

IUT	Robert Schuman	
Institut universitaire de technologie		
Département informatique		
Université de Strasbourg		

Introduction aux probabilités

1 Abrégé de cours

1.1 Espaces probabilisés

1.1.1 Expérience aléatoire, modèle, univers

Pour parler de probabilités, au sens mathématiques il faut

- (i) considérer une expérience *aléatoire* c'est-à-dire dont l'issue n'est pas (totalement) prévisible *a priori*;
- (ii) puis *modéliser* l'ensemble des issues possibles de l'expérience, c'est-à-dire les décrire comme les éléments d'un ensemble : *l'univers* Ω (des issues possibles).

Exemples.

- Lancer d'une pièce de monnaie : $\Omega = \{Pile, Face\}$
- Lancer d'un dé standard : $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Tirage du Loto de la FDJ : 5 nombres choisis au hasard, sans remise, parmi 49, et, *indépendamment*, un numéro entre 1 et 10
 $\Omega = \{\text{choix de 5 nombres parmi 49}\} \times \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ (produit cartésien)
- Choix d'un nombre *au hasard* entre 0 et 1, $\Omega = [0, 1]$

1.1.2 Mesures de probabilités

On note $\mathcal{P}(\Omega)$ l'ensemble des parties (ou des sous-ensembles) de l'ensemble Ω . On rappelle que $(\mathcal{P}(\Omega), \cup, \cap)$ constitue une algèbre de Boole.

Définition 1. Soit Ω l'ensemble des issues possibles d'une expérience aléatoire.

Une *mesure de probabilité* sur Ω est une **fonction** $P : \mathcal{P}(\Omega) \longrightarrow [0, 1]$ vérifiant

- (i) $P(\Omega) = 1$ (on dit que Ω est l'événement certain);
- (ii) pour toute famille dénombrable de sous-ensembles $A_1, A_2, \dots, A_n, \dots$, deux à deux disjoints, pour lesquels P est défini, on a

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$$

L'ensemble Ω muni de la probabilité P est appelé *espace probabilisé*.

Remarques :

- En fait, l'ensemble de définition de P doit vérifier certaines conditions techniques supplémentaires (essentiellement de stabilité par réunion et intersection), mais on fera mine d'ignorer cette difficulté ;
- Pour des univers finis, la condition (ii) se ramène au cas d'une réunion finie disjointe, puis (argument de récurrence) au cas de deux événements disjoints. Autrement dit, la condition (ii) s'écrit simplement :

$$A_1 \cap A_2 = \emptyset \Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

Une conséquence de (ii) est que, pour A et B des parties de Ω où P est définie,

$$P(B) = P(B \cap A) + P(B \cap \bar{A}),$$

de quoi l'on déduit :

- (i) $P(\bar{A}) = 1 - P(A)$ (où \bar{A} est le complémentaire ensembliste de A dans Ω) ;
- (ii) en particulier, $P(\emptyset) = 0$;
- (iii) si $A \subset B$, alors $P(A) \leq P(B)$;
- (iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Dans un premier temps, on ne va considérer que des univers finis : les concepts probabilistes sont ceux du cas général (infini, dénombrable ou non), mais leur manipulation pose moins de difficultés techniques.

1.1.3 Univers finis, équiprobabilité

Pour définir une probabilité sur un ensemble fini $\Omega = \{\omega_1, \dots, \omega_n\}$, il suffit de préciser $P(\{\omega_i\})$ pour tout *événement élémentaire* $\{\omega_i\} \subset \Omega$: on en déduit facilement la valeur de P sur **tout** sous-ensemble de Ω .

Par exemple si tous les événements élémentaires $\{\omega_1\}, \dots, \{\omega_n\}$ d'un univers fini ont même probabilité, on dit qu'ils sont *équiprobables*, ou encore que l'univers est muni de la probabilité *uniforme*. En situation d'équiprobabilité, la probabilité d'un événement élémentaire (parmi n possibles) vaut donc $1/n$.

Plus généralement, la probabilité d'un événement sera proportionnelle à son cardinal :

$$P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

Deux modèles équiprobables classiques.

- (i) Une suite de n épreuves indépendantes (pour une définition formelle voir la section suivante), chaque épreuve donnant lieu à k issues possibles (on parle souvent de *tirages avec remises*) $\omega_1, \dots, \omega_k$, se modélise comme un ensemble de k^n *n-uplets* (ou *mots de longueur n*) :

$$\Omega = \{(x_1, x_2, \dots, x_n) \mid x_i \in \{\omega_1, \dots, \omega_k\}, \forall i = 1, \dots, n\}$$

Si le résultat de chaque épreuve prise isolément est équiprobable, alors Ω l'est aussi.

- (ii) Un tirage aléatoire de k objets distincts parmi n (*tirages sans remises*)
- si l'on tient compte de l'ordre d'obtention des objets, le nombre de tirages possibles est

$$\frac{n!}{(n-k)!} = n(n-1)(n-2)\dots(n-k+1)$$

- si l'on ne tient pas compte de l'ordre, le nombre de tirages possibles est

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (\text{nombre appelé } \textit{coefficient binomial } (n, k)).$$

Là encore, si chaque objet pris isolément est équiprobable, chaque tirage possible l'est aussi, pour chacun des deux modèles.

1.1.4 Indépendance

Intuitivement l'idée d'événements A et B *indépendants* (la survenue de l'événement A ne modifie pas la probabilité *a priori* de B).

Typiquement, dans les tirages avec remises, le résultat de chaque tirage (successif) est indépendant des autres, au contraire des tirages sans remise, où l'obtention d'un objet à un tirage donné interdit de l'obtenir aux suivants. Formellement :

Définition 2. Soit A et B des parties de Ω . On dit que les événements A et B sont *indépendants* si

$$P(A \cap B) = P(A) \times P(B).$$

Remarques.

- Il ne faut pas confondre « A et B sont indépendants » : $P(A \cap B) = P(A) \times P(B)$, et « A et B sont disjoints » : $P(A \cup B) = P(A) + P(B)$.
En fait des événements disjoints (non vides) ne sont jamais indépendants !
- Si l'intuition suggère que les résultats de deux expériences aléatoires sont *indépendants* (lancers successifs de pièces, ou de dés, par exemple), le choix du modèle devrait refléter cette indépendance. Mais il faut noter qu'en dehors de quelques situations simples, la notion d'indépendance est difficile à appréhender intuitivement.
- On dit que trois événements A , B et C sont *indépendants (dans leur ensemble)*, s'ils sont indépendants (définition précédente) *deux à deux*, et que de plus

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C).$$

- On peut définir de façon analogue (et même récursive !) l'indépendance d'un nombre quelconque d'événements.

1.2 Variable aléatoire (cas discret)

1.2.1 Définition et exemples

Définition 3. Une *variable aléatoire (réelle)*, abrégée en *v.a.*, X sur un espace probabilisé (Ω, P) est simplement une fonction $X : \Omega \rightarrow \mathbb{R}$.

En fait, pour qu'une fonction soit une v.a., elle doit aussi vérifier l'importante condition technique d'être *mesurable*, qui garantit en quelque sorte qu'elle est « compatible » avec la probabilité P . Nous ignorerons ces difficultés.

