Calcul des probabilités, Deug deuxième année

Gérard Letac*

I L'espace de probabilités $(\Omega, \mathcal{A}, P)$.

Le calcul des probabilités est la science qui modélise les phénomènes aléatoires. Une modélisation implique donc certainement une simplification des phénomènes, mais cette simplification conduit à une quantification, donc à la possibilité de faire des calculs et à prédire. Le jet d'un dé, le tirage du Loto pourraient être analysés par les lois de la mécanique, mais ce serait trop compliqué pour être utile. La modélisation du calcul des probabilités a été inventée par A. N. Kolmogorov dans un livre paru en 1933. Cette modélisation est faite à partir de 3 objets $(\Omega, \mathcal{A}, P)$ que nous allons décrire.

1.1 L'espace des observables Ω.

Nous conviendrons que effectuer une expérience, c'est sélectionner par un procédé quelconque un élément ω dans un ensemble Ω: jeter un dé est sélectionner un élément de $\Omega = \{1, 2, 3, 4, 5, 6\}$; jeter ensemble deux dés rouge et vert est sélectionner un élément de l'ensemble $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ des couples ordonnés (i, j) avec $1 \leq i \leq 6$ et $1 \leq j \leq 6$ (ici Ω a 36 points). Plus délicat: jeter ensemble deux dés indiscernables est sélectionner un élément de l'ensemble Ω des couples (i, j) avec $1 \leq i \leq j \leq 6$ (ici Ω a $6 + \frac{1}{2} 6 \times 5 = 21$ points). Observer la durée de vie d'une ampoule de 100 watts est sélectionner un élément de $\Omega = [0, +\infty[$. Mesurer la durée de vie de 12 ampoules de 100 watts est sélectionner un élément de $\Omega = [0, +\infty[^{12}$.

Cet ensemble Ω est appelé l'espace des observables. On dit aussi dans la littérature l'espace échantillon, l'espace des évènements- élémentaires, l'expérimental ou encore l'évènementiel. Ses points ω sont appelés observables ou évènements- élémentaires. Il est très important qu'il soit clairement défini. On peut s'exercer à définir Ω dans les 2 cas suivants : jeter 12 fois de suite la même pièce de monnaie, jeter en même temps 12 pièces de monnaie identiques (on admet que la pièce tombe sur pile ou sur face, et jamais sur la tranche).

1.2 La tribu des évènements $\mathcal{A}$.

Les questions qu'on se pose sur le résultat d'une expérience sont systématiquement du type suivant: on choisit un sous ensemble A de l'espace d'observables Ω et on se demande:

*Université Paul Sabatier, 31062 Toulouse.

le résultat ω de l'expérience va-t-il tomber dans A ou non? Les parties de Ω pour lesquelles on se pose ce genre de question sont appelées des évènements. Un des premiers points délicats de la théorie est que on ne va pas toujours considérer tous les sous ensembles de Ω comme des évènements. Dans l'exemple de la lampe de 100 watts, il parait inintéressant de se demander si sa durée de vie, mesurée en heures, est un nombre irrationnel, et intéressant de se demander si elle tombe dans l'intervalle $[300, 400]$. L'idée de Kolmogorov est que l'ensemble $\mathcal{A}$ des évènements a une structure de tribu, c'est-à-dire satisfait aux trois axiomes suivants

1. Si $A \in \mathcal{A}$, alors son complémentaire $A^c = \Omega \setminus A$ est aussi dans $\mathcal{A}$.

2. Si on a une suite finie ou dénombrable $A_1, \ldots, A_n, \ldots$ d'éléments de $\mathcal{A}$, alors leur réunion $\bigcup_{n \geq 1} A_n$ est aussi dans $\mathcal{A}$.

3. L'ensemble vide $\emptyset$ est dans $\mathcal{A}$.

Tirons quelques conséquences de ces axiomes.

Proposition 1.1: Soit $\mathcal{A}$ une tribu de parties de l'ensemble Ω. Alors $\Omega \in \mathcal{A}$. De plus, si on a une suite finie ou dénombrable $A_1, \ldots, A_n, \ldots$ d'éléments de $\mathcal{A}$, alors leur intersection $\bigcap_{n \geq 1} A_n$ est aussi dans $\mathcal{A}$.

Démonstration: En appliquant les axiomes 1 et 3, on a le premier résultat. Pour le second, il suffit de se rappeler que le complémentaire d'une réunion finie ou infinie d'ensembles est l'intersection des complémentaires ("Loi de Morgan"). Donc

$$\bigcap_{n \geq 1} A_n = (\bigcup_{n \geq 1} A_n^c)^c,$$

et le deuxième membre de cette égalité est donc dans $\mathcal{A}$: on applique successivement l'axiome 1, puis 2, puis 1 à nouveau.

Le langage de la théorie des ensembles permet des calculs systématiques sur les évènements. Toutefois, il faut savoir que le langage courant, que nous utilisons dans une première étape pour décrire des évènements a sa traduction ensembliste. Voici un petit dictionnaire :

Ensemble Ω:	évènement certain
Ensemble vide:	évènement impossible
$A \cup B$:	A *ou* B sont réalisés ("ou" non exclusif)
$A \cap B$:	A *et* B sont réalisés
A et B sont disjoints:	les évènements A et B sont incompatibles
$A^c = \Omega \setminus A$:	évènement contraire de A.

Le fait que on ne sorte pas de la famille des évènements intéressants à considérer en prenant une intersection ou une réunion d'évènements est raisonnable si ceux ci sont en nombre fini. Le fait de se permettre ceci également quand on en a une infinité est plus subtil: les mathématiques ne maniant que des ensembles finis sont élémentaires mais les résultats exacts auquels elles conduisent sont trop compliqués pour être utilisables.

Le passage à l'infini est le passage de l'algèbre à l'analyse, donc à des approximations maniables et à de puissantes techniques issues du calcul différentiel et intégral. Quant au fait que dans ce passage à l'infini, on se limite à une infinité *dénombrable* d'évènements, c'est un point technique qu'on ne justifiera que dans un cours de 3 ème année d'université. Rappelons qu'un ensemble E avec une infinité d'éléments est dit dénombrable si il existe une bijection entre E et l'ensemble $\mathbf{N}$ des entiers positifs: l'ensemble $\mathbf{Q}$ des nombres rationnels est dénombrable, le segment $[0,1]$ ne l'est pas, comme nous l'avons vu en première année.

Finalement, ce point délicat: "on ne considère pas nécessairement tout sous ensemble A de Ω comme un élément de la tribu $\mathcal{A}$ des évènements" ne jouera pas un grand rôle dans la suite. Typiquement, nous envisagerons deux cas particuliers importants:

- Le cas où Ω lui même est dénombrable, et nous prendrons comme tribu $\mathcal{A}$ la famille $\mathcal{P}(\Omega)$ de tous les sous ensembles de Ω.

- Le cas où Ω est la droite réelle $\mathbb{R}$. Nous prendrons alors pour tribu $\mathcal{A}$ la tribu $\mathcal{B}$ (dite tribu de Borel, dont les éléments sont appelés des *boréliens*) qui est la plus petite tribu qui contient tous les intervalles de $\mathbb{R}$. On peut laborieusement démontrer que $\mathcal{B} \neq \mathcal{P}(\mathbb{R})$; toutefois, une description complète des éléments de $\mathcal{B}$ n'est pas possible, et en fait pas très utile en pratique: les seuls boréliens que nous aurons à manipuler seront les intervalles (attention, $\mathbb{R}$ ou une demi droite sont aussi des intervalles) ou des réunions finies, ou plus rarement, dénombrables, d'intervalles.

Ce ne sont pas les seuls espaces de probabilité utilisés: on verra le schéma Succès Echec à la section 2 et le cas $\Omega = \mathbb{R}^2$ ou $\mathbb{R}^n$ plus tard.

1.3 La probabilité P.

Etant donnés un espace d'observables Ω et une tribu d'évènements $\mathcal{A}$ formée de certains sous ensembles de Ω, une probabilité P est une application de $\mathcal{A}$ dans $[0,1]$, donc un procédé qui associe à tout évènement A un nombre $P(A)$ compris entre 0 et 1 appelé probabilité de A, et qui satisfait aux axiomes suivants

- L'évènement certain est de probabilité 1: $P(\Omega) = 1$.

- Axiome d'additivité dénombrable:

  pour toute suite $A_1, A_2, \ldots, A_n \ldots$ d'évènements de $\mathcal{A}$ qui sont de plus deux à deux disjoints, c'est à dire tels que $A_k \cap A_j = \emptyset$ si $k \neq j$, alors la série

$$\sum_{k=1}^{\infty} P(A_k)$$

  converge et a pour somme $P(\bigcup_{k\geq 1} A_k)$.

Le triplet $(\Omega, \mathcal{A}, P)$ est alors appelé un espace de probabilité. Voici quelques conséquences immédiates des axiomes.

Théorème 1.2 Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Alors
1) $P(\emptyset) = 0$.
2) Si $A_1, A_2, \ldots, A_n$ dans $\mathcal{A}$ sont deux à deux disjoints, alors

$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n);$$

en particulier $P(A^c) = 1 - P(A)$. Si les $A_1, A_2, \ldots, A_n$ ne sont pas nécessairement deux à deux disjoints, alors

$$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n).$$

3) Si A et B sont dans $\mathcal{A}$ et si $A \subset B$ alors $P(A) \leq P(B)$.
4) Si A et B sont dans $\mathcal{A}$, mais ne sont pas nécessairement disjoints, alors

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Si les $A_1, A_2, \ldots, A_n$ dans $\mathcal{A}$ ne sont pas nécessairement deux à deux disjoints, alors

$$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n).$$

5) Continuités croissante et décroissante: Soit une suite $B_1, B_2, \ldots, B_n \ldots$ d'évènements de $\mathcal{A}$ qui soit ou bien croissante (c'est à dire que pour tout $n \geq 1$ on a $B_n \subset B_{n+1}$) ou bien décroissante (c'est à dire que pour tout $n \geq 1$ on a $B_n \supset B_{n+1}$). Alors, dans le cas croissant:

$$\lim_{n \to +\infty} P(B_n) = P(\bigcup_{n \geq 1} B_n);$$

et dans le cas décroissant:

$$\lim_{n \to +\infty} P(B_n) = P(\bigcap_{n \geq 1} B_n).$$

6) Sous additivité dénombrable: Soit une suite $B_1, B_2, \ldots, B_n \ldots$ d'évènements de $\mathcal{A}$. Alors ou bien la série $\sum_{k=1}^{\infty} P(B_k)$ diverge; ou bien elle converge et dans ce cas sa somme est $\geq P(\bigcup_{n \geq 1} B_n)$.

Démonstration: 1) L'axiome d'additivité dénombrable est appliquable à la suite constante définie par $A_n = \emptyset$, qui est effectivement formée d'évènements deux à deux disjoints. La série dont le terme général $P(\emptyset)$ est constant ne peut converger que si ce terme général est 0.

2) Sa première partie se démontre en appliquant l'axiome d'additivité dénombrable à $A_1, A_2, \ldots, A_n$ continuée par $\emptyset = A_{n+1} = A_{n+2} = \cdots$, et en utilisant le 1). Appliquer ça à $n = 2$, $A_1 = A$ et $A_2 = A'$ fournit $1 = P(\Omega) = P(A) + P(A^c)$ en utilisant le premier axiome d'une probabilité.

3) On écrit $B = A \cup (B \setminus A)$ comme réunion de deux ensembles disjoints (notez que $B \setminus A = B \cap A'$ est bien dans $\mathcal{A}$), et on applique le 2): $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

4) De même on écrit

$$P(B) = P(A \cap B) + P(B \setminus A), \; P(A) = P(A \cap B) + P(A \setminus B),$$

puis on écrit $A \cup B = (A \cap B) \cup (B \setminus A) \cup (A \setminus B)$ comme réunion de trois ensembles deux à deux disjoints et on applique le 2):

$$P(A \cup B) = P(A \cap B) + P(B \setminus A) + P(A \setminus B) =$$
$$P(A \cap B) + (P(B) - P(A \cap B)) + (P(A) - P(A \cap B)) = P(A) + P(B) - P(A \cap B);$$

Pour terminer le 4) on démontre le résultat par récurrence sur n. C'est trivial pour $n = 1$. Si c'est démontré pour n, appliquons la première partie de ce 4) à $A = A_1 \cup \cdots \cup A_n$ et à $B = A_{n+1}$. On obtient, à l'aide de l'hypothèse de récurrence

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B) \leq (\sum_{k=1}^{n} P(A_k)) + P(A_{n+1}).$$

5) Dans le cas croissant, posons $A_1 = B_1$ et, pour $n \geq 2$, $A_n = B_n \setminus B_{n-1}$. Les $A_1, A_2, \ldots, A_n \ldots$ sont alors deux à deux disjoints. La série $\sum_{k=1}^{\infty} P(A_k)$ est donc convergente. D'après le 2) on a

$$P(B_n) = P(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} P(A_k)$$

Passons à la limite dans l'égalité ci dessus; on obtient

$$\lim_{n \to +\infty} P(B_n) = \sum_{k=1}^{\infty} P(A_k).$$

Or d'après l'axiome d'additivité dénombrable, le second membre est $P(\bigcup_{k \geq 1} A_k)$, qui est aussi par définition des A_n égal à $P(\bigcup_{n \geq 1} B_n)$.

Dans le cas décroissant, on se ramène au cas précédent par passage au complémentaires, à l'aide de la loi de Morgan: le complémentaire d'une union est l'intersection des complémentaires:

$$\lim P(B_n) = 1 - \lim P(B_n^c) = 1 - P(\cup_{n \geq 1} B_n^c) = 1 - (1 - P(\cap_{n \geq 1} B_n)) = P(\cap_{n \geq 1} B_n).$$

6) La suite d'évènements définie par $C_n = B_1 \cup \cdots \cup B_n$ est croissante et on peut lui appliquer le 5). En utilisant aussi la sous additivité finie du 2) on a donc

$$P(\bigcup_{n \geq 1} B_n) = \lim_{n \to +\infty} P(C_n) \leq \lim_{n \to +\infty} (P(B_1) + \cdots + P(B_n)) = \sum_{k=1}^{\infty} P(B_k).$$

Exercices sur la section 1.

1. Soit A, B, C trois évènements d'un espace de probabilité. Montrer à l'aide du Th. 1.2 4) que

   $$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C).$$

   Etablir une formule de ce genre pour une réunion de 4 évènements.

2. Soit $\mathcal{A}$ une tribu d'évènements sur Ω, et soit f une fonction positive sur $\mathcal{A}$ ayant les propriétés suivantes: $f(\Omega) = 1$, $f(A \cup B) = f(A) + f(B)$ si A et B sont des évènements disjoints et , si (B_n) est une suite décroissante de $\mathcal{A}$ telle que $\cap_{n \geq 1} B_n = \emptyset$ alors
$$\lim_{n \to +\infty} f(B_n) = 0.$$
Montrer qu'alors f est une probabilité. Méthode: si (A_n) est une suite d'évènements deux à deux disjoints, considérer
$$B_n = \cup_{k \geq n+1} A_k.$$

II Quatre espaces de probabilité importants.

2.1 L'espace Ω est fini ou dénombrable.

Dans ce cas on suppose habituellement que la tribu des évènements $\mathcal{A}$ est $\mathcal{P}(\Omega)$, l'ensemble de toutes les parties de Ω. Par exemple, si Ω est formé de 2 éléments notés a et b, alors $\mathcal{P}(\Omega)$ est constitué des 4 sous ensembles suivants: l'ensemble vide $\emptyset$, les deux singletons $\{a\}$ et $\{b\}$ et $\Omega = \{a, b\}$ lui même. Plus généralement, on a le fait suivant:

Proposition 2.1: Si un ensemble Ω a un nombre fini N d'éléments, alors $\mathcal{P}(\Omega)$ a 2^N éléments.

Démonstration: On procède par récurrence sur N. C'est trivial pour $N = 1$ ou 0. Si c'est vrai pour N, considérons
$$\Omega = \{a_1, \ldots, a_N, a_{N+1}\} \quad \text{et} \quad \Omega' = \{a_1, \ldots, a_N\}.$$
Les parties de Ω se partagent en deux catégories:

Catégorie 1: celles qui ne contiennent pas a_{N+1}.

Catégorie 2: celles qui contiennent a_{N+1}.

Il est clair que la catégorie 1 est égale à $\mathcal{P}(\Omega')$ et que la catégorie 2 est en bijection avec $\mathcal{P}(\Omega')$, la bijection étant obtenue en ajoutant a_{N+1} aux éléments de $\mathcal{P}(\Omega')$. Comme d'après l'hypothèse de récurrence $\mathcal{P}(\Omega')$ a 2^N éléments, on en conclut que $\mathcal{P}(\Omega)$ a $2^N + 2^N = 2^{N+1}$ éléments, et la récurrence est étendue.

Proposition 2.2: Si Ω est infini dénombrable, alors $\mathcal{P}(\Omega)$ est infini non dénombrable.

Démonstration: La démonstration est analogue à la démonstration de Cantor. Sans perte de généralité on suppose Ω égal à l'ensemble $\mathbf{N}$ des entiers positifs ou nuls. Si $X \subset \mathbf{N}$, on lui associe la fonction indicatrice $\mathbf{1}_X$ définie sur $\mathbf{N}$ et à valeurs 0 ou 1 par $\mathbf{1}_X(k) = 1$ si $k \in X$ et $\mathbf{1}_X(k) = 0$ si $k \notin X$. Remarquons aussi qu'inversement, si une fonction f définie sur $\mathbf{N}$ est à valeurs 0 ou 1, alors c'est une indicatrice d'ensemble, c'est-à-dire qu'il existe X tel que $f = \mathbf{1}_X$: il s'agit de $X = \{k \in \mathbf{N}; f(k) = 1\}$.

Montrons alors la proposition par l'absurde en supposant que $\mathcal{P}(\mathbf{N})$ soit dénombrable, c'est-à-dire qu'il existe une application bijective $n \mapsto X_n$ de $\mathbf{N}$ sur $\mathcal{P}(\mathbf{N})$. Alors la fonction f définie sur $\mathbf{N}$ et à valeurs 0 ou 1 par
$$f(k) = 1 - \mathbf{1}_{X_k}(k)$$

est l'indicateur de quelque sous ensemble X_n de $\mathbf{N}$ et donc pour tout k de $\mathbf{N}$ on a

$$\mathbf{1}_{X_n}(k) = 1 - \mathbf{1}_{X_k}(k),$$

ce qui est une contradiction si $k = n$.

Les probabilités sont alors décrites par le résultat suivant

Proposition 2.3: Soit Ω un ensemble fini ou dénombrable. Soit $x \mapsto p_x$ une application de Ω dans les réels ≥ 0 telle que

$$\sum_{x\in\Omega} p_x = 1.$$

Pour tout $A \subset \Omega$, notons alors

$$P(A) = \sum_{x\in A} p_x.$$

Alors $(\Omega, \mathcal{P}(\Omega), P)$ est un espace de probabilité. Inversement, toute probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ est du type précédent, avec $p_x = P(\{x\})$.

Remarque: Si Ω est fini, la proposition est évidente. Si Ω est dénombrable, les sommes ci dessus quand A est dénombrable ont la signification suivante: puisque A est dénombrable, on peut numéroter ses éléments, c'est-à-dire qu'il existe une application bijective $n \mapsto x_n$ de $\mathbf{N}$ sur A. $P(A)$ est alors défini rigoureusement comme la somme de la série $\sum_{n=0}^{\infty} p_{x_n}$. Toutefois, ce nombre ne dépend que de A, et non de la numérotation particulière de A choisie par $n \mapsto x_n$, grâce au théorème suivant sur les séries, que nous admettrons, ainsi que la proposition elle même:

Théorème 2.4: Si la série $\sum_{n=0}^{\infty} u_n$ est absolument convergente de somme S, et si $n \mapsto \sigma(n)$ est une bijection de $\mathbf{N}$ sur lui même, alors $\sum_{n=0}^{\infty} u_{\sigma(n)}$ est aussi absolument convergente et de somme S.

Exercices sur 2.1.

1. Soit $\lambda > 0$. Soit P la probabilité définie sur $(\mathbf{N}, \mathcal{P}(\mathbf{N}))$ par

   $$P(\{n\}) = e^{-\lambda}\frac{\lambda^n}{n!}.$$

   Soit A l'ensemble des nombres pairs. Calculer $P(A)$. Soit N un entier, montrer que

   $$P(\{0, 1, \ldots, N\}) = 1 - \int_0^{\lambda} e^{-t}\frac{t^N}{N!}dt$$

   (Méthode: considérer les deux membres comme des fonctions de λ dont on montrera qu'elles ont même valeur pour $\lambda = 0$ et même dérivée).

2. Soit P la probabilité définie sur $(\mathbf{N}^*, \mathcal{P}(\mathbf{N}^*))$ par $P(\{n\}) = 2^{-n}$. Calculer la probabilité de tirer un nombre $n > 3$; un nombre n multiple de 3; un nombre dont le reste est 3 si on le divise par 4.

2.2 Le cas équiprobable.

Considérons le cas particulier de la Proposition 2.3 où Ω a un nombre fini $N = |\Omega|$ d'éléments et où tous les p_x sont égaux (et donc égaux à $1/N$.) Dans ce cas, si $A \subset \Omega$ on a

$$P(A) \ = \frac{|A|}{|\Omega|} = \ \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Pour exploiter cette égalité, il est nécessaire de posséder quelques principes généraux de dénombrement d'ensembles et de fonctions contenus dans les deux prochains théorèmes. Si E et F sont des ensembles, on note par $E \times F$ leur produit cartésien, c'est-à-dire l'ensemble des couples (x, y) tels que $x \in E$ et $y \in F$. On note par F^E l'ensemble des fonctions f définies sur E et à valeurs dans F. Si E est fini et est de taille $n = |E|$ et si k est un entier avec $0 \leq k \leq n$ on note par $\mathcal{P}_k(E)$ l'ensemble des parties de E de taille k.

Théorème 2.5: a) Si E et F sont des ensembles finis, alors $|E \times F| = |E| \times |F|$. Plus généralement, si $F_1, \ldots, F_n$ sont des ensembles finis: $|F_1 \times \cdots \times F_n| = |F_1| \times \cdots \times |F_n|$. Ensuite $|F^E| = |F|^{|E|}$. Enfin, si $p = |F| \geq n = |E|$, le nombre de fonctions *injectives* de E vers F est $p(p-1)(p-2)\cdots(p-n+1)$. En particulier, le nombre de fonctions bijectives de E vers E, appelées permutations de E, est égal à $n!$

b) Si E est fini et est de taille $n = |E|$ et si k est un entier avec $0 \leq k \leq n$ alors

$$|\mathcal{P}_k(E)| = C_n^k = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}.$$

Démonstration: a) La première formule est évidente : si $e_1, \ldots, e_n$ et $f_1, \ldots, f_p$ sont les éléments de E et F, le nombre de couples (e_i, f_j) est np. L'extension à n facteurs est immédiate également. Cette extension est ensuite appliquée au cas particulier où tous les ensembles F_j sont égaux au même ensemble F. Si $|E| = n$, il y a alors bijection entre F^E et $F \times \cdots \times F$ (n fois). D'où $|F^E| = |F| \times \cdots \times |F| = |F|^n = |F|^{|E|}$. Quant au nombre de fonctions injectives, la formule donnée se justifie facilement: on identifie E à $(1, 2, \ldots, n)$, et l'image de 1 peut occuper p positions, l'image de 2 peut occuper une des $p-1$ positions restantes, l'image de 3 une des $p-2$ positions restantes, etc. Faire $E = F$ pour le nombre de permutations de E (on rappelle que si $|E| = |F|$ avec E fini, alors une fonction f de E vers F est injective si et seulement si elle est surjective).

Pour la partie b), rappelons la formule de Pascal:

Proposition 2.6: Si k est un entier avec $1 \leq k \leq n$ on a

$$C_n^{k-1} + C_n^k = C_{n+1}^k.$$

Démonstration:

$$C_n^{k-1} + C_n^k = \frac{n!}{(k-1)!(n-k)!}\left[\frac{1}{n-k+1} + \frac{1}{k}\right] = \frac{(n+1)!}{k)!(n+1-k)!} = C_{n+1}^k.$$

Pour montrer b), on observe que c'est trivial pour $k = 0$, puis on fixe $k > 0$ et on montre b) par récurrence sur n. C'est trivial pour $n = k$. Supposons enfin b) vrai pour n

et supposons que E ait $n+1$ éléments, qu'on prend égaux à $1, 2, \ldots, n+1$ sans perte de généralité. Soit aussi E' l'ensemble des n premiers entiers. On partage alors les éléments de $\mathcal{P}_k(E)$ en deux catégories:

Catégorie 1: ceux qui ne contiennent pas $n+1$.

Catégorie 2: ceux qui contiennent $n+1$.

La catégorie 1 est égale à $\mathcal{P}_k(E')$ et a donc C_n^k éléments par l'hypothèse de récurrence. La catégorie 2 est en bijection avec $\mathcal{P}_{k-1}(E')$ (enlever $n+1$ à un membre de la catégorie 2 pour avoir un élément de $\mathcal{P}_{k-1}(E')$) et donc par l'hypothèse de récurrence a C_n^{k-1} éléments. La formule de Pascal montre alors que $\mathcal{P}_k(E)$ a C_{n+1}^k éléments et la récurrence est étendue.

Voici un exemple d'application du théorème précédent.

Anniversaires. n personnes sont réunies. Quelle est la probabilité que au moins deux d'entre elles aient le même anniversaire?

On formalise le problème en le simplifiant un peu: on ignore d'abord le problème du 29 février, et on postule donc que l'espace des observables est $\Omega = F^E$ où E est l'ensemble des personnes et où F est l'ensemble des $p = 365$ jours de l'année: on observe donc la fonction $f \in \Omega$ qui à chaque personne associe son anniversaire. On postule ensuite qu'on est dans le cas équiprobable, ce qui n'est qu'une approximation: il y a plus d'enfants conçus au printemps et en été qu'en novembre sous nos climats. Finalement, il est plus facile de calculer la probabilité du complémentaire A^c de l'évènement A "deux personnes au moins ont le même anniversaire", car c'est la probabilité que la fonction f soit injective. D'après le théorème 2.5 a), c'est

$$P(A^c) = \frac{1}{365^n} 365(365-1)\cdots(365-n+1) = \prod_{k=1}^{n-1}(1-\frac{k}{365}) = \exp\sum_{k=1}^{n-1}\log(1-\frac{k}{365}).$$

Si n n'est pas grand, une évaluation approximative de cette somme se fait en remplaçant $\log(1-x)$ par $-x$ et en utilisant la somme d'une progression arithmétique étudiée en Terminale

$$\sum_{k=1}^{n-1} k = \frac{1}{2}n(n-1) \sim n^2/2,$$

qui conduit à l'approximation $P(A^c) \sim \exp(-n^2/730)$. Pour voir par exemple pour quel n on a $P(A^c) \sim 1/2$ on prend $n \sim \sqrt{730\log 2} \sim 23$. Pour un calcul plus sérieux, on peut utiliser l'encadrement pour $0 < x < 1$:

$$-x - \frac{x^2}{2(1-x)} < \log(1-x) < -x - \frac{x^2}{2};$$

La majoration de droite se déduit du développement en série entière, celle de gauche se montre en étudiant la fonction $x + \frac{x^2}{2(1-x)} + \log(1-x)$. On a aussi besoin de la somme des premiers carrés:

$$\sum_{k=1}^{n-1} k^2 = \frac{1}{6}n(2n-1)(n-1) \sim n^3/3,$$

qui s'établit par récurrence. Si $x \leq (n-1)/365$, alors $-1/(1-x) \geq -365/(365-n+1)$. D'où l'encadrement :

$$-\frac{n(n-1)}{2}\frac{1}{365} - \frac{n(n-1)(2n+1)}{6}\frac{1}{2\times 365^2}\frac{365}{365-n+1} <$$

$$\log P(A^c) < -\frac{n(n-1)}{2}\frac{1}{365} - \frac{n(n-1)(2n+1)}{6}\frac{1}{2\times 365^2}.$$

Par exemple, si $n = 35$ on trouve $P(A^c) = 0,186...$ Si 35 personnes sont réunies, la probabilité que deux d'entre elles au moins aient le même anniversaire est donc $0,813...$

Le prochain théorème sert en particulier à résoudre le problème plus difficile du calcul du nombre de fonctions *surjectives*.

