice de performance du procédé** et qui sont définis de la façon suivante :

$$Pp = \frac{T_S - T_I}{6\sigma} \text{ et } Ppk = \inf\left(\frac{\mu - T_I}{3\sigma}, \frac{T_S - \mu}{3\sigma}\right).$$

La probabilité qu'un exemplaire soit non admissible est inférieure à $2 - 2\Pi(3Ppk)$ mais supérieure à $2 - 2\Pi(3Pp)$. Si $\mu = \frac{T_I + T_S}{2}$ alors $Pp = Ppk$ et la probabilité qu'un exemplaire soit non admissible est égale à $2 - 2\Pi(3Pp)$.

Les indices Pp et Ppk mesurent la **capabilité opérationnelle** de la production.

Plus les indices Pp et Ppk sont grands, meilleure est la qualité de la production. Dans l'industrie automobile on impose $Pp = 1,66$ et $Ppk = 1,33$: avec ces valeurs la probabilité qu'un exemplaire produit pésente un défaut est d'au plus de 7 sur 100000.

7.3 Carte de contrôle et capabilité du procédé

Il existe de multiples cartes de contrôle. Nous allons nous limiter à la description de deux d'entre elles, celles qui permettent de contrôler la stabilité de la valeur moyenne d'une cote au cours de la production d'un objet et celle qui permet de contrôler la stabilité de l'étendue.

Pour contrôler la variation de la valeur moyenne dans le temps on prélève à intervalles réguliers un nombre fini k (par exemple $k = 20$) de petits échantillons de taille n ($n = 5$ exemplaires successifs de l'objet produit par exemple) et on consigne dans un tableau pour chaque échantillon, les valeurs $x_{i,1},\dots,x_{i,n}$, leur moyenne $\overline{x_i}$, l'étendue R_i et l'écart-type σ_i de l'échantillon. On reporte sur un graphique les points $(i,\overline{x_i})$: ce graphique est la carte de contrôle. On note $\overline{x}$ la moyenne des moyennes $\overline{x_i}$ des échantillons, $\overline{R}$ la moyenne des étendues des échantillons et $\overline{\sigma}$ la moyenne des écarts-types des échantillons. Alors on peut estimer l'écart-type $\hat{\sigma}$ de la production par

$$\hat{\sigma} = \frac{\overline{R}}{D_2} \text{ ou par } \hat{\sigma} = \frac{\overline{\sigma}}{C_4}$$

où les nombres D_2 et C_4 sont fournis par le tableau suivant :

n	2	3	4	5	6	7	8	9	10
C_4	0,8	0,89	0,92	0,94	0,95	0,96	0,97	0,97	0,97
D_2	1,13	1,69	2,06	2,33	2,53	2,7	2,85	2,97	3,08

On peut alors encore définir les indices Cp et Cpk de **capabilité du procédé** en posant

$$Cp = \frac{T_S - T_I}{6\hat{\sigma}} \text{ et } Cpk = \inf\left(\frac{\mu - T_I}{3\hat{\sigma}}, \frac{T_S - \mu}{3\hat{\sigma}}\right).$$

L'industrie impose souvent $1,33$ et $1,66$ comme valeurs minimales de Cpk et Cp.

Puisque le calcul de Cp ou de Cpk est fait à partir d'un écart-type estimé nous obtenons seulement des indices estimés $\hat{C}p$ et $\hat{C}pk$. Il existe des tables qui permettent de garantir des valeurs minimales de Cp et Cpk en fonction de $\hat{C}p$ ou de $\hat{C}pk$ et du nombre de contrôles avec un risque fixé.

On dit que la valeur moyenne de la cote est controlée si $|\overline{x} - \overline{x_i}| \leq A_2\overline{R}$ où le nombre A_2 est fourni par le tableau suivant :

n	2	3	4	5	6	7	8	9	10
A_2	1,88	1,02	0,73	0,58	0,48	0,42	0,37	0,34	0,31

Les coefficients A_2 et D_2 vérifient

$$D_2 = \frac{3}{\sqrt{n}A_2}.$$

On peut aussi contrôler de façon analogue la stabilité de l'étendue. L'étendue est dite controlée si $D_3\overline{R} \leq R_i \leq D_4\overline{R}$ où D_3 et D_4 sont fournis pas le tableau suivant :

n	2	3	4	5	6	7	8	9	10
D_3	0	0	0	0	0	0,076	0,136	0,184	0,223
D_4	3,267	2,575	2,282	2,115	2,004	1,924	1,864	1,816	1,777

Le calcul des constantes C_4, D_2, D_3 et D_4 repose sur les éléments de statistique qui sont hors du cours.

Exemple 7.1 On produit un objet dont l'unique cote doit être comprise entre $T_S = 1,2$ et $T_I = 0,8$ et de cote nominale égale à 1. Pour contrôler la capabilité du procédé on prélève 20 échantillons formés chacun de 5 exemplaires produits successivement. On obtient les relevés suivants :

échantillon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$x_{i,1}$	1	1,03	1,07	0,96	1,01	1,08	1,03	1,02	0,99	1,01	0,99	1,02	0,96	1,08	0,99	1,05	1,02	0,98	1,01	1,08
$x_{i,2}$	0,99	1,06	1,02	1,08	1,02	1,04	1,07	1,05	0,98	1,04	0,95	1,05	1,03	1,01	1	1,03	0,96	1,04	1,03	1,07
$x_{i,3}$	0,98	1,02	0,99	1,06	1,07	0,98	1	0,95	1,08	1,06	1,01	0,99	1,02	1,07	0,98	1,06	1,05	1	0,99	1,01
$x_{i,4}$	1,01	0,99	1,05	0,96	0,95	1,05	0,96	0,99	1,05	1,07	0,96	0,98	0,98	1,06	1,04	0,99	0,99	1,03	1,02	0,96
$x_{i,5}$	1,02	1,01	0,97	1,04	1,03	1,06	0,99	1,04	1,03	1,02	0,98	1,01	1,05	0,98	1,08	1	1,01	1,07	1,04	0,98
$\overline{x_i}$	1	1,02	1,02	1,02	1,02	1,04	1,01	1,01	1,03	1,04	0,98	1,01	1,01	1,04	1,02	1,03	1,01	1,02	1,02	1,02
R_i	0,04	0,07	0,1	0,12	0,12	0,1	0,11	0,1	0,1	0,06	0,06	0,07	0,09	0,1	0,1	0,07	0,09	0,09	0,05	0,12
σ_i	0,02	0,03	0,04	0,06	0,04	0,04	0,04	0,04	0,04	0,03	0,02	0,03	0,04	0,04	0,04	0,03	0,03	0,04	0,02	0,05

Pour faire les cacul de capabilité et pour déterminer si la moyenne et l'étendue sont contrôlées on prend

$$C_4 = 0,94,\ A_2 = 0,58,\ D_3 = 0 \text{ et } D_4 = 2,115.$$

Le calcul donne $\overline{x} = 1,018$, $\overline{R} = 0,088$, $\overline{\sigma} = 0,036$, $\hat{\sigma} = 0,038$ (calculé avec la formule $\hat{\sigma} = \frac{\overline{\sigma}}{C_4}$).

On en déduit que

$$Cp = 1,75 \text{ et } Cpk = 1,6.$$

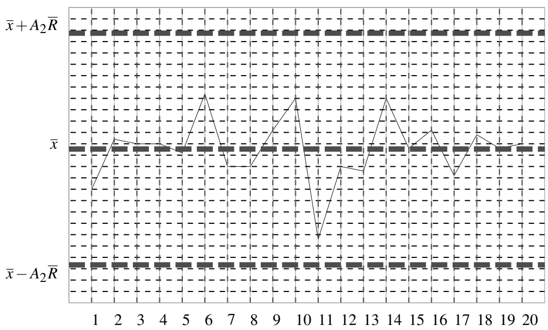
Ces résultats correspondent à une bonne capabilité.

De plus

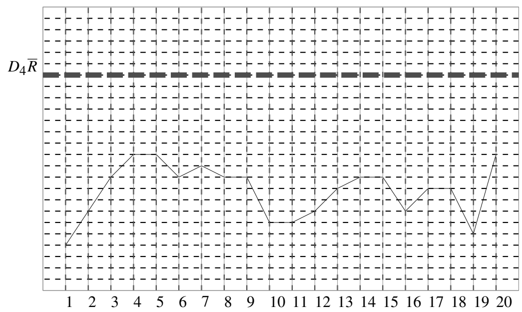
$$|\overline{x_i} - \overline{x}| \leq A_2\overline{R} = 0,05 \text{ et } D_3\overline{R} = 0 \leq R_i \leq D_4\overline{R} = 0,19.$$

Ainsi la moyenne et l'étendue sont controlées

On obtient les cartes de contôle suivantes sur lesquelles figurent les bandes données définies par $|\overline{x_i} - \overline{x}| \leq A_2$ et par $R_i \leq D_4\overline{R}$.



Carte de contrôle de la moyenne



Carte de contôle de l'étendue

Feuille d'exercices 1
Probabilités et statistique
DUT GMP, IUT de Rennes
première année

Exercice 1

Ce tableau donne la répartition suivant les classes d'age des membres d'un club d'échecs.

classe d'âge	$[0,10[$	$[10,20[$	$[20,30[$	$[30,40[$	$[40,50[$	$[50,60[$	$[60,70[$	$[70,80[$	$[80,90]$
effectif	19	37	55	22	16	14	25	10	2

On suppose qu'à l'intérieur de chaque classe la population est répartie uniformément.
1) Dire quel caractère est étudié, donner sa nature, donner un encadrement de l'étendue et calculer l'effectif de la population.
2) Donner sous forme de tableau les fréquences de chaque classe ainsi que les effectifs et fréquences cumulés des classes.
3) Calculer les quartiles et dire quelle est la médiane.
4) Calculer la moyenne, la variance et l'écart-type.
5) Faire une représentation graphique.

Exercice 2

Ce tableau donne la répartition des surfaces (en %) suivant la quantité de matière sèche accumulée par heure à l'hectare des champs de maïs (en $kg/(h*hec)$) d'une exploitation agricole.

mati. sèc./h/hec.	$[0,30[$	$[30,35[$	$[35,40[$	$[40,45[$	$[45,50[$	$[50,55[$	$[55,60[$	$[60,65[$
surface en %	1	5	4	10	19	38	18	5

L'exploitation comprend 127 hectares de maïs et on suppose qu'à l'intérieur de chaque classe la quantité de matière sèche accumulée par heure et à l'hectare varie uniformément.
1) Donner sous forme de tableau les surfaces en hectares de chaque classe ainsi que les surfaces cumulées en hectares et en % pour chaque classe.
2) Calculer les quartiles et dire quelle est la médiane.
3) Calculer la moyenne, la variance et l'écart-type.
4) Faire une représentation graphique.