Définition 4. La *loi*, ou *distribution de probabilité*, d'une v.a. X prenant un nombre (au plus) dénombrable de valeurs (une telle v.a. est dite *discrète*) $x_1, x_2, \dots, x_n, \dots$ est la donnée, pour chaque entier $i = 1, 2, \dots, n, \dots$, de la probabilité $P(X = x_i) = P(X^{-1}(\{x_i\}))$.

Exemples.

- On lance une pièce de monnaie, et on considère la v.a. qui vaut 0 si l'on obtient *pile* et 1 si on obtient *face*. Une telle v.a. (ne prenant que les valeurs 0 et 1) est dite *v.a. de Bernoulli*. La loi d'une v.a. de Bernoulli est totalement déterminée par la probabilité p d'obtenir *face*.
- On note le résultat obtenu lors du lancer d'un dé standard. La v.a. ainsi obtenue prend ses valeurs dans l'ensemble $\{1, 2, 3, 4, 5, 6\}$, et sa loi est *uniforme* sur cet ensemble si le dé est équilibré.
- La somme des points obtenus lorsqu'on lance deux dés est une v.a. à valeurs entières comprises entre 2 et 12, mais non équiprobables *a priori*.
- Le nombre de lancers effectués lors de l'expérience consistant à lancer une pièce jusqu'à obtenir *face* est une v.a. à valeurs dans \mathbb{N} (le nombre de valeurs possibles n'est pas fini *a priori*). La loi d'une telle v.a. est dite *géométrique de paramètre p* , où p est la probabilité d'obtenir *face* à un lancer donné.

1.2.2 Espérance

Définition 5. L'*espérance mathématique* d'une v.a. X , prenant les valeurs $\{x_1, \dots, x_n\}$ est la moyenne des valeurs prises par X , pondérées par leurs probabilités, ou encore

$$E(X) = \sum_{i=1}^n P(X = x_i) x_i.$$

Remarques.

- L'espérance d'une v.a. est en quelque sorte sa « valeur moyenne ».
- Ce n'est pas forcément une valeur prise par la v.a. (contrairement à la *médiane* ou au *mode*).
- L'espérance d'une v.a. ne dépend **que de sa loi**, et non du résultat d'une expérience aléatoire, c'est une grandeur *déterministe*.
- Pour une v.a. prenant une infinité dénombrable de valeurs, on peut essayer de prolonger la définition en passant à la limite sur des sommes finies, mais ça ne marche pas toujours : toutes les v.a. n'admettent pas une espérance.

Exemples.

- Une v.a. de Bernoulli de paramètre p a pour espérance p (la *probabilité de succès*).
- L'espérance du nombre de points obtenus lorsqu'on lance un dé équilibré est $7/2$.

Théorème 1 (Linéarité de l'espérance). Soit X et Y des v.a. définies sur un même espace (Ω, P) , et soit $a \in \mathbb{R}$, on a :

- (i) $E(aX) = aE(X)$.
- (ii) $E(X + Y) = E(X) + E(Y)$.

Corollaire 1. Soit X_1, \dots, X_n des v.a. sur Ω , alors

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n).$$

Exemple. L'espérance du nombre total de points obtenus lorsqu'on lance n dés équilibrés est $7n/2$.

1.2.3 Schéma de Bernoulli

Soit n un entier et p un réel, $p \in [0, 1]$. On considère l'expérience aléatoire qui consiste à répéter n fois une épreuve, chaque répétition étant indépendante des autres, et donnant lieu à deux issues possibles :

- « succès », noté S , avec probabilité p ;
- « échec », noté E , avec probabilité $1 - p$.

Une telle expérience est appelée un *schéma de Bernoulli*, un exemple typique étant la réalisation de n lancers d'une pièce qui donne « face » avec probabilité p .

L'univers associé à une telle expérience peut être décrit comme l'ensemble des mots binaires de longueur n . Il y a donc 2^n issues possibles, **qui ne sont pas équiprobables**, sauf si $p = 1/2$.

Théorème 2. Soit N la variable aléatoire qui compte le nombre de succès lors de la réalisation d'un schéma de Bernoulli de paramètres n et p . La loi de N est donnée par

$$P(N = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{pour tout entier } k, 0 \leq k \leq n, .$$

On dit que N suit une *loi binomiale de paramètres n et p* , et on écrit $N \sim \mathcal{B}(n, p)$.

Une v.a. de loi binomiale est aussi la somme de n variables de Bernoulli indépendantes, de même paramètre p ; on en déduit (corollaire 1) que l'espérance d'une v.a. binomiale de loi $\mathcal{B}(n, p)$ est np .

1.2.4 Variance

On introduit une nouvelle grandeur déterministe pour rendre compte de la *dispersion* d'une v.a. autour de son espérance.

Définition 6. La *variance* d'une v.a. X , d'espérance μ , est la grandeur notée $\text{Var}(X)$, ou encore σ_X^2 , et définie par

$$\text{Var}(X) = E((X - \mu)^2).$$

L'*écart type* de la v.a. X est défini par $\sigma_X = \sqrt{\text{Var}(X)}$.

Remarques.

- Il est facile d'établir que $\text{Var}(X) = E(X^2) - E(X)^2$, cette seconde formule étant souvent plus pratique pour les calculs.
- Pour une v.a. de Bernoulli X , de paramètre p , $\text{Var}(X) = p(1 - p)$.

Contrairement à l'espérance, la variance n'a pas de propriété de linéarité :

Théorème 3. *Pour toutes v.a. X et Y et tout réel a , on a*

$$(i) \quad \text{Var}(aX) = a^2 \text{Var}(X)$$

$$(ii) \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2[E(XY) - E(X)E(Y)].$$

Remarque. On appelle *covariance* des v.a. X et Y la quantité

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y),$$

qui mesure la « dépendance linéaire » entre X et Y , et aussi le défaut d'additivité de la variance pour la v.a. somme $X + Y$.

Le *coefficient de corrélation linéaire* des v.a. X et Y (non constantes) est une version « normalisée » de la covariance, c'est-à-dire insensible au choix de l'unité de mesure, et dont la valeur, toujours comprise entre -1 et 1, est définie par

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

1.2.5 Indépendance de variables aléatoires

Définition 7. *On dit que deux v.a. discrètes $X, Y : \Omega \rightarrow \mathbb{R}$ sont *indépendantes* si, pour tous réels x, y , les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants, c'est-à-dire si, pour tous réels x et y ,*

$$P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\}) \times P(\{Y = y\}).$$

Remarque.

- Des exemples typiques de v.a. indépendantes sont les v.a. de Bernoulli associées aux lancers successifs d'une pièce de monnaie, ou les valeurs obtenues pour des lancers successifs d'un dé, et plus généralement les v.a. associées aux répétitions indépendantes d'une expérience.
- Les quantités du type $P(\{X = x\} \cap \{Y = y\})$ définissent la *loi conjointe de X et de Y* : la propriété d'indépendance de X et de Y peut se reformuler en disant que la loi conjointe est le produit des lois de X et de Y .
- La définition s'étend au cas de v.a. non discrètes, en considérant l'indépendance de tous les événements du type $\{X \leq x\}$, pour toute valeur x réelle.