Théorème 2.7: (Principe d'inclusion- exclusion) Soit E un ensemble fini et soit f et g des fonctions réelles définies sur $\mathcal{P}(E)$ satisfaisant pour tout $A \subset E$:

$$f(A) = \sum_{B\subset A} g(B).$$

Alors pour tout $A \subset E$:

$$g(A) = \sum_{B\subset A} (-1)^{|A\setminus B|} f(B).$$

Démonstration: Si $C \subset A \subset E$ notons

$$F(A,C) = \sum_{C\subset B\subset A} (-1)^{|A\setminus B|}.$$

Si $|A \setminus C| = n$, puisque il y a C_n^k parties de $A \setminus C$ de taille k on peut écrire $F(A,C) = \sum_{k=0}^n (-1)^k C_n^k$, qui est à son tour $(1+(-1))^n$ à cause de la formule du binôme de Pascal $(a+b)^n = \sum_{k=0}^n a^{n-k} b^k C_n^k$. Donc si $n > 0$, c'est-à-dire si $C \neq A$, on a $F(A,C) = 0$. Si $n = 0$, c'est-à-dire si $C = A$ on a $F(A,C) = 0$. Calculons alors le second membre de l'égalité à démontrer:

$$\sum_{B\subset A} (-1)^{|A\setminus B|} f(B) = \sum_{B\subset A} (-1)^{|A\setminus B|} \sum_{C\subset B} g(C) =$$

$$\sum_{C\subset B} g(C) \sum_{C\subset B\subset A} (-1)^{|A\setminus B|} = \sum_{C\subset B} g(C) F(A,C) = g(A).$$

La première égalité exploite le lien entre f et g, la seconde inverse les sommations par rapport aux indices de sommation B et C, la troisième résulte de la définition de $F(A,C)$, la quatrième du calcul de F précédent et fournit le résultat voulu.

Voici deux applications.

Nombre de fonctions surjectives. Si $|E| = n \geq |F| = p$, quel est le nombre de fonctions surjectives de E vers F? Pour répondre on applique le théorème précédent aux fonctions f et g définies sur $\mathcal{P}(F)$ ainsi: si $A \subset F$, $f(A) = |A|^n$ est le nombre de fonctions de E vers F dont l'image est contenue dans A (on pourrait donc dire tout aussi bien les fonctions de E vers A); et $g(A)$ est le nombre de fonctions de E vers F dont l'image est exactement égale à A (on pourrait dire les fonctions de E vers A qui sont surjectives). On veut donc calculer $g(F)$.

Les hypothèses du théorème sont remplies, on a bien en effet $f(A) = \sum_{B\subset A} g(B)$. Par conséquent

$$g(F) = \sum_{B\subset F} (-1)^{|F\setminus B|} |B|^n = \sum_{k=0}^{p} C_p^k (-1)^{p-k} k^n.$$

Problème des rencontres. Si E a n éléments, combien y a-t-il de permutations σ de E sans point fixe, c'est-à-dire telles que pour tout $j \in E$ on ait $\sigma(j) \neq j$?.

On applique le théorème précédent aux fonctions f et g définies sur $\mathcal{P}(E)$ ainsi: si $A \subset E$, $f(A) = |A|!$ est le nombre de permutations de E telles que pour tout $j \in A^c$ on ait $\sigma(j) = j$, et $g(A)$ est le nombre de permutations de E telles que pour tout $j \in A^c$ on ait $\sigma(j) = j$ et pour tout $j \in A$ on ait $\sigma(j) \neq j$. On veut donc calculer $g(E)$.

Les hypothèses du théorème sont remplies, on a bien en effet $f(A) = \sum_{B \subset A} g(B)$. Par conséquent

$$g(E) = \sum_{B \subset E} (-1)^{|F \setminus B|} |B|! = \sum_{k=0}^{n} C_n^k (-1)^{n-k} k! = n! \sum_{k=0}^{n} (-1)^{n-k} \frac{1}{(n-k)!} = n! \sum_{k=0}^{n} (-1)^k \frac{1}{k!}.$$

Si Ω est l'ensemble des permutations de E et si il est muni de la probabilité équiprobable, la probabilité pour qu'une permutation aléatoire soit sans point fixe est donc

$$\sum_{k=0}^{n} (-1)^k \frac{1}{k!},$$

soit approximativement $e^{-1} = 0,367...$ si $n > 6$.

Exercices sur 2.2.

1. Soit des entiers tels que $2 \leq a \leq b \leq c$. On tire de façon équiprobable une partie de taille a de l'ensemble des $b + c$ entiers > 0. Calculer la probabilité pour que 0 d'entre eux soient $> a$; pour que 2 d'entre eux exactement soient $> a$.

2. Deux dés non pipés sont marqués sur leurs six faces 1,2,2,3,3,4 et 1,3,4,5,6,8 respectivement. On jette une fois ces deux dés et on note par A_k l'évènement "la somme des points i du premier dé et des points j du second est k". Calculer pour $k = 2, 3, \ldots, 12$ le nombre $P(A_k)$.

3. 12 méchantes fées se penchent sur le berceau des quintuplés et attribuent chacune au hasard à un enfant un défaut. Quel est la probabilité qu'il y ait au moins un enfant parfait?

2.3 Le schéma Succès-Echec.

Le schéma Succès-Echec fini. Si une expérience a deux issues, arbitrairement notées succès (S) et échec (E) et si on la répète n fois, ce qu'on observe est une suite de longueur n de S et de E. Pour modéliser cela, on introduit l'espace des observables $\Omega = \{E, S\}^n$ formé des 2^n suites $\omega = (\omega_1, \ldots, \omega_n)$ où les ω_j sont égaux à E ou S. On munit Ω de la tribu $\mathcal{P}(\Omega)$. Quant à la probabilité, on se fixe un nombre p tel que $0 < p < 1$ qui est la probabilité d'un succès si on n'effectue qu'une fois l'expérience. Introduisons alors l'importante quantité $X(w)$ définie ainsi: si $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ alors $X(w)$ désigne le nombre de succès que comprend la suite ω. Par exemple, $X(SSES) = 3$, $X(EEEE) = 0$. Pour $\omega \in \Omega$ tel que $X(\omega) = k$ on définit alors $P(\{\omega\}) = p^k(1-p)^{n-k}$; Comme tout évènement $A \in \mathcal{P}(\Omega)$ est réunion de singletons $\{\omega\}$ deux à deux disjoints, cela suffit à définir $P(A)$ et donc la probablité P sur $(\Omega, \mathcal{P}(\Omega))$.

Parmi ces évènements, les plus importants sont les $\{X = k\}$ (ceci est une sténographie que nous utiliserons souvent pour écrire brièvement l'évènement $\{\omega \in \Omega\ ;\ X(\omega) = k\}$). Voici leur probabilité:

Proposition 2.9: Pour le schéma Succès Echec fini associé à la probabilité p d'un succès, si X est le nombre de succès en n expériences, alors

$$P(X = k) = C_n^k p^k (1-p)^{n-k}.$$

Démonstration: Notons $A = \{\omega \in \Omega\ ;\ X(\omega) = k\}$. Définissons l'application de A dans $\mathcal{P}_k(\{1, 2, \ldots, n\})$ par $\omega \mapsto \{j\ ;\ \omega_j = S\}$. Il est clair que c'est une bijection; donc d'après le Théorème 2.5 b), $|A| = C_n^k$. Enfin puisque tous les $\{\omega\}$ contenus dans A ont la même probabilité $p^k(1-p)^{n-k}$ on obtient

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = |A| p^k (1-p)^{n-k} = C_n^k p^k (1-p)^{n-k}.$$

Le schéma Succès-Echec infini. Il s'agit ensuite de modéliser le cas où on veut effectuer un nombre arbitraire d'expériences: par exemple on peut vouloir répéter les essais jusqu'à ce qu'apparaisse 4 succès consécutifs. Une telle modélisation est impossible avec le schéma fini ci dessus, et on prend alors pour espace Ω des observables l'ensemble $\{E, S\}^{\mathbf{N}_*}$ des suites infinies de S et de E, en notant par $\mathbf{N}_*$ l'ensemble des entiers > 0. Il est clair que Ω est en bijection avec les parties de $\mathbf{N}_*$, et donc d'après la proposition 2.2 Ω n'est pas dénombrable. Cela cause une sérieuse difficulté en ce qui concerne la construction de l'espace de probabilité correspondant. On construit la tribu $\mathcal{A}$ et la probabilité P par un procédé d'approximation que nous décrivons maintenant.

Fixons l'entier n et définissons $\Omega' = \{E, S\}^{\{1,\ldots,n\}}$ et $\Omega'' = \{E, S\}^{\{n+1, n+2, \ldots\}}$, de sorte que $\Omega = \Omega' \times \Omega''$, et définissons la tribu suivante de parties de Ω :

$$\mathcal{A}_n = \{A \times \Omega''\ ;\ A \in \mathcal{P}(\Omega')\}.$$

Intuitivement, les évènements de $\mathcal{A}_n$ sont les évènements ne dépendant que de ce qui s'est passé jusqu'à l'instant n. En particulier, nous avons $\mathcal{A}_n \subset \mathcal{A}_{n+1}$.

Si $\omega' = (\omega_1, \ldots, \omega_n) \in \Omega'$ comprend k succès, définissons la probabilité $P_n(\{\omega'\} \times \Omega'') = p^k(1-p)^{n-k}$. Cela permet donc de définir la probabilité P_n sur $\mathcal{A}_n$. L'espace de probabilité $(\Omega, \mathcal{A}_n, P_n)$ est presque identique à l'espace du schéma Succès Echec fini décrit ci dessus.

Maintenant, notons

$$\mathcal{A}' = \cup_{n \geq 1} \mathcal{A}_n.$$

La famille $\mathcal{A}'$ n'est pas une tribu, car ce n'est pas fermé pour la réunion dénombrable. Voici un contre exemple. Soit A_n l'ensemble des suites ω infinies comprenant au moins un succès à l'instant n ou avant. Alors A_n est dans $\mathcal{A}_n$ et donc dans $\mathcal{A}'$. Pourtant $A = \cup_{n \geq 1} A_n$ n'est pas dans $\mathcal{A}'$. En effet A est l'ensemble des suites ω infinies comprenant au moins un succès. Mais il n'existe pourtant aucun n tel que $A \in \mathcal{A}_n$, et donc $A \notin \mathcal{A}'$. Réaliser cette chose subtile fait progresser dans la compréhension de la théorie. On définit alors la tribu $\mathcal{A}$ sur Ω comme la plus petite tribu contenant $\mathcal{A}'$.

Pour définir enfin la probabilité P sur $\mathcal{A}$, on fait l'observation essentielle suivante: on a non seulement $\mathcal{A}_n \subset \mathcal{A}_{n+1}$, mais de plus la restriction de P_{n+1} au sous ensemble $\mathcal{A}_n$ de $\mathcal{A}_{n+1}$, qui était le domaine de définition de P_{n+1}, coincide avec P_n. Par conséquent, il existe une fonction universelle P' définie sur $\mathcal{A}'$ telle que pour tout $A \in \mathcal{A}'$ on ait $P'(A) = P_n(A)$ pour tous les n tels que $A \in \mathcal{A}_n$. A partir de ce point, les choses cessent d'être élémentaires, et nous sommes obligés d'admettre le théorème suivant, dont la démonstration est donnée en troisième année d'université:

Théorème 2.10: Il existe une et une seule probabilité P sur $\mathcal{A}$ telle que pour tout $A \in \mathcal{A}'$ on ait $P(A) = P'(A)$.

On peut ainsi démontrer l'idée intuitive qu'un évènement de probabilité strictement positive, même petite, finit toujours par arriver. Plus précisément, si A est l'ensemble des $\omega \in \Omega$ comprenant au moins un succès, alors $P(A) = 1$. En effet, si B_n est l'ensemble des $\omega \in \Omega$ comprenant au moins un succès avant l'instant n ou à l'instant n, alors $A = \cup_{n \geq 1} B_n$ et $B_n \subset B_{n+1}$. Par continuité monotone (Th. 1.2, (5)) on a donc $\lim P(B_n) = P(A)$. Comme $P(B^c) = (1-p)^n$ tend vers 0, on a le résultat. Plus généralement on peut montrer que toute séquence a finie donnée à l'avance (par exemple SSEESSEESSEESSEE, ou le codage en binaire d'une fable de La Fontaine) finira par arriver. Plus précisément:

Théorème 2.11: Soit $a = (a_1, \ldots, a_n) \in \{E, S\}^n$ une suite fixée de longueur n de succès et d'échecs, et soit

$$A = \{\omega \in \Omega \ ; \ \text{il existe } N \geq 0 \text{ avec } \omega_{N+1} = a_1, \ldots, \omega_{N+n} = a_n\}.$$

Alors $P(A) = 1$.

Démonstration: Soit k le nombre de S dans la suite a. Notons

$$A_N = \{\omega \in \Omega \ ; \ \omega_{N+1} = a_1, \ldots, \omega_{N+n} = a_n\}.$$

Alors $P(A_N) = p^k(1-p)^{n-k}$ par définition de P. Introduisons $B_m = \cup_{j=0}^{m-1} A_{jn}$. Alors $B_m \subset B_{m+1}$ et

$$A = \cup_{N \geq 0} A_N \supset B = \cup_{m \geq 0} B_m.$$

On a de plus

$$P(B_m^c) = P(\cap_{j=0}^{m-1} A_{jn}^c) = (1 - p^k(1-p)^{n-k})^m \to_{m \to \infty} 0.$$

Par continuité monotone, on a donc $P(B^c) = 0$. D'où $1 = P(B) \leq P(A) = 1$.

2.4 Le cas où $\Omega = \mathbb{R}$.

Ce cas est naturellement le plus important de tous. La tribu mise sur $\mathbb{R}$ est la tribu de Borel $\mathcal{B}$ définie à la section 1 comme la plus petite tribu contenant les intervalles (ouverts, fermés, semi ouverts, demi droites) Parmi ses éléments, les boréliens, les seuls qu'on aura concrètement à manipuler sont les réunions d'intervalles.

Pour décrire les probabilités sur $(\mathbb{R}, \mathcal{B})$, introduisons une définition importante:

Soit F une fonction de $\mathbb{R}$ dans $\mathbb{R}$. On dit que F est une *fonction de répartition* si elle satsifait aux trois propriétés suivantes:

- F est croissante (au sens large);
- $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$;
- F est continue à droite en tout point x, c'est-à-dire $\lim_{h\searrow 0} F(x+h) = F(x)$.

On a alors le théorème fondamental suivant:

Théorème 2.12: Soit P une probabilité sur $(\mathbb{R}, \mathcal{B})$. Soit F_P la fonction réelle définie par

$$F_P(x) = P(]-\infty, x]).$$

Alors F_P est une fonction de répartition. Inversement, si F est une fonction de répartition, alors il existe une et une seule probabilité sur $(\mathbb{R}, \mathcal{B})$ telle que $F_P = F$.

Démonstration: Si $x < y$, alors $A =]-\infty, x] \subset B =]-\infty, y]$, et donc $F_P(x) = P(A) \leq P(B) = F_P(y)$. Ensuite, si (x_n) tend vers $-\infty$ en décroissant et si $A_n =]-\infty, x_n]$, alors $A_n \supset A_{n+1}$ et $\cap_{n\geq 1} A_n = \emptyset$; par continuité monotone $P(A_n)$ tend vers 0. Donc $\lim_{n\to\infty} F_P(x_n) = 0$. Comme ceci est vrai quelle que soit la suite (x_n) tendant vers $-\infty$ en décroissant, on en déduit $\lim_{x\to-\infty} F_P(x) = 0$. De même, si (y_n) tend vers ∞ en croissant et si $B_n =]-\infty, y_n]$, alors $B_n \subset B_{n+1}$ et $\cup_{n\geq 1} B_n = \mathbb{R}$; par continuité monotone $P(B_n)$ tend vers $P(\mathbb{R}) = 1$ et on a $\lim_{y\to+\infty} F_P(y) = 1$.

Enfin, si $h_n \searrow 0$, soit $C_n =]\infty, x + h_n]$. Alors $C_n \supset C_{n+1}$ et $\cap_{n\geq 1} C_n =]\infty, x]$. Par continuité monotone on a donc $\lim_{n\to+\infty} F(x + h_n) = F_P(x)$, d'où la continuité à droite annoncée de la fonction F_P.

Nous admettrons la réciproque, qui est la partie difficile.

Commentaires: Ce résultat est assez rassurant: bien qu'on connaisse mal la tribu $\mathcal{B}$, et donc les probabilités définies dessus, il y a en fait bijection entre l'ensemble de toutes les probabilités sur $\mathbb{R}$ et l'ensemble moins abstrait de toutes les fonctions de répartition. Mais la démonstration complète est réservée à la 3 ème année.

La fonction de répartition permet de calculer les probabilités de tous les intervalles. Pour simplifier, adoptons la notation pour la limite à gauche en x de la fonction croissante F:

$$F(x-0) = \lim_{h\nearrow 0} F(x+h).$$

Proposition 2.13: Soit F la fonction de répartition d'une probabilité P sur $(\mathbb{R}, \mathcal{B})$. Alors

- $P(]-\infty, x[) = F(x-0)$, $P(]x, +\infty[) = 1 - F(x)$, $P([x, +\infty[) = 1 - F(x-0)$.
- Pour $a \leq b$, $P(]a, b]) = F(b) - F(a)$, $P([a, b[) = F(b-0) - F(a-0)$.
- $P(]a, b[) = F(b-0) - F(a)$, $P([a, b]) = F(b) - F(a-0)$ et en particulier

$$P(\{a\}) = F(a) - F(a-0).$$

Démonstration: La première égalité s'obtient en considérant $A_n =]-\infty, x + h_n]$, où h_n est < 0 et croît vers 0. Alors $A_n \subset A_{n+1}$ et $\cup_{n\geq 1} A_n =]-\infty, x[$. Par convergence monotone l'égalité s'ensuit. Les deux suivantes s'obtiennent par passage au complémentaire. La suivante découle de l'égalité

$$]-\infty, b] =]-\infty, a] \cup]a, b],$$

et du fait que au second membre les deux ensembles sont disjoints. De même

$$]-\infty, b[=]-\infty, a[\cup [a, b[$$

fournit l'égalité suivante grâce à la première égalité de la liste. Laissons les dernières en exercice.

Donnons maintenant des exemples de fonctions de répartition

1. **Fonctions de répartition à densité.** Soit f une fonction positive définie sur $\mathbb{R}$ qui ait des discontinuités au plus en un nombre fini de points $a_1 < a_2 < \cdots < a_N$ et qui soit telle que les intégrales $\int_{a_i}^{a_{i+1}} f(x)dx$ convergent et satisfassent

$$\sum_{i=0}^{N} \int_{a_i}^{a_{i+1}} f(x)dx = 1,$$

avec la convention $a_0 = -\infty$ et $a_{N+1} = +\infty$.

Par exemple $f_1(x) = \frac{1}{2}e^{-|x|}$, $f_2(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, $f_3(x) = 0$ si $x \leq 0$ et $f_3(x) = \frac{1}{\sqrt{\pi x}}e^{-x}$ si $x > 0$, qu'il est plus rapide de définir par

$$f_3(x) = \frac{1}{\sqrt{\pi x}} e^{-x} \mathbf{1}_{]0,\infty[}(x),$$

où $\mathbf{1}_E(x) = 1$ si $x \in E$ et $\mathbf{1}_E(x) = 0$ sinon: la fonction $\mathbf{1}_E$ s'appellera désormais *l'indicateur* de l'ensemble E. Dernier exemple:

$$f_4(x) = \mathbf{1}_{[0,1]}(x).$$

Dans ces exemples, $N = 0$ pour f_1 et f_2, $N = 1$ pour f_3 et $N = 2$ pour f_4.

On définit alors la fonction F par $F(x) = \int_{-\infty}^{x} f(t)dt$. Il est clair que F est une fonction de répartition. Ici, elle est de plus continue et, d'après le théorème fondamental du calcul intégral, elle satisfait $F'(x) = f(x)$ pour tout $x \notin \{a_1, \ldots, a_N\}$. La fonction f s'appelle alors la *densité* de la fonction de répartition F. Il est important de ne pas confondre les deux fonctions F et f. Pour les exemples ci dessus de densités, les fonctions de répartition correspondantes seront respectivement

$$F_1(x) = \frac{1}{2}e^{x} \text{ pour } x \leq 0,\ F_1(x) = 1 - \frac{1}{2}e^{-x},$$

$$F_2(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x,$$

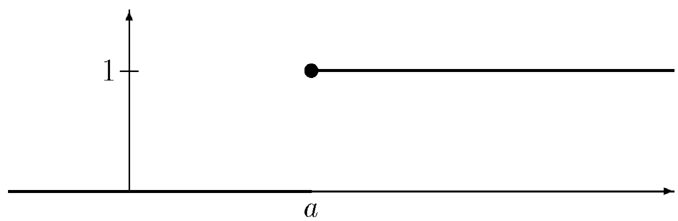
$$F_4(x) = 0 \text{ pour } x \leq 0,\ F_4(x) = x \text{ pour } 0 \leq x \leq 1,\ F_4(x) = 1 \text{ pour } 1 \leq x,$$

($F_3(x)$ ne peut s'exprimer de façon élémentaire).

2. **La probabilité δ_a de Dirac.** Si a est un réel, il s'agit de la probabilité sur $\mathbb{R}$ définie par $\delta_a(A) = 0$ si $a \notin A$, et $\delta_a(A) = 1$ si $a \in A$. Appliquant ceci à $A =]-\infty, x]$, on obtient la fonction de répartition

$$F_{\delta_a}(x) = 0 \text{ pour } x < a, \; F_{\delta_a}(x) = 1 \text{ pour } a \leq x.$$

Voici son graphe



Si $a = 0$, cette fonction s'appelle l'échelon de Heaviside. Les travaux de 1894 de cet ingénieur électricien sont à la source de la théorie moderne des distributions. Cette théorie permet par exemple de donner un sens à la dérivation de la fonction ci dessus: c'est la probabilité de Dirac δ_a qui jouerait alors le rôle de la dérivée.

3. **Probabilité discrète sur un nombre fini de points.** Soit N un entier > 0, soit $a_1 < a_2 < \cdots < a_N$ des réels et soit $p_1, \ldots, p_N$ des nombres positifs tels que $p_1 + \cdots + p_N = 1$. On considère la probabilité sur $\mathbb{R}$ définie par

$$P = p_1 \delta_{a_1} + \cdots + p_N \delta_{a_N}.$$

En d'autres termes, si A est un borélien:

$$P(A) = p_1 \delta_{a_1}(A) + \cdots + p_N \delta_{a_N}(A) = \sum_{j; a_j \in A} p_j.$$

En particulier, si $A =]-\infty, x]$, on obtient la fonction de répartition

$$F_P(x) = \sum_{j; a_j \leq x} p_j,$$

dont le graphe est celui d'une fonction en escalier croissante, où le saut en a_j est égal à p_j. Ce cas revient un peu au cas où Ω n'avait qu'un nombre fini de points, puisqu'ici P est concentrée sur $\{a_1, \ldots, a_N\}$.

4. **Probabilité discrète.** Si on remplace la suite finie précédente par un ensemble dénombrable de $\mathbb{R}$, l'extension est facile. Pour simplifier, supposons que l'ensemble dénombrable soit formé des points d'une suite (a_n) telle que $a_1 < a_2 < \cdots < a_n < \cdots$ et soit p_n des nombres positifs tels que $\sum_1^\infty p_n = 1$. On formera de même la probabilité P définie pour tout Borélien A par

$$P(A) = \sum_1^\infty p_n \delta_{a_n}(A),$$

dont la fonction de répartition est en escalier croissante vec une infinité de points de discontinuités.

5. **Type mixte.** On rencontre un peu rarement des fonctions de répartition de la forme $F = \lambda G + (1-\lambda)H$ où G est une fonction de répartition à densité, comme vu à l'exemple 1, où H est une fonction de répartition d'une probabilité discrète, comme vu aux exemples 2, 3 ou 4, et où $0 < \lambda < 1$. Si H a une discontinuité en a de saut p, alors F a une discontinuité en a de saut $(1-\lambda)p$.

Exercices sur 2.4.

1. Calculer la densité des fonctions de répartition suivantes:

   $F_1(x) = 0$ si $x \leq 0$ et $F_1(x) = 1 - \exp(-x)$ si $x > 0$;

   $F_2(x) = 0$ si $x \leq 1$ et $F_2(x) = 1 - \frac{1}{x^a}$ si $x > 1$ (avec a>0).

2. Calculer la fonction de répartition de la densité suivante:

   $f(x) = 1/2$ si $-2 < x < -1$, $f(x) = 1/2$ si $1 < x < 2$, et 0 ailleurs.

3. On note par $[x]$ la partie entière du nombre réel x, c'est-à-dire l'entier n tel que $n \leq x < n+1$. Par exemple $[\sqrt{2}] = 1$, $[-\sqrt{2}] = -2$, $[3] = 3$. On considère la probabilité discrète de fonction de répartition $F(x) = 0$ si $x < 0$ et $F(x) = 1 - \frac{1}{2^{[x]+1}}$ si $x \geq 0$. Tracer le graphe de F. Calculer les probabilités des évènements suivants:

   $A_1 = \{0\}$, $A_2 = \{1, 2\}$, $A_3 = \{4, 5, \ldots\}$.

III Probabilités conditionnelles et indépendance.

3.1 Conditionnement.

Si $(\Omega, \mathcal{A}, P)$ est un espace de probabilité, soit $B \in \mathcal{A}$ un évènement tel que $P(B) > 0$. On définit alors la nouvelle probabilité P_B sur $\mathcal{A}$ par

$$P_B(A) = \frac{P(A \cap B)}{P(B)},$$

qu'on note aussi $P_B(A) = P(A|B)$, et qui se lit "probabilité de A conditionnée par B", ou "sachant B", ou "sachant que B est réalisé". $(\Omega, \mathcal{A}, P_B)$ est un authentique espace de probabilité, puisque $P_B(\Omega) = P(\Omega \cap B)/P(B) = 1$ et que, si les $(A_n)_{n \geq 1}$ sont deux à deux disjoints et dans $\mathcal{A}$, on a bien

$$P_B(\cup_{n\geq 1} A_n) = \frac{1}{P(B)} P(\cup_{n \geq 1}(A_n \cap B)) = \frac{1}{P(B)} \sum_{n \geq 1} P(A_n \cap B)) = \sum_{n \geq 1} P_B(A_n).$$

Il faut toutefois réaliser que la probabilité P_B est concentrée sur B et ne charge pas B^c.

