Département STPI

Troisième Année IC

Probabilités et Statistiques

C. Boyer, S. Grusea, C. Marteau et S. Thao

Année universitaire 2013-2014

Simona Grusea
Bureau 11, département GMM
simona.grusea@insa-toulouse.fr

Clément Marteau
Bureau 123, département GMM
clement.marteau@insa-toulouse.fr

Table des matières

Chapitre I

Introduction à la théorie des probabilités

I.1 Pourquoi les probabilités ?

Dans de nombreux domaines (science, sociologie, médecine, etc...), on s'intéresse à des phénomènes dans lesquels apparaît l'effet du hasard. Ces phénomènes sont caractérisés par le fait que les résultats des observations varient d'une expérience à l'autre.

Une expérience est appelée "aléatoire" s'il est impossible de prévoir à l'avance son résultat et si, répétée dans des conditions identiques, elle peut donner des résultats différents:

- succession d'appels à un standard téléphonique non surchargé;
- observation de la durée de vie d'un individu anonyme dans une population;
- observation de la durée de fonctionnement sans panne d'appareil;
- jeu de pile ou face.

Voici d'autres exemples de domaines d'applications des probabilités.

Fiabilité. On considère un système formé par plusieurs composants. On s'intéresse à la fiabilité du système: on va chercher à calculer la probabilité que le système fonctionne encore à un instant donné. Il faut pour cela connaître la probabilité que chacun des composants fonctionne à cet instant et tenir compte du fait que les composants ne fonctionnent peut-être pas indépendamment les uns des autres.

Fatigue des matériaux. Les données de fatigue des matériaux sont très dispersées. On fait alors appel à des modélisations probabilistes et à des méthodes statistiques afin, par exemple, de construire des intervalles de confiance pour le nombre moyen de cycles jusqu'à la rupture.

Télécommunications. En télécommunications, on doit souvent tenir compte du "bruit" dans les systèmes. Par exemple, supposons qu'un système émet soit un 0, soit un 1, et qu'il y a un risque p que le chiffre émis soit mal reçu. Il est alors intéressant de calculer la probabilité qu'un 0 ait été émis, sachant qu'un 0 a été reçu, ou encore la probabilité qu'il y ait une erreur de transmission.

I.2 Espace probabilisé

I.2.1 Définitions

Définition 1 *On appelle* ***univers*** *associé à une expérience aléatoire l'ensemble* Ω *de tous les résultats possibles de cette expérience.*

Le choix de l'ensemble Ω comporte une part d'arbitraire. Il dépend de l'idée que l'on a, a priori, sur les résultats de l'expérience aléatoire. Donnons quelques exemples:

1. On lance une pièce de monnaie. Pour l'ensemble Ω, on peut choisir soit $\Omega = \{$ pile, face $\}$, soit $\Omega = \{$ pile, face, tranche $\}$.

2. On s'intéresse à l'état de fonctionnement d'un système. Dans ce cas $\Omega = \{0, 1\}$ avec la convention 0 si le système est en panne et 1 s'il fonctionne.

3. Le résultat de l'expérience aléatoire est le nombre de tirages nécessaires dans un jeu de pile ou face jusqu'à l'obtention du premier "pile". Dans ce cas, $\Omega = \{1, 2, 3, \cdots\} = \mathbb{N}^*$.

4. On considère la succession des appels à un standard téléphonique non surchargé et l'on étudie la répartition des instants où le standard reçoit un appel, à partir d'un instant choisi comme origine (on admet que deux appels ne peuvent se produire rigoureusement au même instant et que le phénomène est limité dans le temps). Une réalisation de cet événement est une suite croissante de nombres réels positifs t_i où t_i désigne l'instant d'enregistrement du ième appel: $\Omega = \{0 < t_1 < t_2 < \cdots < t_n < t_{n+1} < \cdots\}$. L'univers Ω est donc une partie de $(\mathbb{R}^+)^{\mathbb{N}}$.

5. On considère l'expérience aléatoire "durée de vie d'un individu". L'ensemble Ω est soit l'ensemble $\mathbb{N}$, soit $\mathbb{R}^+$ selon le procédé discontinu ou continu de cette mesure.

Nous constatons que Ω peut être fini (exemples 1 et 2), dénombrable (exemples 3 et 5) ou non dénombrable (exemples 4 et 5). Lorsque Ω est fini ou dénombrable, on parle d'univers *discret*. Sinon on parle d'univers *continu*.

Définition 2 *Etant donnée une expérience aléatoire, un* ***événement aléatoire*** *est une partie de l'ensemble des résultats possibles de l'expérience, c'est donc un sous-ensemble* A *de l'univers* Ω*. On dit que l'événement* A *est réalisé si le résultat* ω *de l'expérience appartient à* A*.*

On sait que l'événement A est réalisé seulement une fois l'expérience aléatoire réalisée.

Exemples:

- Si l'on s'intéresse à l'événement suivant: "on a obtenu un chiffre pair lors d'un lancer d'un dé à 6 faces", on introduit $A = \{2, 4, 6\}$, qui est un sous-ensemble de $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Si l'on s'intéresse à l'événement suivant: "la durée de vie du composant est supérieure ou égale à 1000 heures", $A = [1000, +\infty[$ est un sous-ensemble de $\Omega = \mathbb{R}^+$.

L'ensemble $\emptyset$ est appelé l'événement impossible et Ω est appelé l'événement certain.

I.2.2 Opérations sur les événements

Les événements aléatoires étant des ensembles, introduisons les opérations ensemblistes classiques de la théorie des ensembles.

Définition 3 *On appelle événement contraire de A, noté A^C, le complémentaire de A dans Ω:*

$$A^C = \{\omega \in \Omega : \omega \notin A\}.$$

L'événement contraire A^C est réalisé si et seulement si A n'est pas réalisé.

Exemple : Si A est l'événement "la durée de vie du composant est supérieure ou égale à 1000 heures": $A = [1000, +\infty[$, l'événement contraire est l'événement "la durée de vie du composant est strictement inférieure à 1000 heures": $A^C = [0, 1000[$.

Définition 4 *Soient A et B deux événements d'un univers Ω.*

- *L'événement "A et B" est celui qui est réalisé si A et B sont réalisés. C'est l'intersection*
$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\}.$$
- *L'événement "A ou B" est celui qui est réalisé si l'un des deux est réalisé ou si les deux sont réalisés. C'est l'union*
$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}.$$
- *L'inclusion $A \subset B$ signifie que l'événement A ne peut être réalisé sans que B le soit.*

Définition 5 *Deux événements A et B sont dits* ***incompatibles*** *si la réalisation de l'un implique la non-réalisation de l'autre.*

Dans l'espace Ω, deux événements incompatibles sont représentés par deux parties disjointes. Si $A \cap B = \emptyset$, alors A et B sont incompatibles. Il est clair, par exemple que A et A^C sont incompatibles.

I.2.3 Probabilité

Définition 6 *Soit Ω un univers associé à une expérience aléatoire et soit $\mathcal{A}$ l'ensemble des parties de Ω. Une* ***probabilité*** *$\mathbb{P}$ sur l'espace $(\Omega, \mathcal{A})$ est une application de $\mathcal{A}$ dans $[0,1]$ telle que*

1. $\mathbb{P}(\Omega) = 1$.
2. *Si $(A_n)_{n\geq 1}$ est une famille d'événements de $\mathcal{A}$ 2 à 2 incompatibles,*

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé espace de probabilité.

On peut déduire de la définition précédente un certain nombre de propriétés.

Proposition 1 *Soient A et B deux événements aléatoires.*

1. $\mathbb{P}(\emptyset) = 0$.
2. $\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) \leq \sum_{n=1}^{N} \mathbb{P}(A_n)$.
3. *Si $A_1, \ldots, A_N$ sont deux-à-deux incompatibles,*

$$\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mathbb{P}(A_n).$$

4. $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.
5. *Si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.*
6. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
7. *Si Ω est fini ou dénombrable, alors pour tout événement A,*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

I.2.4 Probabilité uniforme

Soit Ω un ensemble fini: $\Omega = \{\omega_1, \ldots, \omega_N\}$. Pour tout $i \in \{1, 2, \ldots, N\}$, on pose $\mathbb{P}(\{\omega_i\}) = \frac{1}{N}$. Alors, pour toute partie A de Ω, on a

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \frac{Card(A)}{N} = \frac{Card(A)}{Card(\Omega)}.$$

Dans le cas du lancer de dé à 6 faces, pour tout $\omega \in \{1, 2, \ldots, 6\}$, $\mathbb{P}(\{\omega\}) = 1/6$.
Si on note l'événement "on a obtenu un chiffre pair" par $A = \{2, 4, 6\}$, alors

$$\mathbb{P}(A) = 3/6 = 1/2.$$

Remarque: Pour un problème donné, il y a souvent plusieurs modélisations possibles, c'est-à-dire que le choix de l'espace de probabilité n'est pas unique.

Remarque: Choisir un élément au hasard signifie que les divers choix possibles sont équiprobables, donc que l'ensemble Ω est muni de la probabilité uniforme. Dans ce cas, tous les calculs sont simples et se ramènent souvent à des calculs d'analyse combinatoire.

Exercice 1. Une urne contient 12 boules numérotées. Les boules numérotées de 1 à 8 sont de couleur blanche, celles portant les numéros de 9 à 12 sont de couleur noire. On tire une boule au hasard dans cette urne. On s'intéresse à la fois au numéro de cette dernière ainsi qu'à sa couleur.

1. Décrire l'espace des issues associé à cette expérience.
2. Quelle est la probabilité de tirer une boule portant un nombre pair?
3. Calculer la probabilité de l'événement "la boule tirée est blanche".
4. Quelle est la probabilité de tirer une boule noire portant un nombre impair?

Exercice 2. Le paradoxe du chevalier de Méré
Voici deux jeux qui étaient fort populaires au XVIIème siècle:

- Dans le premier jeu, les joueurs pariaient sur le fait d'obtenir au moins un as en lançant quatre fois un dé à six faces.
- Dans le deuxième jeu, ils pariaient sur l'apparition d'au moins un double as en lançant 24 fois une paire de dés.

Le Chevalier de Méré, un noble de l'époque, adorait ces jeux et, comme beaucoup d'autres joueurs de son époque, il pensait que la probabilité de gagner était la même pour les deux jeux.
Par contre, son expérience lui avait montré que le premier jeu était plus avantageux que le deuxième. Le Chevalier de Méré demanda alors au fameux mathématicien Blaise Pascal de l'éclaircir sur ce problème. Celui-ci, dans un échange de lettres avec son ami Pierre Fermat, réussit à résoudre le fameux "paradoxe", que vous êtes invités à votre tour à résoudre.

Chapitre II

Probabilités conditionnelles et indépendance

II.1 Probabilités conditionnelles

Dans le chapitre précédent, on a parlé de la probabilité d'un événement sans tenir compte de la réalisation d'autres événements. En pratique, on peut considérer plusieurs événements, certains pouvant avoir une influence sur la réalisation d'autres événements.

Exemple: On lance deux dés. Soient les événements $A = \{ \text{la somme est } \geq 11\}$ et $B = \{ \text{le lancer du 1er dé donne } 6\}$. Il est clair que la réalisation de B influe sur la réalisation de A.

Supposons que l'on s'intéresse à la réalisation d'un événement A , tout en sachant qu'un événement B est réalisé. Si A et B sont incompatibles, alors la question est réglée: A ne se réalise pas. Mais si $A \cap B \neq \emptyset$, il est possible que A se réalise. Cependant, l'espace des événements possibles n'est plus Ω tout entier, mais il est restreint à B. En fait, seule nous intéresse la réalisation de A à l'intérieur de B, c'est-à-dire $A \cap B$ par rapport à B. Ceci justifie la définition suivante.

Définition 7 *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soient A et B deux événements aléatoires tels que $\mathbb{P}(B) \neq 0$. On appelle* ***probabilité conditionnelle*** *de A sachant B la quantité*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Remarque: On a les égalités suivantes:

$$\begin{aligned} \text{Si } \mathbb{P}(B) > 0, \quad \mathbb{P}(A \cap B) &= \mathbb{P}(A|B) \times \mathbb{P}(B). \\ \text{Si } \mathbb{P}(A) > 0, \quad \mathbb{P}(A \cap B) &= \mathbb{P}(B|A) \times \mathbb{P}(A). \end{aligned}$$

Proposition 2 (formule des probabilités totales) *Soit $(A_i)_{i\in I}$ une famille d'événements aléatoires formant une partition de Ω, c'est-à-dire tels que:*

- $\cup_{i\in I} A_i = \Omega$;
- $A_i \cap A_j = \emptyset$ *pour tout* $i \neq j$.