Théorème 4. *Pour des v.a. indépendantes X et Y , on a $E(XY) = E(X)E(Y)$, ou encore $\text{Cov}(X, Y) = 0$, et, par conséquent :*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

On peut définir la notion de *n variables aléatoires mutuellement indépendantes* X_1, X_2, \dots, X_n comme l'indépendance des événements $\{X_1 = x_1\}, \{X_2 = x_2\}, \dots, \{X_n = x_n\}$ dans leur ensemble, pour toutes les valeurs possibles de x_1, \dots, x_n .

Il est facile de déduire du théorème précédent qu'on a dans ce cas

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

On en déduit en particulier que la variance d'une v.a. binomiale de loi $\mathcal{B}(n, p)$ est $np(1-p)$.

1.3 Variables aléatoires à densité

On étend maintenant la notion de v.a. sur un espace probabilisé (Ω, P) au cas où l'ensemble des valeurs possibles n'est plus *discret*, c'est-à-dire au plus dénombrable, mais *continu* (typiquement un intervalle réel). Au prix d'une approche plus abstraite (*théorie de la mesure et de l'intégration*), on pourrait traiter les deux cas à l'intérieur du même cadre mathématique.

1.3.1 Intégration (guide de survie)

L'intégrale comme aire.

La motivation première d'une *théorie de l'intégration* est le calcul d'aire de surfaces « suffisamment régulières » (typiquement le calcul d'aires délimitées par des graphes de fonctions, elles-mêmes « suffisamment régulières »). Il existe plusieurs constructions rigoureuses qui permettent de donner un sens précis et un cadre bien défini à cette question (ce qui diffère d'une théorie à l'autre est essentiellement ce qu'on entend par « suffisamment régulières »), mais elles dépassent les objectifs de ce cours. On se contentera donc d'énoncer et d'admettre quelques propriétés fondamentales de l'intégrale.

D'abord, si f est une fonction « régulière » (on pourra, par exemple, penser à une fonction *continue*, ou à une fonction *continue sauf en un nombre fini de points*, ce qui couvrira tous les cas utiles pour ce cours) et à valeurs positives, et si $a \leq b$ sont deux réels, *l'intégrale de f sur $[a, b]$* , notée

$$\int_a^b f(t)dt \quad \text{ou plus simplement} \quad \int_a^b f$$

est l'aire comprise entre le graphe de f et l'axe des abscisses, pour les abscisses variant de a à b (dans un repère orthonormé); notez en particulier que $\int_a^a f = 0$.

On étend l'énoncé au cas où f prend des valeurs négatives, en décomptant l'aire *algébriquement* c'est-à-dire négativement sur les intervalles où f est négative.

On étend également la notation au cas où $a > b$, en posant $\int_b^a f = -\int_a^b f$ (toutes ces extensions sont cohérentes, et justifiées par la notion d'orientation du plan...).

On « déduit » de l'interprétation en terme d'aires tout un ensemble de propriétés comme par exemple la *Relation de Chasles* :

$$\int_a^b f = \int_a^c f + \int_c^b f \quad (\text{pour tous réels } a, b, \text{ et } c)$$

ou encore la *linéarité de l'intégrale* :

$$\int_a^b (\lambda f + g) = \lambda \int_a^b f + \int_a^b g.$$

Il est important de noter que les constructions/définitions de l'intégrale sont compatibles avec l'intuition du calcul d'une aire, et qu'elles fournissent aussi des techniques de calcul (d'approximation) de la valeur des intégrales (techniques de découpages des surfaces en surfaces de formes simples et d'aires faciles à calculer, rectangles par exemple).

Intégrale et primitive.

On rappelle que F est une *primitive* de f sur l'intervalle I , si la dérivée de F est f ($F' = f$) sur I . Il existe un lien *a priori* surprenant entre les notions d'intégrales et de primitives, qu'on peut « résumer » ainsi : si f est la dérivée continue d'une fonction F sur l'intervalle I , alors pour tous réels $a, b \in I$

$$\int_a^b f = F(b) - F(a).$$

L'intérêt de ce résultat est double : permettre de calculer une primitive au moyen d'un calcul d'aire, et, réciproquement, de calculer des intégrales lorsqu'on reconnaît une dérivée dans la fonction à intégrer.

On observera aussi que la notation $\int_a^x f(t)dt$ est appropriée pour désigner la primitive de f s'annulant en a sur I .

Par exemple, on a

$$\ln(x) = \int_1^x \frac{dt}{t}, \quad \text{ou encore} \quad 1 - \exp(-x) = \int_0^x \exp(-t)dt.$$

Lorsqu'on travaille avec une fonction f présentant des discontinuités, on peut considérer une primitive sur chaque intervalle où f est continue, et « recoller » les morceaux continûment, ce qui est possible car la primitive d'une fonction sur un intervalle est définie à une constante près (considérer par exemple $\int_0^x f(t)dt$ lorsque f est nulle sur $] -\infty, 1[$ et vaut 1 sur $[1, +\infty[$).

Intégrales généralisées.

Lorsque le domaine de définition de la fonction f est \mathbb{R} , l'application $x \mapsto \int_a^b f(t)dt$ est définie pour tous réels a et b , et on peut se poser la question de sa limite lorsque $b \rightarrow +\infty$ ou $a \rightarrow -\infty$ (ce sont des exemples *d'intégrales généralisées*).

Lorsque la fonction f décroît suffisamment vite vers 0 en $\pm\infty$, ces deux limites existent et sont finies (c'est typiquement le cas lorsque f est nulle en dehors d'un intervalle borné, ou que $|f(x)|$ décroît vers 0 plus vite que $1/x^2$, par exemple). On dit alors que les intégrales généralisées en question sont *convergentes*.

Dans le cas d'une fonction à valeurs positives, cela signifie aussi que l'aire totale sous le graphe de f est finie, et on la notera encore par un symbole d'intégrale $\int_{\mathbb{R}} f$ ou $\int_{-\infty}^{+\infty} f$.

1.3.2 Densité et fonction de répartition

Définition 8. On appelle **fonction densité**, ou **distribution de probabilité** une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, positive et intégrable^(*) sur \mathbb{R} , telle que

$$\int_{\mathbb{R}} f(t) dt = 1.$$

En pratique, n'importe quelle intégrale convergente sur \mathbb{R} de fonction positive fournit une densité de probabilité après normalisation de la fonction : si $\int_{\mathbb{R}} g = A$, où g est une fonction positive, et $A > 0$ alors $f = g/A$ est une distribution de probabilité.