Pour énoncer le prochain résultat, il est commode d'introduire un nouveau terme: une suite finie $(B_n)_{n=1}^N$ ou dénombrable $(B_n)_{n=1}^{+\infty}$ d'évènements est appelée une *partition de* Ω si les B_n sont deux à deux disjoints et si leur réunion est égale à Ω.

Théorème 3.1: Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité, soit $(B_n)_{n \geq 1}$ une partition de Ω finie ou dénombrable avec $P(B_n) > 0$ pour tout n, et soit $A \in \mathcal{A}$ tel que $P(A) > 0$.

1. Si $P(B) > 0$, alors $P(A \cup B) = P(A|B)P(B) = P(B|A)P(A)$.
2. (Principe des probabilités totales) $P(A) = \sum_{n\geq 1} P(A|B_n)P(B_n)$.
3. (Formule de Bayes) Pour tout k: $P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{n\geq 1} P(A|B_n)P(B_n)}$.

Démonstration: Cet énoncé est décoré du titre de théorème plutôt par son importance pratique que par la difficulté de sa démonstration: pour le 1), utiliser la définition de $P(A|B)$. Pour le 2) observer que les $A \cap B_n$ sont une partition de A et donc d'après l'axiome d'additivité $P(A) = \sum_{n\geq 1} P(A \cap B_n)$ et terminer en utilisant le 1). Pour le 3) on a

$$P(A|B_k)P(B_k) = P(A \cup B_k) = P(B_k|A)P(A) = P(B_k|A)\sum_{n\geq 1} P(A|B_n)P(B_n),$$

successivement en utilisant deux fois le 1) puis une fois le 2). Le résultat est équivalent au 3).

Exemple: Dans une population le nombre de châtains est de 50%, et le nombre de blonds, de noirs ou d'autres couleurs est égal. La génétique nous apprend que les probabilités conditionnelles pour qu'un enfant soit châtain (évènement A) sachant que son père est blond (évènement B) est $P(A|B) = 0,2$, et que de même, avec des notations évidentes $P(A|C) = 0,7$, $P(A|N) = 0,6$ et $P(A|R) = 0,1$. Calculons $P(A)$ et $P(B|A)$. Les évènements B, C, N, R forment une partition avec $P(B) = P(N) = P(R) = 1/6$ et $P(C) = 1/2$. Les probabilités totales donnent donc $P(A) = 0,2 \times 1/6 + 0,7 \times 1/2 + 0,6 \times 1/6 + 0,1 \times 1/6 = 1/2$ et la formule de Bayes donne $P(B|A) = P(A|B)P(B)/P(A) = 1/15$.

3.2 Indépendance d'évènements.

Parfois A et B sont tels que $P_B(A) = P(A)$: savoir que B est réalisé ne modifie pas la probabilité de A. Ainsi dans le schéma succès échec fini avec $N = 2$, Ω a 4 éléments SS, SE, ES, EE de probabilités respectives $p^2, p(1-p), (1-p)p, (1-p)^2$. Si $B = (SS, SE)$ est l'évènement: "le premier essai est un succès" et $A = (SS, ES)$ est l'évènement: "le second essai est un succès" alors $A \cap B = (SS)$, $P(A) = p^2 + (1-p)p = p$, $P(B) = p^2 + p(1-p) = p$, $P(A \cap B) = p^2$ et donc $P_B(A) = P(A)$. C'est le phénomène essentiel pour les probabilités des évènements indépendants (qu'il ne faut pas confondre avec les évènements disjoints) et que nous allons définir.

Soit $\{A_1, \ldots, A_N\}$ une famille finie d'évènements d'un espace de probabilité $(\Omega, \mathcal{A}, P)$. On dit que c'est une famille indépendante (on dit parfois un "système indépendant d'évènements") si pour toute partie non vide I de $\{1, 2, \ldots, N\}$ on a

$$P(\cap_{i\in I} A_i) = \prod_{i\in I} P(A_i).$$

Par exemple si $N = 2$, la famille d'évènements $\{A, B\}$ est indépendante si et seulement si $P(A \cap B) = P(A)P(B)$; dans le cas où $P(B) > 0$ il serait équivalent de dire $P_B(A) = P(A)$. On a coutume de dire par abus de langage que A et B sont indépendants (abus, car l'adjectif qualificatif "indépendant" n'a de sens que s'il s'applique à la paire) ou plus correctement que A est indépendant de B, expression qui ne rend toutefois pas justice à la symétrie de la définition d' indépendance.

Si $N = 3$ la famille d'évènements $\{A, B, C\}$ est indépendante si et seulement si

$$P(A \cap B) = P(A)P(B),\ P(B \cap C) = P(B)P(C),\ P(C \cap A) = P(C)P(A),$$

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

Notez que la deuxième ligne n'est pas entraînée par la première. Si Ω a 4 points 1,2,3,4 de probabilité $1/4$ chacun, les 3 évènements $A = 1, 2$, $B = 1, 3$ et $C = 1, 4$ satisfont la première ligne et pas la deuxième: ils sont seulement deux à deux indépendants.

Si N est quelconque, il n'y a pour montrer l'indépendance que $2^N - 1 - N$ égalités à vérifier, puisque l'ensemble vide pour I est exclu et que les N cas où I est un singleton sont triviaux. Notez aussi que l'ensemble vide et l'ensemble Ω sont indépendants de n'importe quoi et qu'une sous famille d'une famille indépendante est encore indépendante. Enfin, on convient de dire qu'une famille *infinie* d'évènements est indépendante si toute sous famille *finie* est indépendante.

Comme exemple d'indépendance de N évènements, considérons dans le schéma Succès Echec fini avec N essais un élément particulier $a = (a_1, \ldots, a_n)$ de Ω, c'est-à-dire une suite particulière de succès et d'échecs. Notons $k = X(a)$ le nombre de succès que comprend la suite a. Soit

$$A_j = \{\omega = (\omega_1, \ldots, \omega_N) \in \Omega\ ;\ \omega_j = a_j\}.$$

Alors $\{A_1, \ldots, A_N\}$ est une famille indépendante. En effet $P(A_j) = p$ si $a_j = S$ et $1 - p$ si $a_j = E$. De plus, par définition du schéma, $P(\{a\}) = p^k(1-p)^{n-k}$ Comme $\cap_{j=1}^N A_j = \{a\}$ on a bien $P(\cap_{j=1}^N A_j) = \prod_{j=1}^N P(A_j)$. La démonstration pour n'importe quel sous ensemble I est analogue.

3.3 Indépendance de sous tribus.

La notion précédente d'évènements indépendants a l'avantage d'être élémentaire, et les inconvénients de ne pas être très maniable et de ne pas refléter la réalité: l'intuition nous fait plutôt penser que c'est un groupe d'évènements qui est indépendant d'un autre groupe, plutôt que deux évènements isolés. Par exemple, il est facile de vérifier que si A est indépendant de B, alors A^c est aussi indépendant de B. La bonne notion de "groupe" d'évènements est en fait celle de sous tribu. D'où la définition suivante:

Soit $\{\mathcal{A}_1, \ldots, \mathcal{A}_N\}$ une famille finie de sous tribus d'un espace de probabilité $(\Omega, \mathcal{A}, P)$. On dit que c'est une famille indépendante si pour tous $B_j \in \mathcal{A}_j$ on a

$$P(B_1 \cap B_2 \cap \ldots \cap B_N) = P(B_1) \ldots P(B_N)$$

(plus la peine donc d'examiner tous les sous ensembles I.) En fait, c'est une puissante généralisation de la notion d'évènements indépendants, d'après le théorème suivant:

Théorème 3.2: Soit $A_1, \ldots, A_N$ des évènements. Soit les tribus à quatre éléments engendrées par les A_j:

$$\mathcal{A}_j = \{\emptyset, A_j, A_j^c, \Omega\}.$$

Alors la famille de sous tribus $\{\mathcal{A}_1, \ldots, \mathcal{A}_N\}$ est indépendante si et seulement si la famille d'évènements $\{A_1, \ldots, A_N\}$ est indépendante.

Démonstration: Pour $\Rightarrow$, soit I une partie de $(1, 2, \ldots, N)$. Prenons alors $B_j = A_j$ si $j \in I$ et $B_j = \Omega$ sinon. Alors

$$P(\cap_{i\in I} A_i) = P(B_1 \cap B_2 \cap \ldots \cap B_N) = P(B_1) \ldots P(B_N) = \prod_{i\in I} P(A_i).$$

Bien qu'une démonstration par récurrence soit possible immédiatement pour la réciproque, nous attendons la section 5 pour avoir une démonstration plus simple.

Exercices sur la section 3.

1. Dans le schéma Succès Echec fini à N essais, on suppose $p = 1/2$ et on considère les deux évènements $A=$ que des succès ou que des échecs, et $B=$ pas plus d'un succès. Montrer que A et B sont indépendants si et seulement si $N = 3$.

2. On munit le segment $\Omega = [0, 1]$ de la probabilité P telle que $P([a, b]) = b - a$ pour tout intervalle $[a, b] \subset [0, 1]$. On considère les trois évènements $A = [0, 1/2]$, $B = [1/4, 3/4]$, $C = [3/8, 7/8]$. Quelles sont les paires d'évènements parmi A, B, C qui sont indépendantes?

IV Image d'une probabilité, variables aléatoires

4.1 Fonctions mesurables

Quand en mathématiques une nouvelle structure est introduite, comme celle d'espace vectoriel, ou comme présentement celle d'espace de probabilité, une démarche féconde est de rechercher les transformations qui préservent cette structure. Pour les espaces vectoriels, ce sont les applications linéaires. Pour les espaces de probabilité, ce sont les "fonctions mesurables" qu'on va introduire dans un instant. Le cas particulier important en sera les "variables aléatoires". Auparavant, adoptons la notation suivante: si E et F sont des ensembles quelconques, si f est une fonction définie sur E et à valeurs dans F, et si enfin B est un sous ensemble de F, l'ensemble A des x de E tels que $f(x)$ soit dans B sera désormais noté par $A = f^{-1}(B)$. Nous l'appellerons l'image inverse de B par f. Insistons sur le fait que f n'est pas nécessairement injective ni surjective. On vérifie facilement que si B_1 et B_2 sont des sous ensembles de F alors on a

$$f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2) \text{ et } f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2),$$

et que la même chose est vraie même avec une famille infinie de B.

Soit alors deux espaces Ω et Ω_1, chacun muni d'une tribu $\mathcal{A}$ et $\mathcal{A}_1$, et soit f une fonction définie sur Ω à valeurs dans Ω_1 On dit que f est une *fonction mesurable* si pour tout $B \in \mathcal{A}_1$, alors $A = f^{-1}(B)$ est un élément de $\mathcal{A}$. Dans ces conditions, on voit facilement que l'ensemble des parties A de Ω qui sont de la forme $f^{-1}(B)$, avec $B \in \mathcal{A}_1$, est une tribu. On la note parfois $f^{-1}(\mathcal{A}_1)$. Comme f est mesurable, c'est donc une sous tribu de $\mathcal{A}$.

Montrer qu'une fonction est mesurable est généralement facile grâce au théorème suivant, dont la démonstration est hors programme.

Théorème 4.1: Soit $\mathcal{F}$ une famille de parties de Ω_1 telle que la tribu $\mathcal{A}_1$ soit la plus petite qui contienne $\mathcal{F}$. Soit f une fonction de Ω à valeurs dans Ω_1. Soit $\mathcal{A}$ une tribu sur Ω. Alors f est mesurable pour ce couple de tribus si et seulement si pour tout $B \in \mathcal{F}$ alors $f^{-1}(B) \in \mathcal{A}$.

Illustrons ceci par un exemple important en l'appliquant au cas où

$$(\Omega, \mathcal{A}) = (\Omega_1, \mathcal{A}_1) = (\mathbb{R}, \mathcal{B}),$$

pour montrer que *toute fonction continue f de $\mathbb{R}$ dans $\mathbb{R}$ est mesurable.* Pour cela, on applique le théorème au cas où $\mathcal{F}$ est l'ensemble de tous les intervalles ouverts: par définition de la tribu $\mathcal{B}$ de Borel, l'hypothèse du théorème est vérifiée. Ensuite, on sait d'après les cours d'analyse l'image inverse d'un intervalle ouvert par une fonction continue est une réunion finie ou dénombrable d'intervalles ouverts, et est donc un borélien.

Démonstration: La partie "seulement si" découle des définitions. Pour la partie "si", l'art est de considérer la tribu $\mathcal{T}$ de parties de Ω engendrée par tous les $f^{-1}(B)$ lorsque B parcourt $\mathcal{F}$ ainsi que

$$\mathcal{T}_1 = \{B \subset \Omega_1;\ f^{-1}(B) \in \mathcal{T}\}.$$

A son tour, $\mathcal{T}_1$ est une tribu de parties de Ω_1 (ce point se vérifie directement facilement), et elle contient $\mathcal{F}$, et donc elle contient la tribu $\mathcal{A}_1$. D'où

$$f^{-1}(\mathcal{T}_1) \supset f^{-1}(\mathcal{A}_1) = \mathcal{A}.$$

Mais comme par définition de $\mathcal{T}_1$ on a

$$f^{-1}(\mathcal{T}_1) \subset \mathcal{T},$$

on en tire que $\mathcal{T} = \mathcal{A}$, ce qui est l'égalité cherchée.

4.2 Image d'une probabilité.

Si alors $(\Omega, \mathcal{A})$ est muni d'une probabilité, alors la fonction mesurable f permet de définir de façon naturelle une pro babilité P_1 sur $(\Omega_1, \mathcal{A}_1)$ ainsi: pour tout $B \in \mathcal{A}_1$

$$P_1(B) = P(f^{-1}(B)).$$

Cette fonction P_1 sur $\mathcal{A}_1$ est bien une probabilité. En effet,

$$P_1(\Omega_1) = P(f^{-1}(\Omega_1)) = P(\Omega) = 1;$$

De plus si B_1 et B_2 sont des parties disjointes de Ω_1, alors $f^{-1}(B_1)$ et $f^{-1}(B_2)$ sont alors des parties disjointes de Ω. Cela permet de vérifier facilement l'axiome d'additivité dénombrable pour P_1.

La probabilité P_1 ainsi fabriquée est appelée l'*image* de la probabilité P par la fonction mesurable f. On parle aussi de la probabilité P_1 *transportée* de P par f. On la note traditionnellement $P_1 = f_*P$. D'autres la notent plus correctement Pf^{-1}, mais c'est moins commode.

4.3 Les variables aléatoires réelles et leurs lois.

Nous appliquons les concepts précédents, qui étaient bien abstraits, au cas où l'espace d'arrivée $(\Omega_1, \mathcal{A}_1)$ est $(\mathbb{R}, \mathcal{B})$. Dans ce cadre, une fonction mesurable de Ω à valeurs dans $\mathbb{R}$ prend le nom de *variable aléatoire réelle*, ou de *variable aléatoire* si le contexte est clair (on pourra ensuite considérer des variables aléatoires à valeurs dans $\mathbb{R}^2$ ou dans $\mathbb{R}^n$ quand on aura précisé de quelle tribu équiper $\mathbb{R}^n$). Plutôt que de noter la variable aléatoire f, la tradition est de la noter par une lettre majuscule comme X. En dépit du nom de "variable aléatoire," qu'on garde pour des raisons historiques, X est donc une fonction réelle définie sur Ω. L'avantage de travailler dans $\mathbb{R}$ est que grâce au Théorème 2.12, on connait comment sont faites les probabilités sur $\mathbb{R}$ et donc les probabilités transportées par les variables aléatoires. On abandonne d'ailleurs également pour $P_1 = X_* P$ ce nom de probabilité transportée de P par la variable aléatoire X, on la note plutôt P_X et on l'appelle la *loi de la variable aléatoire* X: c'est une probabilité sur $\mathbb{R}$. Quant à la fonction de répartition F_{P_X}, il est plus simple de la noter F_X. Donc on a

$$F_X(x) = P(\{\omega \in \Omega \ ; \ X(\omega) \leq x\});$$

ici encore, il est plus simple d'écrire $F_X(x) = P(X \leq x)$. Enfin, v.a. est une abréviation courante pour "variable aléatoire".

A propos du schéma Succès Echec fini d'ordre N, nous avons déjà rencontré la variable aléatoire X qui était le nombre de succès en N expériences pour laquelle nous avons vu que $P(X = k) = C_N^k p^k (1-p)^{N-k}$ C'est donc dire que la loi de X est la loi discrète concentrée sur les entiers $0, 1, \ldots, N$ et égale à

$$(1-p)^N \delta_0 + N(1-p)^{N-1} p \delta_1 + \cdots + C_N^k p^k (1-p)^{N-k} \delta_k + \cdots + p^N \delta_N$$

(Rappelons que δ_k est la probabilité de Dirac concentrée en k).

Plus généralement, soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Une variable aléatoire X sur Ω ne prenant qu'un nombre fini de valeurs $a_1 < a_2 < \ldots < a_N$ sera dite *étagée.* Les parties $X^{-1}(\{a_j\}) = A_j$ de Ω sont des éléments de $\mathcal{A}$, puisque les $\{a_j\}$ sont des intervalles, d'un type un peu particulier, et donc des boréliens. Les A_j sont deux à deux disjoints, et si on introduit leurs indicateurs, on peut écrire

$$X = a_1 \mathbf{1}_{A_1} + \cdots + a_N \mathbf{1}_{A_N}.$$

Si $p_j = P(A_j)$ on voit que la loi de X est

$$P_X = p_1 \delta_{a_1} + \cdots + p_N \delta_{a_N}.$$

Une autre manière de dire la même chose est d'écrire $P(X = a_j) = p_j$ pour tout j.

Il y a un certain nombre de lois de probabilités qu'on rencontre souvent dans la nature que nous pourrions présenter maintenant, mais il est préférable de définir quelques caractéristiques des variables aléatoires avant pour pouvoir présenter une carte d'identité plus complète de chacune de ces lois classiques.

Exercices sur la section 4.

1. Soit X une variable aléatoire de densité $ax^{a-1}\mathbf{1}_{]0,1[}(x)$. Calculer l'image de sa loi par $x \mapsto x/(1-x)$. Méthode: calculer la fonction de répartition de $Y = X/(1-X)$ et dériver celle ci.

2. Soit X une variable aléatoire suivant la loi de Cauchy, c'est-à-dire de densité $\frac{1}{\pi(1+x^2)}$. Calculer l'image de sa loi par $x \mapsto 1/x$.

V L'espérance mathématique d'une variable aléatoire.

5.1 Les variables aléatoires étagées.

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Désignons par $\mathcal{E}$ l'ensemble de toutes les variables aléatoires réelles étagées définies sur Ω. A tout élément X de$\mathcal{E}$ nous associons un nombre appelé *espérance mathématique de* X, noté par $\mathbb{E}(X)$, et défini ainsi: si la loi de X est

$$P_X = p_1\delta_{a_1} + \cdots + p_N\delta_{a_N},$$

alors

$$\mathbb{E}(X) = p_1a_1 + \cdots + p_Na_N.$$

En fait, $\mathcal{E}$ est un espace vectoriel et $X \mapsto \mathbb{E}(X)$ est une forme linéaire positive dessus, comme le montre le théorème suivant:

Théorème 5.1: (Linéarité et positivité de l'espérance) Si X et Y sont des v.a. étagées sur Ω alors $\lambda X + \mu Y$, pour des réels λ et μ, est encore une v.a. étagée. De plus $\mathbb{E}(\lambda X + \mu Y) = \lambda\mathbb{E}(X) + \mu\mathbb{E}(Y)$. Enfin $\mathbb{E}(X) \geq \mathbb{E}(Y)$ si $X \geq Y$.

Démonstration: Introduisons les lois de X et Y:

$$P_X = p_1\delta_{a_1} + \cdots + p_N\delta_{a_N},\ P_Y = q_1\delta_{b_1} + \cdots + q_M\delta_{b_M},$$

notons $X^{-1}(\{a_i\}) = A_i$, $Y^{-1}(\{b_j\}) = B_j$ et $C_{ij} = A_i \cap B_j$ et $r_{ij} = P(C_{ij})$. La matrice (r_{ij}) a pour somme des lignes le vecteur ligne $(q_1, \ldots, q_M)$ et pour somme des colonnes le vecteur colonne ${}^t(p_1, \ldots, p_N)$. Les valeurs prises par $Z = \lambda X + \mu Y$ sont les $c_{ij} = \lambda a_i + \mu b_j$ et comme $Z^{-1}(\{c_{ij}\}) = C_{ij} \in \mathcal{A}$, on en déduit que Z est aussi une v.a. Sa loi est

$$P_Z = \sum_{ij} r_{ij}\delta_{c_{ij}},$$

et est donc d'espérance

$$\mathbb{E}(Z) = \sum_{ij} r_{ij}c_{ij} = \sum_{ij} r_{ij}(\lambda a_i + \mu b_j) = \lambda\sum_i a_i\sum_j r_{ij} + \mu\sum_j b_j\sum_i r_{ij} = \lambda\mathbb{E}(X) + \mu\mathbb{E}(Y).$$

Quant à l'inégalité, il suffit d'observer que $\mathbb{E}(X - Y) \geq 0$ par définition de l'espérance et d'appliquer ensuite la linéarité qu'on vient de démontrer.

Variable aléatoire de Bernoulli. Un exemple particulièrement simple et important de v.a étagée est celui où X ne prend que les valeurs 0 et 1, c'est à dire où la loi de X est

$$P_X = (1-p)\delta_0 + p\delta_1,$$

où $p \in [0,1]$. Sa loi est appelée une loi de Bernoulli. Son espérance est p. Si X est définie sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$, soit $A = \{\omega\ ;\ X(\omega) = 1\}$ alors $X = \mathbf{1}_A$ est l'indicateur de A, et on a donc

$$\mathbb{E}(\mathbf{1}_A) = P(A).$$

Inversement, un indicateur a toujours une loi de Bernoulli.

Nous allons utiliser le théorème précédent et les indicateurs pour terminer la démonstration du théorème 4.2. On veut donc montrer que si $B_j \in \mathcal{A}_j = \{\emptyset, A_j, A_j^c, \Omega\}$ et si les A_j sont indépendants, alors

$$P(\cap_{j=1}^N B_j) = \prod_{j=1}^N P(B_j).$$

On le montre en remarquant d'abord que dans les 4 cas possibles pour B_j, il existe deux nombres a_j et b_j tels que

$$\mathbf{1}_{B_j} = a_j + b_j \mathbf{1}_{A_j};$$

on prend en effet $a_j = b_j = 0$ si B_j est vide, $a_j = 1$, $b_j = 0$ si B_j est plein, $a_j = 0$, $b_j = 1$ si $B_j = A_j$, $a_j = 1$, $b_j = -1$ si $B_j = A_j^c$. D'où le calcul:

$$P(\cap_{j=1}^N B_j) = \mathbb{E}(\prod_{j=1}^N \mathbf{1}_{B_j}) = \mathbb{E}(\prod_{j=1}^N (a_j + b_j \mathbf{1}_{A_j})) = \mathbb{E}[\sum_I (\prod_{j\in I^c} a_j)(\prod_{j\in I} b_j \mathbf{1}_{A_j})] =$$

$$\sum_I (\prod_{j\in I^c} a_j)(\prod_{j\in I} b_j)\mathbb{E}(\prod_{j\in I} \mathbf{1}_{A_j}) = \sum_I (\prod_{j\in I^c} a_j)(\prod_{j\in I} b_j) P(\cap_{j\in I} A_j) =$$

$$\sum_I (\prod_{j\in I^c} a_j)(\prod_{j\in I} b_j)(\prod_{j\in I} P(A_j)) = \prod_{j=1}^N (a_j + b_j P(A_j)) = \prod_{j=1}^N \mathbb{E}(\mathbf{1}_{B_j}) = \prod_{j=1}^N P(B_j).$$

Dans cette chaîne de 9 égalités, la première, la cinquième et les 2 dernières s'appuient sur le fait que l'espérance de l'indicateur est la probabilité, la deuxième sur la définition des a_j et b_j, la troisième et la septième sur un développement algébrique; enfin, surtout, la quatrième s'appuie sur le théorème précédent et la sixième sur l'indépendance des A_j.

5.2 Espérance d'une variable aléatoire quelconque.

Toutes les variables aléatoires ne sont pas étagées, mais toutes sont approchables par des v.a. étagées, et cela va permettre de définir l'espérance d'une v.a. quelconque. Plus précisément, on a le théorème suivant:

Théorème 5.2: Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité, et $X : \Omega \to \mathbb{R}$ une variable aléatoire *positive*. Alors

1. Il existe une suite croissante de v.a. étagées (X_n) telle $X = \lim_{n\to+\infty} X_n$.

2. Si la suite (X_n) ci dessus est telle que $\mathbb{E}(X_n)$ soit bornée, alors le nombre $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$ ne dépend que de X et non de la suite particulière (X_n), dans le sens que si (X'_n) a les propriétés demandées à (X_n) au 1), alors la suite $\mathbb{E}(X'_n)$ a la même limite.

3. Si Y est une autre v.a positive sur $(\Omega, \mathcal{A}, P)$ telle que $E(Y)$ existe, et si λ et μ sont des nombres ≥ 0, alors $\mathbb{E}(\lambda X + \mu Y)$ existe et est égale à $\lambda\mathbb{E}(X) + \mu\mathbb{E}(Y)$.

4. Si $0 \leq X \leq Y$ et si $\mathbb{E}(Y)$ existe, alors $\mathbb{E}(X)$ existe et $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

5. Si $X \geq 0$, alors $\mathbb{E}(X) = 0$ si et seulement si la loi de X est la probabilité de Dirac en 0.

Nous omettons la démonstration, bien que celle ci ne soit pas difficile. Il faut insister sur le fait que l'espérance de cette v.a. positive n'existe pas toujours.

Ce théorème définit donc $\mathbb{E}(X)$ pour des v.a positives. Pour passer au signe quelconque, on considère une v.a. X définie sur $(\Omega, \mathcal{A}, P)$ et on écrit cette fonction de ω comme différence de deux fonctions positives $X = X_+ - X_-$, où a_+ signifie $\max(a, 0)$ et $a_- = (-a)_+$ (rappelons que cela implique $a = a_+ - a_-$ et $|a| = a_+ + a_-$). Donc $|X| = X_+ - X_-$. On dira que $\mathbb{E}(X)$ existe si, au sens du théorème 5.2, l'espérance de $|X|$ existe. Dans ces conditions, d'après le 4) du théorème, $\mathbb{E}(X_+)$ et $\mathbb{E}(X_-)$ existent, et on définit enfin l'espérance de X par $\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_+)$. On a alors l'importante extension du théorème de linéarité et de positivité:

Corollaire 5.3: Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité, soit $\mathcal{L}_1$ l'ensemble des variables aléatoires X sur cet espace telles que $\mathbb{E}(X)$ existe (ou, de façon équivalente, telles que $\mathbb{E}(|X|)$ soit finie). Alors $\mathcal{L}_1$ est un espace vectoriel et $X \mapsto \mathbb{E}(X)$ est une forme linéaire sur $\mathcal{L}_1$, telle que de plus $\mathbb{E}(X) \geq \mathbb{E}(Y)$ si $X \geq Y$.

Appliquons cela à deux cas particuliers importants, celui où X est discrète et positive et celui où la loi de X a une densité.

