# ProbaStats - Compléments
# Introduction

Salim Najib

Printemps 2022

## Introduction

Bienvenue à ces compléments de probastats ! J'ai décidé de (tenter de) les faire car, pour bon nombre de notions abordées dans le cours, j'ai réalisé que si l'on m'avait expliqué un point d'une certaine manière, j'aurais gagné beaucoup de temps. Ainsi, je partage ces explications sous la forme de compléments de cours et d'exercices additionnels.
**Ces compléments ne visent pas à remplacer le cours.** Je ne commente que les points où je juge que j'ai quelque chose à rajouter par rapport au cours. Ainsi, il se pourrait que certaines fiches de compléments soient plutôt courtes.
Les exercices se veulent conceptuels et/ou mathématiquement pas forcément aisés, mais d'un niveau plausible à l'examen (voire d'un niveau normal pour le cours de modèles stochastiques de BA5).

Nombre des exos supplémentaires seront pris du livre *Introduction à la théorie des probabilités* de Robert C. Dalang et Daniel Conus, certains pourraient être pris du cours de *Modèles stochastiques pour les communications* enseigné par Patrick Thiran, en BA5.

N'hésitez pas à me contacter si vous trouvez une erreur, si petite qu'elle soit, ou si vous avez des questions. Les probas c'est un sujet riche et portant à confusion (et aux erreurs de calcul...).

Vous souhaitant un excellent semestre,
Salim

# ProbaStats - Compléments
# Bases de probabilité, probabilité conditionnelle et combinatoire

Salim Najib

Printemps 2022

## Introduction

Les attentes en combinatoire dans le cours de ProbaStats sont plus basses que celles d'AICCI, je trouve, dans le sens où les problèmes qui vous seront posés ce semestre sont plus faciles que ceux d'AICCI. Cependant, ces attentes sont plus systématiques : il faudra maîtriser ces bases assez attentivement.

Concernant les axiomes de probabilité, pour une compréhension profonde de la suite, je ne peux pas suffisamment mettre en exergue l'importance de la compréhension voire la mémorisation des résultats fondamentaux de probabilité, i.e les définitions précises d'un espace de probabilité $(\Omega, \mathcal{F}, \mathrm{P})$ et encore plus les propriétés de la mesure de probabilité P qui seront revues et motivées en partie ci-dessous.

Idem, la probabilité conditionnelle est un sujet fondamental pour la suite.

## Sommaire

# 1 Bases de probabilité

## 1.1 Motivation des axiomes

Cette motivation passera par de l'intuition sur la théorie de la mesure.

1. L'univers / sample space $\Omega$ est un ensemble, rien de plus, rien de moins. Nous nous donnons pour but de *mesurer* des sous-ensembles de $\Omega$.

2. La tribu ou sigma-algèbre $\mathcal{F} \subseteq 2^{\Omega}$ et la mesure ou distribution de probabilité $\mathrm{P} : \mathcal{F} \to [0;1]$ donnent un moyen de *mesurer* certains sous-ensembles de $\Omega$, précisément ceux dans $\mathcal{F}$. Ainsi (si vous revoyez les axiomes définissant $\mathcal{F}$) :

   - Pourquoi, si $A \in \mathcal{F}$, est-ce que $A^c \in \mathcal{F}$ aussi ? C'est car, naturellement, si l'on sait mesurer $A$, i.e si $\mathrm{P}(A)$ est un réel bien défini, on aimerait mesurer le complémentaire de $A$, et il se trouve en plus que $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$. C'est plutôt naturel : si vous pouvez mesurer un segment d'une planche de bois, vous voudriez pouvoir mesurer le reste de la planche, de telle sorte à ce que la somme de la longueur du segment mesuré et de la longueur du reste égale la longueur de toute la planche.
   - Pourquoi, si $A \in \mathcal{F}$ et $B \in \mathcal{F}$, alors $A \cap B$ et $A \cup B \in \mathcal{F}$ ? Là aussi, si vous pouvez mesurer deux segments de la planche de bois, pas forcément disjoints, vous aimeriez pouvoir mesurer des "combinaisons" de ces segments, au sens large (des unions, des intersections, des différences $A \setminus B$, etc.).

   Ainsi, nous voyons que $\mathcal{F}$ est *riche*, dans le sens où : si un évènement $A \in \mathcal{F}$, $\mathcal{F}$ contient aussi son complémentaire, les unions de $A$ avec tous les autres éléments de $\mathcal{F}$, les intersections, etc.

3. Le but de ces axiomes est somme toute de poser des définitions naturelles. On pose que $\mathrm{P}(\Omega) = 1$. En fait, cette constante 1 est assez arbitraire, elle définit simplement une échelle conventionnelle de mesure : la mesure de tout l'ensemble $\Omega$ est par convention non pas 10, ni 100, mais 1 car on pourrait plus tard parler de "proportion", de "pourcentage" plus simplement. On pose que si l'ensemble des évènements $\{A_i\}_{i=1}^{\infty}$ sont disjoints 2-à-2 (pairwise disjoint, i.e $A_i \cap A_j = \emptyset$ si $i \neq j$) alors $\mathrm{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)$ car, là aussi, si vous détenez un segment de bois $A_1$ et un autre $A_2$ disjoint du premier, alors la mesure de l'union des deux segments est simplement la somme des 2 longueurs.

## 1.2 Propriétés importantes de la distribution de probabilité P

Après avoir lu la motivation ci-dessus (qui pourrait en fait vous paraître évidente ou pas), repensez aux propriétés fondamentales vues en cours, dont :

- $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
- $A \subset B \implies \mathrm{P}(B \setminus A) = \mathrm{P}(B) - \mathrm{P}(A)$
- ...

# 2 Probabilité conditionnelle et conséquences

## 2.1 Premiers constats et définitions

Si $A, B \in \mathcal{F}$, alors la probabilité conditionnelle de $A$ sachant $B$ est, si $\mathrm{P}(B) \neq 0$ :

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

Deux interprétations existent pour la probabilité conditionnelle :

- La première est celle du redimensionnement : une nouvelle mesure est attribuée à l'évènement $A$ car on le considère dans un univers $B$ plus petit que $\Omega$. Notons que formellement, $(\Omega, \mathcal{F}, \mathrm{P}_B)$ est un espace de probabilité, où $\mathrm{P}_B : \mathcal{F} \to [0;1]$ est définie par $\mathrm{P}_B(A) = \mathrm{P}(A|B)\ \forall A \in \mathcal{F}$. Ceci sera montré plus tard, en exercice, et justifie cette perspective.