On suppose de plus que $\mathbb{P}(A_i) \neq 0$ pour tout $i \in I$. Alors

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|A_i)\mathbb{P}(A_i).$$

Proposition 3 (formule de Bayes) *Sous les mêmes hypothèses que la proposition précédente, on a:*

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i)\mathbb{P}(A_i)}{\sum_{j\in I} \mathbb{P}(A|A_j)\mathbb{P}(A_j)}.$$

La formule de Bayes (publiée après sa mort en 1763) présente un grand intérêt car elle permet de modifier notre connaissance des probabilités en fonction d'informations nouvelles. Cette formule joue donc un rôle très important dans la statistique bayésienne.

II.2 Indépendance

Définition 8 *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, et soient A et B deux événements aléatoires. On dit que A et B sont **indépendants** si*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Remarque: A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$. Cette condition signifie que la probabilité de réalisation de l'événement A n'est pas modifiée par une information concernant la réalisation de l'événement B.

Proposition 4 *Si A et B sont deux événements indépendants alors:*

- A^C *et* B *sont également indépendants;*
- A *et* B^C *sont également indépendants;*
- A^C *et* B^C *sont également indépendants.*

Nous allons maintenant définir l'indépendance de plus de 2 événements aléatoires.

Définition 9 *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Pour $n \geq 2$, soient $A_1, A_2, \ldots A_n$, des événements aléatoires.*

- *Ces événements sont deux à deux indépendants si pour tout couple (i, j) avec $i \neq j$ on a*

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

- *Ces événements sont indépendants (dans leur ensemble) si pour tout $k \in \{2, 3, \ldots, n\}$ et tout choix d'indices distincts $i_1, \ldots, i_k$, on a*

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots P(A_{i_k}).$$

Exercice 1. On s'intéresse à la durée de vie d'un téléphone mobile. Pour tout $t \geq 0$, on note D_t l'évènement $D_t := \{\text{le téléphone fonctionne encore à la date } t\}$, et l'on suppose que $\mathbb{P}(D_t) = \exp(-ct)/(1+t), c > 0$. Calculer la probabilité que le téléphone fonctionne encore au bout de 2 ans sachant qu'il a déjà fonctionné 1 an.

Exercice 2. On considère un système de communication qui émet soit un 0 soit un 1 avec équiprobabilité. À cause du bruit, le signal émis est parfois mal reçu. Soit E_i l'événement "i est émis" et soit R_i l'événement "i est reçu", pour $i = 0$ et 1. On suppose que $\mathbb{P}(R_0/E_0) = 0.7$ et $\mathbb{P}(R_1/E_1) = 0.8$.
a) Calculer $\mathbb{P}(E_0/R_1)$.
b) Calculer la probabilité d'une erreur de transmission.

Exercice 3. On considère un système formé de 5 constituants indépendants. La probabilité qu'un constituant soit défectueux est de $0,2$. Quelle est la probabilité que ce système fonctionne si les constituants sont branchés:

- en série ?
- en parallèle ?

Exercice 4. On lance deux dés et on considère les événements A: "le premier dé donne un nombre pair", B: "le deuxième dé donne un nombre pair" et C: "la somme des dés est un nombre pair". Montrer que les événements A, B et C sont deux à deux indépendants. Sont-ils indépendants dans leur ensemble?

Chapitre III

Variables aléatoires réelles

III.1 Introduction

Dans de nombreuses expériences aléatoires, on n'est pas intéressé directement par le résultat de l'expérience, mais par une certaine fonction de ce résultat. Considérons par exemple l'expérience qui consiste à observer, pour chacune des n pièces produites par une machine, si la pièce est défectueuse ou non. Nous attribuerons la valeur 1 à une pièce défectueuse et la valeur 0 à une pièce en bon état. L'univers associé à cette expérience est $\Omega = \{0,1\}^n$. Ce qui intéresse le fabricant est la proportion de pièces défectueuses produites par la machine. Introduisons donc une fonction de Ω dans $\mathbb{R}$ qui à tout $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ de Ω associe le nombre

$$X(\omega) = \sum_{i=1}^{n} \frac{\omega_i}{n},$$

qui correspond à la proportion de pièces défectueuses associée à l'observation de ω. Une telle fonction X définie sur Ω et à valeurs dans $\mathbb{R}$ s'appelle une variable aléatoire réelle.

III.2 Définitions

III.2.1 Variable aléatoire

Définition 10 *Étant donné un univers Ω, une* ***variable aléatoire réelle*** *(v.a.r.) est une application de Ω dans $\mathbb{R}$:*

$$X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}.$$

III.2.2 Loi de probabilité

Définition 11 *Soit Ω un univers muni d'une probabilité $\mathbb{P}$, et soit X une v.a.r. On appelle* ***loi de probabilité*** *de X, notée $\mathbb{P}_X$, l'application qui à toute partie A de $\mathbb{R}$ associe*

$$\mathbb{P}_X(A) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A\}\right).$$

Remarque : Dans la suite du cours, on utilisera la notation abrégée: $\mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A\}\right) = \mathbb{P}(X \in A)$. De même, on notera $\mathbb{P}(X = x)$ la probabilité $\mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$.

Proposition 5 *L'application $\mathbb{P}_X$ définit une probabilité sur $\mathbb{R}$.*

III.2.3 Fonction de répartition

Définition 12 *La fonction de répartition de la v.a.r. X est définie par*

$$F_X(x) = \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}.$$

Propriétés de la fonction de répartition:

1. $0 \leq F_X(x) \leq 1, \ \forall x \in \mathbb{R}$.
2. F_X tend vers 0 en $-\infty$ et vers 1 en $+\infty$.
3. F_X est croissante.
4. F_X est continue à droite.

Proposition 6 *On a l'identité*

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a), \quad \forall (a, b) \in \mathbb{R}^2, \ a < b.$$

Remarque: On montre facilement que F_X est continue si et seulement si $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$. On parle alors de loi diffuse ou de v.a.r. continue (voir définition 15).

Définition 13 *Soit X une v.a.r. de fonction de répartition F_X supposée strictement croissante de $I \subset \mathbb{R}$ dans $]0, 1[$. Le **quantile d'ordre** $\alpha \in]0, 1[$ de X est le nombre $x_\alpha \in I$ tel que $F_X(x_\alpha) = \alpha$, ce qui signifie que*

$$\mathbb{P}(X \leq x_\alpha) = \alpha.$$

Remarques:

- $x_{1/2}$ est appelé **médiane** de X. La médiane vérifie les deux égalités

$$P(X \leq x_{1/2}) = 1/2 = P(X > x_{1/2}).$$

- Dans le cas où F_X n'est pas strictement croissante mais simplement croissante, on définit le quantile d'ordre α par

$$x_\alpha = \inf \left\{x \in \mathbb{R} : F_X(x) \geq \alpha\right\}.$$

III.3 Différents types de variables aléatoires réelles

III.3.1 Variables aléatoires discrètes

Définition 14 *Une v.a.r. X à valeurs dans un ensemble $\mathcal{X}$ fini ou dénombrable est appelée v.a.r.* ***discrète****. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités:*

$$\mathbb{P}_X(x) = \mathbb{P}(X = x), \quad \forall x \in \mathcal{X}.$$

Ainsi, pour toute partie A de $\mathcal{X}$, on a alors:

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x) \quad \text{et} \quad \mathbb{P}_X(\mathcal{X}) = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x) = 1.$$

Exemple: Supposons que l'on observe la durée de vie T d'une ampoule électrique et que cette durée de vie T, exprimée en heures, satisfait pour tout $0 < a < b$,

$$\mathbb{P}(a < T \leq b) = \exp(-a/100) - \exp(-b/100).$$

On note X le nombre de périodes complètes de 100 heures que dure l'ampoule. Les valeurs possibles de X etant entières, la v.a.r. X est donc discrète. Calculons la fonction de répartition de X. Comme X est positive, on a

$$F_X(x) = \mathbb{P}(X \leq x) = 0, \quad \forall x < 0.$$

De plus, pour tout $n \in \mathbb{N}$,

$$\mathbb{P}(X = n) = \mathbb{P}(100n \leq T < 100(n+1)) = \exp(-n) - \exp\left(-(n+1)\right).$$

Ainsi, on a donc pour tout $x \geq 0$:

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \sum_{n=0}^{[x]} \mathbb{P}(X = n) \\
&= 1 - \exp\left(-([x] + 1)\right).
\end{aligned}$$

On notera que la fonction F_X est une fonction en escalier.

Exercice 1. On lance une paire de dés et on appelle X le minimum et Y le maximum des deux nombres obtenus. Déterminer les lois des variables aléatoires X et Y.

III.3.2 Variables aléatoires continues

Définition 15 *Soit X une v.a.r. qui prend un nombre infini non dénombrable de valeurs. Si F_X est une fonction continue, on dit que X est une v.a.r.* ***continue****. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités $\mathbb{P}(a < X < b)$, pour tout $a < b$.*

Remarque: Notons que l'on peut mettre $<$ ou $\leq$ dans ce qui précède car la variable étant continue, on a $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$.

Exemple : Soit $\lambda > 0$. Une v.a.r. X de fonction de répartition

$$F_X(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

est continue.

Définition 16 *Si l'on peut écrire la fonction de répartition d'une variable continue sous la forme*

$$F_X(t) = \int_{-\infty}^{t} f_X(x)dx,$$

où f_X est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, alors on dit que f_X est la densité de probabilité de la v.a.r. X.

Ceci implique que l'on a pour tout $a < b$:

$$\mathbb{P}(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx.$$

Cette intégrale étant positive pour tout $a < b$, il en résulte que $f_X \geq 0$. De plus, puisque $\lim_{t\to+\infty} F_X(t) = 1$, on a

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1.$$

Une densité de probabilité est donc une fonction positive ou nulle, d'intégrale 1, et qui caractérise la loi d'une v.a.r. continue. De plus, en tout point $x_0 \in \mathbb{R}$ où F_X est dérivable, on a $f_X(x_0) = F'_X(x_0)$.

Exemple: Dans l'exemple de la durée de vie T d'une ampoule électrique, T a pour densité de probabilité

$$f(x) = \begin{cases} \exp(-x/100)/100 & \text{pour tout } x \geq 0 \\ 0 & \text{pour tout } x < 0. \end{cases}$$

Enfin, établir que deux v.a.r. (discrètes ou continues) X et Y ont même loi, c'est démontrer que l'on a l'égalité suivante:

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a < Y \leq b), \quad a, b \in \mathbb{R}.$$

Ainsi, en faisant tendre a vers $-\infty$, on obtient le résultat suivant:

Théorème 1 *Deux v.a.r. à valeurs dans le même ensemble d'arrivée ont la même loi si et seulement si leurs fonctions de répartition sont égales.*

III.4 Exemples fondamentaux

III.4.1 Variables discrètes

Soit X une v.a.r. discrète prenant ses valeurs dans un ensemble $\{x_1, x_2, \ldots, x_n\}$, éventuellement infini. Alors la loi de X est caractérisée par l'ensemble des probabilités $\mathbb{P}(X = x_i)$, c'est-à-dire les nombres réels positifs p_i tels que

$$\mathbb{P}(X = x_i) = p_i \quad \text{avec} \quad 0 \leq p_i \leq 1 \quad \text{et} \quad \sum_{i=1}^{n} p_i = 1.$$

a) Loi de Bernoulli

On dit qu'une v.a.r. X à valeurs dans $\{0, 1\}$ suit une loi de Bernoulli de paramètre $p \in]0, 1[$, notée $\mathcal{B}(p)$, si

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p.$$

Par exemple, cette loi intervient lorsque l'on modélise l'état de fonctionnement d'un système. La probabilité que le système fonctionne vaut p et la probabilité que le système ne fonctionne pas vaut $1-p$. Cette loi s'applique aussi aux jeux de hasard de type binaire comme pile ou face ...

b) Loi binomiale

On dit qu'une v.a.r. X à valeurs dans $\{0, 1, \ldots, n\}$ suit une loi binomiale de paramètres (n, p), notée $\mathcal{B}(n, p)$, si

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad \forall k \in \{0, \ldots, n\}.$$

Cette loi intervient par exemple pour modéliser le nombre de pièces défectueuses dans un lot de n pièces, qui ont chacune une probabilité p d'être défectueuse, indépendamment les unes des autres.

c) Loi géométrique

On dit qu'une v.a.r. X à valeurs dans $\mathbb{N}^*$ suit une loi géométrique de paramètre $p \in]0, 1[$, notée $\mathcal{G}(p)$, si

$$\mathbb{P}(X = k) = p(1 - p)^{k-1}, \quad \forall k \in \mathbb{N}^*.$$

Cette loi permet de modéliser le nombre de réalisations indépendantes d'une expérience à 2 issues (succès-échec), jusqu'à l'obtention du premier succès, si à chaque réalisation la probabilité de succès est p.

d) Loi de Poisson

On dit qu'une v.a.r. X à valeurs dans $\mathbb{N}$ suit une loi de Poisson de paramètre $\lambda > 0$, notée $\mathcal{P}(\lambda)$, si

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad k \in \mathbb{N}.$$

Cette loi intervient comme comportement limite de la loi binomiale lorsque $n \to +\infty$ et $np \to \lambda$.