Proposition 5.4: Soit X une v.a discrète avec

$$P_X = \sum_{j=1}^{\infty} p_j \delta_{a_j}$$

où $\sum_{j=1}^{\infty} p_j = 1$. Alors $E(X)$ existe si et seulement si la série $\sum_{j=1}^{\infty} p_j a_j$ est absolument convergente. S'il en est ainsi, alors

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} p_j a_j.$$

Démonstration: Montrons le d'abord si les a_n sont positifs ou nuls. Alors puisque $X = \sum_{j=1}^{\infty} a_j \mathbf{1}_{A_j}$, où les évènements $A_j = \{X = j\}$ sont deux à deux disjoints dans Ω, il suffit de considérer la v.a. étagée $X_n = \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}$, qui est nulle sur $\cup_{j=n+1}^{\infty} A_j$, et qui définit une suite ayant les propriétés requises au théorème 5.2. Le résultat est alors clair.

Si les a_n ne sont pas positifs on écrit $a_n = (a_n)_+ - (a_n)_-$ et les deux séries $\sum_{j=1}^{\infty} p_j(a_j)_+$ et $\sum_{j=1}^{\infty} p_j(a_j)_-$ convergent si et seulement si $\sum_{j=1}^{\infty} p_j a_j$ est absolument convergente. Cela permet de conclure facilement.

Proposition 5.5: Supposons que la loi de la v.a. X ait une densité f avec un nombre fini de points de discontinuités $a_1 < \ldots < a_N$. Alors $E(X)$ existe si et seulement si $\int_{-\infty}^{\infty} xf(x)dx$ est absolument convergente. S'il en est ainsi, alors

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Démonstration: Contentons nous de donner les idées de la démonstration quand X est positive et quand sa densité f est continue. L'extension aux hypothèses du théorème sera alors standard. On découpe $[0, n]$ en $n2^n$ intervalles égaux par les points $x_k = \frac{k}{2^n}$, avec $k = 0, 1, \ldots, n2^n$, on convient $x_{n2^n+1} = +\infty$ et on définit la variable aléatoire étagée $X_n = x_k$ quand $x_k \leq X < x_{k+1}$. Ceci est bien une suite croissante et on a bien $\lim_{n\to+\infty} X_n = X$.

Si $\int_0^{\infty} xf(x)dx$ converge, notons

$$D_n = \int_0^{\infty} xf(x)dx - \mathbb{E}(X_n) = \sum_{k=0}^{n2^n} \int_{x_k}^{x_{k+1}} (x - x_k)f(x)dx.$$

Soit $\epsilon > 0$. Il existe un entier A tel que $\int_A^{\infty} xf(x)dx \leq \epsilon$. Soit alors K tel que $x_K = A$ et soit F la fonction de répartition de F. On partage alors D_n en deux sommes A_n et B_n, avec

$$A_n = \sum_{k=K}^{n2^n} \int_{x_k}^{x_{k+1}} (x - x_k)f(x)dx \leq 2\int_A^{+\infty} xf(x)dx \leq 2\epsilon,$$

$$B_n = \sum_{k=0}^{K-1} \int_{x_k}^{x_{k+1}} (x - x_k)f(x)dx = -\int_0^A F(x)dx + \sum_{k=0}^{K-1} (x_{k+1} - x_k)F(x_{k+1},$$

la dernière égalité étant obtenue par intégration par parties en posant $u = (x - x_k)$ et $v' = f$. Notons que les symboles x_K et K sont des fonctions de n. Si n tend vers l'infini, (B_n) tend vers zéro, comme suite des différences entre une intégrale et les sommes de Riemann de cette intégrale. On voit donc que (D_n) tend vers 0. Le cas où $\int_0^{\infty} xf(x)dx$ diverge est similaire.

Exercices sur 5.2

1. Calculer l'espérance d'une variable aléatoire de loi

$$\sum_{n=1}^{\infty} \frac{4}{n(n+1)(n+2)} \delta_n.$$

2. Pour quelles valeurs de $a > 0$ la variable aléatoire X ayant pour fonction de répartition $F_X(x) = 1 - \frac{1}{(1+x)^a}$ si x>0, et $F_X(x) = 0$ si $x \leq 0$, possède-t-elle une espérance?

5.3 Théorème du transport.

Il arrive souvent qu'on ait besoin de calculer, non l'espérance de la variable aléatoire X, mais l'espérance d'une fonction $Y = g(X)$ de celle ci. Si on applique la définition de l'espérance, cela suppose qu'on calcule la loi de Y, ce qui peut être très incommode. Le résultat suivant simplifie ce problème.

Théorème 5.6: Soit X une v.a. sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$. Soit $x \mapsto y = g(x)$ une fonction mesurable de $\mathbb{R}$ dans $\mathbb{R}$. Si X est étagée ou discrète et de loi

$$P_X = \sum_{j \geq 1} p_j \delta_{a_j},$$

alors $\mathbb{E}(g(X))$ existe si et seulement si $\sum_{j\geq 1} p_j g(a_j)$ converge absolument et dans ce cas $\mathbb{E}(g(X))$ est égale à cette somme.

Si X a une densité f, alors de même $\mathbb{E}(g(X))$ existe si et seulement si $\int_{-\infty}^{\infty} g(x)f(x)dx$ est absolument convergente, et dans ce cas $\mathbb{E}(g(X))$ est égale à la somme de l'intégrale.

Démonstration: On montre d'abord le résultat quand X est étagée, puis quand X est positive en appliquant la définition de l'espérance d'une variable aléatoire positive, et on passe facilement au cas où X est de signe quelconque.

Exercices sur 5.3

1. Soit une variable aléatoire X de densité $\frac{1}{2}\exp(-|x|)$. Soit z un nombre réel et soit $g(x) = \exp(zx)$. Pour quelles valeurs de z $Y = g(X)$ a-t-elle une espérance? La calculer quand elle existe.

2. X une variable aléatoire de densité $\frac{1}{2}\mathbf{1}_{[-1,1]}(x)$ et soit $Y = \tan(\frac{\pi}{2}X)$. Etudier de deux manières l'existence éventuelle de $\mathbb{E}(Y)$: soit à l'aide du théorème du transport, soit en calculant la densité de Y: pour cela, écrire d'abord la fonction de répartition de Y puis dériver.

5.4 Variables aléatoires indépendantes et espérance du produit.

Soit maintenant $(X_1, \ldots, X_N)$ une suite de v.a. sur $(\Omega, \mathcal{A}, P)$. On se rappelle que si $\mathcal{B}$ est la tribu de Borel, alors par définition des variables aléatoires $X_j^{-1}(\mathcal{B}) = \mathcal{A}_j$ est une sous tribu de $\mathcal{A}$.

Nous dirons que c'est une suite de variables aléatoires indépendantes si la famille de sous tribus $\{\mathcal{A}_1, \ldots, \mathcal{A}_N\}$ est une famille indépendantes. Ceci entraîne un fait simple et utile: si les X_j sont des v.a. indépendantes, et si f_j est une fonction réelle quelconque, alors les $Y_j = f_j(X_j)$ sont des v.a. indépendantes aussi.

Dans le théorème suivant, qui sert à caractériser l'indépendance pratiquement, contentons nous de $N = 2$: la généralisation $N > 2$ est évidente.

Théorème 5.7: Soit X et Y deux variables aléatoires sur $(\Omega, \mathcal{A}, P)$. Alors elles sont indépendantes si et seulement si pour tous x et y réels on a

$$P(X \leq x; Y \leq y) = F_X(x)F_Y(y) = P(X \leq x)P(Y \leq y).$$

En particulier, si elles sont discrètes de lois respectives

$$P_X = \sum_{i \geq 1} p_i \delta_{a_i}, \ P_Y = \sum_{j \geq 1} q_j \delta_{b_j},$$

alors elles sont indépendantes si et seulement si pour tout couple (i, j) on a

$$P(X = a_i; Y = b_j) = p_i q_j = P(X = a_i) P(Y = b_j).$$

Démonstration: Partie $\Rightarrow$. Introduisons les évènements $A = \{X \leq x\} \in X^{-1}(\mathcal{B})$ et $B = \{Y \leq y\} \in X^{-1}(\mathcal{B})$. Par hypothèse ils sont indépendants.

Partie $\Leftarrow$. Elle n'est pas élémentaire et sera montrée en 3 ème année.

Toutefois, dans le cas discret de la seconde partie la démonstration directe est facile.

Voici enfin un théorème d'une importance considérable.

Théorème 5.8: Soit $(X_1, \ldots, X_N)$ une suite de v.a. indépendantes sur $(\Omega, \mathcal{A}, P)$. Alors le produit $X_1 \cdots X_N$ a une espérance si et seulement si chaque X_j a une espérance. Dans ces conditions l'espérance du produit est le produit des espérances:

$$\mathbb{E}(X_1 \cdots X_N) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_N).$$

Démonstration: On le démontre d'abord pour $N = 2$, et une récurrence permet de passer au cas de N quelconque. Pour $N = 2$, notons $X = X_1$ et $Y = X_2$ pour simplifier. On le démontre d'abord dans le cas où X et Y sont étagées. Ceci fait, on suppose ensuite que X et Y sont positives. Il est facile de construire deux suites croissantes (X_n) et (Y_n) de v.a. étagées qui sont de plus *indépendantes*. Comme $(X_n Y_n)$ est à son tour une suite de v.a. qui croit vers XY, on arrive au résultat. Quant au passage au cas où les X et Y ne sont plus positives, il est standard.

Exercices sur 5.4

1. Soit X et Y deux variables aléatoires indépendantes à valeurs dans les entiers ≥ 0 de lois respectives données par $P(X = n) = (1 - p)^n p$ et $P(Y = n) = (1 - q)^n q$, où p et q sont dans $]0, 1[$. Montrer à l'aide de la deuxième partie du Th. 5.7 que $U = X - Y$ et $V = \min(X, Y)$ sont indépendantes.

2. Soit une matrice carrée d'ordre 2 dont les coefficients sont des variables aléatoires indépendantes et de même loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Calculer l'espérance du carré du déterminant de cette matrice.

VI Moments, fonctions génératrices, transformées de Laplace.

6.1 Moments et variance.

Théorème 6.1: Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité, et soit n un entier > 0. Soit $\mathcal{L}_n$ l'ensemble des v.a. X sur cet espace telles que l'espérance $m_n = \mathbb{E}(X^n)$, appelée *moment* d'ordre n, existe. Alors $\mathcal{L}_n$ est un espace vectoriel, et on a

$$\mathcal{L}_1 \supset \mathcal{L}_2 \supset \cdots \supset \mathcal{L}_n.$$

Démonstration: Puisque $f(x) = x^n$ définit une fonction convexe sur la demi-droite positive, on peut écrire pour x et y positif que

$$(\frac{x+y}{2})^n \leq \frac{1}{2}(x^n + y^n),$$

et donc $|X+Y|^n \leq (|X|+|Y|)^n \leq 2^{n-1}(|X|^n + |Y|^n)$. Une autre méthode pour obtenir cette inégalité est de montrer que $g(t) = 2^{n-1}(t^n+1) - (t+1)^n$ atteint son minimum sur $[0, +\infty[$ en $t = 1$ et de considérer $g(x/y)$.

Si maintenant les espérances de $|X|^n$ et de $|Y|^n$ sont finies, on en déduit d'après la fin du théorème 5.2 que l'espérance de $|X+Y|^n$ est finie et que $X+Y$ est dans $\mathcal{L}_n$ quand X et Y y sont. Enfin, pour voir que si l'espérance de $|X|^n$ est finie il en est de même pour $|X|^{n-1}$, on utilise l'inégalité

$$|X|^{n-1} \leq 1 + |X|^n,$$

qu'on vérifie immédiatement en étudiant les cas $|X| \leq 1$ et $|X| \geq 1$. Le fait que $\mathcal{L}_{n-1} \supset \mathcal{L}_n$ s'en déduit.

On utilise parfois aussi le *moment centré* d'ordre n de la variable aléatoire X, défini par $\mathbb{E}[(X-m_1)^n]$ où $m_1 = \mathbb{E}(X)$. Remarquons au passage que si le moment non centré m_n existe, alors le moment centré existe, puisque c'est l'espérance d'un polynôme en X de degré n et qu'on vient de voir que les moments de degré inférieur à n existaient.

Le cas particulier réellement important est le cas où $n = 2$. On appelle le moment centré d'ordre 2 de X la *variance* de X, et sa racine carrée positive l'*écart type* de X, encore appelé déviation standard. On note l'écart type $\sigma(X)$ et la variance $\sigma^2(X)$, ou plus rarement $V(X)$. Insistons sur le fait que l'écart type à la dimension de la variable aléatoire: si celle ci s'exprime en centimètres, l'écart type s'exprime en centimètres et la variance en centimètres carrés. Il faut connaître les deux formules suivantes:

Proposition 6.2: Si X a un moment d'ordre 2, alors pour λ réel

$$\sigma^2(\lambda X) = \lambda^2 \sigma^2(X),$$

et (Formule de Huyghens):

$$\sigma^2(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

En particulier, $(\mathbb{E}(X))^2 \leq \mathbb{E}(X^2)$, avec égalité si et seulement si la loi de X est une probabilité de Dirac.

Démonstration: La première formule est immédiate. Pour Huyghens:

$$\sigma^2(X) = \mathbb{E}(X^2 - 2m_1 X + {m_1}^2) = \mathbb{E}(X^2) - 2m_1\mathbb{E}(X) + {m_1}^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Ici on a utilisé le fait que l'espérance d'une constante est la constante elle même et que $m_1 = \mathbb{E}(X)$. Quant à la dernière inégalité elle vient du fait qu'une variance est toujours positive ou nulle. Si la variance est nulle, alors appliquant le 5) du théorème 5.2 à la v.a. positive $Y = (X-m_1)^2$, alors la loi de Y est δ_0 et celle de X est donc δ_{m_1}.

Il y a également à connaître deux inégalités célèbres:

Proposition 6.3: (Inégalité de Markov) Si Y est une variable aléatoire positive ou nulle dont l'espérance existe, alors pour tout $y > 0$ on a

$$P(Y \geq y) \leq \frac{1}{y}\mathbb{E}(Y).$$

(Inégalité de Tchebychev) Si X est une variable aléatoire ayant un second moment, alors pour tout $t > 0$ on a

$$P(|X - \mathbb{E}(X)| \geq t) \leq \frac{1}{t^2}\sigma^2(X).$$

Démonstration:

$$\mathbb{E}(Y) = \mathbb{E}(Y\mathbf{1}_{Y \geq y} + Y\mathbf{1}_{Y<y}) \geq \mathbb{E}(Y\mathbf{1}_{Y \geq y}) \geq \mathbb{E}(y\mathbf{1}_{Y \geq y}) \geq y\mathbb{E}(\mathbf{1}_{Y \geq y}) = yP(Y \geq y),$$

ce qui est équivalent à l'inégalité de Markov en divisant les extrémités par y.

On applique ensuite Markov à $Y = (X - m_1)^2$ et à $y = t^2$. Comme

$$P(|X - m_1| \geq t) = P((X - m_1)^2 \geq t^2) \leq \frac{1}{t^2}\mathbb{E}((X - m_1)^2) = \frac{1}{t^2}\sigma^2(X),$$

l'inégalité de Tchebychev est aussi démontrée.

Finalement, la variance d'une somme de variables aléatoires indépendantes est la somme des variances. Plus précisément:

Proposition 6.4: Si $X_1, X_2, \ldots, X_N$ sont des variables aléatoires indépendantes ayant un second moment, alors

$$\sigma^2(X_1 + \cdots + X_N) = \sigma^2(X_1) + \cdots + \sigma^2(X_N).$$

Démonstration: Procédons par récurrence sur N. C'est trivial pour $N = 1$. Montrons le pour $N = 2$. Notons pour simplifier $X = X_1 - \mathbb{E}(X_1)$ et $Y = X_2 - \mathbb{E}(X_2)$. Tous deux sont d'espérance nulle. Alors

$$\sigma^2(X_1 + X_2) = \mathbb{E}((X + Y)^2) = \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) = \sigma^2(X_1) + \sigma^2(X_2),$$

car $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) = 0$ en utilisant l'indépendance de X et de Y. Ensuite, supposons le résultat vrai à l'ordre $N - 1$. Alors appliquant le résultat pour $N = 2$ au couple $X = X_1 + \cdots + X_{N-1}$ et $Y = X_N$, puis l'hypothèse de récurrence, on arrive au résultat.

En corollaire, on a donc la *loi faible des grands nombres* qui dit que en un certain sens, si des variables aléatoires sont indépendantes et de même loi, alors leur moyenne arithmétique tend vers leur espérance commune. Plus précisément:

Théorème 6.5: Soit $X_1, X_2, \ldots$ une suite infinie de v.a. indépendantes et de même loi, et possédant un second moment. Alors, pour tout nombre $\epsilon > 0$ fixé on a

$$\lim_{n\to\infty} P\left(|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1)| \geq \epsilon\right) = 0.$$

Démonstration: Notons $S_n = X_1 + \cdots + X_n$. Alors $\mathbb{E}(S_n/n) = \mathbb{E}(X_1)$ et

$$\sigma^2(S_n/n) = \sigma^2(S_n)/n^2 = (\sigma^2(X_1) + \cdots + \sigma^2(X_n))/n^2 = \sigma^2(X_1)/n.$$

Ici on a utilisé successivement les propositions 6.2 puis 6.4, puis le fait que les X_j sont de même loi et ont donc même variance. Appliquons alors l'inégalité de Tchebychev à $X = S_n/n$ et à $t = \epsilon$; on obtient

$$P\left(|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1)| \geq \epsilon\right) \leq \frac{1}{n\epsilon^2}\sigma^2(X_1),$$

qui tend bien vers 0 pour ϵ fixé.

Commentaires: l'importance philosophique de la loi des grands nombres est non négligeable: elle justifie la démarche que nous avons adoptée pour modéliser le calcul des probabilités. L'idée d'expérience décrite au début de ce cours est la sélection d'un point ω dans un espace d'observables Ω, mais par un procédé susceptible d'être répété ad libitum et dans les mêmes conditions. Soit S une partie de Ω, comptons le nombre de fois où S est réalisé en n essais, divisons ce nombre par n et notons par f_n la fraction, ou la fréquence, ainsi obtenue. L'idée de probabilité est basée sur la constatation physique que la suite des f_n converge vers un nombre $P(S)$ qu'on appellera probabilité de S. Si la théorie est bien faite, c'est à dire si les axiomes sont bien choisis, on doit retrouver cette constatation physique quelque part à l'état de théorème dans la théorie développée à partir de ces axiomes. C'est le cas. En effet, le Ω initial décrivant *une* expérience est remplacé par un produit infini $\prod_{j=1}^{\infty} \Omega_j$ où les Ω_j sont identiques à l'Ω initial, et sont les résultats possibles de l'expérience répétée à l'instant j. Les points de ce produit sont donc des suites infinies $\omega = (\omega_j)_{j=1}^{\infty}$. Quant à la probabilité sur le produit, elle est telle que toutes les fonctions $f_j(\omega) = \omega_j$ soient indépendantes. Ceci fait, notons $X_j(\omega) = 1$ si $\omega_j \in S$ et $X_j(\omega) = 0$ sinon. On a une suite de v.a. de Bernoulli indépendantes et de même loi d'espérance $p = P(S)$. La loi faible des grands nombres dit que $f_n = \frac{1}{n}(X_1 + \cdots + X_n)$ converge vers $P(S)$, dans le sens décrit au théorème 6.5. Il existe un théorème avec une conclusion plus précise, appelé *loi forte des grands nombres*, que nous exposons maintenant.

Théorème 6.6: Soit $X_1, \ldots, X_n, \ldots$ des variables aléatoires de Bernoulli indépendantes et de même loi $q\delta_0 + p\delta_1$, avec $0 < p = 1 - q < 1$. Alors

$$\Pr(\lim_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) = p) = 1.$$

Démonstration: Elle s'appuie sur le lemme de Borel: *Si* $(A_n)_{n\geq 1}$ *est une suite d'évènements telle que* $\sum_{n\geq 1} \Pr(A_n)$ *converge, alors* $\Pr(\cap_{k\geq 1} \cup_{n\geq k} A_n) = 0$.

La démonstration de ce lemme est à peu près triviale: Puisque la suite $(r_k)_{k\geq 1}$ des restes de la série convergente tend vers 0 et que pour tout entier k on peut écrire

$$\Pr(\cap_{k\geq 1} \cup_{n\geq k} A_n) \leq \Pr \cup_{n\geq k} A_n) \leq \sum_{n\geq k} \Pr(A_n) = r_k,$$

le résultat s'ensuit en faisant tendre k vers l'infini.

On se fixe ensuite un nombre $\epsilon > 0$ et on note pour simplifier

$$U_n(\epsilon) = U_n = \frac{1}{n}(X_1 + \cdots + X_n) - p - \epsilon, \ \ A_n(\epsilon) = A_n = \{U_n > 0\}, \ \ B(\epsilon) = \{\overline{\lim}_{n\to\infty} U_n > 0\}$$

Le point délicat de la démonstration est de montrer que pour tout $\epsilon > 0$ il existe un nombre $r_\epsilon = r \in]0,1[$ tel que $P(A_n) \leq r^n$. Admettons ce point quelques instants et achevons la démonstration. On remarque d'abord que

$$\cap_{k\geq 1} \cup_{n\geq k} A_n = \{\forall k, \ \exists n \geq k; \ U_n > 0\}.$$

Un point subtil est ensuite l'inclusion d'évènements:

$$\{\overline{\lim}_{n\to\infty} U_n > 0\} \subset \{\forall k, \ \exists n \geq k; \ U_n > 0\} \subset \{\forall k, \ \exists n \geq k; \ U_n \geq 0\} \subset \{\overline{\lim}_{n\to\infty} U_n \geq 0\}.$$

Il n'y a jamais égalité dans ces inclusions: il suffit de penser aux cas $U_n = 1/n$ et $U_n = -1/n$ pour s'en convaincre. Nous n'allons utiliser que la première inclusion. Ayant admis que $\Pr(A_n) < r^n$ avec $r \in]0,1[$, comme la série géométrique de raison r converge, le lemme de Borel est appliquable et on en déduit que $\Pr(B(\epsilon)) = 0$.

Ensuite on observe que si $0 < \epsilon < \epsilon'$ on a $B(\epsilon) \supset B(\epsilon')$. Changeons un peu de notation en écrivant pour N entier $B_N = B(1/N)$. La suite d'évènements $(B_N)_{N\geq 1}$ est donc croissante. Mais comme tous les B_N sont de probabilité nulle, on a encore $\Pr(\cup_{N\geq 1} B_N) = 0$. Analysons alors l'évènement $\cup_{N\geq 1} B_N$. On a

$$\cup_{N\geq 1} B_N = \{\exists N; \ \overline{\lim}_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) > p + \frac{1}{N}\} = \{\overline{\lim}_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) > p\}.$$

Nous avons donc montré que

$$\Pr(\overline{\lim}_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) > p) = 0.$$

Appliquons ce résultat aux variables de Bernoulli $X'_n = 1 - X_n$. Elles sont de loi $p\delta_0 + q\delta_1$ et donc $\Pr(\overline{\lim}_{n\to\infty} \frac{1}{n}(X'_1 + \cdots + X'_n) > q) = 0$. Cependant $\frac{1}{n}(X'_1 + \cdots + X'_n) = 1 - \frac{1}{n}(X_1 + \cdots + X_n)$ et donc

$$\Pr(\underline{\lim}_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) < p) = 0.$$

L'union de deux évènements de probabilité nulle est nulle, le complémentaire de cette union est de probabilité 1. Cela entraîne:

$$\Pr\left(\overline{\lim}_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) \leq p \leq \underline{\lim}_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n)\right) = 1.$$

Donc avec probabilité 1, les limites supérieure et inférieure sont égales à p. C'est le résultat annoncé.

Reste à montrer qu'il existe $r_\epsilon = r \in]0,1[$ tel que

$$\Pr(A_n) = \Pr(\frac{1}{n}(X_1 + \cdots + X_n) > p + \epsilon) \leq r^n.$$

A l'aide d'un nombre $s > 0$ arbitraire, nous donnons d'abord une autre présentation de cet évènement:

$$A_n = \{(\frac{1}{n}(X_1 + \cdots + X_n) > p + \epsilon\} = \{e^{s(X_1+\cdots+X_n)} > e^{sn(p+\epsilon}\}.$$

On applique alors l'inégalité de Markov (proposition 6.3) à $Y = e^{s(X_1+\cdots+X_n)}$ et $y = e^{sn(p+\epsilon)}$. On en tire

$$\begin{aligned}
\Pr(A_n) &\leq \frac{1}{y}\mathbb{E}(Y) \\
&= e^{-sn(p+\epsilon)}\mathbb{E}(e^{s(X_1+\cdots+X_n)}) \\
&= (e^{-s(p+\epsilon)}\mathbb{E}(e^{sX_1}))^n \\
&= (e^{-s(p+\epsilon)}(q + pe^s))^n \\
&= (qe^{-sp-s\epsilon} + pe^{sq-s\epsilon})^n.
\end{aligned}$$

Insistons sur le fait que cette inégalité est valable pour tout $s > 0$. Observons alors qu'il existe des valeurs de s telles que $s \mapsto \varphi(s) = qe^{-sp-s\epsilon} + pe^{sq-s\epsilon}$ soit < 1. Une manière de le voir est de calculer $\varphi(0) = 1$ et $\varphi'(0) = -\epsilon$. Cela entraîne évidemment, puisque $-\epsilon = \varphi'(0) = \lim_{s\to 0}(1 - \varphi(s))/s$, qu'il existe $s_0 > 0$ proche de 0 tel que $r = \varphi(s_0) < 1$. Comme $\varphi > 0$ cela termine la démonstration.

Exercices sur 6.1

1. Soit X une variable aléatoire telles que $0 \leq X \leq 1$. Montrer que $\sigma^2(X) \leq \frac{1}{4}$. Méthode: si $m = \mathbb{E}(X)$, écrire

$$\frac{1}{4} - (X - m)^2 = (\frac{1}{2} - m)^2 + X(1 - X)$$

et prendre l'espérance de chaque membre.

6.2 Les variables aléatoires à valeurs entières.

Nous allons nous concentrer pour un moment sur les variables à valeurs dans l'ensemble **N** des entiers ≥ 0. Dans ce cas les moments seront plus faciles à calculer grâce à l'introduction de la notion de *fonction génératrice* de X :

Théorème 6.6: Soit X une v.a. à valeurs dans **N** de loi $P_X = \sum_{n=0}^{+\infty} p_n\delta_n$. On désigne par $f_X(z)$ la somme de la série entière

$$\sum_{n=0}^{+\infty} p_n z^n$$

de rayon de convergence R. Alors

1. $R \geq 1$ et, pour $|z| \leq 1$ on a $f_X(z) = \mathbb{E}(z^X)$.

2. Pour tout n on a $p_n = \frac{1}{n!}f_X^{(n)}(0)$. En particulier, la connaissance de f_X donne la connaissance de la loi de X.

3. Pour tout n le moment d'ordre n $\mathbb{E}(X^n)$ existe si et seulement si la dérivée à gauche d'ordre n au point 1 de la fonction $z \mapsto f_X(z)$ définie sur $[-1,1]$ existe et est finie. Dans ce cas,

$$\mathbb{E}(X(X-1)\cdots(X-n+1)) = f_X^{(n)}(1) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1)p_n;$$

en particulier $\mathbb{E}(X) = f'_X(1)$, $\mathbb{E}(X^2) = f''_X(1) + f'_X(1)$.