Exercice 3

Des cylindres fabriqués par une machine devraient avoir un diamètre nominal de 35,50mm avec une tolérance de $\pm 0,25mm$. La production d'une journée est contrôlée. Le résultat du contrôle est résumé par le tableau suivant.

diamètre	$[35:35,1[$	$[35,1:35,2[$	$[35,2:35,3[$	$[35,3:35,4[$	$[35,4:35,5[$	$[35,5:35,6[$	$[35,6:35,7[$	$[35,7:35,8[$	$[35,8:35,9[$	$[35,9:36[$
effectif	11	24	43	77	93	101	71	42	28	10

1) Calculer la moyenne, la médiane, la variance et l'écart-type.
2) Calculer le pourcentage mininum de cylindres satisfaisants
3) On suppose qu'à l'intérieur de chaque classe les diamètres des cylindres se répartissent uniformément. Refaire le calcul de la question précédente sous cette hypothèse.

Feuille d'exercices 2
Probabilités et statistique
DUT GMP, IUT de Rennes
première année

Exercice 1

Finir une course en formule 1 tient au volant (le pilote), au moteur et aux pneus. La probabilité de défaillance du volant (du pilote) est de 3%, celle du moteur de 15% et celle des pneus de 10%. Ces probabiltés sont indépendantes. Quelle est la probabilité de finir une course sans défaillance ?

Exercice 2

Un industriel reçoit des devis de deux marchands de bois. Le premier s'élève à 172000 euros et le second à 180000 euros. L'expérience montre que 10% des planches livrées par le premier marchand est inutilisable et 95% de celles livrées par le second est de qualité satisfaisante.
1) Avec ces éléments, quel fournisseur choisir ?
Pour ne pas être dépendant d'un seul fournisseur, l'entrepreneur décide de confier 30% du marché au premier fournisseur et le reste au second.
2) Combien va coûter le stock total ?
3) Quelle est la probabilité qu'une planche prise dans le dépot provienne du premier fournisseur ?
4) Quelle est la probabilité qu'une planche prise dans le dépot possède un défaut ?
5) L'entrepreneur préfère organiser le recyclage du bois défectueux plutôt que de le stocker inutilement. Quelle est la probabilité de trouver une planche du premier fournisseur dans la benne qui contient les planches défectueuses ?

Exercice 3

On hésite entre deux terrains pour implanter un centre commercial. L'accès du premier aux services de secours est possible par deux routes. L'une est coupée aléatoirement en moyenne une heure par jour de façon intermitente et on estime le risque d'une coupure dans une journée de l'autre qui est très peu fréquentée à 10^{-3}. L'accès du second terrain aux services de secours est aussi possible par deux routes qui présentent des risques journaliers de coupure indépendants de 10^{-2}. En fonction du seul critère d'accès aux secours quel terrain convient le mieux ?

Exercice 4

Considérons le jeu suivant. Il se joue en deux tours et à l'issue de chaque tour un joueur est éliminé du jeu et perd tous ses points, ou il gagne 1 point ou encore il gagne 2 points. En adoptant une stratégie risquée pendant un tour la probabilité de gagner 2 points est de $0,45$, celle de gagner 1 point est de $0,2$ et celle d'être éliminé est de $0,35$. Une stratégie sûre pendant un tour donne la probabilité $0,25$ de gagner 2 points, la probabilité $0,50$ de gagner 1 point et la probabilité $0,25$ d'être éliminé.
1) Quelle est, en fonction de la stratégie, l'espérance de gain de points en un tour ?
2) Quelle est, en fonction de la stratégie, la probabilité de ne pas être éliminé du jeu ?
3) Vérifier qu'à l'issue du jeu un joueur non éliminé peut avoir gagné 2, 3 ou 4 points.
4) Quelle est, en fonction de la stratégie, la probabilité de gagner 2 points à la fin du jeu, 3 points, 4 points ?
5) Quelle est, en fonction de la stratégie, l'espérance de points gagnés dans le jeu ?
6) Quels arguments peuvent défendre la stratégie risquée ?
7) Que penser de la stratégie mixte qui consiste à appliquer la stratégie sûre au premier tour et en cas de succès la stratégie risquée au second.

Feuille d'exercices 3
Probabilités et statistique
DUT GMP, IUT de Rennes
première année

Exercice 1

Calculer $p(X = 3)$ si X suit la binomiale $\mathfrak{B}(4\ ,\ 0,3)$ et $p(Y \leq 2)$ si Y suit la binomiale $\mathfrak{B}(5\ ,\ 0,8)$. Déterminer λ pour que $p(Z \leq \lambda) \leq 0,8 \leq p(Z \leq \lambda + 1)$ si Z suit la binomiale $\mathfrak{B}(5\ ,\ 0,2)$.

Exercice 2

1) Le distributeur de café est déréglé et une fois sur 10 le café est froid. Quelle est la probabilité qu'au plus 10 cafés sur les 100 servis dans une journée soient froids.
2) La technologie d'un réseau informatique garantit la transmission entre deux pôles si au moins deux des cinq lignes qui les relient fonctionnent. Dans une journée, chacune des lignes peut tomber en panne de façon indépendante des autres avec la probabilité $\frac{1}{100}$. Quelles sont les probabilités pour que le réseau ne connaisse pas de dysfonctionnement au cours d'une journée, d'une année, de 10 années ?

Exercice 3

On suppose que X est une variable aléatoire qui suit la loi normale centrée réduite.
1) Déterminer

$$\begin{array}{cccc} p(X < 0,3), & p(X < -1,1), & p(X > 0,4), & p(X > -0,7), \\ p(-0,2 < X < 1,1), & p(-1,3 < X < -1), & p(-0,6 < X < 0,6), & p(0,8 < X < 1,4). \end{array}$$

2) Trouver $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ et λ_5 tels que

$$p(X < \lambda_1) = 0,3\ , \quad p(X < \lambda_2) = 0,8\ , \quad p(X > \lambda_3) = 0,1\ , \quad p(X > \lambda_4) = 0,6\ , \quad p(-\lambda_5 < X < \lambda_5) = 0,8.$$

Exercice 4

On suppose que X est une variable aléatoire qui suit la loi normale $\mathcal{N}(5,9)$.
1) Déterminer

$$p(X < 6), \quad p(X < 1), \quad p(X > 7), \quad p(X > 2), \quad p(7 < X < 9), \quad p(1 < X < 4), \quad p(3 < X < 9).$$

2) Trouver $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ et λ_5 tels que

$$p(X < 5 - \lambda_1) = 0,3\ , \quad p(X < \lambda_2) = 0,8\ , \quad p(X > 5 + \lambda_3) = 0,1\ , \quad p(X > \lambda_4) = 0,6\ , \quad p(5 - \lambda_5 < X < 5 + \lambda_5) = 0,8.$$

Exercice 5

Une usine est équipée de machines qui permettent de produire des pièces aux cotes exactes à 0,05 mm près. Deux équipes se relaient dans la journée pour fabriquer des pièces d'un diamètre nominal de 37 mm. La tolérance demandée est de 0,2 mm (i.e $\pm$0,1 mm). On constate que le diamètre des pièces fabriquées par la première équipe suit la loi normale $\mathcal{N}(37,01\ ,\ 4.10^{-4})$ alors que celui des pièces fabriquées par la seconde équipe suit la loi normale $\mathcal{N}(36,98\ ,\ 25.10^{-4})$.
1) Quel est le pourcentage de pièces conformes aux exigences et produites par la première équipe ?
2) Quel est le pourcentage de pièces conformes aux exigences et produites par la seconde équipe ?
3) Pour améliorer la qualité de la production, vaut-il mieux investir dans des machines plus précises ou dans un stage de formation continue ?
4) Quelle loi normale du type $\mathcal{N}(37\ , v)$ doit suivre la production pour que 95% des pièces soient conformes.

Feuille d'exercices 4
Probabilités et statistique
DUT GMP, IUT de Rennes
première année

Exercice 1

1) Montrer que si X est une variable aléatoire positive d'espérance m alors pour tout $a > 0$ on a

$$p(X \geq a) \leq \frac{m}{a} \text{ (Inégalité de Markov (}XIX^e\text{ siècle)).}$$

2) Majorer la probabilité pour qu'une lettre arrive en au moins 10 jours sachant que le délai d'acheminement moyen est de 1,5 jours.

Exercice 2

Un différent oppose une entreprise et son fournisseur. Un expert examine un échantillon de 100 pièces considérées défaillantes par l'entreprise. La distribution des masses des pièces de l'échantillon a une moyenne de 99,5 g et un écart-type de 0,2 g alors que le cahier des charges stipule que la masse d'une pièce acceptable doit être de 100 g.
1) Donner une estimation de la moyenne des masses des pièces considérées défaillantes par l'entreprise avec un risque de 5%.
2) En prenant la moyenne la plus favorable pour le fournisseur et en considérant que l'écart-type des masses des pièces considérées défaillantes par l'entreprise est de 0,2 g, calculer la proportion de pièces qui ont été considérées à tord comme défaillantes.

Exercice 3

Un atelier produit des cylindres dont l'alésage doit être de 82,7 mm avec une tolérance de 0,01 mm. On prélève un échantillon de 200 cylindres. L'alésage moyen observé est de 82,707. L'écart-type de l'échantillon est de 13.10^{-4} mm.
1) Donner une estimation de l'alésage moyen des cylindres produits avec un risque de 5%.
2) En prenant la moyenne la plus défavorable et en considérant que l'écart-type des alésages des cylindres produits est de 13.10^{-4} mm, calculer la proportion de cylindres conformes.
3) Avec quel risque peut on considérer que 99% de la production est conforme ?

Exercice 4

Une association de riverains milite pour la construction d'un contournement routier de leur village. Ils ont relevé tous les jours ouvrables pendant deux ans le nombre de voitures immatriculées hors département et traversant le village. Cette distribution suit approximativement une loi normale de moyenne 2025 véhicules/jour et d'écart-type 100 véhicules/jour.
1) Quelle est la probabilité pour qu'on compte plus de 1800 voitures un jour ouvrable pris au hasard ?
2) Quelle est la probabilité pour que sur un échantillon de 5 jours ouvrables on compte moins de 10000 voitures ?
3) Le conseil général est prêt à défendre le projet de contournement. Selon un expert le projet se justifie économiquement si le débit moyen journalier est supérieur à 2000. Pour s'en assurer il branche des capteurs sur la route pendant trois semaines. Avec quelle probabilité le verdict des capteurs sera acclamé par le village ?