Définition 9. Soit $X : (\Omega, P) \mapsto \mathbb{R}$ une v.a. (réelle). On dit que X est une **v.a. à densité** s'il existe une densité de probabilité f telle que, pour tous réels a et b ^(†),

$$P(a \leq X \leq b) = \int_a^b f(t) dt$$

Bien que la loi de X soit totalement déterminée par sa densité f , il est parfois plus commode de travailler avec la fonction de répartition (cumulative) dont les valeurs admettent une interprétation comme probabilités :

Définition 10. Soit X une v.a. de densité f . La **fonction de répartition** de X est la primitive F de f de limite nulle en $-\infty$:

$$F(x) = \int_{-\infty}^x f(t) dt.$$

On notera qu'on a donc $P(X \leq x) = F(x)$, et qu'en particulier F admet pour limite 1 en $+\infty$.

Exemple. Prenons le cas d'une v.a. de loi uniforme sur $[0, 1]$. Sa densité est la fonction f valant 1 sur $[0, 1]$, et 0 partout ailleurs, et on a dans ce cas :

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } 1 \leq x. \end{cases}$$

1.3.3 Espérance et variance

Définition 11. Soit $X : (\Omega, P) \mapsto \mathbb{R}$ une v.a. de densité f . L'**espérance de X** est définie, lorsque l'intégrale généralisée en jeu est convergente, par

$$E(X) = \int_{\mathbb{R}} t f(t) dt.$$

Dans le cas contraire, on dit que X n'admet pas d'espérance.

(*). C'est-à-dire « assez » régulière, en un sens qu'on ne précisera pas... En pratique, on pourra penser à une fonction continue par morceaux : voir la section précédente.

(†). La définition cache en fait une difficulté : comment est-on passé d'une probabilité sur Ω , où X est définie, à une distribution sur \mathbb{R} ? Mais la théorie garantit que tout se passe bien...

On peut montrer que l'espérance ainsi définie correspond bien à la notion de *valeur moyenne* de X , comme dans le cas discret.

Plus généralement, et comme dans le cas discret, encore, étant donné une v.a. X de densité f , toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ (assez régulière...) définit une nouvelle v.a. $g(X)$ sur (Ω, P) , et on définit

$$E(g(X)) = \int_{\mathbb{R}} g(t)f(t)dt$$

lorsque cette intégrale est convergente. En particulier :

Définition 12. La *variance* de X est la valeur (lorsqu'elle est finie) de l'intégrale

$$E((X - E(X))^2) = \int_{\mathbb{R}} (t - E(X))^2 f(t)dt.$$

Notez en particulier que la question de la variance n'a de sens que pour les v.a. d'espérance finie; on montre par ailleurs que la formule alternative $\text{Var}(X) = E(X^2) - E(X)^2$ est encore valable dans le cas continu.

1.3.4 Lois normales

On présente dans cette section l'exemple de famille de lois à densité le plus important en pratique.

Définition 13. On dit qu'une v.a. X est distribuée selon une *loi normale*^(‡) d'espérance $\mu \in \mathbb{R}$ et de variance $\sigma^2 \in \mathbb{R}_+$, et on écrit $X \sim \mathcal{N}(\mu, \sigma^2)$ si, pour tous nombres réels a et b , $a \leq b$, on a

$$P(a \leq X \leq b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Remarques.

- La décroissance rapide de la fonction $t \mapsto \exp(-t)$ quand $t \rightarrow +\infty$ garantit la convergence de l'intégrale généralisée.
- Le coefficient $1/(\sigma\sqrt{2\pi})$ assure que $P(\mathbb{R}) = 1$.
- Il s'agit d'une distribution symétrique par rapport à la droite verticale $x = \mu$.

L'omniprésence théorique et pratique des lois normales est sans doute la conséquence de leurs remarquables propriétés algébriques de stabilité, propriétés qui font l'objet des deux théorèmes suivants.

Théorème 5. La normalité est conservée par transformation affine d'une v.a. de loi normale. Soit α et β des nombres réels, soit $X \sim \mathcal{N}(\mu, \sigma^2)$, alors :

$$\alpha X + \beta \sim \mathcal{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$$

(‡). Ou loi de Gauß, ou de Laplace-Gauß.

Conséquence immédiate : si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $X' = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$. Observez que la v.a. X' est alors *centrée* (son espérance vaut 0) et *réduite* (son écart type vaut 1).

Comme

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq X' \leq \frac{b - \mu}{\sigma}\right),$$

on voit que la distribution de X peut se déduire de la distribution de X' , et donc qu'il suffit de connaître la distribution $\mathcal{N}(0, 1)$ (celle-ci a été calculé avec précision et tabulée : voir en annexe de ce document).

Théorème 6. Soit $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ des v.a. **indépendantes**, alors

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Ce résultat s'étend par récurrence à la somme d'un nombre quelconque de v.a. normales indépendantes.

1.4 Résultats asymptotiques

1.4.1 Loi (faible) des grands nombres

On commence par un résultat essentiel qui majore la probabilité qu'une v.a. s'éloigne de son espérance, *l'inégalité de Bienaymé-Tchebychev*.

Théorème 7. Soit X une v.a. d'espérance μ et d'écart type σ . Pour tout nombre réel positif c , on a

$$P(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}.$$

Remarques.

- Par exemple, la probabilité qu'une v.a. s'écarte de sa moyenne de plus de 2 écarts types est inférieure à 1/4, de plus de 5 écarts types, est inférieure à 1/25. Évidemment le résultat n'a d'intérêt que pour $c > 1$.
- La majoration est indépendante de la loi de X (à espérance et variance fixées) !
- C'est la meilleure majoration possible « en général », c'est-à-dire indépendamment de la loi particulière considérée, même si la connaissance d'une loi particulière permet d'obtenir une estimation plus fine. Par exemple, pour le QI, le théorème annonce une probabilité inférieure à 1/8 de dépasser la valeur de 130, alors que celle-ci est en fait proche de 1/40.

Soit X_1, X_2, \dots, X_n une suite de v.a. *indépendantes et identiquement distribuées* (c'est-à-dire de même loi), propriété qu'on abrègera dans la suite en « *i.i.d.* », d'espérance μ , et d'écart type σ . Les X_i constituent ainsi un *échantillon* de réalisations de leur loi commune.

Définition 14. La *moyenne empirique de l'échantillon des X_i* est définie par

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

On a alors

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu \quad (\text{par linéarité de l'espérance})$$

et

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n} \quad (\text{par indépendance des } X_i)$$

En appliquant l'inégalité de Tchebychev à \bar{X}_n , dont l'écart-type vaut $\frac{\sigma}{\sqrt{n}}$, on obtient :

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

On en déduit la *loi faible des grands nombres*

Théorème 8. Pour tout $\varepsilon > 0$,

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \longrightarrow 0 \quad \text{lorsque } n \rightarrow +\infty.$$

On dit aussi que \bar{X}_n converge vers μ en probabilité.

Remarques.

- Le théorème dit par exemple que la proportion de *face* lors d'un grand nombre de lancers d'une pièce équilibrée est (probablement) proche de $1/2$, en accord avec l'intuition reliant probabilités *a priori* et fréquences à long terme.
- Mais, attention, ce résultat n'implique pas que le nombre de *face* tende vers la moitié du nombre des lancers ^(§), ni l'existence de phénomènes de « compensations ».

1.4.2 Théorème Central Limite

Soit X_1, X_2, \dots, X_n une suite de v.a. *i.i.d.* d'espérance μ et d'écart type σ .