- La seconde que j'aime appeler "bayésienne" : c'est la probabilité de $A$ après avoir "observé" $B$. Cette interprétation se rapproche donc de celle d'entropie conditionnelle d'AICCII : combien d'incertitude est-ce que $A$ contient encore après avoir observé $B$. Nous verrons sous peu comment la loi de Bayes justifie cette vue.

Notons que la seconde se comprend bien dans le contexte d'évènements indépendants : si $A$ et $B$ sont indépendants, alors $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$, donc si $\mathrm{P}(B) \neq 0$, alors $\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \mathrm{P}(A)$ : observer $B$ ne change rien quant à l'information apportée par $A$. Observons en plus que l'indépendance est symétrique : $\mathrm{P}(B|A) = \mathrm{P}(B)$ aussi.

## 2.2 Loi de Bayes

Si $A$ et $B$ sont deux évènements non vides dans $(\Omega, \mathcal{F}, \mathrm{P})$ alors la loi de Bayes énonce :

$$\mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A)$$

Cette équation s'écrit aussi, tant qu'aucune des probabilités ne vaut 0 :

$$\frac{\mathrm{P}(A|B)}{\mathrm{P}(B|A)} = \frac{\mathrm{P}(A)}{\mathrm{P}(B)} \quad \text{ou encore} \quad \mathrm{P}(A|B) = \mathrm{P}(A)\frac{\mathrm{P}(B|A)}{\mathrm{P}(B)}$$

En nous rappelant de notre intuition AICCIIesque sur l'information et l'incertitude, nous pouvons formuler de manière équivalente que :

- Pour l'équation de gauche : le rapports des incertitudes après avoir observé l'autre évènement est égal au rapport des incertitudes avant observation;
- Pour l'équation de droite : l'incertitude de $A$ après avoir observé $B$ est celle initiale, fois un certain facteur correcteur : c'est celui de l'incertitude de $B$ après observation de $A$, sur celle initiale de $B$ - donc en quelque sorte combien d'information est gagnée ou perdue concernant $B$ après observation de $A$.

Je signale que ce sont que des bribes d'intuition discutables, si elles ne vous parlent pas trop ce n'est pas bien grave, je les pose ici au cas où ça vous parle.

## 2.3 Loi des probabilités totales

Enfin, rappelons cette loi fondamentale des probabilités totales, qui vous suivra partout et que vous devriez garder tout le temps à l'esprit (surtout si vous faites le cours de modèles stochastiques le semestre prochain). Elle énonce :

Soient $\{B_i\}_{i=1}^{\infty}$ des évènements à probabilité non nulle, disjoints 2-à-2 dans $(\Omega, \mathcal{F}, \mathrm{P})$, et soit $A \in \mathcal{F}$ tel que $A \subseteq \cup_{i=1}^{\infty} B_i$. Alors :

$$\begin{aligned}
\mathrm{P}(A) &= \mathrm{P}\left(A \cap (\cup_{i=1}^{\infty} B_i)\right) && \text{car } C \subseteq D \implies C = C \cap D \\
&= \mathrm{P}\left(\cup_{i=1}^{\infty} (A \cap B_i)\right) && \text{par distributivité : } C \cap (D \cup E) = (C \cap D) \cup (C \cap E) \\
&= \sum_{i=1}^{\infty} \mathrm{P}(A \cap B_i) && \text{car } \{B_i\} \text{ disjoints 2-à-2} \implies \{A \cap B_i\} \text{ disjoints aussi} \\
\mathrm{P}(A) &= \sum_{i=1}^{\infty} \mathrm{P}(A|B_i)\mathrm{P}(B_i) && \text{par définition de la probabilité conditionnelle.}
\end{aligned}$$

Cette formule nous donne une méthode pour calculer $\mathrm{P}(A)$ : calculer $\mathrm{P}(A|B_i)$ pour des $\{B_i\}$ bien choisis, i.e pour lesquels on sait calculer $\mathrm{P}(A|B_i)$ et $\mathrm{P}(B_i)$, puis rassembler le tout.

# 3 Exercices

## 3.1 Ex 1 : La probabilité conditionnelle est une mesure de probabilité

Soit $(\Omega, \mathcal{F}, \mathrm{P})$ un espace de probabilité. Soit $B \in \mathcal{F}$ tel que $\mathrm{P}(B) \neq 0$. Définissons encore :

$$\begin{aligned} \mathrm{P}_B : \mathcal{F} &\to [0;1] \\ A &\to \mathrm{P}_B(A) = \mathrm{P}(A|B) \end{aligned}$$

Montrer que $\mathrm{Pr}_B$ est une distribution de probabilité dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathrm{P}_B)$.

## 3.2 Ex 2 : Propriété sur les évènements indépendants

Soient $E_1, ..., E_n$ des évènements indépendants (mutuellement) avec $n \geq 2$. Montrer que :

$$\mathrm{P}\left(\cup_{i=1}^{n} E_i\right) = 1 - \prod_{i=1}^{n}(1 - \mathrm{P}(E_i))$$

## 3.3 Ex 3 : Calcul de probabilité

Soient $A$ et $B$ 2 évènements tels que $\mathrm{P}(A) = \mathrm{P}(B) > 0$ et $\mathrm{P}(A \cap B) = 0$. Donc (plusieurs réponses peuvent être correctes) :

- ☐ $\mathrm{P}(A \cup B \mid B) = \frac{1}{2}$
- ☐ $\mathrm{P}(A \mid A \cup B) = \frac{1}{2}$
- ☐ $\mathrm{P}(A \mid B) = 0$
- ☐ Les 3 autres réponses sont fausses.

# 4 Corrigés

## 4.1 Ex 1

Il faut vérifier les 3 axiomes définissant une distribution de probabilité :

- $0 \leq \mathrm{P}_B(A) \leq 1, \forall A \in \mathcal{F}$
- $\mathrm{P}_B(\Omega) = 1$
- $\{A_i\}_{i=1}^{\infty}$ disjoints 2-à-2 $\implies \mathrm{P}_B(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{P}_B(A_i)$

Ainsi, soient $A \in \mathcal{F}$ et $\{A_i\}_{i=1}^{\infty}$ disjoints 2-à-2 :

- $0 \leq \mathrm{P}(A \cap B) \leq \mathrm{P}(B)$ car $A \cap B \subseteq B$ donc, en divisant par $\mathrm{P}(B) \neq 0$, $0 \leq \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)} = \mathrm{P}(A|B) = \mathrm{P}_B(A) \leq 1$.
- $\mathrm{P}_B(\Omega) = \mathrm{P}(\Omega|B) = \frac{\mathrm{P}(\Omega\cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B)}{\mathrm{P}(B)} = 1$ car $B \subseteq \Omega$ donc $\Omega \cap B = B$.
- Un peu plus de calculs :

  $$\mathrm{P}_B(\cup_{i=1}^{\infty} A_i) = \mathrm{P}(\cup_{i=1}^{\infty} A_i|B) = \frac{\mathrm{P}\left((\cup_{i=1}^{\infty} A_i) \cap B\right)}{\mathrm{P}(B)} = \frac{\mathrm{P}\left(\cup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathrm{P}(B)} = \sum_{i=1}^{\infty} \frac{\mathrm{P}(A_i \cap B)}{\mathrm{P}(B)} = \sum_{i=1}^{\infty} \mathrm{P}_B(A_i)$$

  L'avant dernière égalité suit du fait que les $\{A_i \cap B\}$ sont disjoints 2-à-2.