Feuille d'exercices 5
Probabilités et statistique
DUT GMP, IUT de Rennes
première année

Exercice 1

Un industriel du verre a développé une fibre composite qu'il produit dans deux usines différentes. Sur 150 prélèvements dans la première usine on mesure une densité moyenne de 1,951 g/cm^3 avec un écart-type de 2.10^{-3}. Sur 180 prélèvements dans la seconde usine on mesure une densité moyenne de 1,9495 g/cm^3 avec un écart-type de $2,1.10^{-3}$. Peut on considérer avec 5% de risque que les deux usines produisent une fibre ayant la même densité.

Exercice 2

Un fabricant produit des feuilles d'un alliage d'une épaisseur de 0,95 mm avec une tolérance de $\pm 0,005$. Un client conteste une livraison sous prétexte que si l'épaisseur moyenne des feuilles livrées est de 0,9535, 10% d'entre elles sont non conformes.
1) Calculer l'écart-type de la distribution de feuilles livrées, suivant les données fournies par le client et sous l'hypothèse que celle-ci suit une loi normale.
2) Le fabricant vérifie la qualité du stock d'où est issue la livraison contestée sur un échantillon de 200 feuilles. Il trouve une épaisseur moyenne de 0,952 et un écart-type de 0,001. Ces résultats le font il douter du sérieux de la réclamation du client ?

Exercice 3

Un constructeur propose une garantie de 5 ans au prix de 199 euros pour son produit phare qui vaut 1000 euros. Sous l'hypothèse qu'une réparation revient en moyenne à 220 euros au constructeur et que la période avant le recours au service après-vente suit la loi normale de moyenne 4 ans et d'écart-type 1,5 ans, calculer le bénéfice de cette offre de garantie pour le constructeur.

Exercice 4

Un industriel commercialise sa production sous deux marques, l'une de prestige réservée au premier choix et la seconde pour le second choix. Les 80% d'un premier échantillon de 500 pièces sont destinées à la marque de prestige alors que seulement 75% d'un second échantillon de 400 pièces le sont. Peut-on considérer avec un risque de 5% que la production est d'égale qualité ?

Exercice 5

Un fournisseur d'accès à internet s'est engagé à fournir un débit en réception de 8 Megabit/seconde. Son objectif est d'offrir ce débit minimum à au moins 97,5% des abonnés.
1) En supposant que son réseau est réglé pour fournir un débit moyen théorique de 8,3 Megabit/seconde suivant une loi normale, calculer l'écart-type théorique qui garantit l'objectif.
2) Un service de contrôle procède à un test sur un échantillon de 500 abonnés. Le débit observé vaut en moyenne 8,1 Megabit/seconde avec un écart-type de 0,06 Megabit/seconde. Peut on considérer avec un risque de 5% que le fournisseur procure un débit moyen théorique de 8,3 Megabit/seconde ?
3) À partir de la moyenne et de l'écart-type donnés par le service de contrôle calculer le pourcentage de clients qui ne bénéficient pas du débit de 8 Megabit/seconde.

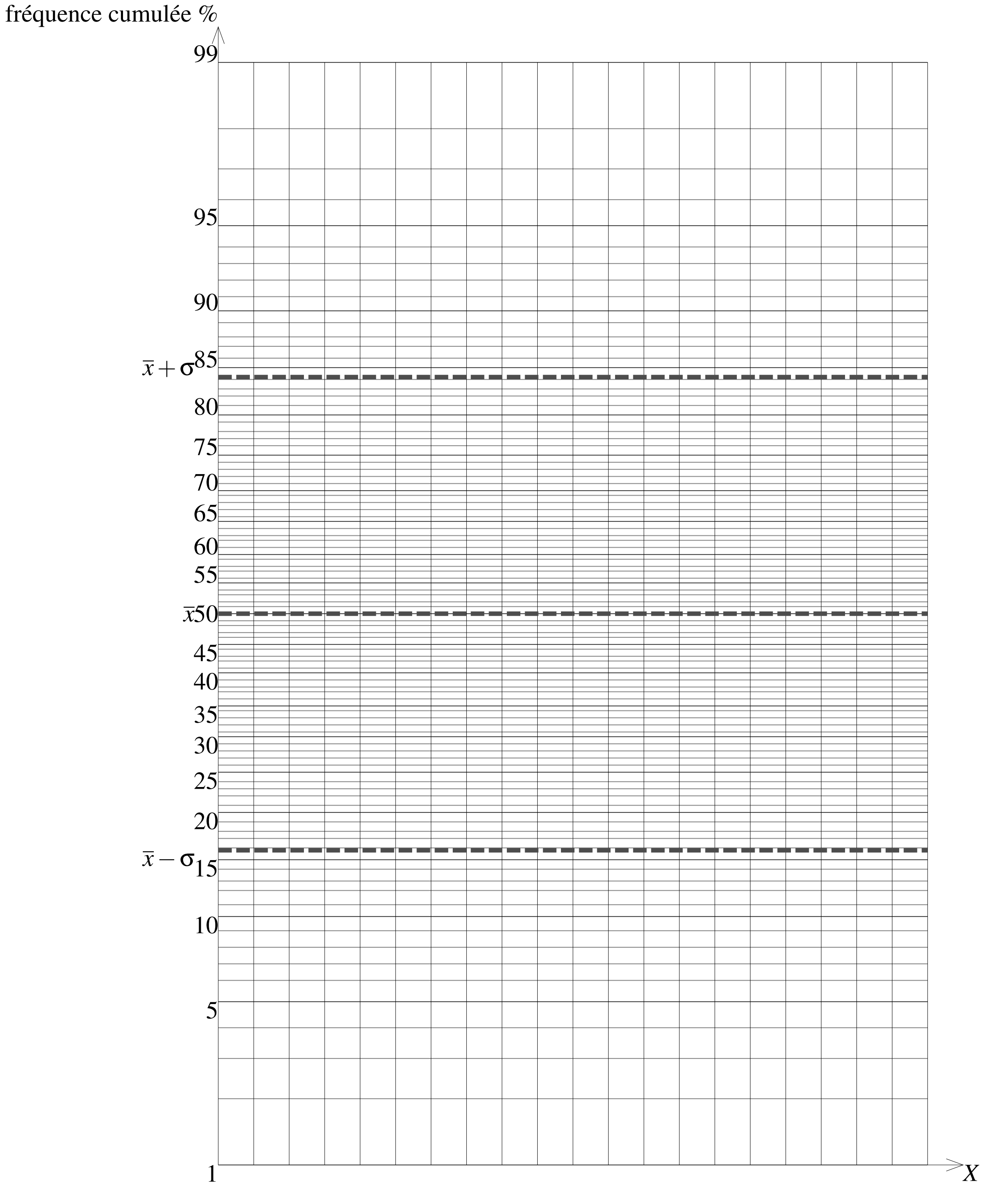
Table de la loi normale

$$\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$$

	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7793	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8906	0,8925	0,8943	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,98260	0,983	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

3	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4	4,5
0,99865	0,99903	0,99931	0,99952	0,99966	0,99977	0,99984	0,999928	0,999968	0,9999966

Papier gausso-arithmétique



Références

Ce texte a été rédigé à l'aide des sources suivantes

— Notes du cours 2005-2006 de Probabilités et Statistique, DUT GMP 1ère année IUT de Rennes, par C. Gentil

— Fascicule de cours et d'exercices de Probabilités et Statistiques, Deug sciences de la vie 1ère année de Rennes 1, par J. Deshayes, J. Memin, M. Peigne, A. Raugi

— Polycopié de Probabilités, second cycle de Mathématiques de Paris VI, par R. Priouret
— Exercices de probabilités, Collection DIA, diffusion Belin, par M. Cotrell, Ch. Duhamel, V. Genon-Catalot
— http://fr.wikipedia.org
— http://www-history.mcs.st-and.ac.uk/history/
— La toile a fourni d'autres sources plus éphémères qui ne sont pal accessibles
— Les conseils éclairés de Max Bauer, Philippe Briand, Hélène Guerin, Florent Malrieu, Daniel Panazzolo et Albert Raugi

Exemples de sujets

On trouvera dans les pages suivantes des sujets typiques de devoirs surveillés.

Nom et Prénom :
Groupe :

Le candidat composera sur cette feuille. Les calculatrices et tous les documents sont autorisés. L'usage du téléphone portable est interdit. La communication avec l'extérieur ou avec les autres candidats est interdite pendant l'épreuve.

Exercice 1

On considère un paquet de copies d'examen réparties de la façon suivante en fonction de leur longueur en pages.

pages	1	2	3	4	5	6
nb. de copies	3	5	36	20	18	18

Quel est le nombre de copie?	100
Quelle est la longueur moyenne d'une copie?	3.99
Quelle est la longueur médiane du paquet de copies?	4
Quelle est la variance de ce paquet de copies?	1.73

Voici la répartition des étudiants reçus et ajournés en fonction de la longueur de leur copie.

pages	1	2	3	4	5	6
reçus	1	2	24	18	12	7
ajournés	2	3	12	2	6	11

Quel est le taux de succès à cet examen ?	64%

Probabilité d'être reçu en fonction de la longueur de la copie.

pages	1	2	3	4	5	6
probabilité d'être reçu	33%	40%	67%	90%	67%	39%

On prend une copie au hasard parmi les copies de 4 pages et plus. Quelle est la probabilité qu'elle appartienne à un étudiant reçu à l'examen ?	66%
On prend une copie au hasard parmi celles des étudiants reçus à l'examen. Quelle est la probabilité qu'elle fasse 4 pages ou plus ?	58%
Est-il vrai qu'on a plus de chance d'être reçu à l'examen si on a fait une copie de 4 pages ou plus ?	OUI

Exercice 2 (dans cet exercice les probabilités sont données en % avec 2 chiffres après la virgule)

Soit X une variable aléatoire qui suit la loi binomiale B(50;0,1).

Calculer la probabilité p(X=0).	0.52%
Calculer la probabilité p(X=1).	2.86%
Calculer la probabilité p(X=2).	7.79%

Soit la variable aléatoire T qui prend les valeurs 10, 4, 1, -1 et qui vérifie $p(T=10)=p(X=0)$, $p(T=4)=p(X=1)$, $p(T=1)=p(X=2)$, $p(T=-1)=p(X>2)$.

Que vaut l'espérance E(T) : -64,43 ou -0,6443 ou 0,6443 ?

-0.6443

Calculer la variance V(T).

1.5246

A l'occasion d'un vide-greniers une loterie est organisée à partir des billets d'entrée. Le prix du billet est d'1 euro et les gagnants ont leur billet remboursé plus un chèque d'un montant de 10 euros, 4 euros ou 1 euro. La variable aléatoire T précédente modélise la probabilité de gagner. On suppose que 1000 visiteurs viennent au vide-grenier.