Soit $S_n = \sum_{i=1}^n X_i$, on rappelle que

- par additivité de l'espérance : $E(S_n) = \sum_{i=1}^n E(X_i) = n\mu$;
- par indépendance : $\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) = n\sigma^2$.

Il s'ensuit immédiatement que la v.a. $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ est *centrée* et *réduite*.

Théorème 9 (Théorème Central Limite). Si n est « assez grand »

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n/n - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0, 1).$$

(Le symbole \approx se lit : « suit approximativement une loi ».)

(§). Non seulement ce dernier énoncé est faux, mais en fait l'écart entre le nombre de *face* et $n/2$ a plutôt tendance à s'accroître et à tendre vers $+\infty$, à la vitesse de \sqrt{n} .

Remarques.

- Noter que l'égalité (à gauche de l'énoncé) n'est qu'une réécriture banale en divisant numérateur et dénominateur par n .
- Le résultat est exact (il ne s'agit pas d'une approximation), quelle que soit la valeur de n , si la loi des X_i est normale, d'après les théorèmes de la section précédente.
- Le TCL dit que la distribution de la moyenne empirique d'un échantillon indépendant, après centrage et réduction, tend vers celle de la loi normale centrée réduite, lorsque la taille de l'échantillon tend vers l'infini, peu importe la loi de la v.a. X_i . C'est un exemple de *convergence en loi*.
- L'approximation est d'autant meilleure que la taille n de l'échantillon est grande. En pratique, on considère que l'approximation est légitime dès lors que $n \geq 30$.
- Dans le cas où les X_i sont des v.a. de Bernoulli de paramètre p , S_n est une loi binomiale, et on obtient

$$\frac{S_n - np}{\sqrt{np(1-p)}} \approx \mathcal{N}(0, 1).$$

Dans ce cas, l'approximation est généralement considérée comme correcte lorsque $np > 5$ et $n(1-p) > 5$.

2 Exercices et Simulations (en C)

Pour chaque exercice, essayez de préciser le modèle (univers et probabilité) considéré.

Les réponses numériques seront, chaque fois que c'est possible, données sous forme de **fractions irréductibles**. Pour obtenir une forme irréductible, il faut donc décomposer numérateurs et dénominateurs en produits de facteurs premiers (il est particulièrement inutile d'effectuer ces produits).

On pourra aussi, si on le souhaite, et pour se faire une idée plus concrète du résultat, donner une valeur décimale approchée des probabilités demandées (essayez d'obtenir une telle valeur sans l'usage de moyens électroniques).

2.1 Modèles équiprobables

Exercice 1. On lance deux dés standard, équilibrés. On considère les événements suivants

- A : le premier dé affiche 1 ;
- B : la somme des valeurs obtenues est 4 ;
- C : le second dé affiche 2 ;
- A_1 : le premier dé donne un résultat pair ;
- A_2 : le second dé donne un résultat pair ;
- A_3 : la somme des valeurs obtenues est paire.

1° Explicitiez les événements $A \cap B$, $A \cup B$, $\overline{A} \cap C$, et $\overline{A} \cap C \cup B$, et déterminer leur probabilité.

2° Parmi les événements A , B , et C , quelles sont les paires d'événements indépendants ?

3° Étudiez l'indépendance des événements A_1 , A_2 , A_3 , deux à deux, puis dans leur ensemble.

Exercice 2. Simulation de lois uniformes.

À partir du générateur pseudo-aléatoire standard `rand()`, simulez

- (a) le tirage au sort d'un nombre réel choisi uniformément entre 0 et 1 ;
- (b) le lancer d'un dé équilibré.

Il n'est pas du tout garanti *a priori* que les valeurs produites par `rand()` se comportent bien comme on l'espère d'une suite « aléatoire » d'entiers, et on se souciera constamment de tester si nos simulateurs ont bien certaines (au moins) des propriétés espérées (on se contentera d'une approche assez naïve tant il est difficile d'explicitier ce qu'il faut entendre par « suite aléatoire »).

Ici, par exemple, on testera les deux simulateurs au moyen d'un exécutable prenant en argument un entier n (la taille de l'échantillon), et calculant la répartition empirique des valeurs obtenues pour n lancers du dé, d'une part, et pour n tirages au sort entre 0 et 1, d'autre part. Dans ce dernier cas, on pourra regrouper les valeurs obtenues par dixième, voire par un nombre de classes (toutes de même largeur) variable passé comme second argument de l'exécutable.

Exemple de présentation de la sortie :

```
Test de loi uniforme modulo 6 :

-----
| 0.164 | 0.169 | 0.168 | 0.171 | 0.160 | 0.168 |
-----

Test de loi uniforme [0, 1], classes de largeur un dixième :

-----
| 0.099 | 0.104 | 0.098 | 0.101 | 0.099 | 0.101 | 0.100 | 0.101 | 0.098 | 0.098 |
-----
```

Évidemment un test aussi naïf ne dit pas grand chose de la qualité du générateur, mais il peut permettre de repérer les aberrations les plus grossières.

Exercice 3. On tire (successivement, sans remise) deux cartes au hasard dans un jeu de 52 cartes (*).

- 1° Quelle est la probabilité d'obtenir une *paire* (deux cartes de même hauteur) ?
- 2° Quelle est la probabilité d'obtenir un *black jack*, c'est-à-dire un as accompagné d'un dix, d'un valet, d'une dame ou d'un roi ?
- 3° Quelle est la probabilité que la première carte soit de hauteur strictement supérieure à la seconde ?

Exercice 4. Une urne contient 4 boules blanches et 6 rouges. On en prélève 5 (sans remise). Quelle est la probabilité d'obtenir

(*). Un jeu de 52 cartes est composé de cartes de 4 *couleurs* ($\diamond, \heartsuit, \spadesuit, \clubsuit$), chaque couleur étant constituée de cartes de 13 *hauteurs* différentes, du 2 au 10, puis Valet, Dame, Roi, As (dans l'ordre croissant).

- (a) 4 boules blanches ?
- (b) au moins une boule blanche ?
- (c) des boules des deux couleurs ?
- (d) 3 rouges et 2 blanches ?
- (e) au moins 3 rouges ?

Exercice 5. Le Chevalier de Méré était un homme instruit et brillant, et un fameux joueur, mais « n'était pas géomètre » (comprendre : versé dans les mathématiques), selon une lettre de Pascal à Fermat.

1° Le Chevalier avait le sentiment qu'il était avantageux de parier sur la sortie d'un six au moins lors de 4 lancers d'un dé. Est-ce vrai ?

2° De même pensait-il qu'il était avantageux de parier sur la sortie d'au moins un double-six lors de 24 lancers de 2 dés. Qu'en penser ?

Exercice 6. Le comte de Yarborough, deuxième du nom, paria à 1000 contre 1 qu'une main (de 13 cartes) au bridge contiendrait au moins une carte de valeur supérieure ou égale au 10^(†). Est-ce un pari raisonnable ?

Exercice 7. Simulation de tirages avec ou sans remises.