4. Si $X_1, X_2, \ldots, X_N$ sont des variables aléatoires indépendantes à valeurs dans $\mathbf{N}$ et si $S = X_1 + X_2 + \cdots + X_N$ alors pour $|z| \leq 1$:

$$f_S(z) = f_{X_1}(z) \cdots f_{X_N}(z),$$

c'est-à-dire que la fonction génératrice d'une somme est le produit des fonctions génératrices.

Démonstration: Il est clair que la série entière converge pour $z = 1$ puisque $\sum_{n=0}^{+\infty} p_n = 1$ et donc que $f_X(1) = 1$. Donc $R \geq 1$. Ensuite, si $|z| = 1$ la série est absolument convergente. Pour le 2), cela découle du lien entre la formule de Taylor et la somme d'une série entière.

Le 3) est plus délicat. Nous le montrons pour $n = 1$. Le principe pour n quelconque est le même. Supposons d'abord que $\mathbb{E}(X)$ existe, c'est-à-dire, d'après la proposition 5.3, que $\sum_{n=0}^{+\infty} np_n$ converge. Montrons qu'alors la dérivée à gauche en 1 de f_X existe et est finie. Celle ci est définie comme la limite quand z croît vers 1 de la fonction

$$\frac{f_X(z) - f_X(1)}{z-1} = \frac{1 - f_X(z)}{1-z} = \sum_{n=0}^{+\infty} p_n \frac{1-z^n}{1-z} = \sum_{n=0}^{+\infty} p_n (1 + z + \cdots + z^{n-1}).$$

Or si $0 \leq z \leq 1$ on a $1 + z + \cdots + z^{n-1} \leq n$. Comme $\sum_{n=0}^{+\infty} np_n$ converge la série précédente converge normalement et sa limite est pour z tendant vers 1 est $\mathbb{E}(X)$.

Inversement, supposons que la dérivée à gauche en 1, notée $f'_X(1)$ existe. Appliquons le théorème des accroissement finis à l'intervalle $[z,1]$ et à la fonction f_X. Il existe donc $c \in]z,1[$ tel que

$$\frac{1 - f_X(z)}{1-z} = f'_X(c) = \sum_{n=1}^{+\infty} np_n c^{n-1}.$$

Ceci tend vers une limite finie si z croit vers 1 par hypothèse. Il est clair puisque c tend vers 1 avec z, que cette limite est supérieure ou égale à toutes les sommes partielles de la série $\sum_{n=0}^{+\infty} np_n$, ce qui prouve que cette série converge. Enfin, trivialement,

$$\sum_{n=1}^{+\infty} p_n c^{n-1} \leq \sum_{n=1}^{+\infty} np_n,$$

ce qui montre finalement que $f'_X(1) = \mathbb{E}(X)$.

Le 4) est une conséquence immédiate du fait que si les X_j sont indépendants, alors les z^{X_j} sont indépendants, et que l'espérance du produit de variables indépendantes est le produit des espérances:

$$f_S(z) = \mathbb{E}(z^{X_1+\cdots+X_N}) = \mathbb{E}(z^{X_1}\cdots z^{X_N}) = \mathbb{E}(z^{X_1})\cdots\mathbb{E}(z^{X_N}) = f_{X_1}(z)\cdots f_{X_N}(z).$$

Commentaires: la démonstration du 3) n'est pas facile si $R = 1$, comme on l'a vu. Si $R > 1$, c'est simple et immédiat par le théorème de dérivation d'une série entière à l'intérieur de l'intervalle de convergence.

Nous étudions maintenant 4 exemples fondamentaux de lois sur $\mathbf{N}$.

La loi de Bernoulli $B_{1,p}$. Pour $0 < p < 1$ c'est la loi

$$B_{1,p} = (1-p)\delta_0 + p\delta_1.$$

Sa fonction génératrice est $f(z) = (1-p) + pz$, son espérance est p et sa variance est $(1-p)p$.

La loi binomiale $B_{N,p}$. C'est la loi du nombre de succès dans le schéma Succès Echec fini à N essais:

$$B_{N,p} = \sum_{k=0}^{N} C_N^k (1-p)^{N-k} p^k \delta_k.$$

Sa fonction génératrice est d'après la formule du binôme, $f(z) = ((1-p)+pz)^N$. Donc en prenant sa dérivée à l'ordre 1, son espérance est donc Np. Quant à sa variance, c'est $N(1-p)p$. On remarque que si X et Y sont indépendantes et de lois respectives $B_{N,p}$ et $B_{M,p}$, alors la loi de $X+Y$ est $B_{N+M,p}$, comme on le voit par la fonction génératrice.

Un bon moyen de retenir ces résultats sur la loi binomiale est d'observer que si $X_1, \ldots, X_N$ sont des variables aléatoires indépendantes de même loi de Bernoulli $B_{1,p}$, alors $S = X_1 + \cdots + X_N$ est de loi binomiale $B_{N,p}$ comme on le voit par la fonction génératrice f_S.

La loi de Poisson $\mathcal{P}_\lambda$. Pour $\lambda > 0$, c'est la loi définie par

$$\mathcal{P}_\lambda = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n.$$

Sa fonction génératrice est $f(z) = \exp(\lambda(z-1))$, son espérance et sa variance sont toutes deux égales à λ. On remarque que si X et Y sont indépendantes et de lois respectives $\mathcal{P}_\lambda$ et $\mathcal{P}_\mu$, alors la loi de $X+Y$ est $\mathcal{P}_{\lambda+\mu}$, comme on le voit par la fonction génératrice.

La manière la plus courante de rencontrer cette loi de Poisson dans la nature est en tant qu'approximation de la loi binomiale. En effet, la suite de lois $B_{N,\lambda/N}$ tend vers $\mathcal{P}_\lambda$ dans le sens suivant: pour tout entier k on a

$$\lim_{N\to\infty} B_{N,\lambda/N}(\{k\}) = \mathcal{P}_\lambda(\{k\}).$$

Pour le voir, on observe que la suite du premier membre est

$$C_N^k (1-\frac{\lambda}{N})^{N-k} (\frac{\lambda}{N})^k = \frac{N(N-1)\cdots(N-k+1)}{N^k} (1-\frac{\lambda}{N})^{-k} \frac{\lambda^k}{k!} (1-\frac{\lambda}{N})^N.$$

Le premier produit tend vers 1, comme quotient de deux polynômes de N de degré k ayant même terme de plus haut degré. Il est clair que toute l'expression tend vers $\frac{\lambda^k}{k!}e^{-\lambda}$ si N tend vers l'infini, par la formule connue $\lim_{N\to\infty}(1+\frac{x}{N})^N = \exp x$.

La loi de Pascal et la loi négative binomiale. Dans le schéma Succès Echec infini, intéressons nous à la loi du temps d'attente T_1 du premier succès , soit $T_1(\omega) = \inf \{n \ ; \ \omega_j = S\}$. La loi de T_1 se calcule facilement en remarquant que dire que $T_1 > n$ est dire que les n premiers essais ont été des échecs, un évènement de probabilité $(1-p)^n$. Donc, puisque

$$P(T_1 = n) = P(T_1 > n-1) - P(T_1 > n) = (1-p)^{n-1} - (1-p)^n = (1-p)^{n-1}p,$$

la loi de T_1, dite *loi de Pascal*, ou *loi géométrique*, est

$$P_{T_1} = p\delta_1 + (1-p)p\delta_2 + \cdots + (1-p)^{n-1}p\delta_n + \cdots$$

Sa fonction génératrice est la fonction homographique $f_{T_1}(z) = \frac{pz}{1-(1-p)z}$, sa moyenne est $1/p$, un résultat qu'il est bon de retenir. Quant à sa variance, c'est $\sigma^2(T_1) = (1-p)/p^2$.

Si ensuite on s'intéresse au temps d'attente T_k du k ième succès, il est intuitivement clair, bien que pas si facile à montrer rigoureusement, que c'est la somme de k variables aléatoires indépendantes $I_1, \ldots, I_k$, de même loi que T_1: la v.a. I_k représente l'intervalle de temps entre les $k-1$ ième et k ième succès. La fonction génératrice est donc $f_{T_k}(z) = (\frac{pz}{1-(1-p)z})^k$, la moyenne k/p et la variance $k(1-p)/p^2$. Toutefois, la loi de T_k est concentrée sur les entiers supérieurs ou égaux à k, et il y a avantage en vue d'une généralisation à considérer plutôt la loi de $T_k - k$, concentrée sur $\mathbf{N}$, de fonction génératrice

$$f_{T_k-k}(z) = (\frac{p}{1-(1-p)z})^k = \sum_{n=0}^{\infty} \frac{1}{n!}k(k+1)\cdots(k+n-1)p^k(1-p)^n z^n,$$

en développant selon la formule du binôme de Newton. Cela entraîne donc que si $n \geq k$:

$$P(T_k = n) = P(T_k - k = n-k) = \frac{1}{(n-k)!}k(k+1)\cdots(n-1)p^k(1-p)^{n-k} = C_{n-1}^{k-1}p^k(1-p)^{n-k},$$

une formule difficile à retenir.

Maintenant, on peut généraliser la loi de $T_k - k$ en remplaçant le paramètre entier k par le paramètre continu positif λ. L'interprétation probabiliste disparait, mais les formules demeurent. On introduit donc la loi dite négative-binomiale $NB_{\lambda,p}$ définie pour $\lambda > 0$ et $0 < p < 1$ par

$$NB_{\lambda,p} = \sum_{n=0}^{\infty} \frac{1}{n!}\lambda(\lambda+1)\cdots(\lambda+n-1)p^\lambda(1-p)^n\delta_n.$$

Une variable aléatoire X qui suit une telle loi est donc telle que si $n \in \mathbf{N}$:

$$P(X = n) = \frac{1}{n!}\lambda(\lambda+1)\cdots(\lambda+n-1)p^\lambda(1-p)^n,$$

sa fonction génératrice est $f_X(z) = (\frac{p}{1-(1-p)z})^\lambda$, sa moyenne est $\lambda(1-p)/p$ et sa variance est $\lambda(1-p)/p^2$.

Exercices sur 6.2

1. Montrer que si deux dés sont marqués sur leurs faces $1,2,3,4,5,6$ il est impossible de les piper de sorte que la somme $X+Y$ de leur points soit telle que $P(X+Y=n)=\frac{1}{11}$ pour $n=2,3,\ldots,12$. Méthode: montrer que les fonctions génératrices $f_X(z)$ et $f_Y(z)$ sont telles que $f_X(z)/z$ et $f_X(z)/z$ sont des polynômes ayant au moins un zéro réel, et que $f_{X+Y}(z)/z^2$ n'a que des zéros imaginaires.

2. Une fonction génératrice f_X est telle que $f_X(z)=(1-\sqrt{1-z})/z$. Quelle est la probabilité pour que $X=n$? Est ce que $\mathbb{E}(X)$ existe?

3. Soit X et Y deux variables aléatoires indépendantes qui suivent des lois de Pascal pas nécessairement identiques. Soit $Z=\min(X,Y)$. Calculer pour n fixé $P(X>n$, $P(Y>n)$, $P(Z>n)$, $P(Z=n)$. Montrer que Z suit une loi de Pascal. Exprimer sa moyenne en fonction des moyennes de X et Y.

6.3 Transformée de Laplace d'une variable aléatoire.

Théorème 6.7: Soit X une variable aléatoire. Soit I_X l'ensemble des z réels tels que $L_X(z)=\mathbb{E}(e^{zX})$ existe. La fonction $z\mapsto L_X(z)$ définie sur I_X est appelée la *transformée de Laplace* de X. Alors

1. L'ensemble I_X est un intervalle contenant 0.

2. Si 0 est dans l'intérieur de I_X, la transformée de Laplace est développable en série entière et les coefficients de cette série sont les $L_X^{(n)}(0)/n!=\mathbb{E}(X^n)/n!$:
$$L_X(z)=\sum_{n=0}^{\infty}\frac{\mathbb{E}(X^n)}{n!}z^n.$$

3. Si I_X est de longueur positive, la loi de X est caractérisée par sa transformée de Laplace. Plus précisément, si $I_X\cap I_Y$ est de longueur positive et si $L_X=L_Y$ sur cet intervalle, alors X et Y sont de même loi.

4. Si X et Y sont indépendantes, alors $I_{X+Y}=I_X\cap I_Y$ et , pour z dans cet intervalle: $L_{X+Y}(z)=L_X(z)L_Y(z)$.

5. Si a et b sont réels avec $a\neq 0$ alors $I_{aX+b}=\frac{1}{a}I_X$ et $L_{aX+b}(z)=\exp(bz)L_X(az)$.

Démonstration: 1) Il est clair que $0\in I_X$. Si $0<s<z$ ou si $z<s<0$ et si $z\in I_X$, montrons que $s\in I_X$. Cela vient du fait que $\exp(sX)\leq 1+\exp(zX)$, comme on le voit en examinant les 4 cas $X\geq 0$ et $X<0$, $z>0$ et $z<0$.

2) Si $[-a,a]\subset I_X$ avec $a>0$, alors comme $\exp(a|X|)<\exp(aX)+\exp(-aX)$ on en déduit que $\mathbb{E}(\exp(a|X|))$ existe, et donc $\mathbb{E}(\exp|zX|)$ existe pour tout $|z|\leq a$. D'où pour un tel z
$$\left|L_X(z)-\sum_{n=0}^{N}\frac{\mathbb{E}(X^n)}{n!}z^n\right|=\left|\mathbb{E}(\exp(zX)-\sum_{n=0}^{N}\frac{(Xz)^n}{n!}\right|=\left|\mathbb{E}(\sum_{n=N+1}^{\infty}\frac{(Xz)^n}{n!}\right|\leq$$

$$\mathbb{E}\left(\sum_{n=N+1}^{\infty}\frac{|Xz|^n}{n!}\right)=\mathbb{E}\left(\exp|zX|-\sum_{n=0}^{N}\frac{|Xz|^n}{n!}\right)=\mathbb{E}(Y_N).$$

La variable aléatoire Y_N décroit vers 0: un théorème de 3ème année dit que cela suffit pour entraîner que $\lim_{N\to\infty}\mathbb{E}(Y_N)=0$; ce qui achève la démonstration du 2).

La partie 3) est beaucoup plus difficile et nous admettrons ce résultat.

La partie 4) est une conséquence du théorème 5.8 appliqué à $N=2$ et à $(X_1,X_2)=(\exp(zX),\exp(zY))$. La partie 5) est immédiate.

A cause du 2) on appelle parfois la transformée de Laplace la *fonction génératrice des moments*. C'est à éviter, pour ne pas confondre avec la fonction génératrice d'une variable aléatoire X à valeurs dans $\mathbf{N}$. D'ailleurs, pour un tel X, les deux notions sont reliées par $f_X(\exp z)=L_X(z)$ et l'intérieur de I_X est alors $]-\infty,\log R[$ où R est le rayon de convergence de la série entière de somme f_X. Les transformées de Laplace sont surtout utilisées pour caractériser des v.a. à densité. Nous en donnons 3 exemples importants.

La loi normale N_{m,σ^2}. C'est la loi la plus importante du calcul des probabilités. On l'appelle aussi une loi gaussienne, une loi de Laplace-Gauss, ou encore une seconde loi de Laplace. Si $m\in\mathbb{R}$ et si $\sigma>0$, elle est définie par sa densité:

$$\frac{1}{\sigma\sqrt{2\pi}}\exp-\frac{(x-m)^2}{2\sigma^2}.$$

Le fait que ce soit une densité de probabilité n'est pas évident, car il faut vérifier que l'intégrale de cette fonction >0 est 1. Si on l'admet pour le cas $m=0$ et $\sigma=1$, on se ramène facilement à ce cas particulier en posant $x=\sigma y+m$. Cette remarque permet alors de montrer que la transformée de Laplace d'une variable aléatoire Y de loi $N_{0,1}$ est

$$L_Y(z)=\mathbb{E}(e^{zY})=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2}+zy}dy=e^{\frac{z^2}{2}}.$$

Pour voir cette dernière égalité il suffit d'écrire que la densité de $N_{z,1}$ est d'intégrale 1. Remarquons que l'intervalle d'existence est $I_Y=\mathbb{R}$

Ensuite, on remarque que si Y est de loi $N_{0,1}$, alors $X=\sigma Y+m$ est de loi N_{m,σ^2}. Pour le voir, il suffit d'écrire la fonction de répartition de X de la manière suivante:

$$F_X(x)=P(\sigma Y+m\leq x)=P(Y\leq\frac{x-m}{\sigma})=F_Y(\frac{x-m}{\sigma});$$

on dérive alors les deux membres extrêmes de la ligne ci dessus: à gauche on obtient la densité cherchée de X, à droite en utilisant le théorème de dérivation des fonctions composées et le fait que la densité de Y est par hypothèse $\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$: ceci fournit pour X la densité de la loi N_{m,σ^2} comme annoncé.

Enfin, pour avoir la transformée de Laplace de X à partir de Y on utilise le 5) du théorème 6.7 pour obtenir que si X est de loi N_{m,σ^2}, alors

$$L_X(z)=\exp(\frac{\sigma^2z^2}{2}+mz).$$

On déduit du 2) du théorème 6.7 qu'alors $\mathbb{E}(X) = m$ et que $\sigma^2(X) = \sigma^2$. On déduit aussi des 3) et 4) du théorème 6.7 que si X_1 et X_2 sont des variables aléatoires indépendantes et de lois respectives $N_{m_1,\sigma_1{}^2}$ et $N_{m_1,\sigma_2{}^2}$, alors $X_1 + X_2$ est de loi $N_{m_1+m_2,\sigma^2{}_1+\sigma_2{}^2}$.

A propos de fonction de répartition, il faut noter que la fonction de répartition Φ de la loi $N_{0,1}$, soit

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy,$$

n'est pas élémentaire. Elle est tabulée dans tous les ouvrages.

On rencontre la loi $N_{0,1}$ dans la nature comme approximation de bien des lois. La plus ancienne est l'*approximation de Moivre Laplace de la loi binomiale*: Si X est de loi $B_{N,p}$, alors la loi de $\frac{X-Np}{\sqrt{Np(1-p)}}$ tend vers la loi $N_{0,1}$ dans le sens suivant: pour tout intervalle $[a, b]$ on a

$$\lim_{N\to\infty} P\left(a \leq \frac{X - Np}{\sqrt{Np(1-p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}} dy.$$

Une autre présentation de ce théorème de Moivre Laplace est donc

$$\lim_{N\to\infty} P\left(a\sqrt{Np(1-p)} + Np \leq X \leq b\sqrt{Np(1-p)} + Np\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}} dy.$$

C'est dire que $P\left(a\sqrt{Np(1-p)} + Np \leq X \leq b\sqrt{Np(1-p)} + Np\right)$ est approchée par $\Phi(b)-\Phi(a)$. Cette approximation est à la base de la statistique.

La démonstration de ce résultat n'est pas élémentaire. Toutefois, l'usage des transformées de Laplace le rend plausible; avec le théorème 6.7, partie 5):

$$L_{\frac{X-Np}{\sqrt{Np(1-p)}}}(z) = (1 - p + p\frac{z}{\sqrt{Np(1-p)}})^N \exp\frac{-Npz}{\sqrt{Np(1-p)}} \to_{N\to\infty} \exp\frac{z^2}{2},$$

par un calcul de développement limité.

Les lois gamma $\gamma_{p,q}$. La loi exponentielle $\gamma_{1,q}$ de moyenne q est la plus importante des lois à densité après la loi normale. Elle est concentrée sur la demi droite positive, sa fonction de répartition est pour $x > 0$ $F(x) = 1 - \exp(-x/q)$ et en dérivant F, sa densité est

$$\frac{1}{q}\exp(-x/q)\mathbf{1}_{]0,+\infty[}(x).$$

On la rencontre dans la nature car c'est une loi sans mémoire: si X suit une loi exponentielle de moyenne q et si x et y sont > 0, alors

$$P(X > x + y|X > y) = \frac{P(X > x + y)}{P(X > y)} = \frac{1 - F(x + y)}{1 - F(y)} = \exp(-x/q) = P(X > x).$$

Par exemple une ampoule électrique ne s'use pas, et le fait que nous sachions qu'elle a déjà duré un temps y ne nous donne aucune information pour savoir si elle va durer au moins un temps x à partir de maintenant.

La transformée de Laplace d'une variable aléatoire X de loi exponentielle existe sur $I_X =]-\infty, 1/q[$ et est égale à $L_X(z) = \frac{1}{1-qz}$. Ceci montre avec le théorème 6.7, 2), que $\mathbb{E}(X) = q$, $\mathbb{E}(X^2) = 2q^2$ et, par la formule de Huyghens, que $\sigma^2(X) = q^2$.

Si p est un nombre entier positif et si $X_1, \cdots, X_p$ sont des v.a. indépendantes et de même loi $\gamma_{1,q}$, la transformée de Laplace de $X_1 + \cdots + X_p$ est donc $(\frac{1}{1-qz})^p$ sur $]-\infty, 1/q[$. Comme la transformée de Laplace détermine la loi, il suffit de montrer (par une intégration par parties qui permet de faire une récurrence sur p) que

$$\frac{1}{(p-1)!} \int_0^{+\infty} \exp(zx - x/q) q^{-p} x^{p-1} dx = (\frac{1}{1-qz})^p$$

pour en déduire que la densité de $X_1 + \cdots + X_p$ est

$$\frac{1}{(p-1)!} \int_0^{+\infty} \exp(-x/q) q^{-p} x^{p-1} \mathbf{1}_{]0,+\infty[}(x) :$$

c'est la densité de la loi $\gamma_{p,q}$.

En fait, comme pour la loi négative binomiale qui a été obtenue par une interpolation des entiers, il est possible dans la loi $\gamma_{p,q}$ de remplacer le paramètre entier par le paramètre $p > 0$. Pour cela on introduit une importante fonction de p appelée fonction Gamma d'Euler et définie pour $p > 0$ par

$$\Gamma(p) = \int_0^{+\infty} \exp(-x) x^{p-1} dx.$$

Une intégration par parties montre que $\Gamma(p+1) = p\Gamma(p)$. Comme $\Gamma(1) = 1$ on en tire que si p est entier $\Gamma(p) = (p-1)!$: cette fonction Gamma interpole les factorielles.

On définit alors la loi $\gamma_{p,q}$ pour $p > 0$ non nécessairement entier par sa densité :

$$\frac{1}{\Gamma} \int_0^{+\infty} \exp(-x/q) q^{-p} x^{p-1} \mathbf{1}_{]0,+\infty[}(x)$$

qui a pour transformée de Laplace $(\frac{1}{1-qz})^p$. On déduit de cette transformée de Laplace que la moyenne est pq et que la variance est pq^2. On appelle p le paramètre de forme et q le paramètre d'échelon. En effet, on voit facilement, soit avec les fonctions de répartition, soit avec les transformées de Laplace, que si X est de loi $\gamma_{p,1}$ alors qX est de loi $\gamma_{p,q}$. Changer q est un simple changement d'unités de mesure, changer p change de façon importante l'allure de la densité.

La loi uniforme sur $[a, b]$. C'est la loi $U_{[a,b]}$, de densité $\frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$. Sa fonction de répartition $F(x)$ est nulle si $x < a$, égale à $\frac{x-a}{b-a}$ si $x \in [a, b]$ et égale 1 si $x > b$. Il est facile de voir que si X est de loi $U_{[0,1]}$ alors $Y = a + (b-a)X$ est de loi $U_{[a,b]}$ (on dit aussi que Y est *uniformément répartie* sur $[a, b]$). La transformée de Laplace n'est pas spécialement remarquable. Pour $U_{[0,1]}$, c'est $L(z) = \frac{1}{z}(e^z - 1)$ si $z \neq 0$ et $L(0) = 1$ Le moment d'ordre n pour $U_{[0,1]}$ s'obtient directement à partir de la définition : c'est $1/(n+1)$. Les variables uniformes sont intensément utilisées en simulation.

Appendice 1: Grandes déviations

Si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes et de même loi, de moyenne m et telle qu'il existe $\alpha > 0$ avec $\mathbb{E}(e^{\alpha|X_n|}) < \infty$, et si on note $S_n = X_1 + \cdots + X_n$, le théorème suivant calcule pour $a > m$ le nombre $\lim_{n\infty}(\Pr(S_n \geq na))^{1/n}$.

Théorème: Soit μ une mesure positive sur $\mathbb{R}$ non concentrée en un point et telle que l'intervalle des θ réels satisfaisant $L(\theta) = \int_{-\infty}^{\infty} e^{\theta x}\mu(dx) < \infty$ ait un intérieur Θ non vide. On considère la fonction strictement convexe sur Θ égale à $k = \log L$ et l'intervalle ouvert $M = k'(\Theta)$, et on note par $\psi : M \to \Theta$ la fonction réciproque de k'.

Soit $m = k'(\theta)$ fixé dans M et $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes et de même loi $e^{\theta x - k(\theta)}\mu(dx)$. Soit enfin $a \in M$ avec $m < a$ et les nombres

$$\begin{aligned} u_n &= \Pr(\frac{1}{n}(X_1 + \cdots + X_n) \geq a) \\ h(m,a) &= -\int_m^a (a-x)\psi'(x)dx = a(\psi(m) - \psi(a)) + k(\psi(a)) - k(\psi(m)). \end{aligned}$$

Dans ces conditions on a

1. (Inégalité des grandes déviations) $u_n^{1/n} \leq e^{h(m,a)}$.
2. (Théorème des grandes déviations) $\lim_{n\infty} u_n^{1/n} = e^{h(m,a)}$.

Commentaires: 1) Une insupportable confusion règne dans la littérature d'enseignement concernant ce résultat, dû à Cramer (1938), principalement à cause de ses généralisations à des hypothèses plus faibles (et peu intéressantes) dans $\mathbb{R}$ ainsi qu'à $\mathbb{R}^d$, où les résultats n'ont pas l'harmonie du résultat ci dessus.

2) Dans sa présentation, le théorème fait jouer un rôle symétrique à toute la famille de lois de probabilités $e^{\theta x - k(\theta)}\mu(dx)$ quand θ varie dans Θ. Cette famille est appelée une famille exponentielle naturelle engendrée par μ. Attention, μ n'est pas unique: μ' engendre la même famille exponentielle, c'est à dire le même *ensemble* de probabilités, indépendamment du paramétrage, si et seulement si il existe a et b réels tels que $\mu'(dx) = e^{ax+b}\mu(dx)$. Il est clair que la loi d'une variable aléatoire réelle X telle qu'il existe $\alpha > 0$ avec $\mathbb{E}(e^{\alpha|X|}) < \infty$ appartient à une famille exponentielle naturelle: il suffit de prendre pour μ la loi de X. Toutefois, pour la loi de X donnée, souvent avec un paramètre, il n'est pas toujours apparent de relier cette loi avec la famille exponentielle à laquelle elle appartient. Par exemple la loi de Bernoulli $(1-p)\delta_0 + p\delta_1$ appartient à la famille exponentielle engendrée par $\mu = \delta_0 + \delta_1$: prendre $p = \frac{e^\theta}{1+e^\theta}$.

3) Implicitement, l'énoncé utilise des résultats simples comme le fait que Θ soit un intervalle et comme la convexité de k, qui se démontrent comme le 1) et le 6) du théorème 6.7 du cours de Deug. De plus, il est facile de voir que avec les notations du théorème, l'espérance des X_i est $m = k'(\theta)$ et leur variance est $k''(\theta) = 1/\psi'(m)$.