Combien le vide-grenier va certainement rapporter (déduction faite des remboursements et des chèques faits aux gagnants de la loterie) ?

644.3 euros

Comment expliquer ce calcul avec la loi faible des grands nombres?

La loi faible des grands nombres assure que si on répète de façon indépendante une variable aléatoire un grand nombre (N) de fois alors la somme des gains (et des pertes) Σ peut être approximée par l'espérance de cette variable aléatoire multipliée par ce grand nombre. Dans le cas précédent l'espérance de gain pour les organisateurs est de 0.6443 euros par visiteur et il y a 1000 visiteurs. Par conséquent le vide-grenier va certainement rapporter 644.3 euros. Plus précisément si t>0 on a d'après la loi faible des grands nombres p(|Σ-644.3|>t)<1542/t^2. Par exemple avec t=100 on obtient que p(|Σ-644.3|>100)<15.42%.

A l'aide du théorème central limite calculer A tel que le gain obtenu par les organisateurs soit avec 95 chances sur 100 dans un intervalle de largeur 2A et de centre le gain qui vient d'être calculé.

76.6 euros

Expliquer le raisonnement.

On déduit du théorème limite central que le gain pour les organisateurs peut être modélisé par une variable aléatoire Y qui suit une loi normale d'espérance 644.3 (1000 fois l'espérance de -T) et de variance 1524.6 (1000 fois la variance de -T). Par conséquent, en utilisant la table de la loi normale centrée réduite N(0;1) et les formules de passage d'une loi normale à N(0;1), on peut estimer avec un risque de 5% d'erreur que le gain pour les organisateur est de 644.3 euros +/- 76.6 euros.

Exercice 3

On considère une variable aléatoire X qui suit une loi normale.

On suppose que X suit la loi normale N(4;9). Calculer P(X<2).	25.25%
On suppose que X suit la loi normale N(5;4). Calculer P(X>2).	93.32%
On suppose que X suit la loi normale N(2;25). Calculer P(1<X<6).	36.74%
On suppose que X suit la loi normale N(8;49). Calculer P(5<X<11).	33.18%
On suppose que X suit la loi normale N(6;36). Calculer P(X<4).	36.94%

On suppose que X suit la loi normale N(4;25). Trouver A tel que P(X<A)=0,8.

8.21

On suppose que X suit la loi normale N(3;4). Trouver A tel P(X<A)=0,2.	1.32
On suppose que X suit la loi normale N(3;36). Trouver A tel que P(3-A<X<3+A)=0,6.	5.05
On suppose que X suit la loi normale N(4;49). Trouver A tel que P(A<X)=0,3.	7.67
On suppose que X suit la loi normale N(49;4). Trouver A tel P(X<A)=0,7.	50.05

Exercice 4

On prélève 226 pommes de la réserve d'un cidrier. La masse moyenne des pièces de l'échantillon est de 115g et l'écart-type de l'échantillon est de 15g.

Donner A tel que la masse moyenne des pommes de la réserve est dans l'intervalle de la forme]115-A,115+A[avec une probabilité de 50%.

0.674

Expliquer le raisonnement.

On estime la variance V de la population égale à n/(n-1) v ou v est la variance de l'échantillon (et n son effectif) et le théorème de la limite central nous dit que la moyenne des masses sur un échantillon de 226 pommes est une variable aléatoire X qui suit une loi normale centrée en M, la masse moyenne pour la population, et de variance V/(n-1)=v/n. Le nombre A recherché est donc tel que la probabilité que |X-M|<A est égal à 50%. Or d'après ce qui précède X-M suit une loi normale centrée de variance 15*15/225 c'est à dire de variance 1. Avec ces données numériques, on obtient A=Π(0.75)=0.674.

Combien de pommes de l'échantillon ont une masse supérieure à 105g?

168

Pourquoi?

D'après le théorème de la limite central on peut supposer que la répartition des pommes de l'échantillon qui est assez grand suit une loi normale X de moyenne 115 g et d'écart type 15 g (de variance 225 g2). Par conséquent, dans l'échantillon, la proportion de pommes de masse supérieure à 105g est égale à la probabilité p(X>105) qui est égale à 1-Π((105-115)/15) ou encore à Π(10/15) qui se calcule à partir des tables de la loi normale : Π(10/15)=0.747=74.7%. Puisque l'échantillon compte 226 pommes on en déduit que 0.747*226 c'est à dire au moins 168 pèsent 105 grammes.

Exercice 5

On veut comparer les consommateurs de 2 groupes de distribution nommés Super A et Hyper B. Les caractères étudiés sont le genre des consommateurs (femme ou homme) et la durée des achats. Les enquêteurs dispose d'un échantillon de 401 clients de Super A dont 195 hommes et d'un échantillon de 401 clients de Hyper B dont 195 femmes. On suppose que les 401 clients de Super A passent en moyenne 60 minutes au super-marché (avec un écart-type de 10 minutes) et que les 401 clients de Hyper B passent en moyenne 65 minutes au super-marché (avec un écart-type de 15 minutes).

Le genre est un caractère... qualitatif

Le temps passé dans un magasin est un caractère... quantitatif

Peut on considérer avec un risque d'au plus 5% que les clientèles des deux groupes de distribution sont indiscernables par leur genre?

OUI

Justifier la réponse.

On note fA=206/401 et fB=195/401 la proportion de femmes parmi les clients des échantillons relatifs à Super A et Hyper B. Si les clientèles sont indiscernables par leurs genres alors la différence fA-fB de ces proportions suit une loi normale centrée et de variance 2*(206/401)x(195/401)/400=0,00125. Or si une variable aléatoire X suit une telle loi alors la probabilité que X soit dans]-6.9%, 6.9%[est de 5%. Puisque c'est le cas de fA-fB qui vaut 2.8% on peut considérer que les clientèles des deux groupes de distribution sont indiscernables en genre (avec un risque de 5%).

Peut on considérer avec un risque d'au plus 5% que les clientèles des deux groupes de distribution sont indiscernables par la durée de leurs achats? **NON**

Justifier la réponse.

On pose T1=60 mn et T2=65 mn. Si c'est le cas alors la différence T2-T1 des durées moyennes suit une loi normale centrée et de variance 10*10/400+15*15/400=0.81. Or si une variable aléatoire X suit une telle loi alors la probabilité que X soit dans l'intervalle]-1.77 mn, 1.77 mn[est de 95%. Puisque ce n'est pas le cas de T2-T1 qui vaut 5 mn on peut considérer que la durée des achats permet de distinguer les clientèles des deux groupes de distribution.

Exercice 6

On dispose de 17 cartes. Un X est dessiné sur 13 d'entre elles et un O sur les 4 autres. Elles sont alignées sur une table de la façon suivante, de la première à la dernière : XXXOOXOXXXOXXXXXX. On tire au hasard une des seize premières cartes. Calculer la probabilité p1 que ce soit un X et montrer que l'évènement tomber sur un X (parmi les seize premières cartes) est indépendant de l'évènement tomber sur une carte (parmi les seize premières cartes) qui est suivie d'un X.

L'univers est formé des 16 premières cartes. Il comporte 4 O et 12 X donc la probabilité p1 de tomber sur X est de 12/16=0.75. Parmi les 16 premières cartes on en trouve 12 suivies d'un X qui se répartissent en 3 O suivis d'un X et 9 X suivis d'un X. La probabilité p2 de tomber sur X sachant que la carte (parmi les 16 premières) est suivie d'un X est donc de 9/12 c'est à dire p2=0.75, la probabilité p3 d'être suivie d'un X est de 12/16 et la probabilité p4 d'être un X qui est suivi d'un X est de 9/16. Puisque p1=p2=p3/p4, tomber sur un X (parmi les 16 premières cartes) est indépendant d'être une carte suivie d'un X.

Exercice 7

On étudie une variable aléatoire à valeurs dans [0,π/2] et telle que telle que p(0<X<a)=sin(a). Calculer la densité f(x) de X en x dans]0,1[(c'est à dire la dérivée de la fonction de répartition), l'espérance E(X) et la variance V(X).

f(x), x dans]0,1[	cos(x)	E(X)	0.57	V(X)	0.14

Expliquer rapidement les calculs qui mènent aux résultats.

La densité f est la dérivée de la fonction de répartition: c'est donc cos.
L'espérance est $E(X) = \int_0^{\pi/2} x\cos(x)dx$. Par intégration par parties on a
$E(X) = [x\sin(x)]_0^{\pi/2} - \int_0^{\pi/2}\sin(x)dx = \pi/2 - 1$.
La variance est $V(X) = \int_0^{\pi/2} x^2\cos(x)dx - E(X)^2$. En intégrant par parties 2 fois et en utilisant le calcul de $E(X)$ on a

$$\begin{aligned}
V(X) &= [x^2\sin(x)]_0^{\pi/2} - \int_0^{\pi/2} 2x\sin(x)dx - E(X)^2, \\
&= [x^2\sin(x)]_0^{\pi/2} + [2x\cos(x)]_0^{\pi/2} - \int_0^{\pi/2} 2\cos(x)dx - E(X)^2, \\
&= (\pi/2)^2 - 2 - (\pi/2-1)^2, \\
&= \pi - 3.
\end{aligned}$$

Nom et Prénom :
Groupe :

Le candidat composera sur cette feuille. Les calculatrices et tous les documents sont autorisés. L'usage du téléphone portable est interdit. La communication avec l'extérieur ou avec les autres candidats est interdite pendant l'épreuve.

Exercice 1

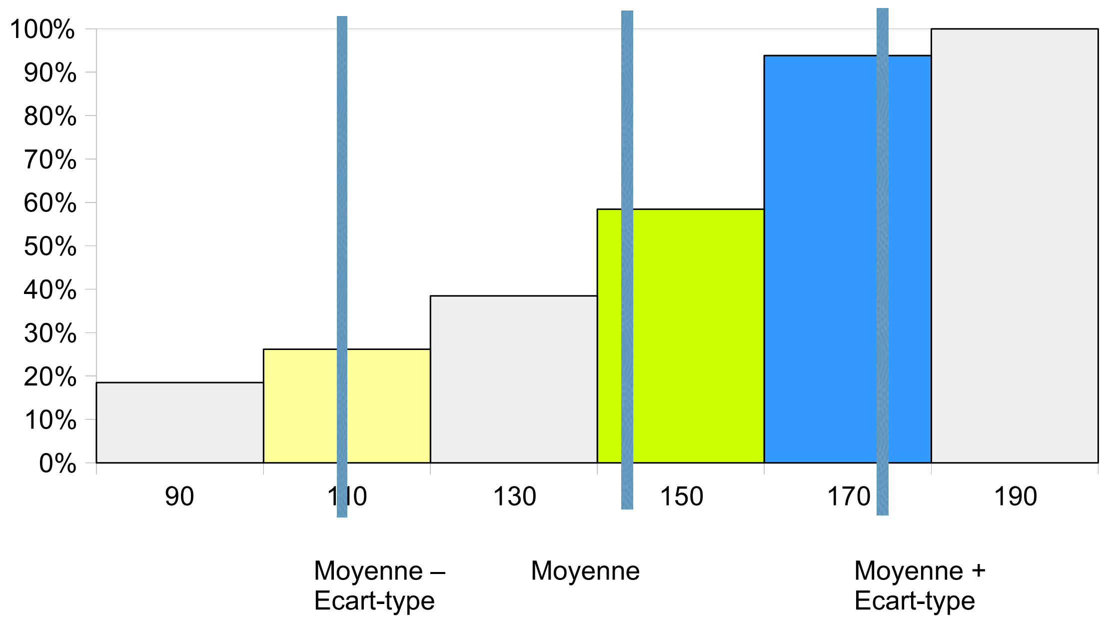
On considère la série quantitative suivante. La population est répartie par classe et à l'intérieur de chaque classe les individus se répartissent uniformément en fonction de la valeur du paramètre quantitatif étudié.