1° Pour réaliser une suite de n tirages « avec remises » parmi N valeurs, ou, ce qui revient au même de n lancers d'un dé à N faces, il suffit de répéter n fois le lancer d'un dé à N faces. On écrira une fonction

```
void with_replacement(unsigned int n, unsigned int N,  
    unsigned int* drawing);
```

simulant cette expérience (le « tableau » `drawing` contient le résultat du tirage, son allocation mémoire est faite par la fonction appelante), les objets du tirage seront représentés par les entiers de 1 à N .

2° Utilisez la fonction précédente pour réaliser une simulation des problèmes posés par le Chevalier de Méré (voir l'exercice 5), et comparez la réponse « expérimentale » à la réponse calculée.

3° Pour réaliser une suite de n tirages « sans remises » parmi N valeurs (avec évidemment $n \leq N$), une idée classique est la suivante :

- on range les N valeurs dans les N cases d'un tableau ;
- puis on tire un nombre i au hasard (*uniformément*) entre 1 et N , et on écrase la valeur écrite dans la i^{e} case du tableau par celle de la dernière case (ce qui garantit la *non remise*) ;
- on recommence en tirant au sort une des $N - 1$ premières cases du tableau, *etc.*

La fonction réalisant un tirage aléatoire sans remise de n valeurs parmi N sera appelée

(†). Autrement dit : une carte au moins parmi les 10, valets, reines, rois, ou as. On appelle aujourd'hui *main de Yarborough* une main sans carte strictement supérieure au 9.

```
void without_replacement(unsigned int n, unsigned int N,  
    unsigned int* drawing);
```

4° Utilisez votre simulateur pour évaluer la probabilité d'une *main de Yarbrough* (voir l'exercice 6).

2.2 Variables aléatoires discrètes

Exercice 8. Les 66 étudiants d'une promotion sont répartis en 3 groupes de TD : un groupe de 27, un groupe de 24 et un groupe de 15.

On tire au sort un des groupes TD, et soit X la v.a. : « nombre d'étudiants dans le groupe choisi ». On tire au sort l'un des étudiants, et soit Y la v.a. : « nombre d'étudiants dans le groupe TD de l'étudiant choisi ».

Calculez les lois, espérances, et variances de X et de Y .

Exercice 9. On lance deux dés équilibrés et on s'intéresse aux variables aléatoires suivantes

- A : « valeur absolue de la différence des nombres obtenus sur les deux dés » ;
- M : « plus grand des deux nombres obtenus ».

1° Déterminez la loi et l'espérance des v.a. A et M (on commencera évidemment par déterminer les valeurs prises).

2° Calculer la variance de M .

3° Les v.a. A et M sont-elles indépendantes ?

Exercice 10. On lance trois fois une pièce de monnaie équilibrée et on note le numéro N du lancer où apparaît pour la première fois *face*, en convenant que $N = 0$, si l'on n'obtient jamais *face*. Déterminez espérance et variance de N .

Exercice 11. Représentez graphiquement l'allure de la distribution de probabilité associée aux v.a. suivantes :

- nombre de *six* lors de 4 lancers d'un dés standard ;
- nombre de *face* lors de 10 lancers d'une pièce équilibrée.

Exercice 12. Un forain propose le jeu suivant : on parie un euro sur un entier compris entre un et six, puis trois dés sont lancés (ou bien, souvent, on lance trois fois une roue portant les nombres de 1 à 6 ...). Si le nombre sur lequel on a misé sort une fois, deux fois, trois fois, le forain rembourse la mise plus un euro, plus deux euros, plus trois euros respectivement, sinon on perd sa mise .

Ce jeu paraît souvent, à première vue, favorable au joueur. Qu'en est-il réellement ? (On pourra mettre en évidence un schéma de Bernoulli...)

Exercice 13. L'arrivée des clients au guichet d'une banque est modélisée de la manière suivante : durant chaque minute, indépendamment les unes des autres, il arrive un client (et un seul) avec probabilité $1/3$, ou bien aucun client (avec probabilité $2/3$).

Soit T la v.a. « temps écoulé (en minutes) jusqu'à l'arrivée du prochain client », toute minute entamée étant comptabilisée.

- 1° Explicitez la loi de T .
- 2° Déterminez pour tout entier positif k la probabilité $P(T > k)$.
- 3° Déterminez espérance et variance de T (cette question est un peu technique).
- 4° Reprenez l'étude précédente pour la v.a. T' où l'on ne comptabilise que les minutes complètement écoulées.

Exercice 14. Simulation de v.a. de lois binomiales et géométriques.

- 1° Commencez par simuler la réalisation d'une v.a. de Bernoulli de paramètre p par une fonction

```
| char bernoulli(double p);
```

renvoyant 0 ou 1 (p est la probabilité de succès, ou encore de renvoyer 1).

Vérifiez la correction de la simulation en calculant la fréquence de succès (qui devrait approcher la valeur du paramètre p pour un grand échantillon).

- 2° Simulez un schéma de Bernoulli (n, p) grâce à une fonction

```
| unsigned int bernoulli_scheme(unsigned int n, double p, char* result);
```

qui enregistre dans le tableau `result` les résultats des épreuves de Bernoulli, et renvoie le nombre de succès.

- 3° Utilisez la fonction précédente pour générer un échantillon des v.a. binomiales de l'exercice 11, puis comparez la distribution expérimentale à la distribution théorique correspondante. Pour générer les distributions binomiales théoriques, écrivez une fonction

```
| double* binomial_distribution(unsigned int n, double p);
```

La fonction renvoie l'adresse du tableau contenant les valeurs calculées de la distribution (c'est évidemment cette même fonction qui gère l'allocation mémoire correspondante).

Exemple de comparaison :

```
Distribution binomiale n=10, p=0.500
Comparaison avec les fréquences empiriques sur un échantillon de taille 1000
-----
| 0.001 | 0.010 | 0.044 | 0.117 | 0.205 | 0.246 | 0.205 | 0.117 | 0.044 | 0.010 | 0.001 |
-----
| 0.003 | 0.005 | 0.031 | 0.123 | 0.245 | 0.230 | 0.188 | 0.125 | 0.042 | 0.008 | 0.000 |
-----
```

(On pourra aussi calculer les différences *relatives* entre fréquences observées et probabilités attendues.

- 4° Simulez une v.a. de loi géométrique au moyen d'une fonction

```
| unsigned int geometric(double p);
```

Là encore, comparez la fréquence des valeurs obtenues sur un échantillon et la distribution théorique (ici l'ensemble des valeurs possibles est infini, donc il faudra « regrouper » les valeurs extrêmes).

Exercice 15. Le casino propose le jeu suivant : on lance une pièce équilibrée **jusqu'à** obtenir « face » ; si « face » est obtenu au n^{e} lancer le joueur gagne 2^n roubles^(‡). Quelle est l'espérance du gain ? Le paradoxe apparaît lorsqu'on pose la question : combien êtes-vous prêt à payer pour jouer une partie ? Que devient l'espérance de gain si l'on fixe un entier N tel que, à partir du N^{e} lancer, on ne gagne que 2^N roubles (après tout, la richesse du monde est finie) ?