III.4.2 Variables continues

Soit X une v.a.r. continue. Alors la loi de X est caractérisée par l'ensemble des probabilités

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx,$$

où f_X est la densité de probabilité de X et a et b sont deux nombres réels, éventuellement infinis. Comme nous l'avons vu plus haut, il suffit de connaître cette densité pour connaître la loi de X.

a) Loi exponentielle

On dit que X suit une loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si la loi de X a pour densité

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

La loi exponentielle est utilisée en fiabilité. Le paramètre λ représente le taux moyen de défaillance alors que son inverse $\theta = 1/\lambda$ est "le temps moyen de bon fonctionnement". La loi exponentielle s'applique bien aux matériels électroniques ou aux matériels subissant des défaillances brutales.

b) Loi Gamma

La loi exponentielle est un cas particulier de la famille des lois Gamma.
Soient $a > 0$ et $\lambda > 0$. On dit que X suit une loi Gamma de paramètres (a, λ), notée $\gamma(a, \lambda)$, si la loi de X a pour densité

$$f_X(x) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} x^{a-1} \exp(-\lambda x) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0, \end{cases}$$

où pour tout $a > 0$, la célèbre fonction gamma est donnée par $\Gamma(a) = \int_0^{+\infty} x^{a-1}\exp(-x)dx$. Pour $a = 1$, on retrouve la loi exponentielle de paramètre λ. Le paramètre a est un paramètre de forme alors que le paramètre λ est un paramètre d'échelle.

Figure III.1: *Tracé de la densité de la loi Gamma*

c) Loi normale de paramètres (m, σ^2)

Soient $m \in \mathbb{R}$ et $\sigma > 0$. On dit que X suit une loi normale de paramètres (m, σ^2), notée $\mathcal{N}(m, \sigma^2)$, si la loi de X a pour densité

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

Figure III.2: *Tracé de la densité de la loi normale de paramètres* $(0, 1)$.

À cause de sa forme, cette courbe est souvent appelée “courbe en cloche”. Elle présente un axe de symétrie vertical pour $x = m$. De plus, comme il n’existe pas d’expression analytique de la fonction de répartition de X, on utilise alors des tables obtenues par des calculs approchés d’intégrales.

Exemple: La loi de la v.a.r. X représentant le diamètre d'un roulement à bille peut être modélisée par une loi $\mathcal{N}(m, \sigma^2)$. Dans ce cas, m représente le diamètre moyen d'un roulement à bille, et σ^2 désigne la précision de cette mesure.

La loi normale s'applique à de nombreux phénomènes, en physique, en économie (erreurs de mesure). Comme on le verra au Chapitre V, elle est la forme limite de nombreuses lois discrètes. Elle peut aussi représenter la fin de vie des dispositifs subissant un phénomène de vieillissement: usure, corrosion, etc ... Il faut cependant remarquer que les variables utilisées dans les domaines technologique ou économique sont bien souvent positives. Pour que la loi normale puisse être représentative d'un tel phénomène, il faut que la probabilité d'obtenir des valeurs négatives de la variable soit très faible. Il faut en particulier éviter d'utiliser cette modélisation pour les queues des distributions.

Chapitre IV

Manipulation de variables aléatoires

IV.1 Caractéristiques des variables aléatoires

IV.1.1 Espérance

Définition 17 *Soit X une v.a.r. et h une application de $\mathbb{R}$ dans $\mathbb{R}$. Donc $h(X)$ est elle aussi une v.a.r.*

- Si X est discrète à valeurs dans un ensemble $\mathcal{X}$, l'espérance de $h(X)$ est la quantité

$$\mathbb{E}(h(X)) = \sum_{x \in \mathcal{X}} h(x)\mathbb{P}(X = x),$$

pourvu que cette série converge (dans le cas où $\mathcal{X}$ est infini).

- Si X est continue et admettant une densité f_X, l'espérance de $h(X)$ est la quantité

$$\mathbb{E}(h(X)) = \int_{-\infty}^{+\infty} h(x) f_X(x) dx,$$

à condition que cette intégrale soit convergente.

Notons que si $h(x) = x$, on obtient $\mathbb{E}(X)$ appelée **espérance mathématique** (ou **moyenne**) de la v.a.r. X. Par ailleurs, pour $A \subset \mathbb{R}$, si l'on définit la v.a.r. suivante:

$$1_{\{X \in A\}} = \begin{cases} 1 & \text{si } X \in A \\ 0 & \text{sinon} \end{cases}$$

qui est appelée fonction caractéristique de l'événement $\{X \in A\}$, alors l'espérance de cette v.a.r. est:

$$\mathbb{E}(1_{\{X \in A\}}) = \mathbb{P}(X \in A) = \mathbb{P}_X(A),$$

d'où le lien étroit entre probabilité et espérance.

Exemple : Une roulette contient 18 cases rouges, 18 cases noires et une case verte. Si on mise sur une couleur et cette couleur sort, on gagne encore une fois la mise, sinon on perd la mise. Soit G le gain obtenu quand on joue une fois en misant 1 euro sur le rouge. La loi de la variable aléatoire G est donnée par $\mathbb{P}(G = -1) = 19/37$, $\mathbb{P}(G = 1) = 18/37$. Le gain moyen est $\mathbb{E}(G) = (-1) \times \frac{19}{37} + 1 \times \frac{18}{37} = -\frac{1}{37}$. Le jeu n'est donc pas équitable et il est en faveur de la banque.

IV.1.2 Propriétés de l'espérance

1. L'espérance est linéaire: pour tout $\alpha, \beta \in \mathbb{R}$, et pour toutes v.a.r. X et Y
$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y).$$
2. Si X est une v.a.r. constante égale à $a \in \mathbb{R}$, c'est-à-dire pour tout $\omega \in \Omega$, $X(\omega) = a$, alors $\mathbb{P}(X = a) = 1$ et $\mathbb{E}(X) = a$.
3. L'espérance d'une v.a.r. positive est positive. En particulier, si $X \geq Y$ (ce qui signifie que pour tout $\omega \in \Omega$, $X(\omega) \geq Y(\omega)$), alors $\mathbb{E}(X - Y) \geq 0$ donc $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

L'espérance d'une v.a.r. X est un indicateur de "localisation" de sa loi:
$$\mathbb{E}(X) \simeq \text{“valeur moyenne de X”}.$$
Néanmoins, la connaissance de l'espérance mathématique donne peu de renseignements sur cette v.a.r. Ainsi, il faut étudier "l'étalement" de sa loi, c'est-à-dire la dispersion de la v.a.r. X autour de sa moyenne $\mathbb{E}(X)$.

IV.1.3 Variance et écart-type

Pour rendre positifs les écarts entre X et son espérance $\mathbb{E}(X)$, un autre outil plus facile à manipuler que la valeur absolue, est à notre disposition: la mise au carré. On ne va donc pas calculer la moyenne des écarts mais la moyenne des écarts au carré. C'est ce qu'on appelle la variance.

Définition 18 *La variance de la v.a.r. X est la quantité:*
$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right].$$

Propriétés :

- $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
- $\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$ pour tout $a, b \in \mathbb{R}$. En particulier, $\mathrm{Var}(X+b) = \mathrm{Var}(X)$.

Afin d'être en mesure de comparer, en termes d'ordre de grandeur, variance et espérance, il faut prendre la racine carrée de la variance. C'est ce qu'on appelle l'écart-type.

Définition 19 *La racine carrée de* $\mathrm{Var}(X)$*, notée σ_X, est appelée écart-type de X.*

Remarque : Si X est une v.a.r. telle que $\mathbb{E}(X) = m$ et $\mathrm{Var}(X) = \sigma^2$, alors la variable $Y = (X - m)/\sigma$ est d'espérance nulle et de variance 1. On dit que Y est centrée (d'espérance nulle) et réduite (de variance 1).

Exemples :

- Loi de Bernoulli $\mathcal{B}(p)$: $\mathbb{E}(X) = p$ et $\mathrm{Var}(X) = p\,(1 - p)$.

- Loi binomiale $\mathcal{B}(n,p)$: $\mathbb{E}(X) = np$ et $\text{Var}(X) = np(1-p)$.
- Loi géométrique $\mathcal{G}(p)$: $\mathbb{E}(X) = \frac{1}{p}$ et $\text{Var}(X) = \frac{1-p}{p^2}$.
- Loi de Poisson $\mathcal{P}(\lambda)$: $\mathbb{E}(X) = \text{Var}(X) = \lambda$.
- Loi normale $\mathcal{N}(m,\sigma^2)$: $\mathbb{E}(X) = m$ et $\text{Var}(X) = \sigma^2$.
- Loi exponentielle $\mathcal{E}(\lambda)$: $\mathbb{E}(X) = \frac{1}{\lambda}$ et $\text{Var}(X) = \frac{1}{\lambda^2}$.

Exercice 1. Soit X une variable aléatoire de densité de probabilité

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

1. Donner les valeurs du paramètre θ pour que f soit bien une densité de probabilité.
2. Calculer l'espérance et la variance de X.

IV.2 Loi d'un couple de variables aléatoires

Pour une expérience donnée, on peut s'intéresser non seulement à la distribution de certaines v.a.r., mais aussi à la relation entre deux ou plusieurs v.a.r. Ainsi, pour étudier la relation entre deux v.a.r. X et Y, on définit la fonction de répartition du couple (X,Y) par

$$F_{X,Y}(x,y) = \mathbb{P}\left(\{X \leq x\} \cap \{Y \leq y\}\right), \quad \forall (x,y) \in \mathbb{R}^2.$$

Il arrive aussi que l'on note par une virgule l'intersection de 2 événements. Le nombre $F_{X,Y}(x,y)$ représente la probabilité que X soit inférieure ou égale à x et Y inférieure ou égale à y. En particulier, les fonctions de répartition des variables marginales sont déterminées par:

- $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\{X \leq x\} \cap \{Y \in \mathbb{R}\}\right)$,
- $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(\{X \in \mathbb{R}\} \cap \{Y \leq y\}\right)$.

IV.2.1 Cas des variables discrètes

Si X et Y sont deux v.a.r. discrètes ayant pour valeurs possibles respectives: $x_1, \ldots, x_n$ et $y_1, \ldots, y_p$, alors la loi du couple (X,Y) est déterminée par l'ensemble des probabilités:

$$p_{X,Y}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j), \quad i = 1, \ldots, n, \quad j = 1, \ldots, p.$$

Remarque: Si p_X et p_Y sont respectivement les lois de X et Y, alors

$$p_X(x_i) = \mathbb{P}(X = x_i) = \mathbb{P}\left(\cup_{j=1}^{+\infty}\{X = x_i, Y = y_j\}\right) = \sum_{j=1}^{+\infty} p_{X,Y}(x_i, y_j),$$

et de même, $p_Y(y_j) = \sum_{i=1}^{+\infty} p_{X,Y}(x_i, y_j)$.

IV.2.2 Cas des variables continues

On dit que le couple de v.a.r. (X, Y) est *continu* s'il existe une densité de probabilité $f_{X,Y} : \mathbb{R}^2 \to [0, +\infty)$ (donc d'intégrale 1 sur $\mathbb{R}^2$) telle que

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x, y)dxdy, \quad \forall A, B \subset \mathbb{R}.$$

Remarques :

- $\forall (a, b) \in \mathbb{R}^2$, $F_{X,Y}(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)dxdy$.
- $\mathbb{P}(X \leq Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{y} f_{X,Y}(x, y)dxdy$.
- Densités marginales: $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy$ et $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dx$.