## 4.2 Ex 2

Commençons le calcul :

$$\begin{aligned}
\mathrm{P}\left(\cup_{i=1}^{n} E_i\right) &= \mathrm{P}\left[\left((\cup_{i=1}^{n} E_i)^c\right)^c\right] && \text{car } (D^c)^c = D \\
&= 1 - \mathrm{P}\left[(\cup_{i=1}^{n} E_i)^c\right] && \text{car } \mathrm{P}(D^c) = 1 - \mathrm{P}(D) \\
&= 1 - \mathrm{P}\left[\cap_{i=1}^{n} (E_i)^c\right] && \text{par De Morgan}
\end{aligned}$$

On aimerait alors que $\mathrm{P}\left[\cap_{i=1}^{n} (E_i)^c\right] \stackrel{?}{=} \prod_{i=1}^{n} \mathrm{P}[(E_i)^c] = \prod_{i=1}^{n} (1 - \mathrm{P}(E_i))$.
Montrons que, si $A$ et $B$ sont indépendants, alors $A^c$ et $B^c$ le sont aussi :

$$\begin{aligned}
\mathrm{P}(A^c \cap B^c) &= \mathrm{P}((A \cup B)^c) && \text{par De Morgan} \\
&= 1 - \mathrm{P}(A \cup B) \\
&= 1 - \mathrm{P}(A) - \mathrm{P}(B) + \mathrm{P}(A \cap B) \\
&= (1 - \mathrm{P}(A)) - \mathrm{P}(B) + \mathrm{P}(A)\mathrm{P}(B) && \text{par indépendance de } A \text{ et } B \\
&= (1 - \mathrm{P}(A)) - (\mathrm{P}(B) - \mathrm{P}(A)\mathrm{P}(B)) \\
&= (1 - \mathrm{P}(A)) - \mathrm{P}(B)(1 - \mathrm{P}(A)) \\
&= (1 - \mathrm{P}(A))(1 - \mathrm{P}(B)) \\
&= \mathrm{P}(A^c)\mathrm{P}(B^c)
\end{aligned}$$

Ainsi, les $\{(E_i)^c\}_{i=1}^{n}$ sont indépend- **wait a minute**. Tout ce que l'on vient de prouver, c'est que les $(E_i)^c$ sont indépendants **2-à-2**, et pas **mutuellement** ! En prenant $A = E_1$ et $B = E_2$, nous venons en fait de faire l'initialisation d'un récurrence dont voici l'hérédité.
Soit $n > m \geq 2$. Supposons que $\{(E_i)^c\}_{i=1}^{m}$ sont (mutuellement) indépendants. Montrons que $\{(E_i)^c\}_{i=1}^{m+1}$ le sont aussi :

$$\begin{aligned}
\mathrm{P}[\cap_{i=1}^{m+1} (E_i)^c] &= \mathrm{P}[(\cap_{i=1}^{m} (E_i)^c) \cap (E_{m+1})^c] \\
&= \mathrm{P}[(\cap_{i=1}^{m} (E_i)^c)]\mathrm{P}[(E_{m+1})^c] && \text{en utilisant l'initialisation avec } A = \cup_{i=1}^{m} E_i \text{ et } B = E_{m+1} \\
&= \left(\prod_{i=1}^{m} \mathrm{P}[(E_i)^c]\right) \mathrm{P}[(E_{m+1})^c] \\
&= \prod_{i=1}^{m+1} \mathrm{P}[(E_i)^c]
\end{aligned}$$

Ainsi $\{(E_i)^c\}_{i=1}^{m+1}$ sont mutuellement indépendants. Donc :

$$\mathrm{P}\left(\cup_{i=1}^n E_i\right) = 1 - \mathrm{P}\left[\cap_{i=1}^n (E_i)^c\right] = 1 - \prod_{i=1}^n \mathrm{P}[(E_i)^c] = 1 - \prod_{i=1}^n (1 - \mathrm{P}(E_i))$$

Notons qu'en général, avec une démonstration similaire quoique plus verbeuse :

$$\{E_i\}_{i=1}^n \text{ sont (mutuellement) indépendants}$$

$$\iff$$

$$\forall \{B_i\}_{i=1}^n \text{ où } B_i = E_i \vee B_i = (E_i)^c \text{ sont mutuellement indépendants}$$

### 4.3 Ex 3

Calculons les 3 probabilités :

- $\mathrm{P}(A \cup B \mid B) = \frac{\mathrm{P}((A \cup B) \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B)}{\mathrm{P}(B)} = 1$. Intuition check: $A \cup B$ est nécessairement arrivé si $B$ est arrivé.
- En utilisant Bayes :

$$\mathrm{P}(A \mid A \cup B) = \frac{\mathrm{P}(A \cup B \mid A)\mathrm{P}(A)}{\mathrm{P}(A \cup B)} = \frac{1 \times \mathrm{P}(A)}{\mathrm{P}(A) + \underbrace{\mathrm{P}(B)}_{=\mathrm{P}(A)} - \underbrace{\mathrm{P}(A \cap B)}_{=0}} = \frac{\mathrm{P}(A)}{\mathrm{P}(A) + \mathrm{P}(A)} = \frac{1}{2}.$$

- $\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{0}{\mathrm{P}(B)} = 0$.

Les réponses 2 et 3 sont donc vraies, les autres sont fausses.