2.3 Variables aléatoires à densité

Exercice 16. A. et B. ont rendez-vous à 19 heures, mais compte tenu des aléas de la circulation leurs temps d'arrivée respectifs T_A et T_B sont des v.a. indépendantes, uniformément distribuées sur un intervalle d'une heure, centré à l'heure de rendez-vous.

- 1° Calculez $P(T_A \leq x \text{ et } T_B \leq x)$ pour $x \in \mathbb{R}$.
- 2° Déterminez la loi (fonction de répartition puis densité) de la v.a. $T_M = \max(T_A, T_B)$. Représentez graphiquement les fonctions de densité et de répartition.
- 3° Calculez l'espérance de T_M .
- 4° Mêmes questions pour $T_m = \min(T_A, T_B)$.
- 5° Quel est en moyenne le temps d'attente du premier arrivé ?

Exercice 17. Une v.a. X admet une *distribution exponentielle de paramètre λ* si elle est à valeurs positives et de densité

$$f_\lambda(t) = \lambda \exp(-\lambda t).$$

- 1° Calculez la fonction de répartition F_λ d'une telle v.a. X .
- 2° Dessinez l'allure du graphe des fonctions f_λ et F_λ .
- 3° Déduisez-en l'expression de la fonction $F_\lambda^{-1} : [0, 1] \rightarrow \mathbb{R}$.
- 4° Vérifiez que pour tous réels a et b tels que $0 \leq a \leq b$

$$P(a \leq X \leq b) = \exp(-\lambda a) - \exp(-\lambda b).$$

- 5° Calculez $E(X)$ et $\text{Var}(X)$. On utilisera à cet effet la technique dite *d'intégration par parties*^(§), s'appuyant sur la formule

$$\int_a^b f(t)g(t)dt = [f(t)G(t)]_a^b - \int_a^b f'(t)G(t), \text{ où } G \text{ est une primitive de } g.$$

(‡). L'expérience décrite et les questions qu'elle soulève sont connues sous le nom de *Paradoxe de Saint-Petersbourg*. Elle a été imaginée par Nicolas Bernoulli en 1713 pour mettre en évidence certaines difficultés de la théorie alors naissante des probabilités, et a fait l'objet de nombreux débats entre savants depuis le début du dix-huitième siècle.

(§). En fait, c'est simplement la version « intégrale » de la *règle de Leibniz* pour la dérivée d'un produit.

Exercice 18. L'objet de l'exercice est d'expliciter le lien entre les distributions exponentielles et géométriques.

1° Soit $p \in [0, 1]$, soit $n \in \mathbb{N}$; déterminez $P(n \leq X < n + 1)$ lorsque X suit une loi exponentielle de paramètre $\lambda = -\ln(1 - p)$.

2° En déduire un procédé pour simuler une v.a. de loi géométrique de paramètre p à partir d'une v.a. de loi exponentielle (qu'on précisera).

Exercice 19. Méthode d'inversion, exemple de la loi exponentielle.

On présente une méthode élémentaire permettant de simuler certaines v.a. à densité. Plus précisément, pour que cette méthode soit applicable, il faut que l'inverse de la fonction de répartition F (celle de la v.a. qu'on souhaite simuler), soit « calculable » de manière raisonnablement simple. C'est typiquement le cas lorsque F^{-1} est une fonction *classique*, ce qui est une hypothèse très forte ^(¶).

1° L'idée est la suivante :

- on part des valeurs d'une v.a. uniforme U sur $[0, 1]$;
- pour chaque réalisation u de U , on peut calculer $F^{-1}(u) = x \in \mathbb{R}$;
- on a ainsi obtenu une v.a. $X = F^{-1}(U)$.

Calculez le fonction de répartition de X , et montrez que X a la même loi que la v.a. qu'on souhaite simuler.

2° Appliquez cette méthode à la simulation d'une v.a. de loi exponentielle de paramètre λ (voir l'exercice 17).

```
| double exp_density(double lambda);
```

On vérifiera que la moyenne de l'échantillon est proche de l'espérance de la v.a. (calculée en exercice).

3° Utiliser la fonction précédente pour simuler une v.a. géométrique de paramètre p (cf. exercice 18). On écrira à cet effet une fonction

```
| unsigned int geom_via_exp(double p);
```

Là encore, contrôlez la moyenne des valeurs obtenues, qui doit être proche de l'espérance de la v.a. simulée.

4° On dispose maintenant de deux méthodes pour simuler une v.a. géométrique : comparez les temps d'exécution des deux méthodes pour la génération d'un échantillon de taille donnée, et différentes valeurs de p .

On utilisera à cet effet la fonction `clock()` de la librairie `time.h`.

Exercice 20. Un ascenseur est censé pouvoir accueillir jusqu'à 16 personnes. On suppose que le poids de ses utilisateurs potentiels est distribué normalement, avec une moyenne de 72 kg et un écart-type de 15 kg.

Comment doit être fixée la charge maximale autorisée, de sorte que la probabilité qu'elle soit dépassée lorsque l'ascenseur fait le plein d'utilisateurs soit inférieure à 1% ?

(¶). Ce n'est souvent pas le cas en pratique, mais on dispose alors d'une généralisation très puissante de la méthode d'inversion appelée *méthode de rejet*.

Exercice 21. Simulation de v.a. distribuées normalement.

Remarquons d'abord qu'il suffit de savoir simuler une loi normale centrée et réduite, pour pouvoir simuler n'importe quelle loi normale.

Il existe plusieurs méthodes efficaces pour simuler une loi normale $\mathcal{N}(0, 1)$, mais la preuve de la correction de ces méthodes dépasse le cadre de ce cours. On admettra donc le résultat suivant

Théorème 10. *Soit (X, Y) un couple de variables aléatoires, de loi uniforme sur le disque unité (autrement dit : la probabilité que (X, Y) « tombe » dans une partie E du disque unité est proportionnelle à l'aire de E).*

Soit R et Θ les (v.a.) coordonnées polaires du point (X, Y) : $X = R \cos(\Theta)$ et $Y = R \sin(\Theta)$.

Si l'on pose $R' = \sqrt{-4 \ln(R)}$, alors $U = R' \cos(\Theta)$ et $V = R' \sin(\Theta)$ sont deux variables aléatoires indépendantes, de même loi $\mathcal{N}(0, 1)$.

Déduire du théorème précédent un algorithme de simulation d'une loi normale (cette méthode est appelée *méthode polaire*), puis des fonctions

```
void standard_normal(double* sample);
double* normal_sample(double mu, double sigma2, size_t sample_size);
```

qui produisent respectivement la réalisation de deux v.a. indépendantes $\mathcal{N}(0, 1)$ (stockées dans le tableau `sample`), et un échantillon de taille `sample_size` de v.a. indépendantes $\mathcal{N}(\mu, \sigma^2)$.

On testera la simulation en regroupant les valeurs de l'échantillon en 4 classes selon leur distance à la moyenne attendue, mesurée en nombre d'écart-types, puis en comparant les fréquences de ces classes avec les probabilités théoriques propres aux lois normales.