IV.3 Indépendance de variables aléatoires

Définition 20 *Deux v.a.r. X et Y sont dites* ***indépendantes*** *si et seulement si*

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A) \times \mathbb{P}(X \in B), \quad \forall A, B \subset \mathbb{R}.$$

On peut montrer que l'indépendance est équivalente à

$$\mathbb{P}(\{X \leq a\} \cap \{Y \leq b\}) = \mathbb{P}(X \leq a) \times \mathbb{P}(Y \leq b), \quad \forall (a, b) \in \mathbb{R}^2,$$

ou encore en termes de fonctions de répartition:

$$F_{X,Y}(a, b) = F_X(a)F_Y(b), \quad \forall (a, b) \in \mathbb{R}^2.$$

Théorème 2 *Soient X et Y deux v.a.r.*

- *Cas discret: X et Y sont indépendantes si et seulement si pour tout couple $(x, y) \in \mathcal{X} \times \mathcal{X}$, on a $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.*
- *Cas continu: X et Y sont indépendantes si et seulement si pour tout couple $(s, t) \in \mathbb{R}^2$, on a $f_{X,Y}(s, t) = f_X(s)f_Y(t)$.*

Exemple : On lance une paire de dés et on note X (respectivement Y) le minimum (respectivement maximum) des deux nombres sortis. On remarque que X et Y peuvent prendre toutes deux des valeurs de 1 à 6, mais $\mathbb{P}(X = i \cap Y = j) = 0$, pour tout $i > j$. Les variables X et Y ne sont donc pas indépendantes.

Chapitre V

Théorèmes limites

Deux théorèmes mathématiques ont une place particulière en théorie des probabilités et en statistiques: la loi des grands nombres et le théorème central limite. Ils interviennent dans l'étude de phénomènes aléatoires comportant un grand nombre de v.a.r. indépendantes de même loi. Par exemple, pour le premier cité, il apparaît lorsque l'on étudie la proportion de "pile" dans un jeu de pile ou face, ou encore la moyenne de lancers de dé successifs. Quant au second, il nous donne de façon informelle une estimation précise de l'erreur que l'on commet en approchant l'espérance mathématique par la moyenne arithmétique.

V.1 Inégalité de Chebychev

Théorème 3 *(inégalité de Chebychev)*
Soit $\epsilon > 0$ et soit X une v.a.r. admettant une variance. Alors on a:

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \epsilon\right) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

Cette inégalité permet de comprendre la signification de l'écart-type $\sigma_X = \sqrt{\mathrm{Var}(X)}$, au sens où il caractérise la dispersion de la v.a.r. autour de son espérance mathématique:

$$\begin{aligned}
\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \epsilon\sigma_X\right) &\leq \frac{1}{\epsilon^2} \\
\mathbb{P}\left(|X - \mathbb{E}(X)| < \epsilon\sigma_X\right) &\geq 1 - \frac{1}{\epsilon^2}.
\end{aligned}$$

Supposons que $\epsilon = 10$. Alors l'événement $\{|X - \mathbb{E}(X)| \geq 10\sigma_X\}$ a peu de chances de se réaliser car

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq 10\sigma_X\right) \leq \frac{1}{100}.$$

Supposons maintenant que $\sigma_X = 0$. Alors nous obtenons pour tout $\epsilon > 0$

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \epsilon\right) \leq 0.$$

Par conséquent

$$\mathbb{P}\left(|X - \mathbb{E}(X)| > 0\right) = 0,$$

et donc X est presque sûrement égale à $\mathbb{E}(X)$.

On doit cependant remarquer que, malgré son intérêt théorique certain, l'inégalité de Chebychev présente peu d'intérêt en pratique, car ne faisant pas intervenir la loi de probabilité suivie par la v.a.r. considérée, elle donne une majoration de la probabilité beaucoup trop grande.

V.2 Loi (faible) des grands nombres

Théorème 4 *(LGN) Soient $X_1, \ldots, X_n$ des v.a.r. indépendantes, de même loi, et admettant une variance. On note $m = \mathbb{E}(X_1)$. Alors, pour tout $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - m\right| > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

*Dans ce cas, on dit que la moyenne arithmétique $\frac{X_1+\ldots+X_n}{n}$ **converge en probabilité** vers l'espérance mathématique m lorsque n tend vers $+\infty$.*

V.3 Théorème central limite

On a vu que deux v.a.r. ont la même loi si et seulement si leur fonctions de répartition sont égales. Ainsi, la fonction de répartition est souvent utilisée en pratique afin de démontrer l'égalité en loi. On est donc amené à définir la convergence en loi comme la convergence des fonctions de répartition associées.

Définition 21 *Soit $(Y_n)_{n\in\mathbb{N}}$ une suite de v.a.r. et soit Y une v.a.r. On dit que $(Y_n)_{n\in\mathbb{N}}$ **converge en loi** vers Y si pour tout x_0 point de continuité de la fonction de répartition F_Y de Y,*

$$F_{Y_n}(x_0) = \mathbb{P}(Y_n \leq x_0) \underset{n\to+\infty}{\longrightarrow} F_Y(x_0) = \mathbb{P}(Y \leq x_0).$$

On note la convergence en loi $Y_n \xrightarrow{\mathcal{L}} Y$.

La convergence en loi est réalisée aux points de continuité de F_Y. C'est la convergence simple de la suite de fonctions de répartition F_{Y_n}.

Propriété d'additivité de la loi normale: si $X_1, \ldots, X_n$ sont des v.a.r. indépendantes et de même loi $\mathcal{N}(m, \sigma^2)$, alors la v.a.r. $X_1 + \ldots + X_n$ suit la loi $\mathcal{N}(nm, n\sigma^2)$. Ce résultat implique que la v.a.r. centrée réduite $\sum_{i=1}^{n} \frac{X_i - m}{\sigma\sqrt{n}}$ suit la loi normale $\mathcal{N}(0, 1)$.

Que se passe-t-il dans le cas général où les v.a.r. X_i ne sont pas nécessairement normales? Le résultat ci-dessus se transforme alors en un résultat de convergence en loi.

Théorème 5 *(TCL) Soient $X_1, \ldots, X_n$ des v.a.r. indépendantes, de même loi, et admettant une variance. On note $m = \mathbb{E}(X_1)$ et $\sigma^2 = \mathrm{Var}(X_1)$. Alors*

$$\frac{X_1 + \ldots + X_n - nm}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \textit{lorsque} \quad n \to +\infty.$$

V.4 Approximation d'une loi binomiale

On contrôle n pièces et on introduit les v.a.r. $X_1, \ldots, X_n$ définies par $X_i = 1$ si la ième pièce contrôlée est défectueuse, et 0 sinon. On note $Y = X_1 + \ldots X_n$ le nombre total de pièces défectueuses dans le lot. Alors la v.a.r. Y suit une loi binomiale de paramètres (n, p) où p est la probabilité qu'une pièce soit défectueuse.

V.4.1 Approximation par une loi normale

Puisque les v.a.r. X_i sont indépendantes, de même loi, et de variance finie:

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1 - p, \quad \mathbb{E}(X_i) = p \quad \text{et} \quad \mathrm{Var}(X_i) = p(1-p),$$

on peut appliquer le TCL:

$$\frac{X_1 + \ldots + X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1) \quad \text{lorsque} \quad n \to +\infty.$$

On peut donc approcher la loi de la v.a.r. $(Y - np)/\sqrt{np(1-p)}$ par une loi normale $\mathcal{N}(0,1)$. Ceci revient à approcher la loi de Y par une loi $\mathcal{N}(np, np(1-p))$. En pratique, on utilise cette approximation si

$$\min(np, n(1-p)) \geq 5.$$

Ainsi, si $p = 1/2$, l'approximation est correcte pour $n \geq 10$, par contre si $p = 1/100$, il faut $n \geq 500$ pour pouvoir l'utiliser.

Deux remarques importantes doivent être faites. Tout d'abord, une v.a.r. binomiale est une variable discrète à valeurs dans $\{0, \ldots, n\}$, alors qu'une v.a.r. normale est continue et à valeurs dans $\mathbb{R}$. Par ailleurs, dans le cas d'une loi binomiale, un point a une probabilité non nulle alors que dans le cas d'une loi normale, un point est un ensemble de probabilité nulle. Pour ces deux raisons, il faut faire une "correction de continuité" quand on utilise l'approximation d'une loi binomiale par une loi normale.

La représentation graphique de la densité d'une variable aléatoire de loi normale est une courbe en "cloche" alors que celle d'une variable aléatoire binomiale est un diagramme en "bâtons". Une valeur approchée de $\mathbb{P}(X = k)$ est donnée par l'aire comprise entre la courbe en cloche et les droites d'abcisses $k - 0,5$ et $k + 0,5$. On obtient alors

$$\mathbb{P}(X = k) \simeq \mathbb{P}\left(\frac{k - 0,5 - np}{\sqrt{np(1-p)}} < Z < \frac{k + 0,5 - np}{\sqrt{np(1-p)}}\right),$$

ou encore

$$\mathbb{P}(X \leq k) \simeq \mathbb{P}\left(Z < \frac{k + 0,5 - np}{\sqrt{np(1-p)}}\right).$$

V.4.2 Approximation par une loi de Poisson

Lorsque p est très petit, on utilise plutôt l'approximation de la loi binomiale par une loi de Poisson, qui est satisfaisante pour $p < 0,1$ et $n > 50$.

Proposition 7 *Soit Y_n une v.a.r. binomiale de paramètres (n, p). On suppose que $n \to +\infty$ et $p = \lambda/n$, où $\lambda > 0$. Alors, pour tout $k \in \mathbb{N}$,*

$$\lim_{n \to +\infty} P(Y_n = k) = \exp(-\lambda)\frac{\lambda^k}{k!}.$$

Chapitre VI

Estimation

VI.1 Introduction

Le cadre est le suivant: on dispose d'un nombre fini d'observations sur une population donnée. L'objectif est de pouvoir tirer des conclusions sur certaines caractéristiques de cette population à partir de ces observations. On fait alors une hypothèse raisonnable: il existe une loi de probabilité sous-jacente telle que les "valeurs observables" des différents éléments de la population étudiée puissent être considérées comme des variables aléatoires indépendantes ayant cette loi.

Un aspect important de l'inférence statistique consiste à obtenir des "estimations fiables" des caractéristiques d'une population (de grande taille) à partir d'un échantillon (de petite taille) extrait de cette population. C'est un problème de décision concernant des paramètres qui le plus souvent sont:

- une espérance mathématique m;
- une proportion p;
- une variance σ^2.

Ces paramètres sont a priori inconnus car la taille réelle de la population étant très grande, il serait trop coûteux de tester tous les éléments de la population. Ainsi, comme un échantillon ne peut donner qu'une information partielle sur la population, les estimations que l'on obtiendra seront inévitablement entachées d'erreurs qu'il s'agit d'évaluer et de minimiser autant que possible.

En résumé, estimer un paramètre inconnu, c'est en donner une valeur approchée à partir des résultats obtenus sur un échantillon aléatoire extrait de la population sous-jacente.

Exemple: Une usine produit des composants éléctroniques. Le fabriquant désire connaître la proportion p de composants défectueux à l'issue du processus de fabrication. Pour obtenir la valeur exacte de p, il suffirait de tester chacun des composants, opération qui peut se révéler extrêmement couteuse en temps ou en argent (voire de facto impossible). Une alternative intéressante consiste à ne tester qu'un nombre fini n de composants prélevés au hasard sur la chaîne de montage. On obtient un ensemble de n réalisations x_i de v.a.r. indépendantes de Bernoulli X_i définies par:

$$X_i = \begin{cases} 1 \text{ si le composant } i \text{ est défectueux} \\ 0 \text{ sinon.} \end{cases}$$

Il est naturel d'estimer p, la proportion d'objets defectueux dans **la population**, par $\bar{x_n} = \frac{1}{n}\sum_{i=1}^n x_i$, correspondant à la proportion de composants defectueux dans **l'échantillon**. En effet, la LGN nous assure de la convergence en probabilité de la v.a.r. $\frac{1}{n}\sum_{i=1}^n X_i$ vers l'espérance de X_1, c'est-à-dire p. On estimera donc p par $\bar{x_n}$. C'est une estimation dite ponctuelle.

Le but des statistiques n'est pas simplement de donner une estimation ponctuelle d'un paramètre d'intérêt (l'estimateur étant aléatoire... et donc bien souvent faux par nature), mais surtout de donner des garanties sur la précision de l'estimation finale. Dans ce sens, le paragraphe VI.2 liste les propriétés que l'on peut attendre a minima d'un estimateur. La section VI.3 détaille quand à elle la construction d'intervalles de confiance.

VI.2 Estimation ponctuelle

VI.2.1 Propriétés d'un (bon) estimateur

a) Consistance

Définition 22 *Un n-échantillon aléatoire issu d'une v.a.r. X est un ensemble $(X_1, \ldots, X_n)$ de n v.a.r. indépendantes et de même loi que X.*

Soit θ un paramètre associé à la loi de X, par exemple $\theta = \mathbb{E}(X)$ ou $\theta = \mathrm{Var}(X)$. À partir de l'observation d'un échantillon aléatoire $(X_1, \ldots, X_n)$, on souhaite estimer le paramètre θ.