# ProbaStats - Compléments
# Variables aléatoires discrètes

Salim Najib

Printemps 2022

## Introduction

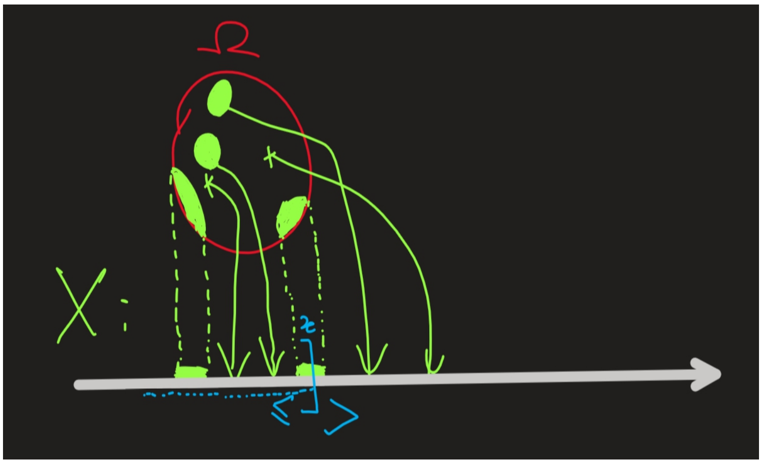
Cette fiche concerne le chapitre 3 des slides de cours. Une grande partie de ce qui sera traité ici est transposable pour les v.a continues, ainsi il vaut mieux s'assurer de bien suivre jusqu'ici. Cette fiche suppose que la fiche précédente sur les bases des probabilités a été lue, notamment dans les retours sur les définitions du cours.

## Contents

# 1 Qu'est-ce qu'une variable aléatoire ?

Soit $(\Omega, \mathcal{F}, \mathrm{P})$ un espace probabilisé. Une variable aléatoire $X : \Omega \to \mathbb{R}$ est une fonction telle que :

$$\forall x \in \mathbb{R} \quad L_x = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$$




La v.a $X$ traduit chaque point de l'univers $\Omega$ en un certain réel, avec en plus l'axiome (peut être non vu en cours d'ailleurs car il sera toujours vérifié par la suite) la fonction $X$ est **mesurable**, i.e il est possible d'attribuer une mesure, une longueur, pour chaque $x$, de $L_x$ i.e l'ensemble des $\omega$ dont l'image par $X$ est telle que $X(\omega) \leq x$. Cette possibilité de mesurer $L_x$ est signifiée formellement par $L_x \in \mathcal{F}$.

Ok, donc on peut poser un threshold i.e une borne supérieure $x$ et donner une taille à l'ensemble des $\omega$ tels que $X(\omega) \leq x$. Quel est l'intérêt ?
C'est que maintenant, au lieu de travailler avec des $\omega \in \Omega$ dont on ne sait a priori rien, nous travaillons avec des $X(\omega) \in \mathbb{R}$, des réels que nous pouvons alors additionner, multiplier, etc. Des réels que nous pouvons également mieux mesurer : une longueur intuitive que nous pouvons attribuer à l'intervalle $]a, b[= \{X(\omega) \mid \omega \in S\}$ pour un certain sous-ensemble $S$ de $\Omega$ serait $b - a$, ou alors nous pouvons aussi choisir un point $x_0 \in \mathbb{R}$ et décider que la longueur de $\{X(\omega) \mid \omega \in S\} = \sigma$ vaut 1 si $x_0 \in \sigma$, 0 sinon. Il y a une infinté de possibilités de mesures de $\mathbb{R}$, que nous retrouvons toutes dans les fonctions de masse, aka densités, aka probability mass functions (PMF) $f_X$ et les fonctions de répartition, aka distributions, aka cumulative distribution functions (CDF) $F_X$.
Ainsi, étant donné $x \in \mathbb{R}$, nous définissons :

$$F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}(\{\omega \in \Omega \mid X(\omega) \leq x\}) = \mathrm{P}(L_x)$$
$$f_X(x) = \mathrm{P}(X = x) = \mathrm{P}(\{\omega \in \Omega \mid X(\omega) = x\})$$

Enfin, différencions entre les variables aléatoires discrètes et les autres variables aléatoires.
Définissons le **support** de $X$ :

$$D_X = \{x \in \mathbb{R} \mid f_X(x) \neq 0\}$$

Si $D_X$ est *dénombrable*, alors on dit que $X$ est une variable aléatoire discrète.

# 2 Moments, espérance

Je rappelle ici seulement un résultat, il y aura davantage à dire dans les exos ci-dessous.

$$E(g(X)) = \sum_{x \in D_X} g(x) f_X(x)$$

# 3 Distributions conditionnelles

Rappelons ici le théorème des probabilités totales pour l'espérance : si $\{B_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ (quantité dénombrable) constituent une partition de $\Omega$ (i.e les $B_i$ sont disjoints et $\Omega = \cup_{i=1}^{\infty} B_i$), alors :

$$E(X) = \sum_{i=1}^{\infty} E(X|B_i)\mathrm{P}(B_i)$$

# 4 Exos

## 4.1 Ex 1 : Calcul d'entropie

Rappelons que si $X$ est une variable aléatoire, son entropie binaire est définie par :

$$H_2(X) = E(-\log_2(f_X(X)))$$

Soient des variables aléatoires $A \sim \text{Bernoulli}(p)$ et $B \sim \text{Geom}(p)$, où $\text{P}(B = k) = p(1-p)^{k-1}$ pour $k \in \mathbb{N}^*$. Calculer $H_2(A)$ et $H_2(B)$.

## 4.2 Ex 2 : Fonction génératrice des probabilités

Soit $X$ une variable aléatoire discrète à support $D_X \subseteq \mathbb{N}$. On définit alors la fonction génératrice des probabilités de $X$ comme suit :

$$G_X(z) = E(z^X)$$

1. Que vaut $G_X(1)$ ?
2. Exprimer $\text{P}(X = k)$ pour $k \in \mathbb{N}$ en fonction de $G_X(z)$ pour un certain $z \in \mathbb{R}$ bien choisi. Hint : Taylor.
3. Exprimer $E(X)$ en fonction de $G_X(z)$ pour un certain $z \in \mathbb{R}$ bien choisi.
4. Soit $Z$ une v.a discrète telle que $D_Z \subseteq \mathbb{N}$ et $G_Z(z) = \frac{1}{(2-z)^5}$, $0 < z < 2$. Soit $Y \sim \text{Geom'}(p)$ avec $0 < p < 1$ indépendante de $Z$ (attention à l'apostrophe : comme dans l'exercice 1, $\text{P}(Y = k) = p(1-p)^y$ pour $y \in \mathbb{N}$).
   Calculer $\text{P}(Z = Y)$. Indice : ne pas calculer $\text{P}(Z = n)$ pour $n \in \mathbb{N}$.

## 4.3 Ex 3 : Calcul de probabilité

Soit $X$ une v.a discrète sur l'espace probabilisé $(\Omega, \mathcal{F}, \text{P})$ à support $D_X \subseteq \mathbb{Z}$. Soit $B \in \mathcal{F}$ tel que $\forall n \in \mathbb{Z}$ $\text{P}(B|X = n) = \frac{1}{4}$. Calculer $\text{P}(B)$. Est-ce que $B$ est indépendant de $X$ ?