2.4 Résultats asymptotiques

Exercice 22.

1° On lance 100 fois une pièce équilibrée. À l'aide de l'inégalité de Tchebychev, donnez un majorant de la probabilité que le nombre de « face » s'éloigne de 50 de plus de 15.

2° Même question pour 10 000 lancers, et pour une distance à l'espérance égale à 3 écarts-types.

3° La distribution de la mesure du *quotient intellectuel standard* dans une population suit, par définition, une loi $\mathcal{N}(100, 225)$. Évaluez la probabilité qu'un individu ait un QI supérieur à 145, et comparer avec le majorant fourni par l'inégalité de Tchebychev.

Exercice 23. On présente deux méthodes d'intégration (*i.e.* de calculs d'intégrales) s'appuyant sur la *loi faible des grands nombres*, constituant deux exemples de mise en œuvre d'un principe de calcul appelé *Méthode de Monte-Carlo* ^(||).

(||). Le problème unidimensionnel présenté est très élémentaire, et dans ce cas la méthode n'a pas d'intérêt pratique, car on dispose d'algorithmes *déterministes* efficaces, mais elle devient intéressante en dimension plus grande (calcul de volumes).

On considère une fonction (disons continue) g définie sur $[0, 1]$ et prenant ses valeurs dans $[0, 1]$, et soit S_g la surface située entre l'axe des abscisses et le graphe de g , on cherche évidemment à évaluer l'aire de S_g .

1° Considérons l'expérience de Bernoulli consistant à tirer au sort (uniformément) un point (X, Y) du carré unité : on dit qu'il y a *succès* si ce point est situé dans S_g , *échec* sinon.

Soit B_n le nombre de succès lors de n répétitions indépendantes de cette expérience de Bernoulli. Quelle est la limite de B_n/n en probabilité ?

En déduire une méthode de calcul (approché) de $\int_0^1 g$.

2° Soit X une v.a. de loi uniforme sur $[0, 1]$, expliciter $E(g(X))$ en fonction de $\int_0^1 g$.

Quelle est la limite en probabilité de la v.a. qui calcule la moyenne d'une suite *i.i.d.* de réalisations de $g(X)$?

En déduire une méthode de calcul (approché) de $\int_0^1 g$.

Exercice 24. Intégration Monte-Carlo.

Mettre en œuvre les deux méthodes de calcul approché d'intégrales présentées à l'exercice 23, sur l'exemple de la fonction définie sur $[0, 1]$ par $g : x \mapsto \sqrt{1 - x^2}$.

Dans ce cas, on connaît l'aire exacte sous la courbe de g : que vaut-elle ?

Essayez de comparer les deux méthodes au moyen d'un exécutable prenant en argument la taille d'échantillon souhaitée.

Exercice 25. On considère un examen constitué de 48 questions auxquelles on répond par « Vrai » ou « Faux ». Un étudiant réussit l'examen s'il répond correctement à 30 questions ou plus.

1° Quelle est la probabilité qu'un étudiant répondant juste avec probabilité $3/4$ passe l'examen avec succès ?

2° Même question pour un étudiant répondant au hasard.

Donnez une « formule » exacte pour le calcul des probabilités demandées, puis des valeurs approchées de celles-ci.

Exercice 26. On s'intéresse aux intentions de vote au sein d'une population ^(**), lors d'un scrutin qui oppose deux candidats A. et B. La proportion (réelle) d'individus votant A. dans la population est $p \in [0, 1]$. Un sondage effectué sur n individus de la population étudiée est vu comme la réalisation d'un schéma de Bernoulli (n, p) , et le nombre d'individus déclarant leur intention de vote pour A. comme une v.a. $S_n \simeq \mathcal{B}(n, p)$.

1° Justifiez que pour n assez grand

$$\frac{S_n - np}{\sqrt{np(1-p)}} = \sqrt{n} \frac{S_n/n - p}{\sqrt{p(1-p)}} \approx \mathcal{N}(0, 1).$$

(**). On remarquera que la taille de la population n'apparaît pas dans le problème.

2° En déduire des réels z_α tels qu'on ait, approximativement,

$$P\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \geq z_\alpha\right) \leq \alpha$$

pour les valeurs α égales à 1%, 5%, 10%.

3° Montrez que p appartient à l'intervalle

$$\frac{S_n}{n} \pm r_\alpha, \text{ où } r_\alpha = \frac{z_\alpha}{2\sqrt{n}}$$

avec une probabilité supérieure à $1 - \alpha$. Cet intervalle *aléatoire* (ces bornes sont des v.a.) est appelé *intervalle de confiance au niveau de confiance $1 - \alpha$ pour p* .

Une réalisation concrète de l'expérience produit un *intervalle de fluctuation pour p* .

4° Déterminez pour chaque valeur de α une valeur de l'entier n telle que $r_\alpha \leq 0,01$.

Exercice 27. Sondages et intervalles de confiance.

Pour une description de la situation étudiée et l'introduction des notations, on renvoie à l'exercice 26. L'objectif du tp est d'essayer de cerner la notion d'intervalle de confiance.

Écrire un programme réalisant la simulation de **nb** sondages, sur un échantillon de **n** individus, destinés à estimer un paramètre $p \in [0, 1]$. Plus précisément pour chaque sondage, on en déduira un intervalle de fluctuation pour le paramètre p au seuil **alpha** (qui est un paramètre du programme). La taille **n** de l'échantillon sera déterminée de sorte que le rayon de l'intervalle de confiance soit égal à un centième, il dépendra donc du niveau de confiance **alpha** visé.

Enfin le programme déterminera le pourcentage de sondages pour lesquels l'intervalle de fluctuation (au seuil **alpha**) calculé contient la vraie valeur du paramètre p .

On aura donc un exécutable dépendant de trois paramètres : **p**, **nb**, et **alpha**, ce dernier pouvant prendre les valeurs 1%, 5%, 10%.

Annexe : Table de la loi normale centrée réduite

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995

Table des matières

1	Abrégé de cours	1
1.1	Espaces probabilisés	1
1.1.1	Expérience aléatoire, modèle, univers	1
1.1.2	Mesures de probabilités	1
1.1.3	Univers finis, équiprobabilité	2
1.1.4	Indépendance	3
1.2	Variable aléatoire (cas discret)	3
1.2.1	Définition et exemples	3
1.2.2	Espérance	4
1.2.3	Schéma de Bernoulli	5
1.2.4	Variance	5
1.2.5	Indépendance de variables aléatoires	6
1.3	Variables aléatoires à densité	7
1.3.1	Intégration (guide de survie)	7
1.3.2	Densité et fonction de répartition	9
1.3.3	Espérance et variance	9
1.3.4	Lois normales	10
1.4	Résultats asymptotiques	11
1.4.1	Loi (faible) des grands nombres	11
1.4.2	Théorème Central Limite	12
2	Exercices et Simulations (en C)	13
2.1	Modèles équiprobables	13
2.2	Variables aléatoires discrètes	16
2.3	Variables aléatoires à densité	18
2.4	Résultats asymptotiques	20
	Annexe : Table de la loi normale centrée réduite	23