Définition 23 *Un estimateur $\hat{\theta}_n$ de θ est une fonction qui dépend uniquement du n-échantillon $(X_1, \ldots, X_n)$. Il est dit* consistant *s'il est "proche" de θ au sens de la convergence en probabilité, i.e. pour tout $\epsilon > 0$,*

$$\mathbb{P}\left(|\hat{\theta}_n - \theta| > \epsilon\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Dans l'exemple de l'introduction, la quantité $\frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur consistant de p et si, par exemple, on a observé 21 pièces défectueuses sur un lot de 1500 pièces prélevées, l'estimation ponctuelle de p obtenue est $\bar{x_n} = 21/1500 = 1,4\%$.

De manière plus générale, pour estimer l'espérance m des variables aléatoires X_i, on utilise souvent la moyenne empirique

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

La LGN garantit en effet la convergence en probabilité vers l'espérance $m = \mathbb{E}(X_1)$.

Exemple: Considérons une v.a.r. X représentant le nombre de grippes attrapées par une personne en un an. On peut supposer que X suit une loi de Poisson de paramètre $\lambda > 0$. Chercher la loi de X, c'est chercher λ, qui n'est autre que l'espérance mathématique de X. Par conséquent, la LGN nous indique que $\bar{X}_n$ est un estimateur consistant de λ: pour tout $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \lambda\right| \geq \epsilon\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

L'inégalité de Chebychev permet quant à elle de démontrer le théorème suivant:

Théorème 6 *Soit $\hat{\theta}_n$ un estimateur de θ. Si l'on a:*

$$\lim_{n\to+\infty} \mathbb{E}(\hat{\theta}_n) = \theta \quad \text{et} \quad \lim_{n\to+\infty} \mathrm{Var}(\hat{\theta}_n) = 0,$$

alors $\hat{\theta}_n$ est un estimateur consistant de θ.

b) Biais

Définition 24 *Soit $\hat{\theta}_n$ un estimateur consistant d'un paramètre θ. On appelle **biais** la quantité $\mathbb{E}(\hat{\theta}_n) - \theta$. L'estimateur $\hat{\theta}_n$ est dit **sans biais** si $\mathbb{E}(\hat{\theta}_n) = \theta$, et **biaisé** sinon.*

Exemple: La moyenne empirique $\bar{X}_n$ est un estimateur consistant et sans biais de l'espérance mathématique m.

c) Écart quadratique moyen

Notons que l'on a

$$\begin{aligned}
\mathbb{E}\left\{(\hat{\theta}_n - \theta)^2\right\} &= \mathbb{E}\left\{(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n) + \mathbb{E}(\hat{\theta}_n) - \theta)^2\right\} \\
&= \mathbb{E}\left\{(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))^2 + (E(\hat{\theta}_n) - \theta)^2 + 2(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))(E(\hat{\theta}_n) - \theta)\right\} \\
&= \mathrm{Var}(\hat{\theta}_n) + (\text{biais})^2,
\end{aligned}$$

car le terme $\mathbb{E}\left\{(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))(\mathbb{E}(\hat{\theta}_n) - \theta)\right\}$ est nul. Ainsi, pour rendre l'écart quadratique moyen $\mathbb{E}\left\{(\hat{\theta}_n - \theta)^2\right\}$ le plus petit possible, il faut que

- $\mathbb{E}(\hat{\theta}_n) = \theta$, donc choisir un estimateur sans biais,
- la variance $\mathrm{Var}(\hat{\theta}_n)$ soit faible.

On choisira donc, parmi les estimateurs consistants et sans biais, celui qui a la variance la plus petite. En d'autres termes, si $\hat{\theta}_n$ est un estimateur consistant et sans biais de θ, on a tout intérêt à ce que $\hat{\theta}_n$ ne varie pas trop autour de sa moyenne. Cette propriété traduit ce que l'on appelle l'efficacité de l'estimateur.

La théorie de l'estimation, qui a émergé au cours du XXème siécle permet, dans des cadres assez généraux, de choisir parmi toutes les statistiques possibles le 'meilleur' estimateur consistant, c'est-à-dire celui qui donnera une estimation ponctuelle la plus proche possible du paramètre d'intérêt et ceci quel que soit l'échantillon.

VI.2.2 Estimateur d'une moyenne m ou d'une proportion

On considère un n-échantillon $(X_1, \ldots, X_n)$ issu d'une loi de moyenne m et de variance σ^2, toutes deux inconnues.

1. d'après la LGN, la moyenne empirique $\bar{X}_n$ est un estimateur consistant de m.
2. l'estimateur $\bar{X}_n$ est sans biais.
3. par indépendance: $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.
4. loi de $\bar{X}_n$:
   - si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\bar{X}_n \sim \mathcal{N}(m, \sigma^2/n)$.
   - lorsque n est grand, d'après le TCL, la loi de $\bar{X}_n$ est approximée par une loi normale $\mathcal{N}(m, \sigma^2/n)$.

L'estimation d'une proportion p est un cas particulier du précédent, au sens où les v.a.r. X_i considérées sont de Bernoulli de paramètre p.

VI.2.3 Estimateur de la variance

Définition 25 *La variance empirique associée à un n-échantillon* $(X_1, \ldots, X_n)$ *est définie par*

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

Définition 26 *Soit* $(Y_1, \ldots, Y_n)$ *un n-échantillon de v.a.r. de loi* $\mathcal{N}(0,1)$. *On appelle* ***loi du chi-deux*** *à* n *degrés de liberté la loi de la v.a.r.* $\sum_{i=1}^n Y_i^2$, *et on la note* $\chi^2_{(n)}$.

Propriétés de la variance empirique:

1. S_n^2 est un estimateur consistant de la variance σ^2.
2. S_n^2 est sans biais.
3. loi de S_n^2: pas de résultat général. Cependant, si $X \sim \mathcal{N}(m, \sigma^2)$, alors la v.a.r. $\frac{n-1}{\sigma^2} S_n^2$ suit une loi du chi-deux à $n-1$ degrés de libertés $\chi^2_{(n-1)}$.

Remarque: Puisque $\mathbb{E}(Y_i) = 0$, on a $\mathbb{E}(Y_i^2) = \mathrm{Var}(Y_i) = 1$. Si V suit une loi $\chi^2_{(n)}$, alors

$$\mathbb{E}(V) = \mathbb{E}(Y_1^2 + \ldots + Y_n^2) = n.$$

Ainsi on retrouve le fait que S_n^2 est un estimateur consistant et sans biais de σ^2:

$$\mathbb{E}(S_n^2) = \frac{\sigma^2}{n-1}\mathbb{E}(\chi^2_{n-1}) = \sigma^2.$$

VI.3 Estimation par intervalle de confiance

Les estimations ponctuelles n'apportent pas d'information sur la précision des résultats, c'est-à-dire qu'elles ne tiennent pas compte des fluctuations d'échantillonnage. Pour évaluer la confiance que l'on peut avoir en une valeur, il est nécessaire de déterminer un intervalle contenant, avec une certaine probabilité fixée au préalable, la vraie valeur du paramètre: c'est l'estimation par intervalle de confiance.

Exemple. On considére une échantillon d'observations $X_1, \ldots, X_n$ tel que $X_i \sim \mathcal{N}(m, \sigma^2)$ pour tout $i \in \{1, \ldots, n\}$, le paramètre m étant inconnu. On a vu que $\bar{X}_n \sim \mathcal{N}(m, \sigma^2/n)$. En particulier, on a donc $P(\bar{X}_n = m) = 0$: la probabilité que notre estimation soit juste... est égale à 0!! Pourtant, on sent bien que $\bar{X}_n$ se rapproche de m quand n croît: il s'agit donc d'essayer de quantifier le crédit que l'on peut accorder à notre estimateur.

Soit $(X_1, \ldots, X_n)$ un n-échantillon aléatoire et θ un paramètre inconnu de la loi des X_i.

Définition 27 *Soit $\alpha \in]0, 1[$. S'il existe des v.a.r. $\theta_{\min}(X_1, \ldots, X_n)$ et $\theta_{\max}(X_1, \ldots, X_n)$ telles que*

$$\mathbb{P}\left(\theta \in [\theta_{\min}(X_1, \ldots, X_n), \theta_{\max}(X_1, \ldots, X_n)]\right) = 1 - \alpha,$$

on dit alors que $[\theta_{\min}(X_1, \ldots, X_n), \theta_{\min}(X_1, \ldots, X_n)]$ est un intervalle de confiance pour θ, avec coefficient de sécurité (ou de niveau de confiance) $1 - \alpha$. On le note $IC_{1-\alpha}(\theta)$.

Dans la pratique, on peut prendre par exemple $\alpha = 5\%$, ce qui nous donne un IC à 95%. Cela signifie qu'il y a 95% de chance que la valeur inconnue θ soit comprise entre $\theta_{\min}(x_1, \ldots, x_n)$ et $\theta_{\max}(x_1, \ldots, x_n)$.

VI.3.1 Intervalle de confiance pour la moyenne d'un échantillon gaussien (variance connue)

Soit $(X_1, \ldots, X_n)$ un n-échantillon de v.a.r. de loi $\mathcal{N}(m, \sigma^2)$. Afin d'estimer m, on utilise la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ qui a pour loi $\mathcal{N}(m, \sigma^2/n)$. Il en résulte que

$$\sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right) \sim \mathcal{N}(0, 1),$$

et que

$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right) \leq z_{1-\alpha/2}\right) = 1 - \alpha,$$

où $z_{1-\alpha/2}$ désigne le $1-\alpha/2$-quantile d'une variable gaussienne centrée réduite. Ceci équivaut à

$$\mathbb{P}\left(\bar{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

On obtient donc un IC pour l'espérance m avec coefficient de sécurité $1-\alpha$ dans le cas où σ est connu: il s'agit de l'intervalle aléatoire

$$\left[\bar{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Ainsi, dans les calculs, l'IC est donné par

$$IC_{1-\alpha}(m) = \left[\bar{x_n} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x_n} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

où $\bar{x}_n$ est l'estimation ponctuelle de m associée à la réalisation du n-échantillon $(X_1, \ldots, X_n)$.

VI.3.2 Intervalle de confiance pour la moyenne d'un échantillon gaussien (variance inconnue)

On estime la variance σ^2, supposée inconnue, par la variance empirique

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

Pour la construction d'un intervalle de confiance dans ce cadre, il est alors nécessaire de remplacer dans les formules précédentes la variance par sa version emprique S_n^2. Il faut donc considérer non plus la quantité $\sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right)$ mais plutot

$$\sqrt{n}\left(\frac{\bar{X}_n - m}{S_n}\right),$$

qui ne suit plus une loi normale mais une loi dite **de Student** à $n-1$ degrés de liberté, que l'on note $\mathcal{T}_{n-1}$. La densité de la loi de Student est une fonction paire, comme la loi normale $\mathcal{N}(0,1)$. On dispose de tables pour obtenir les quantiles de cette loi. On en déduit donc que

$$\mathbb{P}\left(-t_{1-\alpha/2} \leq \sqrt{n}\left(\frac{\bar{X}_n - m}{S_n}\right) \leq t_{1-\alpha/2}\right) = 1-\alpha,$$

ce qui équivaut à

$$\mathbb{P}\left(\bar{X}_n - t_{1-\alpha/2}\frac{S_n}{\sqrt{n}} \leq m \leq \bar{X}_n + t_{1-\alpha/2}\frac{S_n}{\sqrt{n}}\right) = 1-\alpha.$$

On obtient donc un IC pour m avec coefficient de sécurité $1-\alpha$, dans le cas où la variance σ^2 est inconnue: il s'agit de l'intervalle aléatoire

$$\left[\bar{X}_n - t_{1-\alpha/2}\frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{1-\alpha/2}\frac{S_n}{\sqrt{n}}\right].$$

Ainsi, dans les calculs, l'IC est donné par

$$IC_{1-\alpha}(m) = \left[\bar{x_n} - t_{1-\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x_n} + t_{1-\alpha/2}\frac{s_n}{\sqrt{n}}\right],$$

où $\bar{x_n}$ et s_n^2 sont les estimations ponctuelles respectives de la moyenne m et de la variance σ^2.