Note : on dit qu'un évènement $B$ est indépendant d'une v.a discrète $X$ si et seulement si les évènements $B$ et $\{\omega \in \Omega \mid X(\omega) = x\} = \{X = x\}$ sont indépendants $\forall x \in \mathbb{R}$ .

## 4.4 Ex 4 : Calcul d'espérance

Considérons un datacenter comportant $n \geq 1$ étagères de 2 ordinateurs chacune. Chaque étagère fonctionne si et seulement si les 2 ordinateurs la composant fonctionnent.
Un feu détruit $0 \leq k \leq 2n$ ordinateurs uniformément. En moyenne, combien d'étagères fonctionnent encore après l'incident ?

## 4.5 Ex 5 : Preuve de la loi des petits nombres

Je conseille de n'entreprendre cet exercice qu'après avoir tenté et corrigé l'exercice 2.

1. Soient deux variables aléatoires discrètes $X, Y$ sur $(\Omega, \mathcal{F}, \text{P})$ à supports $D_X, D_Y$ dans $\mathbb{N}$. Considérons $G_X$ et $G_Y$ leurs fonctions génératrices de probabilités. Montrer que $X$ et $Y$ suivent la même distribution $F$ si et seulement si $G_X(z) = G_Y(z)$ $\forall z \in \mathbb{R}$. Ainsi nous montrons que la fonction génératrice de probabilité d'une distribution la caractérise entièrement.
2. Montrer ainsi, en utilisant les fonctions génératrices de probabilité, que si $B_n \sim \text{Bin}(n, p_n)$ et que $\lim\limits_{n\to\infty} np_n = \lambda \in \mathbb{R}$, alors $\lim\limits_{n\to\infty} B_n \sim \text{Pois}(\lambda)$.

# 5 Correction

## 5.1 Ex 1

Concernant l'entropie d'une Bernoulli, comme échauffement :

$$\begin{aligned}
H_2(A) &= -\sum_{x\in D_A} \log_2(f_A(x))f_A(x) \\
&= -\sum_{x\in\{0,1\}} \log_2(f_A(x))f_A(x) \\
&= -\left[f_A(1)\log_2(f_A(1)) + f_A(-1)\log_2(f_A(-1))\right] \\
&= -\left[p\log_2(p) + (1-p)\log_2(1-p)\right]
\end{aligned}$$

Puis, plus conséquent, voici le calcul de l'entropie d'une géométrique :

$$\begin{aligned}
H_2(B) &= -\mathbb{E}[\log_2(f_B(B))] \\
&= -\sum_{k\in D_B} \log_2(f_B(k))f_B(k) \\
&= -\sum_{k=1}^{\infty} \log_2(f_B(k))f_B(k) \\
&= -\sum_{k=1}^{\infty} \log_2((1-p)^{k-1}p)(1-p)^{k-1}p \\
&= -\sum_{k=1}^{\infty} \left[(k-1)\log_2(1-p) + \log_2(p)\right](1-p)^{k-1}p, \text{ en utilisant } \log(a^c b) = c\log(a) + \log(b) \\
&= -\left(p\log_2(1-p)\sum_{k=1}^{\infty}(k-1)(1-p)^{k-1} + p\log_2(p)\sum_{k=1}^{\infty}(1-p)^{k-1}\right) \\
&= -\left(p\log_2(1-p)\sum_{m=0}^{\infty} m(1-p)^m + p\log_2(p)\sum_{m=0}^{\infty}(1-p)^m\right) \\
&= -\left(p\log_2(1-p)\sum_{m=0}^{\infty} mq^m + p\log_2(p)\sum_{m=0}^{\infty} q^m\right), \text{ en notant } q = 1-p
\end{aligned}$$

Ici, notons que, vu que $|q| = |1-p| < 1$ :

$$\sum_{m=0}^{\infty} q^m = \frac{1}{1-q} = \frac{1}{p}$$

Ensuite, pour calculer $\sum_{m=0}^{\infty} mq^m$, utilisons le trick de la dérivée présenté en cours pour le calcul de moments de variables aléatoires géométriques. Commençons par noter que :

$$\frac{d}{dq}[q^m] = mq^{m-1}$$

Donc :

$$\sum_{m=0}^{\infty} mq^{m-1} = \sum_{m=0}^{\infty}\frac{d}{dq}[q^m] = \frac{d}{dq}\left[\sum_{m=0} q^m\right] = \frac{d}{dq}\left[\frac{1}{1-q}\right] = \frac{1}{(1-q)^2}$$

Ainsi :

$$\begin{aligned}
H_2(B) &= -\left(p\log_2(1-p)q\sum_{m=0}^{\infty} mq^{m-1} + p\log_2(p)\sum_{m=0}^{\infty} q^m\right) \\
&= -\left(p\log_2(1-p)\frac{q}{(1-q)^2} + p\log_2(p)\frac{1}{p}\right) \\
&= -\left(\log_2(1-p)\frac{1-p}{p} + \log_2(p)\frac{p}{p}\right) \\
&= -\frac{(1-p)\log_2(1-p) + p\log_2(p)}{p}, \text{ remarquons que } H_2(B) = \frac{H_2(A)}{p}.
\end{aligned}$$

### 5.2 Ex 2

Tâchons de mieux comprendre $G_X$ d'abord :

$$G_X(z) = E(z^X) = \sum_{k \in D_X} z^k f_X(k) \overset{D_X \subseteq \mathbb{N}}{=} \sum_{k=0}^{\infty} z^k \mathrm{P}(X = k)$$

Ainsi, $G_X$ est la **série de MacLaurin** (i.e de Taylor autour de 0) $\sum z^k p_k$ dont les coefficients $p_k$ sont les probabilités $p_k = \mathrm{P}(X = k)$. Par unicité du développement de Taylor :

$$G_X(z) = \sum_{k=0}^{\infty} \frac{G_X^{(k)}(0)}{k!} z^k = \sum_{k=0}^{\infty} \mathrm{P}(X = k) z^k$$

Ainsi, par identification des coefficients :

$$\forall k \in \mathbb{N} \ \mathrm{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}$$

Répondons à présent aux questions posées.

1. $G_X(1) = E(1^X) = E(1) = 1$; en outre remarquons que $G_X(1) = \sum_{k \in D_X} f_X(k)$ qui vaut 1 car $f_X$ est une PMF.