Compléter le tableau suivant et calculer, l'effectif total, la moyenne, la variance et la classe qui contient la valeur médiane.

classe	[80,100[	[100,120[	[120,140[	[140,160[	[160,180[	[180,200[
effectif (ni)	12	5	8	13	23	4
ci (milieu de la classe)	90	110	130	150	170	190
ni.ci	1080	550	1040	1950	3910	760
ni.ci^2	97200	60500	135200	292500	664700	144400

effectif total	moyenne	variance	classe du 1er quartile	classe de la médiane	classe du 3e quartile
65	142,92	1026,84	[100,120[	[140,160[	[160,180[

Faire un histogramme des fréquences cumulées sur lequel apparaîtront les éléments caractéristiques calculés.



Exercice 2

On considère une variable aléatoire X qui suit une loi binomiale de paramètres (n,p). Calculer p(X=k) en fonction des valeurs suivantes de n,p et k.

n	3	5	8	10	12	15
p	0,1	0,2	0,5	0,8	0,7	0,5
k	2	3	4	3	7	5
p(X=k)	2,70E-002	5,12E-002	2,73E-001	7,86E-004	1,58E-001	9,16E-002

Exercice 3

L'INSEE dans son étude annuelle du niveau de vie, regroupe les personnes en dix classes de même effectifs ordonnée en fonction du revenu. Voici à huit ans d'intervalle les résultats obtenus.

année	revenu moyen groupe 1	revenu moyen groupe 2	revenu moyen groupe 3	revenu moyen groupe 4	revenu moyen groupe 5	revenu moyen groupe 6	revenu moyen groupe 7	revenu moyen groupe 8	revenu moyen groupe 9	revenu moyen groupe 10
2003	8230	11550	13730	15630	17500	19540	21950	25170	30330	49890
2011	8000	11890	14280	16370	18460	20690	23320	26880	32710	58700

Compléter le tableau suivant (choisir la bonne réponse ou compléter la case vide).

Entre 2003 et 2011 les 10% les plus riches ont vu leurs revenus augmenter de	3,6%	12,3%	**17,7%**	autre
Entre 2003 et 2011 les 10% les plus pauvres ont vu leurs revenus baisser de	0,5%	**2,8%**	3,9%	autre
Le revenu moyen en 2003 est de	19540	**21352**	17500	autre
Le revenu moyen en 2011 est de	**23130**	18460	20690	autre
Quel est le pourcentage de variation du revenu moyen en huit ans ?	8,33%			
Approximation du revenu médian 2003 (faire la moyenne de deux valeurs bien choisies)	18520			
Approximation du revenu médian 2011 (faire la moyenne de deux valeurs bien choisies)	19575			
Quel est le pourcentage de variation du revenu médian en huit ans ?	5,70%			

Exercice 4

On considère une variable aléatoire X qui suit une loi normale centrée réduite.

A l'aide de la table de la loi normale ou avec la calculatrice calculer p(X<k).

k	1	1,5	2,5	-1	-2	4
p(X<k)	0,84	0,93	0,99	0,16	0,02	1

A l'aide de la table de la loi normale ou avec la calculatrice rechercher k connaissant p(X<k) ou p(X>k).

p(X<k)	0,3	0,8
k	-0,52	0,84

p(X>k)	0,4	0,1
k	0,25	1,28

Exercice 5

Tableau de répartition de clés et de cadenas dans une boîte

clé simple	clé de sécurité	cadenas simple	cadenas de sécurité
7	18	10	15

Calculer le nombre de pièces.	50

Calculer les probabilités suivantes (en %, avec deux chiffres).

On prend une pièce au hasard. Probabilité pour que ce soit une clé.	50%
On prend une pièce au hasard. Probabilité pour que ce soit un pièce de sécurité.	66%
On prend une pièce au hasard. Probabilité pour que ce soit une clé de sécurité.	36%
On prend une pièce simple au hasard. Probabilité pour que ce soit une clé.	41%
On prend une clé au hasard. Probabilité pour que ce soit une pièce simple.	28%
On prend une pièce de sécurité au hasard. Probabilité pour que ce soit un cadenas.	45%

Changer le nombre de clés simples pour que « prendre une pièce simple » et « prendre une clé » soient indépendants.	12

Exercice 6

Dans un jeu on avance son pion de 2 si le dé lancé tombe sur 6 et on avance de 1 sinon.

Probabilité d'avancer de 3 en 3 coups	57,87%	Probabilité d'avancer de 4 en 3 coups	34,72%
Probabilité d'avancer de 5 en 3 coups	6,94%	Probabilité d'avancer de 6 en 3 coups	0,46%

Dpt GMP - IUT de Rennes
Université de Rennes 1

Nom et Prénom :
Groupe :

Le candidat composera sur cette feuille. Les calculatrices et tous les documents sont autorisés. L'usage du téléphone portable est interdit. La communication avec l'extérieur ou avec les autres candidats est interdite pendant l'épreuve.

Exercice 1

On considère un stock formé de pièces réparties de la façon suivante en fonction de leur prix.

prix (euros)	10	15	25	40	70	100
nb.de pièces	10	18	24	5	11	32

Quel est le nombre de pièces?	100
Quelle est le prix moyen d'une pièce?	51.4
Quelle est le prix médian?	25
Quelle est la variance de ce stock de pièces?	1377.54

Voici la répartition de la couleur (rouge ou noire) des pièces en fonction de leur prix.

prix (euros)	10	15	25	40	70	100
noire	1	2	14	2	5	7
rouge	9	16	10	3	6	25

Quelle est la proportion de pièces noires ?	31%

Probabilité d'être rouge en fonction du prix.

prix (euros)	10	15	25	40	70	100
probabilité d'être rouge	90%	89%	42%	60%	55%	78%

On prend une pièce au hasard parmi les pièces de 30 euros ou plus. Quelle est la probabilité qu'elle soit rouge ?	71%
On prend une pièce au hasard parmi les pièces rouges. Quelle est la probabilité que son prix soit supérieur ou égal à 30 euros ?	49%
Est-il vrai qu'une pièce a plus de chance d'être rouge si son prix est supérieur ou égal à 30 euros ?	OUI

Exercice 2

Calculer la probabilité p(X>=2) si X suit la loi binomiale B(4;0,3).	34.83%
Calculer la probabilité p(X=3) si X suit la loi binomiale B(40;0,2).	2.05%
Calculer la probabilité p(X=0) si X suit la loi binomiale B(90;0,01).	40.47%
Calculer la probabilité p(Y=2) si Y suit la loi binomiale B(200;0,01).	27.20%

Calculer la probabilité p(Y<=3) si Y suit la loi binomiale B(150;0,02).	64.72%
Calculer la probabilité p(Y<=499) si Y suit la loi binomiale B(999;0,5).	50.00%

Exercice 3

Les deux tableaux suivants indiquent pour deux pays les taux de chômage parmi les actifs par tranche d'âge ainsi que l'effectif d'actifs de chaque tranche d'âge.

Pays A

tranche d'âge	[20,30[	[30,40[	[40,50[	[50,60[	[60,70[
tx de chômage	25.00%	6.00%	5.00%	20.00%	8.00%
effectif d'actifs	300000	400000	500000	300000	100000

Pays B

tranche d'âge	[20,30[	[30,40[	[40,50[	[50,60[	[60,70[
tx de chômage	24.00%	5.00%	4.00%	19.00%	7.00%
effectif d'actifs	410000	300000	300000	400000	50000

Calculer les taux de chômage de chacun de ces deux pays.

Taux de chômage moyen du pays A	12.00%
Taux de chômage moyen du pays B	14.03%

Expliquer pourquoi, alors que pour chaque classe le taux de chômage du pays A est supérieur à celui du pays B on observe dans le pays A un taux de chômage moyen inférieur à celui du pays B.

Les classes à taux de chômage élevé ([20,30[et [50,60[) représentent 49% des actifs du pays B et 44% des actifs du pays A. C'est l'inverse pour les trois autres classes à taux de chômage bas qui représentent 56% des actifs de A et 51% de ceux de B. Cette surreprésentation dans le pays B des classes à taux de chômage élevé (et l'inverse dans A) compense le fait que pour chaque classe le taux de chômage est 1% inférieur dans le pays B à celui de la classe correspondante du pays A.

Exercice 4

On considère une variable aléatoire X qui suit une loi normale.

On suppose que X suit la loi normale N(3;9). Calculer P(X<1).	25.25%
On suppose que X suit la loi normale N(5;4). Calculer P(X>3).	84.13%
On suppose que X suit la loi normale N(1;25). Calculer P(1<X<7).	38.49%
On suppose que X suit la loi normale N(4;49). Calculer P(1<X<6).	27.83%
On suppose que X suit la loi normale N(5;9). Calculer P(X<3).	25.25%

On suppose que X suit la loi normale N(3;9). Trouver A tel que P(X<A)=0,9.	6.84
On suppose que X suit la loi normale N(5;4). Trouver A tel P(X<A)=0,3.	3.95

On suppose que X suit la loi normale N(1;25). Trouver A tel que P(1-A<X<1+A)=0,8.	6.41
On suppose que X suit la loi normale N(4;49). Trouver A tel que P(A<X)=0,2.	9.89
On suppose que X suit la loi normale N(5;9). Trouver A tel P(X<5-A)=0,3.	1.57

Exercice 5

On prélève 122 pièces d'un container. La masse moyenne des pièces de l'échantillon est de 107g et l'écart-type de l'échantillon est de 2g.

Donner A tel que la masse moyenne des pièces du container est dans l'intervalle de la forme]107-A,107+A[avec une probabilité de 5% d'erreur.