VI.3.3 Intervalle de confiance pour une proportion

Revenons à l'exemple de l'introduction de ce chapitre: on cherche à estimer la proportion p de composants défectueux produits par une usine. On prélève un lot de n composants et on note X_i la v.a.r. qui vaut 1 si la pièce i est défectueuse, et 0 sinon. On estime p par la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Les v.a.r. X_i étant de Bernoulli, on peut alors utiliser l'approximation donnée par le TCL. Soit $Z \sim \mathcal{N}(0,1)$ et $z_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$. Par le TCL,

$$\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1), \text{ lorsque } n \to +\infty.$$

Ceci implique que

$$\mathbb{P}\left(-z_{1-\alpha/2} \le \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \le z_{1-\alpha/2}\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) = 1-\alpha,$$

c'est-à-dire

$$\mathbb{P}\left(\bar{X}_n - z_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}} \le p \le \bar{X}_n + z_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha.$$

Ceci ne fournit pas un IC pour p car les bornes de l'intervalle dépendent de p. Mais on peut montrer que l'on a le même résultat de convergence, en remplaçant p dans les bornes de l'intervalle par son estimateur consistant $\bar{X}_n$. On obtient alors

$$\mathbb{P}\left(\bar{X}_n - z_{1-\alpha/2}\frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} \le p \le \bar{X}_n + z_{1-\alpha/2}\frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha.$$

On dit que l'intervalle

$$\left[\bar{X}_n - z_{1-\alpha/2}\frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha/2}\frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}\right]$$

est un IC asymptotique pour le paramètre p, de coefficient de sécurité $1-\alpha$.
Pour $\alpha = 5\%$, on lit dans les tables $z_{1-\alpha/2} = z_{97,5\%} = 1,96$. Ainsi, on en déduit que si l'on a observé en pratique 21 pièces défectueuses sur 1500 (c'est-à-dire que l'on remplace dans les calculs la moyenne empirique aléatoire $\bar{X}_n$ par l'estimation ponctuelle

$\bar{x_n} = 21/1500$), l'intervalle de confiance asymptotique pour p est $[0.0081, 0.0199]$.

Remarque: Si les v.a.r. $X_1, \ldots, X_n$ ne sont pas gaussiennes mais que n est assez grand (en pratique supérieur à 30), alors le TCL nous garantit que la moyenne empirique suit approximativement une loi gaussienne. Ceci permet de considérer un très large éventail de situations. Dans le cas particulier où l'intervalle de confiance dépendrait d'une quantité inconnue, il est toujours possible de remplacer cette dernière par un estimateur consistant: la convergence en loi est conservée (lemme de Slutsky).

VI.3.4 Intervalle de confiance pour la variance d'une population gaussienne

On sait que

$$S_n^2 \sim \frac{\sigma^2}{n-1}\chi^2_{(n-1)} \text{ et que } \mathbb{E}(S_n^2) = \frac{\sigma^2}{n-1}\mathbb{E}(\chi^2_{(n-1)}) = \sigma^2,$$

c'est-à-dire que S_n^2 est un estimateur sans biais de σ^2. Qui plus est, la LGN nous donne la consistance de S_n^2. De plus, on lit dans des tables les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ de la loi du $\chi^2_{(n-1)}$, respectivement notés $v_{\alpha/2}$ et $v_{1-\alpha/2}$ (il est normal que les quantiles qui nous intéressent ne soient pas opposés car la densité de cette loi n'est pas paire, à l'inverse de la loi normale centrée réduite). On obtient alors

$$\mathbb{P}\left(v_{\alpha/2} \leq \frac{n-1}{\sigma^2}S_n^2 \leq v_{1-\alpha/2}\right) = 1-\alpha.$$

Ceci équivaut à

$$\mathbb{P}\left(\frac{(n-1)S_n^2}{v_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{v_{\alpha/2}}\right) = 1-\alpha.$$

On obtient donc un IC pour la variance σ^2 avec coefficient de sécurité $1-\alpha$: il s'agit de l'intervalle aléatoire

$$\left[\frac{(n-1)S_n^2}{v_{1-\alpha/2}}, \frac{(n-1)S_n^2}{v_{\alpha/2}}\right].$$

Ainsi, dans les calculs, l'IC est donné par

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s_n^2}{v_{1-\alpha/2}}, \frac{(n-1)s_n^2}{v_{\alpha/2}}\right],$$

où s_n^2 est l'estimation ponctuelle de σ^2 associée à la réalisation du n-échantillon $(X_1, \ldots, X_n)$:

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x_n})^2.$$

Chapitre VII

Tests d'hypothèses

La théorie des tests propose une démarche générale permettant:

- de confronter une hypothèse avec la réalité, ou plus exactement, avec ce que l'on perçoit de la réalité à travers les observations à disposition;
- de prendre une décision à la suite de cette confrontation.

Si les problèmes traités par l'estimation (ponctuelle ou par intervalle de confiance) sont de type quantitatif, i.e. conduisent à un résultat numérique, ceux traités par les tests d'hypothèses sont d'ordre qualitatif, i.e. conduisent à une réponse du type rejet/acceptation de l'hypothèse statistique effectuée.

VII.1 Exemple introductif

Des relevés effectués pendant de nombreuses années ont permis d'établir que le niveau naturel des pluies dans la Beauce X en millimètres par an suit une loi normale $\mathcal{N}(600, 100^2)$. Des entrepreneurs, surnommés faiseurs de pluie, prétendaient pouvoir augmenter de $50mm$ le niveau moyen de pluie, ceci par insémination des nuages au moyen d'iodure d'argent. Leur procédé fut mis à l'essai entre 1951 et 1959 et on releva les hauteurs de pluies suivantes:

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959
mm	510	614	780	512	501	534	603	788	650

Que pouvait-on en conclure? Deux hypothèses s'affrontaient: ou bien l'insémination était sans effet, ou bien elle augmentait réellement le niveau moyen de pluie de $50mm$. Ces hypothèses pouvaient se formaliser comme suit, si m désigne l'espérance mathématique de X, v.a.r. égale au niveau annuel de pluie.

$$\left\{ \begin{array}{l} H_0 : m = 600mm \\ H_1 : m = 650mm. \end{array} \right.$$

Les agriculteurs hésitaient à opter pour le procédé forcément onéreux des faiseurs de pluie. Ainsi, il fallait donc que l'expérience puisse les convaincre, c'est-à-dire que les faits observés contredisent nettement l'hypothèse H_0, dite "hypothèse nulle" (H_1 s'appelle "hypothèse alternative"). Les agriculteurs n'étaient donc décidés à abandonner H_0 qu'en présence de faits expérimentaux traduisant une éventualité improbable compte-tenu de H_0.

Par essence, l'objectif d'un test n'est pas de déterminer si H_0 est fondamentalement vraie ou non, mais plutôt de voir si H_0 est une hypothèse cohérente avec les données observées. On souhaite donc voir H_0 rejetée uniquement dans le cas où les observations la rendent invraisemblable. Ce qui amène généralement à la pratique suivante: on fixe une valeur $\alpha \in]0;1[$, appelée risque de première espèce, et l'on impose au test d'être tel que si H_0 est vraie, la probabilité de rejeter à tort H_0 soit inférieure à α.

Dans notre exemple, les agriculteurs choisirent $\alpha = 0,05$, c'est-à-dire qu'ils assumaient le risque de se tromper dans 5 cas sur 100, en croyant les promesses des faiseurs de pluie, alors que ces faiseurs de pluie étaient en réalité des charlatans.

Comment décider? Puisqu'il s'agit de "tester" la moyenne m, il est naturel de s'intéresser à la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ que l'on sait consistante vers m par la LGN. Pour ce test, $\bar{X}_n$ est appelée *la statistique de test*. Alors, si H_0 est vraie, comme l'expérience porte sur $n = 9$ années, on a $\bar{X}_n \sim \mathcal{N}(600, 100^2/9)$.

Les données relevées indiquent que $\bar{x}_n = 610,2mm$ (c'est la moyenne obtenue sur toutes les années). L'hypothèse H_1 proposant une moyenne m alternative supérieure à 600, il est naturel de prendre la règle de décision suivante:

- Si $\bar{X}_n$ est plus grande qu'un seuil t (à déterminer), on rejette H_0,
- Si $\bar{X}_n < t$, on accepte H_0.

Reste maintenant à choisir le paramètre t de telle sorte que l'erreur de première espèce soit inférieure ou égale à 5%, autrement dit

$$P_{H_0}(\text{rejeter } H_0) = P_{H_0}(\bar{X}_n > t) \leq 5\%.$$

Il est alors facile de calculer cette dernière probabilité car sous H_0, la v.a.r. $Z = \frac{\bar{X}_n - 600}{100/3}$ suit une loi normale $\mathcal{N}(0,1)$:

$$\begin{aligned}
\mathbb{P}_{H_0}(\bar{X}_n > t) &= \mathbb{P}\left(Z > \frac{t-600}{100/3}\right) = 5\%, \\
&\Leftrightarrow \frac{t-600}{100/3} = 1,64, \\
&\Leftrightarrow t = 600 + 1,64 \times \frac{100}{3} \simeq 655.
\end{aligned}$$

Conclusion: Si la valeur observée de $\bar{X}_n$ est supérieure à 655, on pourra en conclure que l'amélioration est significative d'un point de vue statistique (et on a au maximum 5% de chances de se tromper en prenant cette décision). Si par contre, $\bar{x}_n < 655$, on pourra emettre des doutes sur l'évolution de la pluviométrie. Sur les données de l'énoncé, on trouve $\bar{x}_n = 610,2$: on ne rejette donc pas H_0.

VII.2 Démarche générale pour la construction d'un test

Dans un test d'hypothèse statistique, il y a deux manières de se tromper:

1 - la possibilité de rejeter l'hypothèse H_0 alors qu'elle est vraie. Le α donné précédemment borne la probabilité de se tromper dans ce sens, c'est le risque de première espèce.

2 - la possibilité d'accepter H_0 alors qu'elle est fausse. La probabilité de se tromper dans ce sens est le risque de deuxième espèce, et est notée β.

Ainsi, dans notre exemple, la première façon de se tromper est de croire les faiseurs de pluie, alors qu'ils ne sont pour rien dans le résultat obtenu (rejeter H_0 alors qu'elle est vraie), tandis que la seconde manière de se tromper est de ne pas croire les faiseurs de pluie, alors que leur méthode est bonne et que seul le hasard (malencontreux pour eux), dû au faible nombre d'observations, a donné des résultats insuffisants pour convaincre les agriculteurs (accepter H_0 alors qu'elle est fausse).

Supposons maintenant que les faiseurs de pluie ont raison. Alors $\bar{X_n}$ suit une loi normale $\mathcal{N}(650, 100^2/9)$, c'est-à-dire que la moyenne en millimètres n'est plus 600 mais 650, sous l'hypothèse H_1. Ainsi, la probabilité de se tromper est:

$$\beta = \mathbb{P}_{H_1}(\bar{X_n} < 655) = \mathbb{P}\left(Z < \frac{655 - 650}{100/3}\right) = \mathbb{P}(Z < 0,15) = 0,56,$$

ce qui est considérable.

Il faut remarquer au cours de cet exemple le rôle particulier joué par H_0. Si la forme de la région de rejet est indiquée par la nature de H_1 (650 plus grand que 600), la valeur de t ne dépend que de H_0. De plus, les deux hypothèses ne jouent pas des rôles symétriques, t étant déterminé par H_0 et α, tandis que β est déterminé par la considération supplémentaire de H_1.

Enfin il faut remarquer que les deux types d'erreur possibles ne sont en général pas du tout de la même importance. En effet, rejeter l'hypothèse H_0 alors qu'elle est vraie (risque de première espèce, qui est maîtrisé) est beaucoup plus coûteux que de la conserver à tort (risque de deuxième espèce, non maîtrisé).

Exemple: Soit m la moyenne du niveau de radioactivité en picocuries par litre. La valeur $m = 5$ est considérée comme la valeur critique entre eau potable et non potable. On peut donc souhaiter tester H_0 : "$m \geq 5$" contre H_1 : "$m < 5$". L'erreur de première espèce a alors pour conséquence de laisser boire de l'eau toxique, alors que l'erreur de deuxième espèce conduit seulement à jeter de l'eau potable ... Dans tous les cas, il faut trouver une procédure de test qui minimise ces deux types d'erreur.

De manière très générale, on retiendra le schéma suivant lors de l'élaboration d'un test statistique:

1. Choix de H_0 et de H_1. Fixer α.
2. Détermination de la statistique de test.
3. Construction d'une règle de décision **cohérente avec les hypothèses**.
4. Calibrage de cette règle en fonction de α et H_0.
5. Calcul de la valeur observée de la statistique de test.
6. Conclusion: rejet ou acceptation de H_0 au risque α.
7. Calcul de la puissance: $1-\beta$.