2. Déjà faite ci-dessus.

3. Nous voulons faire apparaître $E(X)$ dans l'expression de quelque chose qui dépend de $G_X$. Notons que :
$$\frac{dG_X}{dz}(z) = \frac{d}{dz}\left[E(z^X)\right] = E\left(\frac{d}{dz}\left[z^X\right]\right) = E(Xz^{X-1})$$
Ainsi :
$$\frac{dG_X}{dz}(1) = E(X \times 1^{X-1}) = E(X)$$

4. Utilisons le théorème des probabilités totales, les évènements $\{Y = y\}_{y \in D_Y}$ formant une partition de l'univers :
$$\begin{aligned}
\mathrm{P}(Z = Y) &= \sum_{y \in D_Y} \mathrm{P}(Z = y \mid Y = y)\mathrm{P}(Y = y) \\
&= \sum_{y=0}^{\infty} \mathrm{P}(Z = y)\mathrm{P}(Y = y) && \text{par indépendance de } Z, Y \\
&= \sum_{y=0}^{\infty} \mathrm{P}(Z = y) p(1-p)^y && \text{car } Y \sim \text{Geom'}(p) \\
&= p\sum_{y=0}^{\infty} (1-p)^y \mathrm{P}(Z = y) \\
&= pG_Z(1-p) \\
&= \frac{p}{(2-(1-p))^5} \\
&= \frac{p}{(1+p)^5}.
\end{aligned}$$

### 5.3 Ex 3

Calculons, toujours à l'aide du théorème des probabilités totales :

$$\begin{aligned}
\mathrm{P}(B) &= \sum_{n \in D_X} \mathrm{P}(B|X=n)\mathrm{P}(X=n) \\
&= \sum_{n \in D_X} \frac{1}{4}\mathrm{P}(X=n) \\
&= \frac{1}{4} \sum_{n \in D_X} \mathrm{P}(X=n) \\
&= \frac{1}{4} \times 1 = \frac{1}{4}.
\end{aligned}$$

Vérifions l'indépendance de $B$ et $X$. Soit $x \in \mathbb{R}$.

- Si $x \notin D_X$, $\mathrm{P}(X=x) = 0$ donc $\mathrm{P}(B|X=x)$ n'est pas bien définie.
  Cependant, $\mathrm{P}(B \cap \{X=x\}) \leq \mathrm{P}(X=x) = 0$ car $B \cap \{X=x\} \subseteq \{X=x\}$, donc

  $$\mathrm{P}(B \cap \{X=x\}) = 0 = 0 \times \mathrm{P}(B) = \mathrm{P}(X=x)\mathrm{P}(B)$$

  Alors $B$ et $\{X=x\}$ sont indépendants.
- Si $x \in D_X$, alors $\mathrm{P}(B|X=x) = \frac{1}{4} = \mathrm{P}(B)$. Alors $B$ et $\{X=x\}$ sont indépendants.

### 5.4 Ex 4

Dénotons par $I_j \sim \text{Bernoulli}(p)$ pour $j = 1, \dots n$ la variable aléatoire de Bernoulli indiquant si l'étagère $j$ a survécu au feu (avec probabilité $p = \mathbb{P}(I_j = 1) = (I_1 = 1)$, par uniformité des cibles de l'incendie) ou pas. Notons ensuite par $N = \sum_{j=1}^{n} I_j$ le nombre d'étagères ayant survécu à l'incendie, en faisant attention au fait que malgré que les $I_j$ soient identiquement distribuées, elles ne sont **pas** indépendantes - si l'on sait qu'une étagère donnée a survécu, cela réduit les chances pour une autre étagère qu'elle survive - ainsi $N$ n'est **pas** une variable aléatoire binomiale.
Calculons son espérance par linéarité :

$$\begin{aligned}
\mathbb{E}[N] &= \mathbb{E}\left[\sum_{j=1}^{n} I_j\right] \\
&= \sum_{j=1}^{n} \mathbb{E}[I_j] \\
&= \sum_{j=1}^{n} \mathbb{E}[I_1] && \text{car les } I_j \text{ suivent la même distribution} \\
&= n\mathbb{E}[I_1] \\
&= n\left(1 \times \mathbb{P}(I_1 = 1) + 0 \times \mathbb{P}(I_1 = 0)\right) \\
&= n\mathbb{P}(I_1 = 1)
\end{aligned}$$

Reste alors à calculer $\mathrm{P}(I_1 = 1)$, i.e la probabilité qu'étant donné une étagère, elle fonctionne encore correctement après l'incendie.

Supposons que cette étagère 1 est composée des ordinateurs $c_1$ et $c_2$. Il faut donc que ni $c_1$ ni $c_2$ ne fassent partie des $k$ ordinateurs touchés :

$$\begin{aligned}
\mathrm{P}(I_1 = 1) = {} & \mathrm{P}(\text{pas la 1ère cible})\mathrm{P}(\text{pas la 2nde cible} \mid \text{pas la 1ère cible})\mathrm{P}(\text{pas la 3ème cible} \mid \text{pas la cible 1 ni 2}) \dots \\
& \mathrm{P}(\text{pas la } n\text{ème cible} \mid \text{pas la cible 1 ni 2 ni ... ni } n-1) \\
= {} & \frac{2n-2}{2n}\frac{2n-3}{2n-1}\frac{2n-4}{2n-2}\cdots\frac{2n-(k+1)}{2n-(k-1)}
\end{aligned}$$

$$\mathbb{P}(I_1 = 1) \overset{\text{simplifier}}{=} \frac{(2n-k)(2n-(k+1))}{2n(2n-1)}$$

Ainsi :

$$E(N) = n\mathrm{P}(I_1 = 1) = \frac{(2n-k)(2n-k-1)}{2(2n-1)}$$

## 5.5 Ex 5

Nous verrons que prendre le temps de commencer à comprendre les fonctions génératrices de probabilités sera utile plus tard, pour les variables aléatoires continues. Tel est le but des exercices 2 et 5.