4) La partie 2) du théorème est plus difficile. La partie 1) est comme on va le voir amusante et élémentaire. Elle fournit une démonstration de poche de la loi forte des grands nombres qui affirme que si $X_1, \ldots, X_n, \ldots$ sont des variables aléatoires indépendantes et de même loi de moyenne m, alors $\Pr(\lim_{n\infty} S_n/n = m) = 1$. Si on fait l'hypothèse

suplémentaire de l'existence de moments exponentiels, c'est à dire qu'il existe $\alpha > 0$ avec $\mathbb{E}(e^{\alpha|X_n|}) < \infty$, alors l'inégalité des grandes déviations et le critère de Cauchy, du fait que $h(m,a) < 0$, entraine que la série $\sum u_n$ converge, et on procède alors comme au Théorème 6.6 du cours pour conclure avec le lemme de Borel.

5) Travaux pratiques: Voici quelques mesures μ classiques, et les lois et les fonctions $h(m,a)$ qui vont avec.

Loi de Bernoulli: $\mu = \delta_0 + \delta_1$, $L(\theta) = 1 + e^\theta$, $\Theta = \mathbb{R}$, $k(\theta) = \log(1+e^\theta)$, $k'(\theta) = \frac{e^\theta}{1+e^\theta}$, $M =]0,1[$, $\psi(m) = \log\frac{m}{1-m}$, $k(\psi(m)) = -\log(1-m)$ et

$$h(m,a) = a\log\frac{a}{m} + (1-a)\log\frac{1-a}{1-m}.$$

Loi de Poisson: $\mu = \sum_{n=0}^\infty \frac{1}{n!}\delta_n$, $L(\theta) = \exp e^\theta$, $\Theta = \mathbb{R}$, $k(\theta) = e^\theta$, $k'(\theta) = e^\theta$, $M =]0,\infty[$, $\psi(m) = \log m$, $k(\psi(m)) = m$ et

$$h(m,a) = a\log\frac{m}{a} + a - m.$$

Loi gamma: Soit $\alpha > 0$ fixé. $\mu = \frac{1}{\Gamma(\alpha)}x^{\alpha-1}\mathbf{1}_{]0,\infty[}(x)dx$, $L(\theta) = \frac{1}{(-\theta)^\alpha}$ si $\theta \in \Theta =]-\infty, 0[$, $k(\theta) = \alpha\log(-\theta)$, $k'(\theta) = \frac{\alpha}{-\theta}$, $M =]0,\infty[$, $\psi(m) = \frac{\alpha}{-m}$, $k(\psi(m)) = \alpha\log\frac{m}{\alpha}$ et

$$h(m,a) = \alpha - \alpha\frac{a}{m} + \alpha\log\frac{a}{m}.$$

Loi normale: Soit $\sigma > 0$ fixé. $\mu = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{x^2}{2\sigma^2})dx$, $L(\theta) = \exp(\frac{\sigma^2\theta^2}{2})$, $\Theta = \mathbb{R}$, $k(\theta) = \frac{\sigma^2\theta^2}{2}$, $k'(\theta) = \sigma^2\theta$, $M = \mathbb{R}$, $\psi(m) = \frac{m}{\sigma^2}$, $k(\psi(m)) = \frac{m^2}{2\sigma^2}$ et

$$h(m,a) = -\frac{(a-m)^2}{2\sigma^2}.$$

Démonstration de l'inégalité des grandes déviations : Notons $S_n = X_1 + \cdots + X_n$. Pour tout $t > 0$ tel que $\theta + t \in \Theta$ l'astuce est d'observer que les deux évènements $\{S_n/n \geq a\}$ et $\{e^{tS_n} \geq e^{nta}\}$ sont les mêmes (comme à la Prop. 6.6). On écrit, à l'aide de l'inégalité de Markov (voir cours de Deug, Prop. 6.2) appliquée à $Y = e^{tS_n}$ et à $y = e^{nta}$:

$$u_n = \Pr(e^{tS_n} \geq e^{nta}) \leq e^{-nta}\mathbb{E}(e^{tS_n}) = [e^{-ta}\mathbb{E}(e^{tX_1})]^n = [e^{-ta}\frac{L(\theta+t)}{L(\theta)}]^n.$$

Donc $u_n^{1/n} \leq e^{-ta+k(\theta+t)-k(\theta)}$. Observons ensuite que $t \mapsto -ta + k(\theta+t) - k(\theta)$ est convexe sur l'intervalle ouvert $(-\theta+\Theta)\cap]0,\infty[$ et que sa dérivée s'y annule au point $t = \psi(a) - \psi(m)$, c'est à dire tel que $k'(\theta+t) = a$. La valeur de $-ta + k(\theta+t) - k(\theta)$ en ce point est exactement $h(m,a)$ et le résultat est montré.

Démonstration du théorème des grandes déviations : On pose désormais $\tau = \psi(a) > \theta$. Avec cette notation, on remarque que

$$h(m,a) = -a(\tau-\theta) + k(\tau) - k(\theta).$$

L'astuce de Harald Cramer ici est d'introduire les variables aléatoires $Y_1, \ldots, Y_n$ indépendantes et de même loi $e^{\tau x - k(\tau)}\mu(dx)$ Si on a lu le commentaire 2, on remarque que cette loi appartient à la même famille exponentielle naturelle que la loi des X_i. L'espérance de Y_i est $a = k'(\tau)$. On pose ensuite $U_n = Y_1 + \cdots + Y_n$, $V_n = U_n - na$ et $v_n = \mathbb{E}[e^{-(\tau-\theta)V_n}\mathbf{1}_{V_n \geq 0}]$. L'espérance de V_n est zéro. On montre alors l'identité remarquable

$$u_n = e^{nh(m,a)}v_n.$$

Pour le voir, on introduit la mesure positive μ_n sur $\mathbb{R}$ égale à la n ième puissance de convolution μ^{*n}, c'est à dire de transformée de Laplace $L(\theta)^n$. La loi de S_n est donc $e^{\theta s - nk(\theta)}\mu_n(ds)$, comme on le vérifie en calculant la transformée de Laplace de cette loi et en voyant qu'elle est égale à celle de S_n, soit

$$\mathbb{E}(e^{tS_n}) = \left(\frac{L(\theta+t)}{L(\theta)}\right)^n = e^{nk(\theta+t)-nk(\theta)}$$

pour tout $t \in -\theta + \Theta$. De même, la loi de U_n est $e^{\tau u - nk(\tau)}\mu_n(du)$. Par conséquent

$$\begin{aligned} & e^{nh(m,a)}\mathbb{E}[e^{-(\tau-\theta)V_n}\mathbf{1}_{V_n \geq 0}] \\ = \; & e^{n[h(m,a)+(\tau-\theta)a]}\mathbb{E}[e^{-(\tau-\theta)U_n}\mathbf{1}_{U_n \geq na}] \\ = \; & e^{n[k(\tau)-k(\theta)]}\int_{na}^{\infty} e^{-(\tau-\theta)u}e^{\tau u - nk(\tau)}\mu_n(du) \\ = \; & \int_{na}^{\infty} e^{\theta u - nk(\theta)}\mu_n(du) = \Pr(S_n \geq na) = u_n, \end{aligned}$$

et l'identité annoncée $u_n = e^{nh(m,a)}v_n$ est montrée. On peut remarquer qu'elle nous donne au passage une seconde démonstration, moins élémentaire, de la partie 1, puisque trivialement $v_n < 1$. Cette partie algébrique étant faite, pour voir que la limite de $u_n^{1/n}$ est $e^{h(m,a)}$, il suffit de montrer que $\lim_{n\to\infty} v_n^{1/n} = 1$. C'est la partie plus difficile.

Commencons par un lemme classique:

Lemme: Si f est une variable aléatoire positive alors l'ensemble des $s \in \mathbb{R}$ tels que $\mathbb{E}(f^s) < \infty$ est un intervalle I et $s \mapsto [\mathbb{E}(f^s)]^{1/s}$ est croissante sur $]0, \infty[\cap I$.

Démonstration du lemme: On pourrait utiliser une inégalité classique de Hölder. Utilisons plutôt ici l'outil familier de la convexité du logarithme de la transformée de Laplace. Soit $(1-p)\delta_0 + p\nu(df)$ la loi de f avec $\nu(df)$ probabilité sur $]0, \infty[$ et $0 < p \leq 1$ (tout est trivial si $p = 0$). Soit $\mu(dx)$ l'image de $\nu(df)$ par $f \mapsto x = \log f$. Soit T la transformée de Laplace de μ, soit I son domaine de finitude et soit $t(s) = \log T(s)$. Sur I on a $\mathbb{E}(f^s) = pe^{t(s)}$. Enfin si $0 < s < s_1$ sont dans I, comme t est convexe on a

$$t(s) = t(\frac{s}{s_1}s_1 + (1 - \frac{s}{s_1})0) \leq \frac{s}{s_1}t(s_1) + (1 - \frac{s}{s_1})t(0).$$

Comme $t(0) = 0$ (car μ est une probabilité) on obtient que $\frac{1}{s}t(s) \leq \frac{1}{s_1}t(s_1)$. Comme

$$p^{\frac{1}{s}} = \exp(\frac{1}{s}\log p) \leq \exp(\frac{1}{s_1}\log p) = p^{\frac{1}{s_1}},$$

le lemme est montré.

Achevons alors la démonstration du théorème. On pose

$$A_n = \{V_n \geq 0\}, \quad B_n = \{e^{-(\tau-\theta)V_n/n^{3/4}} \geq \frac{1}{2}\}.$$

On a alors

$$v_n^{n^{-3/4}} \overset{(1)}{\geq} \mathbb{E}[e^{-(\tau-\theta)V_n/n^{3/4}}\mathbf{1}_{V_n\geq 0}] \overset{(2)}{\geq} \frac{1}{2}\Pr(A_n \cap B_n) \overset{(3)}{\geq} \frac{1}{2}(\Pr(A_n) - \Pr(B_n)).$$

Dans cette chaîne d'inégalités, (1) vient du lemme appliqué à $f = e^{-(\tau-\theta)V_n}\mathbf{1}_{V_n\geq 0}$ et au couple $s_1 = 1$ et $s = 1/n^{3/4}$, (2) est l'inégalité de Markov appliquée à $Y = f^{n^{-3/4}}$ et $y = 1/2$, et (3) vient du fait que si A et B sont deux évènements alors $A \subset (A \cap B) \cup B$ et donc $\Pr(A \cap B) \geq \Pr(A) - \Pr(B)$. Faisons alors tendre n vers l'infini. D'après le théorème central limite, la loi de $V_n/\sqrt{n}$ tend vers une loi normale centrée. On en déduit que $\Pr(A_n)$ tend vers $1/2$ et, puisque B_n est aussi $B_n = \{\frac{V_n}{\sqrt{n}} \geq \frac{\log 2}{(\tau-\theta)}n^{1/4}\}$, on en déduit que $\Pr(B_n)$ tend vers 0. Par conséquent, la limite inférieure de $v_n^{n^{-3/4}}$ est $\geq 1/4$. Mais $\liminf \frac{1}{n^{3/4}} \log v_n \geq -\log 4$ entraine naturellement que $\liminf \frac{1}{n} \log v_n \geq 0$. Comme $\log v_n \leq 0$ la limite de $\frac{1}{n}\log v_n$ est bien 0 et le théorème des grandes déviations est démontré.

Appendice 2: Convergence des lois binomiales vers la loi de Poisson

Cet appendice montre une chose peu connue: c'est que la suite des lois binomiales de paramètres convenables converge vers une loi de Poisson, non seulement faiblement, mais aussi au sens de la convergence en norme de mesures. Cet appendice peut intéresser aussi les étudiants d'agrégation qui ont à traiter du sujet "lois binomiales, lois de Poisson".

Adoptons les notations suivantes: δ_a désigne la masse de Dirac en a; si $m > 0$, on définit la loi de Poisson de moyenne m par

$$p_m(dx) = \sum_{n=0}^{\infty} e^{-m}\frac{m^n}{n!}\delta_n(dx)$$

et si $0 < p < 1$ on définit la loi de Bernoulli de moyenne p par

$$b_{1,p}(dx) = (1-p)\delta_0(dx) + p\delta_1(dx).$$

Si n est un entier > 0, on définit la loi binomiale $b_{n,p}$ comme la nième puissance de convolution de la loi de Bernoulli:

$$b_{n,p}(dx) = (b_{1,p})^{*n}(dx) = ((1-p)\delta_0 + p\delta_1)^{*n}(dx) = \sum_{k=0}^{n} C_n^k(1-p)^{n-k}p^k\delta_k(dx).$$

C'est un résultat simple et important que de constater que la suite de probabilités $(b_{n,m/n})_{n>m}$ converge faiblement vers p_m. En effet si $n \geq k$, alors

$$C_n^k\left(1-\frac{m}{n}\right)^{-k}\left(\frac{m}{n}\right)^k$$

est une fraction rationnelle en n et l'examen des termes de plus haut degré au numérateur et au dénominateur montre que

$$\lim_{n\to+\infty} C_n^k\left(1-\frac{m}{n}\right)^{n-k}\left(\frac{m}{n}\right)^k = e^{-m}\frac{m^k}{k!}.$$

Toutefois, un résultat plus fort est vrai, puisque en fait $(b_{n,m/n})_{n>m}$ converge fortement vers p_m. S'agissant ici de probabilités concentrées sur l'ensemble $\mathbf{N}$ des entiers, cette convergence forte est une convergence dans $l_1(\mathbf{N})$ et revient à affirmer que

$$\|b_{n,m/n} - p_m\| = \sum_{k=0}^{n}\left|C_n^k\left(1-\frac{m}{n}\right)^{n-k}\left(\frac{m}{n}\right)^k - e^{-m}\frac{m^k}{k!}\right| + \sum_{k=n+1}^{\infty} e^{-m}\frac{m^k}{k!}$$

tend vers 0 si $n \to +\infty$. Nous allons montrer ce résultat de deux manières. Celle de Le Cam(1960) est courte et utilise une ingénieuse idée de couplage. Celle de Prohorov (1963) donne plus d'informations en montrant que $\|b_{n,m/n} - p_m\|$ est équivalente à un $\phi(m)/n$ et calcule explicitement $\phi(m)$.

Théorème 1: (Le Cam) Si $n > m > 0$ on a

$$\|b_{n,m/n} - p_m\| \leq \frac{4m^2}{n}.$$

Démonstration: Posons pour simplifier $p = m/n \in]0,1[$ et considérons des variables aléatoires $(X_1, Y_1), \ldots, (X_n, Y_n)$ de $\mathbf{N}^2$ indépendantes et de même loi m_p définie par

$$\begin{aligned} m_p(0,0) &= e^{-p} - p + pe^{-p} \\ m_p(0,1) &= p - pe^{-p} \\ m_p(1,1) &= pe^{-p} \\ m_p(n,0) &= \frac{p^n}{n!} e^{-p} \text{ si } n \geq 2, \end{aligned}$$

et $m_p(a,b) = 0$ ailleurs. Alors on constate facilement que X_i suit une loi de Poisson et que Y_i suit une loi de Bernoulli, toutes deux de moyenne p. Elles ne sont pas indépendantes, et satisfont à l'inégalité

$$\Pr(X_i = Y_i) = m_p(0,0) + m_p(1,1) = e^{-p} - p + 2pe^{-p} \geq 1 - 2p^2,$$

héritée du fait que $e^{-p} \geq 1 - p$ pour tout réel p. Notons pour simplifier $X = X_1 + \cdots + X_n$, qui est donc de loi de Poisson p_m, et $Y = Y_1 + \cdots + Y_n$, de loi binomiale $b_{n,p}$. Donc

$$\Pr(X \neq Y) \leq \Pr(\cup_{i=1}^n (X_i \neq Y_i) \leq \sum_{i=1}^n \Pr(X_i \neq Y_i) \leq 2np^2.$$

Ensuite, si A est une partie de $\mathbf{N}$, on a l'inclusion d'évènements

$$(X \in A) = (X = Y \in A) \cup (Y \neq X \in A) \subset (Y \in A) \cup (X \neq Y)$$

qui entraine $\Pr(X \in A) - \Pr(Y \in A) \leq P(X \leq Y)$. Le raisonnement fait en échangeant les rôles de X et Y donne finalement

$$|\Pr(X \in A) - \Pr(Y \in A)| \leq 2np^2 = \frac{2m^2}{n}.$$

Pour terminer, on applique cette inégalité à l'ensemble $E = \{k \in \mathbf{N}; p_m(k) \geq b_{n,p}(k)\}$ puis à son complémentaire $E' = \mathbf{N} \setminus E$. Comme

$$\|b_{n,m/n} - p_m\| = (\Pr(X \in E') - \Pr(Y \in E')) + (\Pr(Y \in E) - \Pr(X \in E)) \leq \frac{4m^2}{n},$$

le résultat est montré.

Théorème 2: (Prohorov) Soit X une variable aléatoire de loi de Poisson p_m. Soit

$$\frac{\phi_n(m)}{n} = \|b_{n,m/n} - p_m\|.$$

Alors

$$\lim_{n \to +\infty} \phi_n(m) = \phi(m) = \frac{1}{2} \mathbf{E}(|X - (X-m)^2|).$$

Démonstration: Notons $p_k = e^{-m}\frac{m^k}{k!}$ pour simplifier les notations. Observons d'abord que le résultat n'est pas si surprenant, car pour k fixé, en notant

$$a_k(n) = \frac{n}{p_k}\left(C_n^k\left(1-\frac{m}{n}\right)^{n-k}\left(\frac{m}{n}\right)^k - p_k\right) \quad \text{si} \quad 0 \le k \le n, \tag{6.1}$$

$$a_k(n) = -n \quad \text{si} \quad n < k,$$

alors

$$\lim_{n\to+\infty} a_k(n) = \frac{1}{2}(k-(k-m)^2) \tag{6.2}$$

par un calcul standard et laborieux de développement limité (voir le détail de ce calcul dix lignes ci dessous). On est donc fondé de penser que

$$\lim_{n\to+\infty}\sum_{k=0}^{\infty} p_k|a_k(n)| = \frac{1}{2}\sum_{k=0}^{\infty} p_k|k-(k-m)^2|. \tag{6.3}$$

Le point délicat est alors de justifier cette interversion de limites. On va le faire par convergence dominée. L'idée pour cela est de considérer pour k fixé $\frac{1}{p_k}C_n^k\left(1-\frac{m}{n}\right)^{n-k}\left(\frac{m}{n}\right)^k$ comme une fonction f_k de $1/n$, en introduisant donc

$$f_k(h) = (1-h)(1-2h)\ldots(1-(k-1)h)(1-mh)^{-k}\exp[m+\frac{1}{h}\log(1-mh)].$$

Cette fonction f_k est définie sur $h > -1/m$, et une reformulation de (2) est d'affirmer que $f_k'(0) = \frac{1}{2}(k-(k-m)^2)$; on le voit ainsi:

$$f_k(h) = [1-(1+2+\cdots+k-1)h+o(h)][1+kmh+o(h)]\exp[m+\frac{1}{h}(-mh-m^2h^2/2+o(h^2))]$$

$$= [1-\frac{k(k-1)}{2}h+o(h)][1+kmh+o(h)][1-\frac{m^2}{2}h+o(h)] = 1+\frac{1}{2}(k-(k-m)^2)h+o(h).$$

Fixons désormais $k_0 > m$. Pour $k > k_0$, soit

$$M_k = \max_{0\le h\le \frac{1}{k}} f_k(h).$$

Montrons que $M = \sup_{k>k_0} M_k$ est fini. Pour cela, notons

$$K = \max_{0\le h\le \frac{1}{k_0}} \exp[m+\frac{1}{h}\log(1-mh)],$$

qui existe comme maximum d'une fonction continue sur un compact. Ensuite, si $0 \le h \le \frac{1}{k} \le \frac{1}{k_0} < \frac{1}{m}$, alors $0 \le 1-jh \le 1$ si $j = 1,\ldots,k-1$ et $(1-mh)^{-k} \le (1-\frac{m}{k})^{-k}$. Donc

$$M_k \le K(1-\frac{m}{k})^{-k}.$$

Or $\lim_{k\to+\infty}(1-\frac{m}{k})^{-k} = \exp(-m)$ est finie. Donc la suite $(M_k)_{k>k_0}$ est bornée et M est fini.

Soit maintenant
$$M'_k=\max_{0\leq h\leq \frac{1}{k}}|f'_k(h)|.$$
Montrons que $M'=\sup_{k>k_0}k^{-3}M_k$ est fini. Notons $G(h)=\frac{1}{h}\log(1-mh)$ si $h\neq 0$ et $G(0)=-m$: G est donc continûment dérivable. Soit
$$K'=\max_{0\leq h\leq \frac{1}{k_0}}|G'(h)|.$$
Alors
$$\frac{f'_k(h)}{f_k(h)}=\frac{km}{1-mh}+G'(h)-\sum_{j=1}^{k-1}\frac{j}{1-jh}.$$
Ensuite, si $0\leq h\leq \frac{1}{k}\leq\frac{1}{k_0}<\frac{1}{m}$, alors
$$\left|-\sum_{j=1}^{k-1}\frac{j}{1-jh}\right|\leq\sum_{j=1}^{k-1}\frac{j}{1-\frac{j}{k}}\leq k\sum_{j=1}^{k-1}j=\frac{k^2(k-1)}{2},$$
et
$$\frac{km}{1-mh}\leq\frac{km}{1-\frac{m}{k}}\leq\frac{k^2m}{k_0-m}.$$
Donc
$$\left|\frac{f'_k(h)}{f_k(h)}\right|\leq\frac{k^2m}{k_0-m}+K'+\frac{k^2(k-1)}{2}.$$
Comme dans cet intervalle f_k est dans $]0,M]$, on en déduit
$$k^{-3}M'k\leq\frac{M}{2}(1-1/k)+\frac{Mm}{k(k_0-m)}+\frac{MK'}{k^3}\leq\frac{M}{2}+\frac{Mm}{k_0(k_0-m)}+\frac{MK'}{k_0^3}.$$
M' est donc fini.

On peut alors terminer la démonstration du théorème: on a donc par la formule des accroissements finis pour $0\leq k\leq n$:
$$|a_k(n)|=|n(f_k(1/n)-f_k(0)|=|f'_k(\theta/n)|\leq M'k^3$$
et pour $k>n$ $|a_k(n)|=n\leq k^3$. Soit M'' le maximum de 1 et M'. Observons que $\sum_{n=0}^{\infty}M''k^3p_k$ converge. On est donc dans les conditions d'application du théorème de la convergence dominée et donc (3) est démontré.

Voici quelques raffinements intéressants sur la fonction de Prohorov ϕ du théorème précédent:

Proposition 2: Pour $m > 0$ on note $p_k = e^{-m}\frac{m^k}{k!}$ et $p_{-1} = 0$. Soit r et R les racines du trinôme $P(x) = x - (x-m)^2$, avec $0 < r < R$, et soient a et A les parties entières de r et R. Alors

$$\phi(m) = m^2(p_a - p_{a-1} + p_{A-1} - p_A).$$

De plus, les deux fonctions de m définies par $q(m) = m(p_a - p_{a-1})$ et $Q(m) = m(p_{A-1} - p_A))$ sont continues, positives et tendent vers $\frac{1}{\sqrt{2\pi e}}$ à l'infini.

Démonstration: Remarquons que les racines r et R existent, car le discriminant simplifié de P est $m + 1/4 > 0$, et qu'elles sont > 0 car de somme $2m + 1 > 0$ et de produit $m^2 > 0$. Si X est une variable aléatoire de Poisson de moyenne m, alors $\mathbf{E}(X - (X-m)^2) = 0$. Donc, puisque $P(x)$ est positif si et seulement si $r < x < R$, on a

$$\phi(m) = \frac{1}{2}\mathbf{E}(|P(X|) = \frac{1}{2}\mathbf{E}(|P(X)| + P(X)) = \sum_{r<k<R} p_k P(k) = \sum_{k=a+1}^{A} p_k P(k)$$

$$= m^2 \sum_{k=a+1}^{A} (-p_{k-2} + 2p_{k-1} - p_k) = m^2 \left(-\sum_{k=a-1}^{A-2} p_k + 2\sum_{k=a}^{A-1} p_k - \sum_{k=a+1}^{A} p_k \right)$$

$$= m^2(p_a - p_{a-1} + p_{A-1} - p_A).$$

Définissons la fonction $m \mapsto r(m) = m + \frac{1}{2} - \sqrt{m + \frac{1}{4}}$ sur $[0, +\infty)$. Elle est continue, et sa dérivée $r'(m) > 0$. r est une bijection croissante de $[0, +\infty)$ sur lui même de fonction réciproque $m = r^{-1}(x) = x + \sqrt{x}$. Donc $a \leq r < a + 1$ implique que

$$a + \sqrt{a} = r^{-1}(a) \leq m < r^{-1}(a+1) = a + 1 + \sqrt{a+1}.$$

Donc sur l'intervalle

$$I_n = \{m;\ a = n\} = [n + \sqrt{n}, n + 1 + \sqrt{n+1})$$

la fonction q prend la valeur

$$q(m) = e^{-m}\frac{m^n(m-n)}{n!}.$$

Elle est donc bien positive. Sa limite à l'extrémité droite de I_n est bien $q(n + 1 + \sqrt{n+1})$ ce qui montre sa continuité. Quant à sa limite à l'infini, c'est évident avec la formule de Stirling. La démonstration pour Q est entièrement analogue: sur l'intervalle

$$J_N = \{m;\ A = N\} = [N - \sqrt{N}, N + 1 - \sqrt{N+1}),$$

la fonction Q prend la valeur

$$Q(m) = e^{-m}\frac{m^N(N-m)}{N!}.$$

Sur les intervalles I_n et J_N, les fonctions q et Q sont concaves.

Commentaires: Il y a de nombreuses références sur cette question, (voir la bibliographie ci dessous, qui conduit à d'autres références) mais pas très accessibles un jour de concours d'agrégation. Dans Letac (1981), Problème IV 3 4ème et 5ème, on trouve la démonstration de Le Cam (1960).

Références:

LE CAM, L. "An approximation theorem for the Poisson binomial distribution". *Pacific J. Math. 10,* 1181-1197.

LETAC, G. (1982), *Intégration et Probabilités, Analyse de Fourier, Exercices corrigés.* Masson, Paris. (Seconde édition 1997).

PROHOROV, Ju. V. (1953), " Asymptotic behavior of the binomial distribution" (en russe). *Uspehi Matematiceskih Nauk. 8*, 135-142.

VERVAAT, W. (1969), " Upper bounds for the distance in total variation between the binomial and the Poisson distribution". *Statistica Neerlandica, 23,* 79-86.

JOHNSON, N. J. and SIMMONS, G. (1971), "On the convergence of the binomial to Poisson distribution". *Annals of Math. Statist. 49,* 1735-1736.

Appendice 3: Annales des problèmes de probabilités de Deug et de licence.

Mia 03, Université Paul Sabatier, Devoir 4, à remettre au premier TD de la semaine du 15 au 19 décembre 1997.

Exercice 1

Soit $0 < p < 1$, soit X une v.a. à valeurs dans $\mathbf{N}$ telle que pour tout $x \in \mathbf{N}$ on ait $P(X = x) > 0$ et soit enfin une v.a. Y sur le même espace de probabilité telle que pour tout x et pour tout $y = 0, 1, \ldots, x$ on ait $P(Y = y|X = x) = C_x^y(1-p)^{x-y}p^y$. On pose $Z = X - Y$.