VII.3 Test sur la moyenne d'un échantillon gaussien (variance connue)

On suppose que l'on observe les réalisations d'un n-échantillon $(X_1,\dots,X_n)$ issu d'une loi $\mathcal{N}(m,\sigma^2)$ avec m inconnue et σ^2 connue. On se donne donc l'hypothèse nulle H_0 : "$m=m_0$", m_0 étant une valeur donnée par l'énoncé ou provenant de connaissances théoriques. Ensuite, il existe trois types d'hypothèses alternatives H_1:

- H_1 : "$m>m_0$", test unilatéral à droite;
- H_1 : "$m<m_0$", test unilatéral à gauche;
- H_1 : "$m\neq m_0$", test bilatéral.

VII.3.1 Test unilatéral à droite

H_1 : "$m>m_0$" (cas de l'exemple introductif).

- Statistique de test: la moyenne empirique $\bar{X}_n=\frac{1}{n}\sum_{i=1}^n X_i$.
- Règle de décision
$$\text{Rejet de } H_0 \Longleftrightarrow \bar{X}_n>t,$$
cette dernière étant donnée par la forme de H_1.
- Calcul de $\bar{x}_n$, la valeur observée de $\bar{X}_n$.
- Détermination de t. Sous H_0, on a que $\bar{X}_n\sim\mathcal{N}(m_0,\frac{\sigma^2}{n})$. On définit alors $Z=\frac{\overline{X}_n-m_0}{\sigma/\sqrt{n}}$, qui suit $\mathcal{N}(0,1)$. Ainsi on a:
$$\mathbb{P}_{H_0}(\bar{X}_n>t)=\mathbb{P}\left(Z>\frac{\sqrt{n}(t-m_0)}{\sqrt{\sigma^2}}\right).$$
On trouve la valeur de t à partir de α et de la table de la loi $\mathcal{N}(0,1)$.
- Conclusion par comparaison de t et de $\bar{x_n}$.

VII.3.2 Test unilatéral à gauche

H_1 : "$m < m_0$". Même méthode que précédemment, sauf que l'on remplace l'événement $\{\bar{X}_n > t\}$ par $\{\bar{X}_n < t\}$.

VII.3.3 Test bilatéral

H_1 : "$m \neq m_0$". Même méthode que précédemment, sauf que l'on remplace l'événement $\{\bar{X}_n > t\}$ par $\{|\bar{X}_n - m_0| > t\}$.

VII.4 Test sur la moyenne d'un échantillon gaussien (variance inconnue)

C'est le test de Student. On suppose que l'on observe les réalisations d'un n-échantillon $(X_1, \ldots, X_n)$ issu d'une loi $\mathcal{N}(m, \sigma^2)$ avec m et σ^2 inconnues. Pour cette section, on considère uniquement la construction d'un test bilatéral, les autres cadres suivant exactement le même principe.

On teste H_0 : "$m = m_0$" contre H_1 : "$m \neq m_0$". Comme σ^2 est inconnue, on l'estime par la variance empirique S_n^2. Par conséquent, la règle de décision est la suivante:

$$\text{Rejet de } H_0 \Longleftrightarrow |\bar{X}_n - m_0| > t \Longleftrightarrow \frac{|\bar{X}_n - m_0|}{S_n/\sqrt{n}} > \frac{t}{S_n/\sqrt{n}}.$$

Sous H_0, on sait que $T_n = \frac{\bar{X}_n - m_0}{S_n/\sqrt{n}}$ suit la loi de Student à $n-1$ degrés de liberté. Ainsi, on peut trouver la valeur de t d'après les tables de la loi de Student et la valeur de α.

Les tests unilatéraux se font de la même manière.

VII.5 Cas d'une population non gaussienne

Si le nombre n d'observations est suffisamment grand, on se ramène à une loi normale grâce au TLC. La loi de $\bar{X}_n$ est approximée par $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$. Les procédures de test sont alors les mêmes.

Chapitre VIII

Annexe 1: Tables statistiques

Chapitre IX

Annexe 2: Sujets des années précédentes

INSA de Toulouse - PO IC3 — Vendredi 16 novembre 2012
Probabilités et Statistiques

Evaluation 1

Durée 1 heure.

Les calculatrices et les documents ne sont pas autorisés.
Le barême sur 14 est indicatif.

Exercice 1 (9 pts.)
Un jeu vidéo est constitué de N niveaux successifs. Lorsque le joueur commence un niveau (ce qui suppose qu'il ait réussi les niveaux précédents), la probabilité qu'il le réussisse est 2/3. Le jeu s'arrête dès que le joueur échoue à un niveau.

Soit X la variable aléatoire qui est égale au nombre de niveaux réussis par le joueur.

1. Quelles sont toutes les valeurs possibles pour X?
2. Calculer $\mathbb{P}(X = N)$. Que représente cette probabilité?
3. Calculer $\mathbb{P}(X = 0)$, $\mathbb{P}(X = 1)$ puis $\mathbb{P}(X = k)$ pour tout $k \in \{1, \ldots, N-1\}$.
4. Vérifier que
$$\sum_{k=0}^{N} \mathbb{P}(X = k) = 1.$$
5. Pour tout entier $0 \leq k \leq N$, montrer que
$$\mathbb{P}(X \geq k) = \left(\frac{2}{3}\right)^k.$$
6. On suppose pour cette question que le jeu a $N = 4$ niveaux. Montrer que la relation
$$\mathbb{E}(X) = \sum_{k=1}^{N} \mathbb{P}(X \geq k). \qquad \text{(IX.1)}$$
est vraie pour $N = 4$.
Indication: Vous pourriez utiliser la relation $\mathbb{P}(X \geq k) = \sum_{i=k}^{4} \mathbb{P}(X = i)$, pour tout $k \leq 4$.
7. On admettra dans cette question que l'équation (IX.1) est vraie pour tout entier N.
Montrer alors que
$$\mathbb{E}(X) = 2 - 3\left(\frac{2}{3}\right)^{N+1}.$$

Exercice 2 (5 pts.)
Une urne contient 15 boules, dont 6 blanches et 9 rouges. On tire au hasard et sans remise une boule de l'urne. Si la boule est blanche, on ajoute 4 boules blanches dans l'urne, sinon on ajoute 10 boules rouges. Puis on tire de nouveau une boule. Pour $i \in \{1, 2\}$, on considère l'événement

$$B_i : \text{ “tirer une boule blanche au } i\text{-ème tirage”.}$$

1. Calculer $\mathbb{P}(B_1)$ et $\mathbb{P}(B_2|B_1)$.
2. Quelle est la probabilité d'obtenir successivement deux boules blanches?
3. Quelle est la probabilité d'obtenir une boule blanche au second tirage?
4. Quelle est la probabilité d'avoir tiré une boule blanche au premier tirage, sachant que la boule tirée au second tirage est blanche?

INSA de Toulouse - 3A IC — Vendredi 11 janvier 2013

Contrôle final de Probabilités et Statistique

Durée 1h30

Les documents, les calculatrices et les téléphones portables sont interdits.
Le barème sur 20 est indicatif.

Quelques données numériques utiles pour l'ensemble du sujet:

Soit Z une variable aléatoire de loi $\mathcal{N}(0,1)$ et F_Z sa fonction de répartition. On a les valeurs approchées suivantes:

$$F_Z(1) = 0.841,\ F_Z(1.65) = 0.95,\ F_Z(1.96) = 0.975,\ F_Z(3) = 0.9987.$$

Exercice 1 (2,5 points)

1. Énoncez l'inégalité de Chebychev.
2. Soit X et Y des variables aléatoires indépendantes, de lois respectives $\mathcal{N}(1,8)$ et $\mathcal{N}(2,8)$. Soit $S = X + Y$. Donnez la loi de S, ainsi que son espérance et sa variance. Calculez ensuite $\mathbb{P}(S < -1)$.

Exercice 2 (9 points)

Soit X une variable aléatoire de loi exponentielle de paramètre 1.

1. Donnez la densité et la fonction de répartition de X.
2. Soit $\mu > 0$. On pose $Y = \frac{X}{\mu}$. Calculez la fonction de répartition de Y et donnez sa loi.
3. Soit $0 < p < 1$ et Z une variable aléatoire, indépendante de X, telle que

$$\mathbb{P}(Z = 1) = p,\ \mathbb{P}(Z = -1) = 1 - p.$$

On pose $T = XZ$.

(a) Calculez l'espérance et la variance de Z.

(b) Pour $t \in \mathbb{R}$, calculez $\mathbb{P}(T > t\ \cap\ Z = 1)$ et $\mathbb{P}(T > t\ \cap\ Z = -1)$, puis $\mathbb{P}(T > t)$, en distinguant les cas $t \geq 0$ et $t < 0$.

(c) Montrez que la fonction de répartition de T est donnée par

$$F_T(t) = \begin{cases} 1 - pe^{-t} & \text{si } t \geq 0, \\ (1-p)e^t & \text{si } t < 0. \end{cases}$$

(d) Déterminez la densité de T.

(e) Calculez l'espérance de T et vérifiez que $\mathbb{E}(T) = \mathbb{E}(X)\mathbb{E}(Z)$.

(f) Montrez que les variables T et Z ne sont pas indépendantes.

Indication: Une variable aléatoire continue U et une variable aléatoire discrète V sont indépendantes si et seulement si $\mathbb{P}(\{U \in [a,b]\} \cap \{V = v\}) = \mathbb{P}(U \in [a,b])\, \mathbb{P}(V = v)$, pour tous $a, b \in \mathbb{R}$ et v valeur possible de V.

Exercice 3 (8,5 points)
Dans une promotion INSA de 1ère année formée de 400 étudiants, on suppose que chaque étudiant souhaite aller en pré-orientation IC avec probabilité p, et ce indépendamment des autres étudiants.

1. Pour estimer p, on interroge n étudiants de la promotion sur leur choix de pré-orientation. Pour $1 \leq i \leq n$ on pose $X_i = 1$ si le $i$ème étudiant interrogé souhaite aller en IC et 0 sinon. On définit ensuite

$$\hat{p}_n = \frac{X_1 + \ldots + X_n}{n}.$$

   (a) Montrez que $\hat{p}_n$ est un estimateur consistant et sans biais de p.

   (b) Justifiez le fait que, pour n grand, la loi de $\hat{p}_n$ peut être approximée par une loi normale, dont vous préciserez les paramètres.

   (c) Construisez un intervalle de confiance asymptotique pour p de niveau de confiance 95%. Justifiez les différentes étapes de la construction.

   (d) On suppose que l'on a interrogé 100 étudiants de la promotion, et que 20 étudiants parmi ceux interrogés souhaitent aller en IC. Donnez une estimation ponctuelle de p, ainsi qu'un intervalle de confiance asymptotique de niveau 95%.

2. On suppose dans la suite que $p = 0.2$. On appelle N le nombre total d'étudiants, parmi les 400 étudiants de la promotion, qui souhaitent aller en IC.

   (a) Quelle loi suit N? Donnez son espérance et sa variance.

   (b) Par quelle loi, dont on précisera les paramètres, peut-on approximer la loi de N? Justifiez.

   (c) On suppose que la pré-orientation IC peut accueillir au maximum 104 étudiants. Donnez une valeur approchée de la probabilité que la pré-orientation IC soit obligée de refuser des étudiants.

INSA de Toulouse - PO IC3 Vendredi 25 novembre 2011
Probabilités et Statistiques

Evaluation 1

Durée 1 heure.

Les calculatrices et les documents ne sont pas autorisés.
Le barême sur 14 est indicatif.

Exercice 1 (4 pts.)
On interroge des étudiants sur la signification du sigle “ADN”. A chacun, on propose trois réponses différentes: A, B ou C, la réponse correcte étant la réponse A. Tout étudiant connaissant la réponse correcte la donne, sinon il choisit au hasard une des trois réponses proposées. On note θ la probabilité qu'un étudiant connaisse la bonne réponse et p_A la probabilité qu'un étudiant interrogé donne la réponse A.

1. Exprimer p_A en fonction de θ.
2. Quelle est la probabilité qu'une personne ayant choisi la réponse A connaisse réellement la signification du sigle “ADN”?
3. Sur un groupe de 24 étudiants, on note N le nombre de bonnes réponses. Quelle loi suit N? Justifier.
4. Quelle est l'espérance du nombre de bonnes réponses obtenues dans ce groupe?

Exercice 2 ($4,5$ pts.)
Pour valider une épreuve de saut en hauteur, un sportif doit réussir deux fois le même saut (mais pas forcément à des essais consécutifs). On suppose qu'à chaque essai il a une probabilité p de réussir le saut, indépendamment d'un essai à un autre. On note S le nombre d'essais que le sportif doit faire pour valider l'épreuve.