0.356

Expliquer le raisonnement.

Trouver A revient à faire une estimation de la moyenne à 95% (5% d'erreur). Puisque l'écart-type de l'échantillon de 122 pièces est de 2g et que la moyenne sur l'échantillon est 107g il suffit de considérer une variable aléatoire X qui suit une loi normale N(107, (2/11)^2) [puisque 11*11=122-1] et de chercher A tel p(107-A<X<107+A)=0,95 c'est à dire p(X<107+A)=0,975. On considère alors Y qui suit N(0,1). On a p(X<107+A)=0,975=p(Y<B) avec A=B*2/11. On trouve B=1,96 et A=0,356.

Combien de pièces de l'échantillon ont une masse supérieure à 105g?

102

Pourquoi?

On suppose que les 122 pièces de l'échantillon se répartissent selon la loi normale de moyenne 107 et d'écart-type 2. On considère donc une variable aléatoire X qui sui une telle loi et une variable aléatoire Y qui suit la loi normale N(0;1). On a p(X>105)=p(Y>(105-107)/2)=p(Y>-1)=p(Y<1)=0,84. Cette probabilité correspond à la proportion de pièces de l'échantillon de masse supérieure à 105g. On trouve donc 122*0,84 c'est à dire environ 102 pièces de masse supérieure à105g.

Exercice 6

On compare l'autonomie de deux vélos « électriques ». On dispose de 101 exemplaires du premier modèle. Leur autonomie moyenne est de 85km avec un écart-type de 5km. L'autonomie moyenne de 121 exemplaires du second modèle est de 98km avec un écart-type de 10km.

Peut on considérer avec un risque d'au plus 5% que les deux modèles de vélos « électriques » peuvent être distingués par leur autonomie?

OUI

Justifier la réponse.

On pose m1=85km et m2=98km. Si l'autonomie ne permet de distinguer alors la différence m2-m1 des distances moyennes suit une loi normale centrée et de variance 5*5/100+10*10/120=1,08. Or si une variable aléatoire X suit une telle loi alors la probabilité que X soit dans l'intervalle]-2,04km, 2,04km[est de 95%. Puisque ce n'est pas le cas de m2-m1 qui vaut 13km on peut considérer que l'autonomie permet de distinguer avec un risque de 5% les deux modèles de vélos électriques.

Exercice 7

Dans une région la pluie obéit à certaines règles probabilistes. S'il pleut un jour la probabilité qu'il pleuve le lendemain est de 20%. S'il ne pleut pas un jour la probabilité qu'il pleuve le lendemain est de 50%. Quelle est la probabilité qu'il pleuve un jour quelconque (entourer la bonne réponse)?

40.00%	50.00%	60.00%	70.00%	**autre**

Justifier la réponse.

On note p la probabilité qu'il pleuve et q celle qu'il ne pleuve pas un jour quelconque. On a p,q dans $[0,1]$, $p+q=1$ (événements complémentaires) et $p=20\%.p+50\%.q$. Donc $q=(1-p)$ et $p=20\%.p+50\%-50\%.p$. Par conséquent $(1+50\%-20\%).p=50\%$ c'est à dire $1{,}3.p=0{,}5$ et donc $p=0{,}5/1{,}3=38{,}5\%$. Aucune des valeurs proposées ne correspond donc à la probabilité de 38,5% qu'il pleuve.

Exercice 8

On prélève 100 ampoules produites par une usine et on observe que la durée de vie moyenne des ampoules de cet échantillon est de 1020 heures avec un écart-type de 15 heures. Donner une estimation à 5% près de la durée de vie moyenne de la production de l'usine (expliquer la réponse).

On va faire une estimation en moyenne avec un risque de 5%. Puisque l'écart-type de l'échantillon de 100 ampoules est de 15h et que la moyenne sur l'échantillon est 1020h il suffit de considérer une variable aléatoire X qui suit une loi normale $N(1020, 225/99)$ et de chercher A tel $p(1020-A<X<1020+A)=0{,}95$ c'est à dire $p(X<1020+A)=0{,}975$. On considère alors Y qui suit une loi normale centrée réduite. On a $p(X<1020+A)=0{,}975=p(Y<B)=$ avec $A=B*15/\text{racine}(99)$. On trouve $B=1{,}96$ et $A=2{,}95$. La moyenne de la production a donc 95% de chance d'être dans l'intervalle $]1020-2{,}95;1020+2{,}95[$.

On suppose que la durée de vie des ampoules de l'échantillon suit la loi normale $N(1020;225)$. Quel est le nombre d'ampoules de cet échantillon qui cassent en moins de 1000 heures (détailler les calcul)?

On note X une variable aléatoire qui suit $N(1020;225)$. La proportion d'ampoules de l'échantillon qui durent moins de 1000h est $p=p(X<1000)$. On considère une variable aléatoire Y qui suit la loi normale $N(0;1)$. On a $p(X<1000)=p(Y<(1000-1022)/15)=p(Y<-1{,}47)=1-p(Y<1{,}47)=0{,}071$. On trouve donc $100*0{,}071$ c'est à dire environ 7 ampoules qui cassent en moins de 100 heures.

Exercice 9

On place un capital initial de K0=100000 euros. Le placement dure deux ans. Au début du placement la banque indique que ce capital pourra augmenter annuellement suivant certaines règles probabilistes. Après un an le capital sera devenu K1 et un an encore après il devient K2 et la probabilité que K1=K0 est de 50%, la probabilité que K1=K0+10% est de 50%, la probabilité que K2=K1 est de 50% et la probabilité que K2=K1+10% est de 50%. De plus l'augmentation éventuelle de K1 à K2 est indépendante de l'augmentation éventuelle de K0 à K1.

Quelles sont les deux valeurs possibles de K1 ?	100000	110000	
Quelles sont les trois valeurs possibles x1<x2<x3 de K2 ?	X1=100000	X2=110000	X3=121000
Calculer les probabilités p1=p(K2=x1), p2=p(K2=x2) et p3=p(K2=x3).	p1=25%	p2=50%	p3=25%
Calculer le capital moyen E et sa variance V au bout de deux ans.	E=110250	V=55187500	

Nom et Prénom :
Groupe :

Le candidat composera sur cette feuille. Les calculatrices et tous les documents sont autorisés. L'usage du téléphone portable est interdit. La communication avec l'extérieur ou avec les autres candidats est interdite pendant l'épreuve.

Exercice 1

On considère un stock formé de pièces réparties de la façon suivante en fonction de leur prix.

prix (euros)	8	13	17	19	24	30
nb.de pièces	12	7	13	11	11	15

Quel est le nombre de pièces?	69
Quelle est le prix moyen d'une pièce?	19.29
Quelle est la classe médiane?	19
Quelle est la variance de ce stock de pièces?	55.66

Voici la répartition de la couleur (rouge ou noire) des pièces en fonction de leur prix.

prix (euros)	8	13	17	19	24	30
noire	1	2	2	3	8	11
rouge	11	5	11	8	3	4

Quelle est la proportion de pièces noires (en % sans chiffre après la virgule) ?	39%

Probabilité d'être rouge en fonction du prix (en % sans chiffre après la virgule).

prix (euros)	8	13	17	19	24	30
probabilité d'être rouge	92%	71%	85%	73%	27%	27%

On prend une pièce au hasard parmi les pièces de 15 euros ou plus. Quelle est la probabilité qu'elle soit rouge (en % sans chiffre après la virgule) ?	52%
On prend une pièce au hasard parmi les pièces rouges. Quelle est la probabilité que son prix soit supérieur ou égal à 15 euros (en % sans chiffre après la virgule) ?	62%
Est-il vrai qu'une pièce a plus de chance d'être rouge si son prix est supérieur ou égal à 15 euros que si elle est prise au hasard dans l'ensemble des pièces ?	NON

Exercice 2 Les probabilités demandées sont données en % sans chiffre après la virgule.

Calculer la probabilité p(X>=2) si X suit la loi binomiale B(5;0,3).	47%
Calculer la probabilité p(X=3) si X suit la loi binomiale B(30;0,2).	8%
Calculer la probabilité p(Y=2) si Y suit la loi binomiale B(100;0,1).	0%
Calculer la probabilité p(Y<=388) si Y suit la loi binomiale B(777;0,5).	50%

Exercice 3

Les deux tableaux suivants indiquent pour deux entreprise les taux de sportifs parmi les salariés par tranche d'âge ainsi que l'effectif des salariés de chaque tranche d'âge.

entreprise A

tranche d'âge	[20,40[	[40,60[
tx de sportifs salariés	25.00%	5.00%
effectif des salariés	100	900

entreprise B

tranche d'âge	[100,120[	[120,140[
tx de sportifs salariés	20.00%	4.00%
effectif des salariés	900	100

Calculer les taux de sportifs salariés de chacune de ces deux entreprises.

Taux de sportifs salariés de l'entreprise A (en % sans chiffre après la virgule)	7%
Taux de sportifs salariés de l'entreprise B (en % sans chiffre après la virgule)	18%

Expliquer pourquoi, alors que pour chaque classe le taux de sportifs salariés de l'entreprise A est supérieur à celui de l'entreprises B on observe dans l'entreprise A un taux de sportifs salariés inférieur à celui de l'entreprise B.

La classe à taux de sportifs élevé ([20,40[) représente 10% des salariés de l'entreprise A et 90% des salariés de l'entreprise B. C'est l'inverse pour la classe à taux de sportifs faible ([40,50[) qui représente 90% des salariés de A et seulement 10% de ceux de B. Cette surreprésentation dans l'entreprise B de la classe à taux de sportifs élevé (et l'inverse dans A) compense le fait que pour chaque classe le taux de sportifs est un peu inférieur dans l'entreprise B à celui de la classe correspondante de l'entreprise A.

Exercice 4

En observant depuis de longues années les performances d'une équipe on conclut que si l'équipe vient de gagner un match la probabilité de gagner le suivant est 60% et si elle vient de ne pas gagner la probabilité de gagner le suivant est de 40%. Quelle est la probabilité qu'elle gagne un match quelconque (entourer la bonne réponse)?

40%	$\boxed{50\%}$	60%	70%	autre

Justifier la réponse.

On note p la probabilité de gagner et q celle de ne pas gagner un match quelconque. On a p,q dans [0,1], p+q=1 (événements complémentaires) et p=60%.p+40%.q. Donc q=(1-p) et p=60%.p+40%-40%.p. Par conséquent (1+40%-60%).p=40% c'est à dire 0,8.p=0,4 et donc p=0,4/0,8=50%.

Exercice 5

Deux commerçants vendent des poires. Le premier les vend 2,20 euros/kg mais seulement 95% des poires achetées sont correctes et le second les vend 2,10 euros/kg mais seulement seulement 90% des poires achetées sont correctes. Quel commerçant choisiriez-vous ? Expliquez votre choix.