1. On suppose dans cette question que X suit une loi de Poisson de moyenne λ. Montrer qu'alors Y suit une loi de Poisson de moyenne $p\lambda$ (Méthode: on peut ou bien calculer $P(Y = y)$ par le principe des probabilités totales, ou bien calculer la fonction génératrice f_Y). Montrer que Y et Z sont indépendantes (Méthode : calculer $P(Y = y; Z = z)$). Montrer que Z suit une loi de Poisson de moyenne $(1-p)\lambda$.

2. Trouver la loi de X si on sait que Y et Z sont indépendantes. Méthode: si $0 < z < 1$ et $0 < s < 1$, montrer que $\mathbf{E}(s^Y z^Z) = f_X((1-p)z + ps)$, et chercher à quelle condition $g((1-p)z + ps) = \log f_X((1-p)z + ps)$ est la somme d'une fonction de s seul d'une autre fonction de z seul: penser à introduire la dérivée seconde de g.

Exercice 2

Soit N un entier fixé, et soit Ω l'ensemble des parties de taille N de l'ensemble $\{1, 2, \ldots, 2N\}$ des 2N premiers entiers. On le munit de la tribu $\mathcal{P}(\Omega)$ et de la probabilité équiprobable. Si $n = 1, \ldots, 2N$, on note pour $\omega \in \Omega$ $X_n(\omega) = 1$ si $n \in \omega$ et $X_n(\omega) = -1$ sinon. On note $S_0 = 0$ et $S_n = X_1 + \cdots + X_n$.

1. Soit u_N le nombre d'éléments de Ω. Calculer u_N. Quelle est la valeur de S_{2N}? Les variables aléatoires X_n sont elles indépendantes? Sont elles de même loi? Quelle est la loi de X_n?

2. Pour k entier entre 0 et N on note A_k l'évènement $S_{2k} = 0$. Calculer $P(A_k)$. Soit $Y = \sum_{k=0}^{N} \mathbf{1}_{A_k}$ Soit $(v_n)_{n \geq 0}$ le terme général de la série qui est le produit de Cauchy de la série de terme général u_n par elle même. Montrer que $\mathbf{E}(Y) = \frac{v_N}{u_N}$.

3. Calculer les sommes des séries entières suivantes à l'intérieur de leur intervalle de convergence $\sum_{k=0}^{\infty} u_n z^n$, $\sum_{k=0}^{\infty} v_n z^n$ (Méthode: donner une présentation du développement en série entière de $(1-x^2)^{-1/2}$ en termes de (u_n)). En déduire une approximation de $\mathbf{E}(Y)$ si N est grand à l'aide de la formule de Stirling.

4. Dans un jeu de 52 cartes, je tire une carte; avant de la regarder, je devine sa couleur: rouge ou noir, et je marque un point si j'ai deviné juste. Je recommence avec le paquet des 51 cartes restantes, et ainsi de suite jusqu'à épuisement des 52 cartes. Je joue avec la meilleure stratégie, qui choisit la couleur la mieux représentée dans le paquet restant et choisit au hasard en cas d'égalité. Quelle est la moyenne des points

marqués? (Méthode: observer que cette stratégie garantit au moins 26 points, et que les points supplémentaires arrivent une fois sur 2 durant les Y_{26} fois où il y a égalité).

Université Paul Sabatier. Examen blanc de Deug MIA 03, Dec.97
(Durée: 3 heures. Aucun document.)

Exercice 1

Soit X et Y deux variables aléatoires indépendantes et de même loi à valeurs dans les entiers ≥ 1 telles que pour $|z| \leq 1$ leur fonction génératrice soit égale à $f_X(z) = 1 - \sqrt{1-z}$. Calculer f_{X+Y}, $P(X = n)$ pour $n \geq 1$ et $P(X + Y = n)$ pour $n \geq 2$.

$\mathbb{E}(X)$ existe-t-elle?

Exercice 2

Soit X et Y deux variables aléatoires indépendantes à valeurs dans les entiers ≥ 0 telles que pour $|z| \leq 1$ leurs fonctions génératrices soient respectivement $f_X(z) = 2 - \sqrt{2-z}$ et $f_Y(z) = 1/(2-z)$. Pour $n \geq 0$ calculer

$$P(X = n),\ P(Y = n),\ P(X + Y = n).$$

Calculer $\mathbb{E}(X)$, $\mathbb{E}(X(X-1))$, et le second moment et la variance de X. Calculer de même la variance de Y et en déduire la variance de $X + Y$.

Exercice 3

Soit X et Y deux variables aléatoires strictement positives, indépendantes et de même loi. Soit $Z = X/Y$. Montrer que $1 \leq \frac{1}{2}(Z + \frac{1}{Z})$. En prenant l'espérance des deux membres de cette égalité, montrer que $1 \leq \mathbb{E}(X)\mathbb{E}(\frac{1}{X})$.

On suppose maintenant que la fonction de répartition F de X est égale à $(x-1)$ si $1 \leq x \leq 2$. Quelles sont les valeurs de F en dehors de cet intervalle? Calculer $\mathbb{E}(X)$ et $\mathbb{E}(\frac{1}{X})$.

Université Paul Sabatier. Examen de Deug MIA 03, 6 Janv.98

(Durée: 4 heures. Aucun document. Faites les calculs lentement et au brouillon. Le correcteur peut déduire des points pour le manque de soin ou les absurdités. Les trois exercices sont indépendants.) On désigne par $\log x$ le logarithme népérien de x.

Exercice 2 (*6,5 points*)

1. Soit $0 < p \leq q < 1$. Développer en série entière les fonctions de z suivantes
$$\frac{1}{(1-pz)(1-qz)} \quad , \quad \frac{z^2}{(1-pz)(1-qz)},$$
et préciser leurs rayons de convergence (on distiguera les cas $p = q$ et $p < q$). (*2,5 points*)

2. Soit $0 < p \leq 1/2$ et $q = 1 - p$. On considère les variables aléatoires indépendantes et à valeurs dans les entiers ≥ 1 de lois respectives
$$P_X = \sum_{n=1}^{\infty} q^{n-1} p \delta_n, \quad \text{et} \quad P_Y = \sum_{n=1}^{\infty} p^{n-1} q \delta_n,$$
c'est-à-dire que pour $n \geq 1$ entier on a $P[X = n] = q^{n-1}p$ et $P[Y = n] = p^{n-1}q$. Calculer pour $|z| \leq 1$ les fonctions génératrices $f_X(z) = \mathbb{E}(z^X)$, $f_Y(z)$ et $f_{X+Y}(z)$. Développer f_{X+Y} en série entière et préciser son rayon de convergence (on distinguera les cas $p = 1/2$ et $p < 1/2$). Déduire du résultat $P[X + Y = n]$ pour $n \geq 2$ entier. Calculer également la moyenne et la variance des trois variables aléatoires X, Y et $X + Y$. (*3,5 points*)

3. Dans le schéma Succès Echec où la probabilité d'un succès est p, soit Z la variable aléatoire qui prend la valeur n si pour la première fois on a un succès au rang $n - 1$ suivi d'un échec au rang n. Montrer que Z est de même loi que la variable aléatoire $X + Y$ considérée à la question précédente. (*0,5 points*)

Exercice 3 (*5,5 points*)

On rappelle que $\mathrm{sh}x = \frac{1}{2}(e^x - e^{-x})$, $\mathrm{ch}x = \frac{1}{2}(e^x + e^{-x})$, et que pour $x \neq 0$
$$\mathrm{ch}x = \frac{\mathrm{sh}(2x)}{2\,\mathrm{sh}x}.$$

1. Soit U une variable aléatoire de fonction de répartition $F_U(x) = P[U \leq x]$ telle que $F_U(x) = 0$ si $x \leq -1$, $F_U(x) = \frac{1+x}{2}$ si $-1 \leq x \leq 1$ et $F_U(x) = 1$ si $1 \leq x$. Tracer le graphe de F_U ainsi que celui de la densité de U. Calculer ensuite $L_U(z) = \mathbb{E}(e^{zU})$ pour z réel non nul. (*1 point*)

2. Si z est réel non nul, montrer par récurrence sur n que

$$\sum_{k=1}^{n} \log \mathrm{ch}(\frac{z}{2^k}) = \log \frac{\mathrm{sh}(z)}{2^n \mathrm{sh}(\frac{z}{2^n})}.$$

En déduire que la série $\sum_{k=1}^{\infty} \log \mathrm{ch}(\frac{z}{2^k})$ converge, et calculer sa somme en fonction de z. (*2 points*)

3. Soit $X_1, X_2, \ldots$ une suite infinie de variables aléatoires indépendantes et de même loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, c'est-à-dire telles que $P[X_1 = 1] = P[X_1 = -1] = \frac{1}{2}$. On forme la nouvelle variable aléatoire

$$S_n = \sum_{k=1}^{n} \frac{X_k}{2^k}.$$

Calculer à l'aide du 2) pour z réel non nul $L_{S_n}(z) = \mathbb{E}(e^{zS_n})$. Montrer que

$$\sum_{k=1}^{\infty} \frac{X_k}{2^k}$$

est convergente. On désigne par S la somme de cette série. On admet que $L_S(z) = \mathbb{E}(e^{zS}) = \lim_{n\to\infty} L_{S_n}(z)$. A l'aide du 2) et du 1), comparer L_S et L_U. (*2,5 points*)

Corrigé de l'examen de Mia 03 du 6 janvier 1998.

Exercice 2: Si $p = q$, d'après la formule du binôme de Newton on a $(1-pz)^{-2} = \sum_{n=0}^{\infty}(n+1)p^n z^n$ et donc

$$z^2(1-pz)^{-2} = \sum_{n=0}^{\infty}(n+1)p^n z^{n+2} = \sum_{n=2}^{\infty}(n-1)p^{n-2}z^n.$$

Ces deux séries entières convergent si et seulement si $|pz| < 1$ et donc sont de rayon de convergence $R = 1/p$.

Si $p < q$ on décompose la première fraction rationnelle en éléments simples:

$$\frac{1}{(1-pz)(1-qz)} = \frac{1}{q-p}\left(\frac{q}{1-qz} - \frac{p}{1-pz}\right) = \frac{1}{q-p}\left(q\sum_{n=0}^{\infty}q^n z^n - p\sum_{n=0}^{\infty}p^n z^n\right) = \sum_{n=0}^{\infty}\frac{q^{n+1}-p^{n+1}}{q-p}z^n.$$

D'après le théorème sur le rayon de convergence de la somme de deux séries entières, le rayon de convergence est ici le plus petit des deux nombres $1/p$ et $1/q$, soit donc $R = 1/q$. Quant à la deuxième série, il suffit de tout décaler de deux:

$$\frac{z^2}{(1-pz)(1-qz)} = \sum_{n=2}^{\infty}\frac{q^{n-1}-p^{n-1}}{q-p}z^n.$$

2) D'après la définition d'une fonction génératrice on a immédiatement pour ces deux lois de Pascal

$$f_X(z) = \sum_{n=1}^{\infty}q^{n-1}pz^n = \frac{pz}{1-qz},\ f_Y(z) = \sum_{n=1}^{\infty}p^{n-1}qz^n = \frac{qz}{1-pz}.$$

Donc, les variables aléatoires X et Y étant indépendantes, d'après le théorème du cours on a donc

$$f_{X+Y}(z) = f_X(z)f_Y(z) = \frac{pqz^2}{(1-pz)(1-qz)}.$$

Si $p = q = 1/2$ d'après la première partie de 1) on a donc

$$f_{X+Y}(z) = \sum_{n=2}^{\infty}\frac{n-1}{2^n}z^n,$$

avec pour rayon de convergence 2, et $P[X+Y=n] = \frac{n-1}{2^n}$ si $n \geq 2$.

Si $p < 1/2 < q$, d'après la seconde partie de 1) on a donc

$$f_{X+Y}(z) = \sum_{n=2}^{\infty}\frac{pq^n - qp^n}{q-p}z^n,$$

avec pour rayon de convergence $1/q$, et $P[X+Y=n] = \frac{pq^n-qp^n}{q-p}$ si $n \geq 2$.

Enfin, on calcule $f'_X(z) = p/(1-qz)^2$, $f''_X(z) = 2pq/(1-qz)^3$, et on en tire, puisque le rayon de convergence est > 1 :

$$\mathbb{E}(X) = f'_X(1),\ \mathbb{E}(X(X-1)) = f''_X(1) = 2q/p^2, \sigma^2(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 = q/p^2.$$

Puisque p et q jouent des rôles symétriques, $\mathbb{E}(Y) = 1/q$ et $\sigma^2(Y) = p/q^2$. Enfin $\mathbb{E}(X+Y) = 1/p + 1/q$ par la linéarité de l'espérance, et $\sigma^2(X+Y) = q/p^2 + p/q^2$ par l'indépendance de X et Y.

3) Une suite de succès et d'échecs telle que $Z = n \geq 2$ est nécessairement telle que ses n premiers termes $(\omega_1, \ldots, \omega_n)$ soient de la forme $EE...EESS...SSE$, c'est-à-dire que il existe un entier $1 \leq x \leq n-1$ avec $\omega_i = E$ si $i < x$, $w_j = S$ si $x \leq j \leq n-1$ et $\omega_n = E$. Si X est le temps d'attente du premier succès, si Y est le temps d'attente du premier échec après qu'on ait eu le premier succès, alors X et Y sont indépendantes et suivent les lois de Pascal ci dessus, et de plus $Z = X + Y$. D'où le résultat demandé.

Exercice 3: 1) La fonction F est continue et est dérivable sauf aux points 1 et -1 et sa dérivée est $\frac{1}{2}\mathbf{1}_{]-1,1[}(x)$. Les graphes sont immédiats. Ensuite

$$L_U(z) = \frac{1}{2}\int_{-1}^{1} e^{zx}dx = \frac{\mathrm{sh} z}{z}.$$

2) Pour $n = 1$ la formule est triviale. Supposons la vraie à l'ordre n et montrons qu'elle est vraie à l'ordre $n+1$. On a donc par cette hypothèse de récurrence, puis par la formule rappelée appliquée à $x = z/2^n$:

$$\sum_{k=1}^{n+1} \log \mathrm{ch}(\frac{z}{2^k}) = \log \frac{\mathrm{sh}(z)}{2^n \mathrm{sh}(\frac{z}{2^n})} + \log \mathrm{ch}(\frac{z}{2^{n+1}}) = \log \frac{\mathrm{sh}(z)}{2^n \mathrm{sh}(\frac{z}{2^n})} + \log \frac{\mathrm{sh}(\frac{2z}{2^{n+1}}))}{2\,\mathrm{sh}(\frac{z}{2^{n+1}})} = \log \frac{\mathrm{sh}(z)}{2^{n+1}\mathrm{sh}(\frac{z}{2^{n+1}})},$$

et la récurrence est étendue. Ensuite, on sait que le premier terme du développement limité de $\mathrm{sh} x$ est x, donc la limite de $2^n \mathrm{sh}(\frac{z}{2^n})$ quand n tend vers l'infini est z. Les sommes partielles de la série considérée tendent donc vers $\log \frac{\mathrm{sh}(z)}{z}$, c'est dire que la série converge et a pour somme $\log \frac{\mathrm{sh}(z)}{z}$.

3) Puisque les v.a. X_k sont indépendantes, il en est de même pour les v.a. $\exp(zX_k/2^k)$ et on a donc, à l'aide de la première partie du 2):

$$\mathbb{E}(\exp zS_n) = \mathbb{E}(\prod_{k=1}^{n} \exp(zX_k/2^k)) = \prod_{k=1}^{n} \mathbb{E}(\exp(zX_k/2^k)) = \prod_{k=1}^{n} \mathrm{ch}(\frac{z}{2^k}) = \frac{\mathrm{sh}(z)}{2^n \mathrm{sh}(\frac{z}{2^n})}.$$

La série $\sum_{k=1}^{\infty} \frac{X_k}{2^k}$ est absolument convergente car $|X_n| = 1$, que la série géométrique de terme général $1/2^n$ est de raison < 1. La deuxième partie du 2) permet d'affirmer que

$$L_S(z) = \lim_{n\to\infty} L_{S_n}(z) = \frac{\mathrm{sh}(z)}{z} = L_U(z).$$

D'après un théorème admis du cours, on peut remarquer d'ailleurs que cela entraîne que S et U sont de même loi, c'est à dire que F_U est la fonction de répartition de S.

Université Paul Sabatier. Licence de mathématiques fondamentales, contrôle intermédiaire du 13 avril 2000.

Durée: 2heures. Aucun document. Affichage des résultats le 5 mai à 14:00.

Exercice 1. Soit $n \geq 2$, et soit $U_1, U_2, \dots, U_n$ des variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. On note $X = \min(U_1, U_2, \dots, U_n)$ et $Y = \max(U_1, U_2, \dots, U_n)$. On fixe a et b tels que $0 \leq a \leq b \leq 1$.

1. Dessiner l'ensemble du plan
$$E_{a,b} = \{(x,y); 0 \leq a \leq x \leq y \leq b \leq 1\}.$$
2. Calculer $G(a,b) = \Pr(0 \leq a \leq X \leq Y \leq b \leq 1)$.
3. Montrer que $\Pr(X \leq a \ ; \ Y \leq b) = G(0,b) - G(a,b)$, et en déduire la densité de (X,Y).
4. Soit $D = Y - X$. Quelle est la densité de la loi jointe de (X,D)?

Exercice 2. Soit a et b fixés dans $]0,1[$. Soit X et Y des v.a. indépendantes, à valeurs dans l'ensemble $\mathbf{N}$ des entiers ≥ 0 et de lois respectives données par $\Pr(X = x) = (1-a)a^x$ et $\Pr(Y = y) = (1-b)b^y$. Soit $M = \min(X,Y)$ et $D = X - Y$. Pour $m \in \mathbf{N}$, calculer

$$\Pr(X \geq m), \ \Pr(Y \geq m), \ \Pr(M \geq m), \ \Pr(M = m).$$

Pour $m \in \mathbf{N}$ et pour d dans l'ensemble $\mathbf{Z}$ des entiers relatifs, calculer $\Pr(M = m; D = d)$. En déduire $\Pr(D = d)$. Les v.a. M et D sont elles indépendantes?

Exercice 3. Les tables montrent que

$$\int_{-\infty}^{0,5} e^{-\frac{z^2}{2}} \frac{dz}{\sqrt{2\pi}} = 0,6914....$$

Soit X une variable aléatoire normale de moyenne -1 et d'écart-type 2. Calculer les nombres suivants:

$$\Pr(X \geq 0), \ \Pr(-2 \leq X \leq -1), \ \Pr(X \notin [-2,0]).$$

Université Paul Sabatier. NT 07, Licence de mathématiques fondamentales, Examen du 23 juin 2000.

Durée: 2heures. Aucun document. Affichage des résultats le 27 juin à 14:00.

Question de cours. Enoncer sans démonstration la loi forte des grands nombres pour une suite de variables aléatoires réelles indépendantes et de même loi, avec sa réciproque.

Problème. Dans tout le problème, si W est une variable aléatoire (v.a.) réelle, on note par φ_W sa transformée de Fourier, définie pour z réel par $\varphi_W(z) = \mathbb{E}(e^{izW})$.

a) Soit U et V deux v.a. indépendantes et de même loi uniforme sur $[0,1]$, et soit $Z = U - V$. Montrer que la densité de Z est $f_Z(z) = (1-|z|)_+$ (où la notation a_+ signifie $a_+ = 0$ si $a \leq 0$ et $a_+ = a$ si $a \geq 0$). On pourra pour cela considérer la fonction $z \mapsto \Pr(U-V \leq z)$ et sa dérivée.

b) Les notations étant celles du a), calculer $\varphi_Z(t)$ (Méthode: calculer φ_U et en déduire $\varphi_{-V} = \varphi_{-U}$). La formule d'inversion

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\varphi_Z(t)dt$$

est elle applicable?

c) On considère une v.a. X réelle telle que $\varphi_X(z) = (1-|z|)_+$. Donner sa densité à l'aide du b).

d) On considère des v.a. $X_1, \ldots, X_n, \ldots$ indépendantes et de même loi que X, définie au c), et on pose $S_n = X_1 + \cdots + X_n$. Calculer $\varphi_{S_n}(z)$, $\varphi_{S_n/n}(z)$ et $\lim_{n\to\infty} \varphi_{S_n/n}(z)$. En déduire que la suite des lois des S_n/n converge vers une loi limite. A l'aide de la formule d'inversion de Fourier, donner la densité de cette loi limite.

e) La suite S_n/n du d) converge t-elle presque-sûrement?

Barême: Q=3, a=3, b=4, c=2, d=6, e=2.

NL 07, corrigé de l'examen du 23 juin 2000.

Question de cours. Soit $X_1, \dots, X_n, \dots$ des v.a. réelles indépendantes et de même loi. Alors la suite $\frac{1}{n}(X_1+\dots+X_n)$ converge presque-sûrement si et seulement si $\mathbb{E}(|X_1|) < \infty$. Dans ces conditions,

$$\lim_{n\to\infty} \frac{1}{n}(X_1+\dots+X_n) = \mathbb{E}(X_1).$$

Question a). On sait que la fonction de répartition $F = F_U$ est $F(u) = u$ si $0 < u < 1$ et est égale à 1 pour $1 \leq u$ et à 0 pour $u \leq 0$. Donc

$$F_Z(z) = \Pr(U \leq z+V) = \mathbb{E}(F(z+V)) = \int_0^1 F(z+v)dv = \int_z^{z+1} F(u)du,$$

ce qui donne pour $-1 \leq z \leq 0$:

$$F_Z(z) = \int_0^{z+1} udu = \frac{1}{2}(z+1)^2,$$

puis pour $0 \leq z \leq 1$:

$$F_Z(z) = \int_z^1 udu = \frac{1}{2}(1-z^2),$$

et naturellement $F_Z(z) = 0$ si $z \leq -1$ et $F_Z(z) = 1$ si $z \geq 1$. En dérivant on a la densité $F'_Z = f_Z$ annoncée.

Question b). $\varphi_U(t) = \mathbb{E}(\exp itU) = 1$ si $t = 0$ et $(\exp(it) - 1)/t$ sinon. Ensuite $\varphi_{-V}(t) = \varphi_{-U}(t) = \varphi_U(-t)$. Puis que U et V sont indépendantes on en déduit que si $t \neq 0$ on a

$$\varphi_Z(t) = \varphi_U(t)\varphi_{-V}(t) = (\exp(it)-1)(\exp(-it)-1)/t^2 = 2(1-\cos t)/t^2,$$

avec trivialement $\varphi_Z(0) = 1$. En appliquant le critère de Riemann pour les intégrales impropres, on voit que la fonction $t \mapsto \varphi_Z(t)$ est intégrable à l'infini, puisque $|\varphi_Z(t)| = |2(1-\cos t)/t^2| \leq 4t^{-2}$. On est donc dans les conditions d'application de la formule d'inversion de Fourier et la formule de l'énonçé est correcte.

Question c). La formule du b) s'écrit explicitement

$$(1-|z|)_+ = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izt}\frac{2(1-\cos t)}{t^2}dt.$$

Faisons y le changement de variable $x = -t$. On obtient

$$(1-|z|)_+ = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{izx}\frac{1-\cos x}{x^2}dt.$$

Cette formule montre que la densité de X est $f_X(x) = \frac{1-\cos x}{\pi x^2}$.

Question d). Puisque les v.a. sont indépendantes et de même loi que X on a $\varphi_{S_n}(z) = (\varphi_X(z))^n = (1-|z|)_+^n$, et donc $\varphi_{S_n/n}(z) = \varphi_{S_n}(z/n) = (1-\frac{|z|}{n})_+^n$, et donc $\lim_{n\to\infty} \varphi_{S_n/n}(z) =$

$e^{-|z|}$. On remarque $z \mapsto \exp -|z|$ est continue. D'après le théorême de Paul Lévy, cela garantit la convergence en loi de la suite $(S_n/n$. On remarque ensuite que la fonction $z \mapsto \exp -|z|$ est intégrable. D'après la formule d'inversion de Fourier, la densité de la loi limite est donc $\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izx-|z|}dz =$

$$\frac{1}{2\pi}\int_{-\infty}^{0} e^{-izx+z}dz + \frac{1}{2\pi}\int_{0}^{\infty} e^{-izx-z}dz = \frac{1}{2\pi}(\frac{1}{1-ix}+\frac{1}{1+ix}) = \frac{1}{\pi(1+x^2)}.$$

Question e). Si la suite S_n/n convergeait presque-sûrement, elle convergerait en loi, et la loi limite aurait la densité $\frac{1}{\pi(1+x^2)}$. Or la loi forte des grands nombres affirme que si S_n/n converge presque sûrement, ce ne peut être que vers une constante, qui serait d'ailleurs l'espérance de X_1. Il y a donc une contradiction, et donc S_n/n ne converge pas presque-sûrement. On vérifie d'ailleurs directement que X_1 ne satisfait pas $\mathbb{E}(|X_1|) < \infty$ et donc que $\mathbb{E}(X_1)$ n'existe pas. En effet, on connait la densité de X_1 par la question c) et il est clair que

$$\int_{-\infty}^{\infty} |x|\frac{1-\cos x}{\pi x^2}dx = 2\int_{0}^{\infty} \frac{1-\cos x}{\pi x}dx = \infty.$$

Université Paul Sabatier. NL 12, Licence de mathématiques pour l'enseignement, Examen du 23 juin 2000.

Durée: 2heures. Aucun document. Affichage des résultats le 27 juin à 14:00.

Question de cours. Soit X et Y deux variables aléatoires définies sur un même espace de probabilité et possédant des variances finies non nulles. Donner la définition du coefficient de corrélation $r(X,Y)$ de (X,Y). Si (X,Y) est de loi normale dans $\mathbb{R}^2$, expliquer pourquoi X et Y sont indépendantes lorsque $r(X,Y)=0$.

Problème. On admet la formule suivante: pour $p>0$ et $t>0$ on a

$$(*) \qquad \int_0^\infty u^{-3/2} e^{-\frac{p^2}{2u}-tu} du = \frac{\sqrt{2\pi}}{p} e^{-p\sqrt{2t}}.$$

a) Si $p>0$ et $\theta>0$, soit U une variable aléatoire (v.a) de loi

$$P_\theta^{(p)}(du) = \frac{p}{\sqrt{2\pi}} u^{-3/2} e^{-\frac{p^2}{2u}-\frac{\theta u}{2}+p\sqrt{\theta}} \mathbf{1}_{]0,\infty[}(u) du.$$

A l'aide de (*), montrer que si $z<\theta/2$ alors

$$\mathbb{E}(e^{zU}) = \exp(p\sqrt{\theta} - p\sqrt{\theta-2z}).$$

En déduire $\mathbb{E}(Ue^{zU})$ par dérivation ainsi que $\mathbb{E}(U)$. Calculer de même $\mathbb{E}(U^2e^{zU})$, $\mathbb{E}(U^2)$ et la variance de U.

b) Soit de plus $q>0$ et V une v.a. indépendante de U et de loi $P_\theta^{(q)}$. A l'aide du a), montrer que $U+V$ est de loi $P_\theta^{(p+q)}$.

c) Soit $X_1,\dots X_n$ des v.a. indépendantes et même loi $P_\theta^{(1)}$, où θ est un paramètre positif inconnu. Donner à l'aide du b) la loi de $S=X_1+\cdots+X_n$.

d) On considère alors $(X_1,\dots,X_n)$ comme un échantillon permettant d'estimer θ et on rappelle que, le modèle étant exponentiel, S est donc une statistique exhaustive. Montrer à l'aide du a) que S/n est un estimateur non biaisé de $g(\theta)=1/\sqrt{\theta}$ et donner son risque quadratique. Connaissant S, calculer l'estimateur $\hat{\theta}_0(S)$ du maximum de vraisemblance pour θ.

e) Les tables montrent que $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{1}\exp(-\frac{z^2}{2})dz = 0,8413$. En déduire une valeur approchée pour n grand de $\Pr(S \geq \frac{n}{\sqrt{\theta}} + \frac{\sqrt{n}}{(\sqrt{\theta})^{3/2}})$, en justifiant votre réponse.