1. Montrer que $\mathbb{P}(S = 2) = p^2$, $\mathbb{P}(S = 3) = 2(1 - p)p^2$ et $\mathbb{P}(S = 4) = 3(1 - p)^2p^2$.
2. Déterminer $\mathbb{P}(S = k)$, pour tout $k \geq 2$.
3. Vérifier que la somme des probabilités trouvées à la question précédente vaut 1.

Exercice 3 (5,5 pts.)
Soit f la fonction définie par

$$f(x) = \begin{cases} 1/2 & \text{si } x \in [-1, 0], \\ a & \text{si } x \in]0, 1], \\ 0 & \text{sinon.} \end{cases}$$

1. Déterminer la valeur de a pour que f définisse bien une densité de probabilité.
2. Soit X une variable aléatoire ayant pour densité f. Calculer $\mathbb{P}(-1/2 < X < 1/2)$.
3. On pose $Y = X^2$. Quelles sont les valeurs possibles de Y?
4. Montrer que la fonction de répartition de Y est donnée par

$$F_Y(y) = \begin{cases} \sqrt{y} & \text{si } y \in [0, 1], \\ 0 & \text{si } y < 0, \\ 1 & \text{si } y > 1. \end{cases}$$

5. Déterminer la densité de la variable aléatoire Y.
6. Calculer l'espérance de Y.

INSA de Toulouse - PO IC3 Mercredi 18 janvier 2012
Probabilités et Statistiques

Evaluation 2

Durée 1h30

Tous les documents, calculatrices et téléphones portables sont interdits.
Le barème sur 20 est approximatif.

Exercice 1 (6 points)
Des électriciens interviennent dans un immeuble pour poser des tableaux électriques. Le fournisseur des ces tableaux a été choisi au hasard (probabilité 0.5) par le commanditaire des travaux parmi deux fabriquants U et V. Une fois le fournisseur choisi, **tous les tableaux proviennent du même fabriquant**.

Les tableaux du fournisseur U sont défectueux avec probabilité 0.1, ceux provenant du fournisseur V avec probabilité 0.3. Pour chaque fournisseur, ce mode de fonctionnement est indépendant d'un tableau à l'autre. On considère les événements suivants:
U= "tous les tableaux proviennent du fournisseur U",
V= "tous les tableaux proviennent du fournisseur V" et A_i= "le i-ème tableau est défectueux".

1. Donner les valeurs de $\mathbb{P}(A_1|U)$, $\mathbb{P}(A_1 \cap V)$ et montrer que $\mathbb{P}(A_1 \cap A_2|U) = (0.1)^2$.
2. En déduire $\mathbb{P}(A_1)$ et $\mathbb{P}(A_1 \cap A_2)$.
3. Les événements A_1 est A_2 sont-ils indépendants? (on pourra utiliser le fait que $\mathbb{P}(A_1) = \mathbb{P}(A_2)$).
4. Pour tout entier $n \in \mathbb{N}$, montrer que $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \frac{3^n+1}{2\times 10^n}$. (Rappel: tous les tableaux proviennent du même fournisseur).
5. Sachant que les n premiers tableaux sont défaillants, quelle est la probabilité que le $(n+1)$-ème le soit également?
6. Sachant que les n premiers tableaux sont défaillants, quelle est la probabilité qu'ils proviennent du fournisseur U?

Exercice 2 (7 points)
Soit X une variable de loi exponentielle de paramètre 1. On pose $Y = \exp(X/2)$. On suppose que Y modélise le temps d'attente d'un bus (en minutes).

1. Montrer que la variable Y est toujours supérieure ou égale à 1.

2. Quelle est la probabilité d'attendre le bus plus de 20 minutes?

3. Déterminer la fonction de répartition de Y.

4. Déduire que la variable aléatoire Y admet comme densité de probabilité la fonction

$$f_Y(y) = \begin{cases} 2y^{-3} & \text{si } y \geq 1, \\ 0 & \text{si } y < 1. \end{cases}$$

5. Calculer le temps moyen d'attente du bus.

6. Sachant que l'on a déjà attendu le bus plus de 10 minutes, quelle est la probabilité que le bus n'arrive pas dans les 10 prochaines minutes?

Exercice 3 (7 points)
On observe une suite de variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi telle que pour tout $i \in \{1, \ldots, n\}$

$$P(X_i = 1) = p, \text{ et } P(X_i = -1) = 1 - p,$$

avec $p \in]0, 1[$ un paramètre inconnu. L'objectif de cet exercice est d'estimer p.

1. Calculer l'espérance et la variance de X_1.

2. On considère l'estimateur de p suivant:

$$\hat{p}_n = \frac{\bar{X}_n + 1}{2}, \text{ avec } \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   Calculer l'espérance et la variance de l'estimateur $\hat{p}_n$.

3. Montrer que pour tout nombre réel $c > 0$ on a

$$\mathbb{P}(|\hat{p}_n - p| > c) \leq \frac{p(1-p)}{nc^2}.$$

4. Montrer que $\hat{p}_n$ est un estimateur consistant et sans biais de p.

5. Énoncer le Théorème de la limite centrale. Donner la loi asymptotique de $\bar{X}_n$ et en déduire celle de $\hat{p}_n$.

6. Montrer que la variable aléatoire

$$D_n = (\hat{p}_n - p)\sqrt{\frac{n}{p(1-p)}}$$

   suit approximativement la loi $N(0,1)$ pour n grand.

7. Construire un intervalle de confiance de niveau 95% pour p.

   Indication: On pourra utiliser le fait que $p(1-p) \leq \frac{1}{4}$, pour tout $p \in]0,1[$. On rapelle également que $F_Z(1,96) = 0,975$ où F_Z désigne la fonction de répartition d'une loi $N(0,1)$.

INSA de Toulouse - PO IC3 — Vendredi 3 décembre 2010
Probabilités et Statistiques

Evaluation 1

Durée 45 minutes

Exercice 1 (4,5 points)
Un étudiant de la PO IC3 va en cours avec une probabilité 0.5 quand il fait beau et 0.8 quand il pleut. Sachant qu'à Toulouse, il pleut 3 jours sur 5 en novembre, et que l'étudiant était en classe le 19, quelle est la probabilité ce jour là ait été un jour de pluie? (On considèrera qu'il fait beau lorsqu'il ne pleut pas!)

Exercice 2 (4,5 points)
Soient X et Y deux variables aléatoires indépendantes telles que:

$$P(X=1)=P(X=-1)=\frac{1}{2} \text{ et } P(Y=1)=P(Y=0)=\frac{1}{2}.$$

1. On pose $Z=XY$. Quelle est la loi de la variable aléatoire Z?
2. Calculer $\mathbb{E}[Z]$ et $\mathrm{Var}(Z)$.
3. Montrer que $P(\{Y=0\}\cap\{Z=1\}) \neq P(Y=0)\times P(Z=1)$. Que peut-on en conclure sur l'indépendance de Y et Z?

Exercice 3 (5 points)
Soit X une variable aléatoire de loi géométrique de paramètre p définie par

$$\mathbb{P}(X=k)=p(1-p)^{k-1}, \text{ pour tout } k\in\mathbb{N}^\star$$

Soit X' une autre variable aléatoire, indépendante de X et de loi géométrique de paramètre p'. On note par M le minimum de X et X'. On cherche à déterminer la loi de M.

1. Pour $k\in\mathbb{N}^\star$, calculer $\mathbb{P}(X>k)$ et $\mathbb{P}(X'>k)$.
2. Pour $k\in\mathbb{N}^\star$, exprimez l'événement $\{M>k\}$ en fonction des variables aléatoires X et X'. Montrer ensuite que

$$\mathbb{P}(M>k)=(1-(p+p'-pp'))^k.$$

3. Quelle est la loi de M? (Indication: on pourra montrer que si Y_1 et Y_2 sont deux variables aléatoires à valeurs dans $\mathbb{N}$ telles que $P(Y_1 > k) = P(Y_2 > k)$ pour tout $k \in \mathbb{N}$, alors Y_1 et Y_2 ont même loi).

Tous les documents, calculatrices et téléphones portables sont interdits. Le barême sur 14 est approximatif.

INSA de Toulouse - PO IC3
Probabilités et Statistiques

Vendredi 21 janvier 2011

Evaluation 2

Durée 1h30

Tous les documents, calculatrices et téléphones portables sont interdits.
Le barème est approximatif.

Quelques rappels utiles :

- Densité de la loi uniforme $\mathcal{U}([0,1])$:

$$f_U(x) = \begin{cases} 1 & \text{si } x \in [0,1], \\ 0 & \text{si } x \notin [0,1]. \end{cases}$$

- Fonction de répartition d'une variable aléatoire X : $F_X(x) = \mathbb{P}(X \leq x)$.
- Quantile d'une variable aléatoire Z suivant la loi $\mathcal{N}(0,1)$: $F_Z(1,96) = 0,975$.

Exercice 1 (7,5 points)
Soit U une variable aléatoire de loi uniforme $\mathcal{U}([0,1])$ et soit $p \in]0,1[$.

1. Déterminer la loi de la variable aléatoire

$$X = \begin{cases} 1 & \text{si } U \leq p, \\ 0 & \text{si } U > p. \end{cases}$$

2. Soit $U_1, \ldots, U_n$ des variables indépendantes de même loi que U.
   (a) Soit Y le nombre de variables $U_i, 1 \leq i \leq n$ qui sont inférieures à p, i.e.

   $$Y = |\{i : U_i \leq p\}|,$$

   où $|.|$ dénote le cardinal. Quelle est la loi de Y? Justifier.
   (b) On considère les variables aléatoires $V = \max(U_1, ..., U_n)$ et $W = \min(U_1, ..., U_n)$. Calculer $\mathbb{P}(V \leq p)$ et $\mathbb{P}(W \leq p)$.
3. Soit $\lambda > 0$ et la variable aléatoire $T = -\frac{\log(U)}{\lambda}$.
   Calculer la fonction de répartition F_T et la densité f_T de T. En déduire la loi de T.

Exercice 2 (7 points)
Soit le couple de variables aléatoires (X, Y) de densité jointe

$$f(x,y) = \begin{cases} \frac{1}{2\sqrt{xy}} & \text{si } (x,y) \in D, \\ 0 & \text{si } (x,y) \notin D, \end{cases}$$

où $D = \{(x,y) \in \mathbb{R}^2 : 0 < x \leq y < 1\}$.

1. Montrer que f est bien une densité de probabilité sur $\mathbb{R}^2$.
2. Montrer que la densité marginale de X est donnée par la formule

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{x}} - 1 & \text{si } x \in]0,1[, \\ 0 & \text{si } x \notin]0,1[. \end{cases}$$

   Calculer la densité marginale f_Y de Y et montrer que Y suit la loi uniforme $\mathcal{U}([0,1])$.

3. Les variables X et Y sont-elles indépendantes?
4. Calculer l'espérance et la variance de X et de Y.

Exercice 3 (7 points)
On s'intéresse à la concentration μ d'un adjuvant dans un certain type de béton. On réalise pour cela n mesures $Y_1, \ldots, Y_n$, ces mesures étant supposées indépendantes et de même loi. Par ailleurs, pour tout $i \in \{1, \ldots, n\}$, on suppose que $Y_i = X_i + \epsilon_i$, où

$$X_i \sim \mathcal{N}(\mu, \sigma_1^2) \text{ et } \epsilon_i \sim \mathcal{N}(0, \sigma_2^2),$$

X_i et ϵ_i étant elle-mêmes indépendantes. Les variables X_i modélisent les fluctuations d'adjuvant d'une mesure à l'autre, les variables ϵ_i représentant quand à elles des erreurs liées à l'instabilité du produit chimique permettant la mise en évidence de l'adjuvant. L'objectif est d'estimer μ, les termes σ_1^2 et σ_2^2 étant supposés connus.

1. Quelle est la loi de Y_1?
2. Soit $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$. Calculer $\mathbb{E}[\bar{Y}_n]$ et $\mathrm{Var}(\bar{Y}_n)$. Donner la loi de $\bar{Y}_n$.
3. Construire un intervalle de confiance à 95% pour la moyenne μ. Commenter la contribution des erreurs de mesure $(\epsilon_i)_{i=1,\ldots,n}$ dans cet intervalle.
4. On donne $\sigma_1^2 = \sigma_2^2 = 1/2$. A l'issue de 100 mesures, on obtient l'estimation $\bar{y}_n = 0.3 g/m^3$. Calculer numériquement l'intervalle de confiance précédent.
5. Enoncer le théorème centrale limite et donner la loi asymptotique de $\bar{Y}_n$ quand les X_i et ϵ_i ne sont pas de loi normale.