Le premier commerçant vend les poires à 2,20 euros/kg avec 95% de poires correctes. Chez lui, le prix de revient d'un kilo de poires correctes est donc de 2,20/0,95 euros c'est à dire de 2,32 euros.Avec ce raisonnement on obtient que le prix de revient d'un kilo de poires correctes chez le second commerçant est de 2,10/0,90 euros c'est à dire de 2,33 euros. Le premier commerçant propose donc des poires d'un (très légèrement) meilleur rapport qualité/prix que le second. On choisit donc le premier commerçant.

Exercice 6

Voici deux phrases.

Phrase 1 : «QUE NUL N'ENTRE S'IL N'EST GEOMETRE».
Phrase 2 : «SCIENCE SANS CONSCIENCE N'EST QUE RUINE DE L'AME».
Dans ces phrases les lettres A, E, I, L, N, O, R, S, T, U rapportent 1 point, les lettres D, G, M 2 points, B, C, P 3 points, les lettres F, H, V 4 points, J, Q 8 points et K, W, X, Y, Z 10 points.

Compléter le tableau suivant.

points	1	2	3	4	8	10
nb lettres ds la phrase 1	24	2	0	0	1	0
nb lettres ds la phrase 2	31	2	5	0	1	0
nb lettres ds les 2 phrases	55	4	5	0	2	0

Total des points de la phrase 1	36
Total des points de la phrase 2	58

Valeur moyenne d'une lettre des deux phrases	1.42

Les probabilités demandées seront données en % sans chiffre après la virgule.

On prend une lettre au hasard dans l'une des deux phrases. Probabilité qu'elle provienne de la phrase 1.	41%
On prend une lettre au hasard dans l'une des deux phrases. C'est un E. Proba. qu'il vienne de la phrase 1.	44%
On prend une lettre au hasard dans la phrase 1. Probabilité que ce soit un E.	26%
On prend une lettre au hasard dans la phrase 2. Probabilité que ce soit un E.	23%

Exercice 7

Donner en justifiant un exemple d'une série statistique dont la moyenne est 3 et l'écart-type 1.

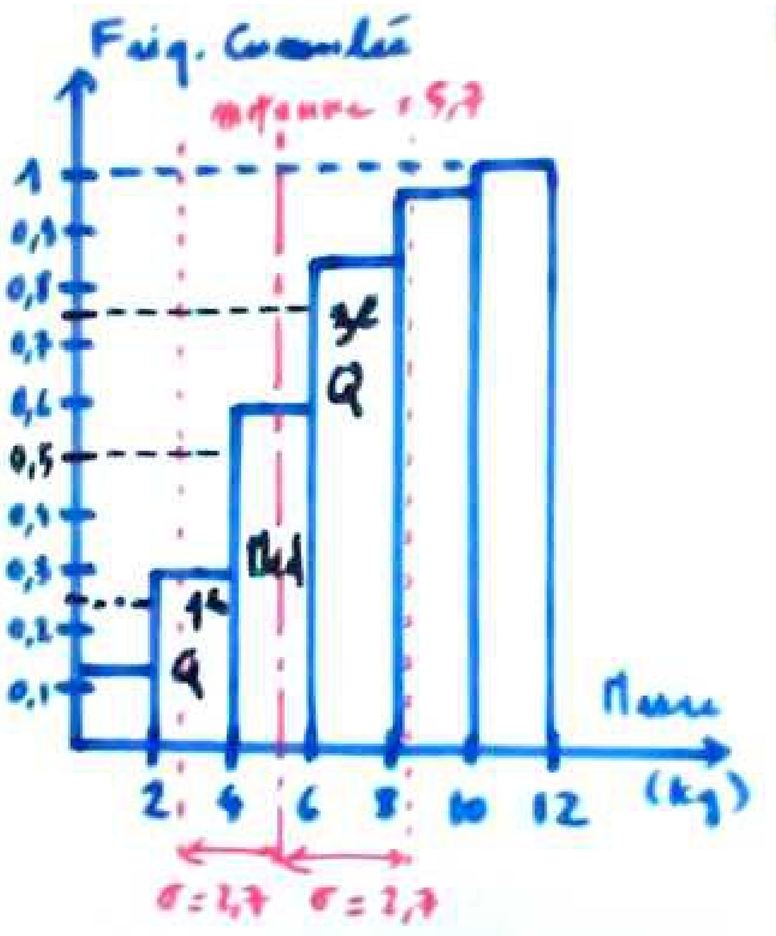
La série statistique x=(x1,x2)$ avec x1=2 et x2=4 convient. En effet la moyenne qui est m= (x1+x2)/2 c'est à dire (2+4)/2 vaut bien 3. La variance qui est V=((x1-m)^2+(x2-m)^2)/2 c'est à dire ((1-2)^2+(3-2)^2)/2 vaut bien 1. L'écart-type qui est la racine carrée de la variance est bien égal à 1.

Exercice 8

On considère la série statistique des colis d'un livreur répartis par classe de masse de 2 en 2 kg.

Classe	[0,2]	]2,4]	]4,6]	]6,8]	]8,10]	]10,12]
Effectif	10	15	25	20	10	5

Faire un histogramme sur lequel on pourra lire la moyenne, l'écart-type ainsi que les classes qui correspondent à la médiane et aux 1er et 3e quartiles.



Nom et Prénom :
Groupe :

Le candidat composera sur cette feuille. Les calculatrices et tous les documents sont autorisés. L'usage du téléphone portable est interdit. La communication avec l'extérieur ou avec les autres candidats est interdite pendant l'épreuve.

Toutes les réponses numériques sont arrondies au dixième inférieur. De plus, lorsqu'il s'agit d'une proportion ou d'une probabilité la réponse est donnée en pourcentage. Par exemple si le calcul donne 37.54 euros on écrit comme réponse 37.5 et si le calcul donne une proportion ou une probabilité de 0.3754 on écrit comme réponse 37.5%.

Exercice 1

On considère une variable aléatoire X qui suit une loi binomiale B(N,r). Compléter le tableau suivant.

N	10	25	70	35	80	5
r	0.1	0.2	0.3	0.5	0.8	0.9
k	8	3	2	1	79	2
p(X=k)	0.0%	13.5%	0.0%	0.0%	0.0%	0.8%
p(X<=k)	99.9%	23.3%	0.0%	0.0%	99.9%	0.8%
p(X>=k)	0.0%	90.1%	99.9%	99.9%	0.0%	99.9%

Exercice 2

On considère une variable aléatoire X qui suit une loi normale N(m,v). Compléter les tableaux suivants.

m	10	25	35	80	5
v	1	9	36	64	4
k	8	31	29	82	1
l	12	34	32	86	2
p(X>=k)	97.7%	2.2%	84.1%	40.1%	97.7%
p(X<=l)	97.7%	99.8%	30.8%	77.3%	6.6%
p(k<=X<=l)	95.4%	2.1%	14.9%	17.4%	4.4%

m	10	25	35	80	5
v	1	9	36	64	4
p(X>=k)	20%	40%	60%	25%	10%
p(X<=l)	10%	30%	20%	75%	20%
p(m-h<=X<=m+h)	20%	30%	50%	10%	80%
k	10.8	25.8	33.5	85.4	7.6
l	8.7	23.4	30.0	85.4	3.3
h	0.3	1.2	4.0	1.0	2.6

Exercice 3

On joue à un jeu en deux tours. A chaque tour on a le choix de la stratégie. Avec la stratégie S les probabilités de gagner 0, 100 et 200 en un tour sont de 10%, 60% et 30% et avec la stratégie R elles sont de 40%, 10% et 50%. Le joueur A choisit de jouer deux fois de suite la stratégie S. Le joueur B choisit de jouer deux fois de suite la stratégie R et le joueur C choisit de prendre la stratégie S au premier tour et de la conserver au second tour s'il a gagné 0 ou 200 au premier tour mais de prendre la stratégie R au second tour s'il a gagné 100 au premier tour. Calculer l'espérance de gain, la probabilité de gagner au moins 300 et la probabilité de gagner 400 pour chacun des 3 joueurs.

	Joueur A	Joueur B	Joueur C
Espérance de Gain au jeu	**240.0**	220.0	234.0
Probabilité de gagner au moins 300 au jeu	45.0%	35.0%	**57.0%**
Probabilité de gagner 400 au jeu	9.0%	**25.0%**	9.0%

Exercice 4

Un jeu de hasard a pour espérance 1000 euros et pour écart-type 500 euros. En utilisant la loi faible des grands nombre montrer que la probabilité que le gain moyen d'un million de joueurs soit compris entre 995 et 1005 euros est d'au moins 99%.

La loi faible des grands nombres dit que si X1,...,Xn sont des variables aléatoires indépendantes et de même loi, d'espérance m et de variance v, alors la variable aléatoire S=(X1+...+Xn)/n vérifie p(|S-m|>=ε)=<V/(nε^2) quel que soit ε>0. Dans la situation étudiée, n=1000000, X1,...,X100000 représentent les gains d'un million de joueurs et S leur gain moyen. On a donc m=1000 et V=500^2. En prenant ε=5 la loi faible des grands nombres indique que la probabilité que le gain moyen s'écarte de plus de 5 euros de 1000 est d'au plus 500^2/(1000000x5x5)=1/100=1%. Par conséquent la probabilité que le gain moyen d'un million de joueurs soient compris entre 995 et 1005 euros est d'au moins 1-1% c'est à dire d'au moins 99%.

Exercice 5

Voici les masses des différentes pièces d'une urne qui contient 700g de pièces.

1 centime	10 centimes	2 euros
2,30g	4,10g	8,50g

La masse moyenne d'une pièce de l'urne est de 7g. Montrer que l'urne contient 100 pièces.

La masse moyenne d'une pièce est m=M/n où M est la masse totale et n l'effectif. Par conséquent l'effectif n est égal à M/m. Ici M=700g et m=7g. L'effectif est donc 700/7 c'est à dire 100. L'urne contient 100 pièces.

Soit p la proportion réelle de pièces de 2 euros dans l'urne. On trouve une proportion de 70% de pièces de 2 euros après avoir fait 10001 tirages dans l'urne avec remise. La théorie de l'estimation de la fréquence nous permet donc de supposer que la fréquence F d'un échantillon pris au hasard par 10001 tirages dans l'urne avec remise suite la loi normale N(p;V) où V est une variance qui se calcule à partir de 70% et de 10001. Que vaut la variance V (entourer la bonne réponse) ?

2.0E-006	1.2E-005	1.7E-005	**2.1E-005**	2.8E-004	3.5E-004	4.3E-004

Peut affirmer avec un risque d'erreur de 5% que l'urne contient 70 pièces de 2 euros (expliquer la réponse) ?