1. $\Longrightarrow$ : Si $X$ et $Y$ suivent la même distribution, alors $f_X(k) = f_Y(k)\ \forall k \in \mathbb{N}$, donc :

$$G_X(z) = E(z^X) = \sum_{k=0}^{\infty} z^k f_X(k) = \sum_{k=0}^{\infty} z^k f_Y(k) = E(z^Y) = G_Y(z)$$

$\Longleftarrow$ : Soit $k \in \mathbb{N}$. Si $G_X = G_Y$, alors, par l'exercice 2 :

$$\mathrm{P}(X = k) = \frac{G_X^{(k)}(0)}{k!} = \frac{G_Y^{(k)}(0)}{k!} = \mathrm{P}(Y = k)$$

2. Calculons les fonctions génératrices $G_{B_n}$ et $G_P$ pour $P \sim \mathrm{Pois}(\lambda)$. D'abord $G_{B_n}$ :

$$\begin{aligned}
G_{B_n}(z) &= E(z^{B_n}) \\
&= \sum_{k=0}^{n} z^k \mathrm{P}(B_n = k) \\
&= \sum_{k=0}^{n} z^k \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (p_n z)^k (1-p_n)^{n-k} \\
&= (p_n z + (1-p_n))^n && \text{par la formule du binôme de Newton} \\
&= (1 + p_n(z-1))^n
\end{aligned}$$

Puis $G_P$ :

$$\begin{aligned}
G_P(z) &= \sum_{k=0}^{\infty} z^k \mathrm{P}(P = k) \\
&= \sum_{k=0}^{\infty} z^k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} \\
&= e^{-\lambda} e^{\lambda z} && \text{car } e^x = \sum_{m=0}^{\infty} \frac{x^m}{m!} \\
&= e^{\lambda(z-1)}
\end{aligned}$$

Ainsi, en utilisant $\lim_{m\to\infty} \left(1 + \frac{a}{m}\right)^m = e^a$ :

$$\lim_{n\to\infty} G_{B_n}(z) = \lim_{n\to\infty} (1 + p_n(z-1))^n = \lim_{n\to\infty} \left(1 + \frac{(np_n)(z-1)}{n}\right)^n$$

Donc :

$$\lim_{n\to\infty} G_{B_n}(z) = \exp\left(\lim_{n\to\infty} np_n(z-1)\right) = e^{\lambda(z-1)} = G_P(z)$$

Les fonctions génératrices de probabilités étant identiques, les variables aléatoires $\lim_{n\to\infty} B_n$ et $P$ suivent alors la même distribution (par la question), qui est $\mathrm{Pois}(\lambda)$.

# ProbaStats - Compléments
# Variables aléatoires continues

Salim Najib

Printemps 2022

## Introduction

Pour une explication de l'intuition sur les variables aléatoires, voir la fiche sur les variables aléatoires discrètes. Il n'y a en fait pas grand chose à rajouter conceptuellement dans le cas continu **avec une seule variable aléatoire**, cette fiche a pour but principal ses exercices pour découvrir certaines propriétés et certains tricks.
La fiche sur les variables aléatoires **multiples** sera *nettement* plus fournie en théorie, suivant la forme du cours (un peu particulière franchement) de probabilités et statistiques pour IC.

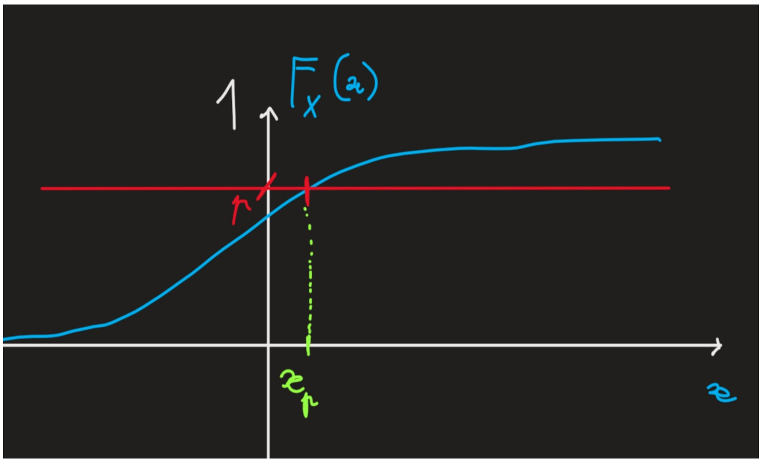
## Contents

# 1 Quantiles

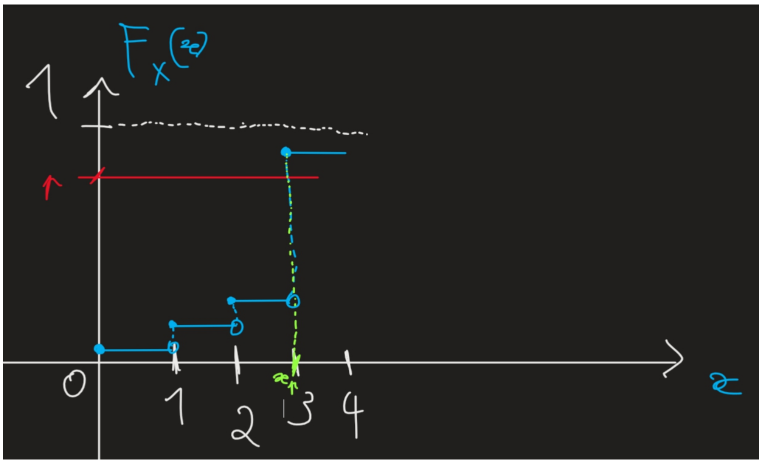
Soit $X$ une v.a. Nous définissons alors le $p-$quantile de $X$, pour $0 < p < 1$ par :

$$x_p = \inf\{x \mid F_X(x) \geq p\}$$

Rappelons que $F_X$ est croissante. Ainsi $x_p$ est "le premier" point $x$ où $F_X$ atteint ou dépasse le seuil $p$. Dans le cas continu, souvent, $F_X$ est strictement croissante donc bijective, alors $x_p = F_X^{-1}(p)$ :




Dans le cas discret, $F_X$ n'est jamais strictement croissante donc jamais injective :




Il y aura besoin des quantiles plus tard dans le cours, gardez alors en tête l'intuition pour vous rappeler de la définition. Se rappeler que dans le cas continu $x_p = F_X^{-1}(p)$ est honnêtement suffisant dans presque tous les cas.

# 2 Transformations

Voici la transformation générale introduite en cours dans le cas continu. Soit $X$ une v.a continue, et soit $g$ une fonction bijective et dérivable. Alors, en posant $Y = g(X)$, en remarquant une structure très similaire à celle de la dérivée d'une composition de fonctions :

$$f_Y(y) = \left|(g^{-1})'(y)\right| f_X(g^{-1}(y))$$

Notons que, par un théorème d'analyse 1 :

$$(g^{-1})'(y) = \frac{1}{g'(g^{-1}(y))}$$

Et donc :

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

Cette formule pourrait être plus pratique parfois. Par exemple, si $X \sim \exp(\lambda)$, posons $Y = e^X$, alors, en notant que $g(y) = e^y$ donc $g'(y) = g(y) = e^y$ et $g^{-1}(y) = \ln(y)$ pour $y > 0$ : $f_Y(y) = \frac{f_X(y)}{|g'(g^{-1}(y))|} = \frac{f_X(y)}{|e^{\ln(y)}|} = \begin{cases} \frac{\lambda}{y} e^{-\lambda y} & \text{si } y > 0; \\ 0 \quad \text{sinon.} \end{cases}$

# 3 Exos

## 3.1 Ex 1 : Variables aléatoires sans mémoire

Nous nous intéressons, dans cet exercice, aux variables aléatoires sans mémoire. **Cet exercice est difficile, il n'y a pas de mal à le skip.** :) Ceci dit, attendez-vous à des exercices conceptuels de ce type (quoique pas tous aussi difficiles non plus) le semestre prochain en Modèles Stochastiques, et surtout à des notions étudiées en cours dont le traitement sera similaire - repensez à cet exercice quand vous étudierez le processus de Poisson.