Barême: Q=4, a=2+2+2, b=1, c=1, d=1+1+3, e=3.

NL 12, corrigé de l'examen du 23 juin 2000.

Question de cours. Si $a = \mathbb{E}(X)$ et $b = \mathbb{E}(Y)$ notons $\mathrm{cov}(X,Y) = \mathbb{E}((X-a)(Y-b))$, $\sigma(X) = (\mathbb{E}((X-a)^2))^{1/2}$ et $\sigma(Y) = (\mathbb{E}((Y-b)^2))^{1/2}$. Alors le coefficient de corrélation est

$$r(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Si de plus (X,Y) est de loi normale $N_{(a,b),\Sigma}$, et si $r(X,Y) = 0$, alors $\mathrm{cov}(X,Y) = 0$ et donc

$$\Sigma = \begin{bmatrix} \sigma^2(X) & 0 \\ 0 & \sigma^2(Y) \end{bmatrix}.$$

Cela entraîne que X et Y sont indépendantes. Cela peut se voir de deux manières:

- Ou bien en considérant la densité de (X,Y) qui en général quand $\det \Sigma \neq 0$ est

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp -\frac{1}{2}(x-a, y-b)\Sigma^{-1} \begin{pmatrix} x-a \\ y-b \end{pmatrix}.$$

  Dans le cas particulier qui nous occupe, on a alors

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma(X)\sigma(Y)} e^{-\frac{(x-a)^2}{2\sigma^2(X)} - \frac{(y-b)^2}{2\sigma^2(Y)}} \\ &= \frac{1}{\sqrt{2\pi}\sigma(X)} e^{-\frac{(x-a)^2}{2\sigma^2(X)}} \times \frac{1}{\sqrt{2\pi}\sigma(Y)} e^{-\frac{(y-b)^2}{2\sigma^2(Y)}} \end{aligned}$$

  La densité $f_{X,Y}(x,y)$ étant le produit d'une fonction de x seul et de y seul est donc la densité d'un couple de v.a. indépendantes.

- Ou bien en considérant la transformée de Laplace ou de Fourier de (X,Y). Procédons par exemple avec la transformée de Fourier. Elle est en général

$$\varphi_{X,Y}(t,s) = \exp(iat + ibs - \frac{1}{2}(t,s)\Sigma \begin{pmatrix} t \\ s \end{pmatrix}).$$

  Dans le cas particulier qui nous occupe, on a alors

$$\begin{aligned} \varphi_{X,Y}(t,s) &= \exp(iat + ibs + \frac{1}{2}(\sigma^2(X)t^2 + \sigma^2(Y)s^2)) \\ &= \exp(iat - \frac{1}{2}\sigma^2(X)t^2) \times \exp(ibs - \frac{1}{2}\sigma^2(Y)s^2) \end{aligned}$$

  La transformée de Fourier $\varphi_{X,Y}(t,s)$ étant le produit d'une fonction de t seul et de s seul est donc la transformée de Fourier d'un couple de v.a. indépendantes.

Question a). Pour calculer $\mathbb{E}(\exp(zU) = \int e^{zu}P_\theta^{(p)}(du)$, il suffit de prendre $t = \frac{\theta}{2} - z$ dans la formule (*) pour avoir le résultat demandé. On remarque que en faisant $z = 0$ on obtient 1, ce qui prouve que $P_\theta^{(p)}(du)$ est bien une loi de probabilité. On remarque que pour tout $t > 0$ et pour tout entier $n \geq 0$ l'intégrale

$$\int_0^\infty u^n u^{-3/2} e^{-\frac{p^2}{2u} - tu} du$$

converge, ce qui entraîne que la fonction définie par l'intégrale $z \mapsto \int e^{zu}P_\theta^{(p)}(du)$ est indéfiniment dérivable sous le signe somme dans l'intervalle $]-\infty, \frac{\theta}{2}[$. Donc

$$\begin{aligned}
\mathbb{E}(U\exp(zU)) &= \frac{d}{dz}\exp(p\sqrt{\theta} - p\sqrt{\theta - 2z}) \\
&= \frac{p}{\sqrt{\theta - 2z}}\exp(p\sqrt{\theta} - p\sqrt{\theta - 2z})
\end{aligned}$$

et donc en faisant $z = 0$ on obtient $\mathbb{E}(U) = \frac{p}{\sqrt{\theta}}$. De la même manière:

$$\begin{aligned}
\mathbb{E}(U^2\exp(zU)) &= \frac{d^2}{dz^2}\exp(p\sqrt{\theta} - p\sqrt{\theta - 2z}) \\
&= (\frac{p^2}{\theta - 2z} + \frac{p}{(\sqrt{\theta - 2z})^3})\exp(p\sqrt{\theta} - p\sqrt{\theta - 2z}).
\end{aligned}$$

Donc en faisant $z = 0$, on obtient $\mathbb{E}(U^2) = (\frac{p^2}{\theta} + \frac{p}{(\sqrt{\theta})^3})$ et en utilisant la formule d'Huyghens on obtient la variance de U :

$$\sigma^2(U) = \mathbb{E}(U^2) - (\mathbb{E}(U))^2 = \frac{p}{(\sqrt{\theta})^3}.$$

Questions b) et c). Puisque U et V sont indépendantes, pour $z < \theta/2$ on a

$$\begin{aligned}
\mathbb{E}(\exp(z(U + V)) &= \mathbb{E}(\exp(zU))\mathbb{E}(\exp(zV)) \\
&= \exp(p\sqrt{\theta} - p\sqrt{\theta - 2z})\exp(q\sqrt{\theta} - q\sqrt{\theta - 2z}) \\
&= \exp((p + q)(\sqrt{\theta} - \sqrt{\theta - 2z}))
\end{aligned}$$

Comme la transformée de Laplace caractérise la loi, cela montre que celle de $U + V$ est $P_\theta^{(p+q)}$. Il est clair alors que si n est un entier >0, et que si les $(X_i)_{1\leq i\leq n}$ sont indépendantes et de même loi $P_\theta^{(1)}$ alors $S = X_1 + \cdots + X_p$ est de loi $P_\theta^{(n)}$. Cela peut se voir très rigoureusement par une récurrence facile sur n.

Question d). D'après le a) on sait que $\mathbb{E}(S) = n/\sqrt{\theta}$. Donc $\mathbb{E}(S/n) = 1/\sqrt{\theta}$. Comme cet estimateur de $1/\sqrt{\theta}$ est non biaisé, son risque quadratique est égal à sa variance. Or la variance de S a été calculée en a) et est $n/(\sqrt{\theta})^3$. Comme en général $\sigma^2(\lambda X) = \lambda^2\sigma^2(X)$, le risque quadratique est donc

$$\sigma^2(S/n) = \frac{1}{n(\sqrt{\theta})^3}.$$

Pour calculer le maximum de vraisemblance connaissant S, rappelons que la loi de S est

$$P_\theta^{(n)}(ds) = \frac{n}{\sqrt{2\pi}} s^{-3/2} e^{-\frac{n^2}{2s} - \frac{\theta s}{2} + n\sqrt{\theta}} \mathbf{1}_{]0,\infty[}(s) ds.$$

C'est dire que si on prend pour mesure de référence la mesure

$$\nu(ds) = \frac{n}{\sqrt{2\pi}} s^{-3/2} e^{-\frac{n^2}{2s}} \mathbf{1}_{]0,\infty[}(s) ds,$$

qui ne dépend pas de θ, alors $P_\theta^{(n)}(ds) = e^{-\frac{\theta s}{2} + n\sqrt{\theta}} \nu(ds)$. Le calcul du maximum de vraisemblance $\hat{\theta}_0(S)$ pour S connu se réduit à la recherche du θ qui maximise sur $]0, +\infty[$ la fonction

$$\theta \mapsto e^{-\frac{\theta S}{2} + n\sqrt{\theta}},$$

ou encore la fonction

$$\theta \mapsto l_S(\theta) = -\frac{\theta S}{2} + n\sqrt{\theta}.$$

L'étude des variations de l_S sur $]0, +\infty[$ est facile: sa dérivée est $l_S'(\theta) = \frac{1}{2}(\frac{n}{\sqrt{\theta}} - S)$ et s'annule seulement en $\theta = n^2/S^2$. Cette dérivée est >0 avant et <0 après. L'estimateur du maximum de vraisemblance est donc $\hat{\theta}_0(S) = n^2/S^2$.

Question e). Notons plutôt $S = S_n$. Le théorème central limite affirme que la suite des lois des v.a.

$$\frac{(S_n - \frac{n}{\sqrt{\theta}})}{\frac{\sqrt{n}}{(\sqrt{\theta})^{3/2}}}$$

converge vers $N_{0,1}$. Donc la probabilité pour que cette v.a. soit ≥ 1 est approximativement $1 - 0,8413$.

Université Paul Sabatier. NT 07, Licence de mathématiques fondamentales, Examen du 13 septembre 2000.

Durée: 2heures. Aucun document. Affichage des résultats le 20 septembre à 14:00.

Problème.

A) Soit $Y_1, \dots, Y_k, \dots$ une suite de variables aléatoires (v.a.) réelles indépendantes et de même loi telle que $\mathbb{E}(|Y_1|) < \infty$. On pose $S_k = Y_1 + \dots + Y_k$. On rappelle que la loi des grands nombres affirme que $\lim_{k\to\infty} S_k/k = \mathbb{E}(Y_1)$ presque sûrement. En déduire que $\lim_{k\to\infty} Y_k/k = 0$ presque sûrement.

B) Dans toute la suite, on considère une suite $(A_n, B_n)_{n\geq 1}$ de v.a. de $]0,\infty[^2$ indépendantes et de même loi (les v.a. A_1 et B_1 peuvent être dépendantes entre elles). On suppose de plus que

$$\mathbb{E}(|\log A_1|) < \infty,\ \mathbb{E}(\log A_1) < 0,\ \mathbb{E}(|\log B_1|) < \infty.$$

En appliquant la loi des grands nombres, montrer que

$$\lim_{n\to\infty} (A_1 A_2 \dots A_n)^{1/n} = \exp(\mathbb{E}(\log A_1)),$$

et que $\lim_{n\to\infty}(A_1A_2\dots A_n) = 0$. En appliquant la question A), montrer que $\lim_{k\to\infty}(B_k)^{1/k} = 1$. De ces résultats, déduire en particulier que la série à termes aléatoires

$$B_1 + \sum_{k=2}^{\infty} A_1 A_2 \dots A_{k-1} B_k$$

est convergente (méthode: lui appliquer le critère de Cauchy $u_n^{1/n}$ de convergence des séries à termes positifs).

C) Pour $n \geq 1$, on considère les transformations F_n, Z_n et W_n affines aléatoires de $\mathbb{R}$ définies par $F_n(x) = A_n x + B_n$, par $Z_n = F_1 \circ F_2 \circ \dots \circ F_n$ et par $W_n = F_n \circ F_{n-1} \circ \dots \circ F_1$. Montrer par récurrence sur n que

$$Z_n(x) = A_1 A_2 \dots A_n x + B_1 + \sum_{k=2}^{n} A_1 A_2 \dots A_{k-1} B_k.$$

Si X est une variable aléatoire positive indépendante des $(A_n, B_n)_{n\geq 1}$, dire pourquoi les v.a. $W_n(X)$ et $Z_n(X)$ sont de même loi.

D) Montrer à l'aide de la question B) que la suite de v.a. $(Z_n(X))_{n\geq 1}$ converge presque sûrement vers une v.a. qu'on note Z. Pourquoi la v.a Z est elle la même quelle que soit la v.a. X? Pourquoi la suite de v.a. $(Z_n(X))_{n\geq 1}$ converge t-elle en loi? A l'aide de la question C), montrer que la suite des lois des v.a. $(W_n(X))_{n\geq 1}$ converge vers la loi de Z.

E) On suppose maintenant de plus que la loi de X est telle que X et $F_1(X) = A_1 X + B_1$ sont de même loi. En déduire qu'alors X et Z sont de même loi (méthode: montrer par récurrence sur n que la v.a. $W_n(X)$ est de même loi que X, et appliquer la question D)).

F) *(Exemple)* On rappelle (cours) que si $\alpha > 0$ et $\beta > 0$ alors

$$\int_0^{+\infty} \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{6.4}$$

On fixe deux nombres $p > 0$ et $q > 0$ et on suppose que la v.a. X a pour loi

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{x^{p-1}}{(1+x)^{p+q}} \mathbf{1}_{]0,\infty[}(x)dx,$$

que la v.a. A_1 a pour loi

$$\frac{\Gamma(2p+q)}{\Gamma(p)\Gamma(p+q)} \frac{a^{p-1}}{(1+a)^{2p+q}} \mathbf{1}_{]0,\infty[}(a)da,$$

et qu'enfin $B_1 = A_1$. On suppose toujours X et A_1 indépendantes. En appliquant (6.4) à des α et β convenables, montrer que si $-p < s < q$ on a

$$\mathbb{E}(A_1^s) = \frac{\Gamma(p+s)\Gamma(p+q-s)}{\Gamma(p)\Gamma(p+q)},\ \mathbb{E}((1+X)^s) = \frac{\Gamma(p+q)\Gamma(q-s)}{\Gamma(p+q-s)\Gamma(q)},\ \mathbb{E}(A_1^s(1+X)^s) = \mathbb{E}(X^s).$$

En déduire que $F_1(X) = A_1(X+1)$ et X sont de même loi. On admet alors sans démonstration que $\mathbb{E}(\log A_1)$ existe et est < 0. En appliquant tout ce qui précède, donner la loi de

$$Z = \sum_{k=1}^{\infty} A_1 A_2 \dots A_k.$$

En particulier, si $q > 0$ et si $\Pr(A_1 \geq a) = \frac{1}{(1+a)^{1+q}}$ pour tout $a \geq 0$, calculer $\Pr(Z \geq z)$ pour tout $z \geq 0$ (méthode: calculer la densité de A_1, puis celle de Z).

Barême: A=2 points, B=4, C=3, D=3, E=3, F=7.

Université Paul Sabatier. NT 07, Licence de mathématiques fondamentales, Examen du 13 septembre 2000, Corrigé.

A)

$$\frac{Y_k}{k} = \frac{S_k}{k} - \frac{k-1}{k}\frac{S_{k-1}}{k-1} \to_{k\to\infty} \mathbb{E}(Y_1) - 1.\mathbb{E}(Y_1) = 0.$$

B) On applique la loi des grands nombres à la suite $Y_k = \log A_k$ et on obtient

$$(A_1 \dots A_n)^{1/n} = \exp(\frac{1}{n}(\log A_1 + \cdots + \log A_n)) \to_{n\to\infty} \exp(\mathbb{E}(\log A_1)).$$

De plus, on sait que $\mathbb{E}(\log A_1) < 0$ et donc que $\exp(\mathbb{E}(\log A_1)) < 1$. Si $u_n = A_1 \dots A_n$, puisque $\lim_{n\to\infty} u_n^{1/n} < 1$ la série $\sum_{n=1}^{\infty} u_n$ converge, son terme général tend donc vers 0 et on a bien $\lim_{n\to\infty} A_1 \dots A_n = 0$ presque sûrement. De la même facon, on applique le A) à $Y_k = \log B_k$ et on en tire que $\lim_{k\to\infty} \frac{1}{k}\log B_k = 0$ et donc $\lim_{k\to\infty}(B_k)^{1/k} = 1$. Considérons enfin $u_k = A_1 A_2 \dots A_{k-1} B_k$. Alors, d'après les résultats précédents on a

$$u_k^{1/k} = ((A_1 A_2 \dots A_{k-1})^{1/k-1})^{(k-1)/k}(B_k)^{1/k} \to_{k\to\infty} \exp(\mathbb{E}(\log A_1)).1 < 1,$$

et d'après le critère de Cauchy de convergence des séries on a que $\sum_{k=1}^{\infty} u_k$ converge presque sûrement, ce qu'il fallait démontrer.

C) La formule est vraie trivialement pour $n = 1$. Supposons la vraie pour $n \geq 1$. Alors

$$\begin{aligned}
Z_{n+1}(x) &\overset{(1)}{=} Z_n(F_{n+1}(x)) \\
&\overset{(2)}{=} A_1 A_2 \dots A_n(A_{n+1}x + B_{n+1}) + B_1 + \sum_{k=2}^{n} A_1 A_2 \dots A_{k-1} B_k \\
&\overset{(3)}{=} A_1 A_2 \dots A_n A_{n+1} x + B_1 + \sum_{k=2}^{n+1} A_1 A_2 \dots A_{k-1} B_k,
\end{aligned}$$

où (1) vient de la définition de $Z_{n+1}(x)$, (2) de l'hypothèse de récurrence et (3) d'un réarrangement. La récurrence est donc étendue.

Puisque les (A_n, B_n) sont de même loi, il est clair que les fonctions affines Z_n et W_n sont de même loi. Leur évaluation en une v.a. X indépendante des (A_n, B_n), et donc indépendante de Z_n et W_n sont donc des v.a. de même loi.

D) On a vu au B) que la série $B_1 + \sum_{k=2}^{\infty} A_1 A_2 \dots A_{k-1} B_k$ converge. Notons par Z sa somme. On a également vu au B) que $\lim_{n\to\infty} A_1 \dots A_n = 0$ presque sûrement: cela entraîne que $\lim_{n\to\infty} Z_n(X) = Z$ presque sûrement. Par définition, Z ne dépend pas de X. La convergence presque sûre entraînant la convergence en loi, on en déduit que la suite des lois de $Z_n(X)$ converge vers la loi de Z. On a vu à la question C) que $W_n(X)$ et $Z_n(X)$ sont de même loi. On en déduit que la suite des lois de $W_n(X)$ converge vers la loi de Z.

E) Montrons par récurrence sur n que $W_n(X)$ et X sont de même loi. C'est vrai par hypothèse pour $n = 1$. Supposons ce résultat vrai pour n. On sait que $W_{n+1}(X) =$

$F_{n+1}(W_n(X))$. De plus $W_n(X)$ est indépendante de F_{n+1}, car $W_n(X)$ est une fonction de $X, A_1, \dots, B_n$ et F_{n+1} dépend de (A_{n+1}, B_{n+1}). Enfin F_{n+1} est de même loi que F_1 par définition, et $W_n(X)$ est de même loi que X, par hypothèse de récurrence. Donc $F_{n+1}(W_n(X))$ est de même loi que $F_1(X)$, qui est de même loi que X par hypothèse. La récurrence est donc étendue.

Or on sait d'après D) que la suite des lois de $W_n(X)$ converge vers la loi de Z. Comme toutes les lois des $W_n(X)$ sont identiques à celle de X, on en déduit que X et Z sont de même loi.

F) Par définition, puis en appliquant (6.4) à $\alpha = s + p > 0$ et $\beta = 2p + q - (s + p) = p + q - s > 0$, on a

$$\mathbb{E}(A_1^s) = \frac{\Gamma(2p+q)}{\Gamma(p)\Gamma(p+q)} \int_0^\infty \frac{a^{s+p-1}}{(1+a)^{2p+q}} da = \frac{\Gamma(p+s)\Gamma(p+q-s)}{\Gamma(p)\Gamma(p+q)}.$$

De même en appliquant (6.4) à $\alpha = p > 0$ et $\beta = q - s > 0$, on a

$$\mathbb{E}((1+X)^s) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^\infty \frac{x^{p-1}}{(1+x)^{p+q-s}} dx = \frac{\Gamma(p+q)\Gamma(q-s)}{\Gamma(p+q-s)\Gamma(q)}.$$

Finalement, en appliquant (6.4) à $\alpha = p + s > 0$ et $\beta = q - s > 0$, on a

$$\mathbb{E}(X^s) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^\infty \frac{x^{s+p-1}}{(1+x)^{p+q}} dx = \frac{\Gamma(p+s)\Gamma(q-s)}{\Gamma(p)\Gamma(q)}.$$

On voit donc que $\mathbb{E}(A_1^s)\mathbb{E}((1+X)^s) = \mathbb{E}(X^s)$. Or A_1 et X sont indépendantes, et donc A_1^s et $(1+X)^s$ le sont également. Comme l'espérance d'un produit de v.a. indépendantes est le produit des espérances, on en déduit que $\mathbb{E}(A_1^s(1+X)^s) = \mathbb{E}(X^s)$. Ce résultat dit que les v.a. $\log(A_1(1+X))$ et $\log X$ ont la même transformée de Laplace. D'après le cours, elles sont donc de même loi, et donc leurs exponentielles $F_1(X) = A_1(1+X)$ et X sont de même loi. On est donc dans les conditions de la question E) et on en déduit que Z et X sont de même loi. Puisqu'ici $A_n = B_n$, alors $Z = \sum_{k=1}^\infty A_1 A_2 \dots A_k$.

Si $\Pr(A_1 \geq a) = \frac{1}{(1+a)^{1+q}}$ pour tout $a \geq 0$, alors la fonction de répartition de A_1 est 0 si $a < 0$ et $1 - \frac{1}{(1+a)^{1+q}}$ si $a \geq 0$. En dérivant par rapport à a on voit que la densité de A_1 est $\frac{1+q}{(1+a)^{2+q}} \mathbf{1}_{]0,\infty[}(a)$, c'est à dire du type de l'exemple avec $p = 1$. La densité de Z est donc $\frac{q}{(1+z)^{1+q}} \mathbf{1}_{]0,\infty[}(z)$, et donc pour tout $z \geq 0$: $\Pr(Z \geq z) = \int_z^\infty \frac{q}{(1+x)^{1+q}} dx = \frac{1}{(1+z)^q}$.

Complétons ce corrigé en montrant le point admis $\mathbb{E}(\log A_1) < 0$. Puisque $\frac{d}{ds}\mathbb{E}(A_1^s) = \mathbb{E}(A_1^s \log A_1)$ on a

$$\mathbb{E}(\log A_1) = \left[\frac{d}{ds} \log \mathbb{E}(A_1^s)\right]_{s=0} = \frac{\Gamma'(p)}{\Gamma(p)} - \frac{\Gamma'(p+q)}{\Gamma(p+q)}.$$

Le problème est donc de montrer que $t \mapsto \frac{\Gamma'(t)}{\Gamma(t)}$ est croissante sur $]0, \infty[$, ou de montrer que $t \mapsto \log \Gamma(t)$ est strictement convexe. Cela vient de

$$(\log \Gamma(t))'' = \frac{1}{\Gamma(t)} \int_0^\infty [\log x - \frac{\Gamma'(t)}{\Gamma(t)}]^2 x^{t-1} e^{-x} dx > 0.$$

Université Paul Sabatier. NT 12, Licence de mathématiques pour l'enseignement, Examen du 13 septembre 2000.

Durée: 2heures. Aucun document. Affichage des résultats le 20 septembre à 14:00.

Question de cours. Soit μ et $(\mu_n)_{n\geq 1}$ des probabilités sur $\mathbb{R}$. On rappelle que on dit que la suite $(\mu_n)_{n\geq 1}$ converge en loi vers μ si pour toute fonction continue bornée réelle ou complexe sur $\mathbb{R}$ on a

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f(x)\mu_n(dx) = \int_{-\infty}^{\infty} f(x)\mu_(dx).$$

(1) Donner sans démonstration, en termes de transformées de Fourier, une condition nécessaire et suffisante de convergence en loi (théorème de Paul Lévy).

(2) Donner sans démonstration, en termes de fonctions de répartition, une condition nécessaire et suffisante de convergence en loi.

Problème. On rappelle que pour $|t| < 1$:

$$1+\sum_{n=1}^{\infty}\frac{\lambda(\lambda+1)\dots(\lambda+n-1)}{n!}t^n = \frac{1}{(1-t)^{\lambda}}.$$

Soit X une variable aléatoire (v.a.) qui suit une loi négative binomiale $NB_{\lambda,p}$, avec $\lambda > 0$ et $0 < p = 1-q < 1$, donc de loi concentrée sur l'ensemble $\mathbf{N}$ des entiers ≥ 0 définie par

$$p^{\lambda}\delta_0 + \sum_{n=1}^{\infty}\frac{\lambda(\lambda+1)\dots(\lambda+n-1)}{n!}p^{\lambda}q^n\delta_n.$$

A. Calculer la fonction génératrice de X, c'est à dire $\mathbb{E}(z^X)$ avec $|z| \leq 1$. Déduire du résultat $\mathbb{E}(X)$, $\mathbb{E}(X^2 - X)$, $\mathbb{E}(X^2)$ et la variance $\sigma^2(X)$ de X. Montrer que $\mathbb{E}(X) < \sigma^2(X)$.

B. Soient $X_1, \dots, X_n$ des v.a. indépendantes et de même loi que X. On pose $\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$. Quelle est la loi de $n\overline{X}_n$? (méthode: considérer sa fonction génératrice et utiliser le A). Calculer $\mathbb{E}(\overline{X}_n)$ et $\mathbb{E}(\overline{X}_n^2)$.

C. Avec les hypothèses de B) on définit la v.a. $S_n \geq 0$ par

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2.$$

Démontrer que $S_n^2 = -\frac{n}{n-1}\overline{X}_n^2 + \frac{1}{n-1}\sum_{k=1}^{n} X_k^2$. En déduire à l'aide du A et du B que la valeur de $\mathbb{E}(S_n^2)$ est la variance de X.

D. Avec les hypothèses de C) on fait tendre n vers l'infini. A l'aide de la loi des grands nombres, calculer les limites presque sures de $\overline{X}_n$, de $\frac{1}{n-1}\sum_{k=1}^{n} X_k^2$, de S_n^2 et de $S_n^2 - \overline{X}_n$ et vérifier que la dernière est >0.

D. On suppose maintenant que λ et p sont des paramètres inconnus, que la suite $(X_1, \ldots, X_n)$ est un échantillon à partir duquel on veut estimer λ et p par la méthode dite des moments. Si $\overline{X}_n$ et S_n^2 sont considérés comme des estimateurs de la moyenne et de la variance de X respectivement, dire si ces estimateurs sont biaisés ou non. On définit les estimateurs $\hat{\lambda}_n$ et $\hat{p}_n = 1 - \hat{q}_n$ par les équations

$$\frac{\hat{\lambda}_n \hat{q}_n}{\hat{p}_n} = \overline{X}_n, \quad \frac{\hat{\lambda}_n \hat{q}_n^2}{\hat{p}_n^2} + \frac{\hat{\lambda}_n \hat{q}_n}{\hat{p}_n} = S_n^2.$$

Montrer $\hat{\lambda}_n > 0$ et $0 < \hat{p}_n < 1$ si et seulement si $S_n^2 - \overline{X}_n > 0$. Au vu de la question D, cette méthode est elle raisonnable?

Barême: QC= 2+2 points, A=1+1+0,5+1+0,5, B=2+1+1, C=1+1, D=1+1+1+0,5+0,5, E=1+1.