La fréquence F d'un échantillon pris au hasard est une variable aléatoire qui suit la loi normale N(p;V) où p est la proportion réelle de pièces et où V=2.1E-5 a été estimée précédemment. Par conséquent X=F-p suit la normale N(0;V). Cherchons donc A tel que la probabilité que |X| soit inférieur à A soit de 95%. On note σ l'écart-type de X (la racine carrée de V) : σ=4.6E-3. Puisque X/σ suit N(0;1) on a p(X/σ<1.96)=97.5% et on en déduit que p(|X/σ|<1.96)=97.5%-(1-97.5%)=95%. Par conséquent on a A=1.96σ, c'est à dire A=1.96x4.6E-3: A=9.0E-3. Or l'échantillon obtenu grâce aux 10001 tirages contient une proportion de 70% de pièces de 2 euros. Par conséquent |70%-p|<9.0E-3<1% avec la probabilité de 95%, c'est à dire avec un risque d'erreur de 5%. On peut donc affirmer avec un risque d'erreur de 5% que la proportion p de pièces de 2 euros dans l'urne est strictement comprise entre 69% et 71%. Or l'urne contient 100 pièces. Donc on peut affirmer que le nombre de pièces de 2 euros est strictement compris entre 69 et 71 c'est à dire vaut 70 avec un risque d'erreur de 5%.

Montrer que si l'urne contient au plus 69 pièces de 2 euros, son montant total est d'au plus 141.1 euros alors que si elle contient au moins 71 pièces de 2 euros son montant total est d'au moins 142.29 euros.

On note A le nombre de pièces de 1 centime, B le nombre de pièces de 10 centimes et C le nombre de pièces de 2 euros. Soit P le montant total. On a P=0.01xA+0.1xB+2xC donc 0.01x(A+B)+2xC=<P=<0.1x(A+B)+C. Comme A+B=100-C (il y a 100 pièces dans l'urne). On obtient que 1+1.99xC=<P=<10+1.90xC. Par conséquent si l'urne contient au plus 69 pièces de 2 euros son montant total est d'au plus 10+1.90x69 c'est à dire 141.1 euros et si l'urne contient au moins 71 pièces de 2 euros son montant total est d'au moins 1+1.99x71 c'est à dire 142.29 euros.

Le montant total que contient l'urne est 142,1 euros. Montrer que l'urne contient 70 pièces de 2 euros, 20 pièces de 10 centimes et 10 pièces de 1 centime et en déduire l'espérance de gain lorsqu'on fait un tirage avec remise.

Puisque l'urne contient 142.1 euros, son montant est strictement inférieur à 142.29 donc elle contient au plus 70 pièces de 2 euros et son montant est strictement supérieur à 141.1 euros donc elle contient au moins 70 pièces de 2 euros. Finalement elle contient 70 pièces de 2 euros. Les pièces de 1 centime et 10 centimes sont au nombre de 30 et représentent donc un montant de 142.10-70*2 c'est à dire de 2.10 euros. Soit A le nombre de pièces de 1 centime et B celui de pièces de 10 centimes. On a A+B=30 et 0.01xA+0.1xB=2.1 donc B=30-A et 3-0.9xA=2.1. Ceci donne A=10 et donc B=20. Il y a 10 pièces de 1 centime et 20 pièces de 10 centimes. Puisqu'il y a 100 pièces on en déduit que la probabilité de tirer 1 centime est de 10%, celle de tirer 10 centimes de 20% et celle de tirer 2 euros de 70%. L'espérance de gain est donc 0.1x0.01+0.2x0.1+0.7x2 euros c'est à dire 1.4 euros (et pour être plus précis de 142.1 centimes).

Exercice 6

Une usine produit des cylindres dont la masse moyenne est de 2014g et l'écart-type 20g. Prélever un échantillon de 1000 cylindres au hasard et mesurer leur masse moyenne correspond à considérer une variable aléatoire X. Dire quelle loi peut on supposer que suit X par application du théorème limite central et la théorie de l'estimation et justifier succinctement la réponse.

Les masses de 1000 cylindres prélevés au hasard représentent 1000 variables aléatoires X1,...,X1000 indépendantes et de même loi, d'espérance m=2014g et de d'écart-type σ=20g et donc de variance v=σ^2=400g^2. La masse moyenne d'un tel échantillon, X=(X1+...+X1000)/1000, est donc une variable aléatoire d'espérance m et de variance V=v/1000=0.4g^2. D'après le théorème limite central la loi que suit X peut être assimilée à une loi normale, c'est donc la loi normale N(m;V) c'est à dire N(2014;0.4). L'écart-type de X (la racine de sa variance) est donc de 6.3E-1.

Donner A tel que la masse moyenne des 1000 cylindres d'un échantillon pris au hasard et dans l'intervalle]2014-A,2014+A[avec une probabilité de 98% (justifier la réponse).

Puisque la masse moyenne X d'un échantillon de 1000 cylindres pris au hasard suit la loi normale N(2014;0.4) d'écart-type 6.3E-1 la variable aléatoire Y=(X-2014)/6.3E-1 suit la loi normale N(0;1). Or le calcul (ou la lecture de la table) montre que p(|Y|<2.33)=98% (car p(Y<2.33)=99%). Par conséquent p(|X-2014|/6.3E-1<2.33)=98% et donc p(|X-2014|<1.5)=98%. Ainsi A=1.5g : la masse moyenne des 1000 cylindres d'un échantillon pris au hasard et dans l'intervalle]2012.5,2015.5[avec une probabilité de 98%

Exercice 7

Soit a,a',b,b' quatre réels strictement positifs. On suppose que a/b=a'/b'. Montrer qu'on a a/a'=b/b' et a/b=(a+a')/(b+b')=a'/b'.

En multipliant par b/a' les deux membres de l'égalité a/a'=b/b' on obtient (a/b)x(b/a')=(a'/b')x(b/a') qui s'écrit aussi (axb)/(bxa')=(a'xb)/(b'xa') et après simplification du premier membre par b et du second par a' on obtient bien a/a'=b/b'. On a donc a/a'+1=b/b'+1. Par conséquent a/b=a/'/b'=(a'x(a/a'+1))/(b'x(b/b'+1))=((a'xa)/a'+a')/((b'xb)/b'+b')=(a+a')/(b+b').

On considère un événement A qui arrive avec la probabilité p(A) et un événement B qui arrive avec la probabilité p(B). Soit B' l'événement complémentaire de B et p(B') sa probabilité. On suppose que p(A), p(B) et p(B') sont tous les trois non nuls. Montrer que A et B sont indépendants si et seulement si p(A∩B)/p(B)=p(A∩B')/p(B').

On a p(B')=1-p(B) et p(A∩B')=p(A)-p(A∩B).
1/ Si A et B sont indépendants alors p(A∩B)=p(A)xp(B) et donc p(A∩B)/p(B)=p(A). Puisque p(A∩B)=p(A)xp(B) on a p(A∩B')=p(A)-p(A∩B)=p(A)-p(A)xp(B)=p(A)x(1-p(B))=p(B'). Donc p(A∩B')=p(A)xp(B') et p(A∩B')/p(B')=p(A)=p(A∩B)/p(B) et donc p(A∩B)/p(B)=p(A∩B')/p(B').
2/ Réciproquement on suppose p(A∩B)/p(B)=p(A∩B')/p(B'). On applique le résultat de la question précédente en posant p(A∩B)=a, p(B)=b, p(A∩B')=a' et p(B')=b'. Ceci donne p(A∩B)/p(B)=(p(A∩B)+p(A∩B'))/(p(B)+p(B')). Puisque p(A∩B)+p(A∩B')=p(A) et p(B)+p(B')=1 on obtient p(A∩B)/p(B)=p(A) c'est à dire p(A∩B)=p(A)xp(B). Ceci signifie que A et B sont indépendants.

Exercice 8

On prélève 50 grumes d'Okan de la cale d'un cargo. On teste leur module d'élasticité longitudinale. On trouve pour cet échantillon de 50 pièces une valeur moyenne de 21130Mpa. Pour être conforme la cargaison devrait avoir un module moyen d'élasticité longitudinale de 22260Mpa avec un écart-type de 3348Mpa. Peut on considérer avec un risque d'au plus 5% que la cargaison n'est pas conforme (justifier la réponse)?

Prélever 50 grumes d'Okan et considérer leurs modules d'élasticité longitudinale revient à considérer 50 variables aléatoires E1,... ,E50 indépendantes et de même loi. Si la production contenue dans le cargo est conforme, la variable aléatoire E=(E1+...+E50) peut être assimilée à une variable aléatoire qui suit une loi normale de moyenne m=22260Mpa et d'écart-type σ=3348/7.7Mpa c'est à dire de 473.4Mpa (car 7.07^2=50). Cherchons A telle que la probabilité que |E-m| soit inférieure à A soit de 95%. Le nombre A est tel que la probabilité que E soit inférieure à m+A est 97.5%. On trouve A=928.0Mpa. Or la moyenne des modules d'élasticité des 50 grumes est M=21130Mpa. On a donc |M-m|=|21130-2260|=1130>928.0=A. L'écart entre la moyenne pour les grumes du cargo et la moyenne attendue est supérieur à A. On peut donc considérer que la cargaison n'est pas conforme avec un risque de 5% d'erreur.

Exercice 9

On compare l'autonomie de deux modèles de téléphone. On dispose de 51 exemplaires du premier modèle. Leur autonomie moyenne est de 96 heures avec un écart-type de 10 heures. L'autonomie moyenne de 33 exemplaires du second modèle est de 99 heures avec un écart-type de 8 heures.

Peut on considérer avec un risque d'au plus 5% que les deux modèles de téléphones peuvent être distingués par leur autonomie (justifier la réponse) ?

On pose T1=96h et T2=99h. Si l'autonomie ne permet pas de distinguer alors la différence T2-T1 des autonomies moyennes suit une loi normale centrée et de variance 10x10/50+8x8/32=4. Or si une variable aléatoire X suit une telle loi alors la probabilité que X soit dans l'intervalle]-3.92, 3.92[est de 95% (car la probabilité que X soit inférieure à 3.92 heures est de 97.5%. Puisque ce est pas le cas de T2-T1 qui vaut 3h on ne peut pas considérer que l'autonomie permet de distinguer avec un risque de 5% les deux modèles de vélos électriques.

Exercice 10 3.92 2.77

Faire un histogramme de la série statistique suivante (avec moyenne, écart-type, classes de la médiane et des 1er et 3e quartiles).

classe	[0;10[	[10;20[	[20;30[	[30;40[	[40;50[	[50;60[
effectif	60	50	40	30	20	10