1. Un exo apparemment déjà fait en série, mais que je laisse ici car... je l'avais mis avant. Montrer que si $X$ est une variable aléatoire discrète sans mémoire à support $D_X = \mathbb{N}^*$, i.e $\forall n, m \in \mathbb{N}^*$ $P(X > n+m \mid X > m) = P(X > n)$, alors $X$ est géométrique. Ceci montre que $X \sim \text{Geom}(p)$ est l'unique variable aléatoire discrète sans mémoire. *Astuce : Calculer $f_X$. Essayer de trouver une formule exprimant $f_X(x)$ en fonction de termes $f_X(i)$ pour des $i < x$, i.e une forme par récurrence.*

2. En vous inspirant de la question ci-dessus, montrer de même que si $X$ est une variable aléatoire continue sans mémoire à support $D_X = \mathbb{R}_+^*$, i.e $\forall t, s \in \mathbb{R}_+^*$ $P(X > s+t \mid X > t) = P(X > s)$, alors $X$ est exponentielle. Ceci montre que $X \sim \exp(\lambda)$ est l'unique variable aléatoire continue sans mémoire. *Astuce : Noter que dans la question 1 nous avons étudié $f_X(x)$, i.e $F_X(x) - F_X(x-1)$, i.e les variations de $F_X$. Peut-être pourrions nous faire de même ici ...?*

## 3.2 Ex 2 : Transformation

Soit $X \sim \mathcal{N}(0,1)$ puis soit $Y = e^X$. Trouver $f_Y$.

## 3.3 Ex 3 : Surface moyenne

Quelle est la surface moyenne d'un disque de rayon $R \sim \text{Unif}(0,1)$ ?

# 4 Correction

## 4.1 Ex 1

1. Soit $x \in \mathbb{N}^*$. A l'aide de l'astuce :

$$\begin{aligned} f_X(x) &= P(X = x) \\ &= P(X > x-1) - P(X > x) \end{aligned}$$

Ok, que savons nous sur $P(X > x-1)$ et $P(X > x)$ ?
Déjà, si $x = 1$, alors $P(X > 1-1) = P(X > 0) = 1$ car $D_X = \mathbb{N}^*$. Alors, $f_X(1) = 1 - P(X > 1)$.
Ensuite, exploitons $P(X > x)$ et la propriété sans mémoire. Il y a plusieurs moyens de le faire, l'idée centrale est d'essayer de se ramener à l'égalité encadrée ainsi qu'à la dernière :

$$\begin{aligned} f_X(x) &= P(X > x-1) - P(X > x) \\ &= P(x-1 < X \leq x) \\ &= \boxed{P(\{X > x-1\} \cap \{X \leq x\})} \\ &= P(X \leq x \mid X > x-1)P(X > x-1) && \text{par déf. de la probabilité conditionnelle} \\ &= (1 - P(X > x \mid X > x-1))P(X > x-1) && \text{par : } P(A^c|B) = 1 - P(A|B) \\ &= P(X > x-1) - P(X > x \mid X > x-1)P(X > x-1) \\ &= P(X > x-1) - P(X > 1)P(X > x-1) && \text{par la propriété sans mémoire} \\ &= P(X > x-1)\left[1 - P(X > 1)\right] \\ &= P(X > x-1)f_X(0) \\ &= \left[1 - P(X \leq x-1)\right] f_X(1) \\ &= \left[1 - P\left(\cup_{i=1}^{x-1}\{X = x\}\right)\right] f_X(1) \\ &= \left[1 - \sum_{i=1}^{x-1} P\left(X = i\right)\right] f_X(1) \\ &= \boxed{\left[1 - \sum_{i=1}^{x-1} f_X(i)\right] f_X(1) = f_X(x)} \end{aligned}$$

Nous avons une formule par récurrence, prouvons enfin par récurrence sur $x \in \mathbb{N}^*$ que :

$$f_X(x) = p(1-p)^{x-1} \text{ avec } p = f_X(1)$$

<u>Cas initial :</u> $p(1-p)^{1-1} = p(1-p)^0 = p = f_X(1)$ par définition de $p$.
<u>Hérédité :</u> Soit $x \in \mathbb{N}^*$. Supposons par hypothèse de récurrence forte que $\forall i \in [[1, x]]\ f_X(i) = p(1-p)^{i-1}$. Montrons que $f_X(x+1) = p(1-p)^{x+1-1} = p(1-p)^x$. En utilisant la formule encadrée ci-dessus :

$$\begin{aligned} f_X(x+1) &= f_X(1)\left[1 - \sum_{i=1}^{x} f_X(i)\right] \\ &= p\left[1 - \sum_{i=1}^{x} p(1-p)^{i-1}\right] \\ &= p\left[1 - p\sum_{i=1}^{x} (1-p)^{i-1}\right] \\ &= p\left[1 - p\sum_{j=0}^{x-1} (1-p)^{j}\right] && \text{où } j = i-1 \\ &= p\left[1 - p\frac{1-(1-p)^x}{1-(1-p)}\right] && \text{en reconnaissant une somme géométrique} \\ &= p\left[1 - p\frac{1-(1-p)^x}{p}\right] \end{aligned}$$

$$\begin{aligned} &= p\left[1 - 1 + (1-p)^x\right] \\ f_X(x+1) &= p(1-p)^x \end{aligned}$$

Ainsi, la seule variable aléatoire discrète sans mémoire à support dans $\mathbb{N}^*$ est la géométrique ! Notons que le nom de "géométrique" provient justement de la somme géométrique rencontrée ci-dessus.

2. Les différences et les récurrences du cas discret vont devenir dérivées et équations différentielles dans le cas continu. Nous allons essayer de suivre les mêmes étapes conceptuelles que dans la question 1, en trouvant une équation différentielle cette fois-ci - pourquoi ? Nous avons vu à la question 1 qu'étudier les variations de $F_X$, i.e étudier $f_X$, était payant. Ici, en plus, notez que pour une variable aléatoire $X \sim \exp(\lambda)$, $1 - F_X(x) = e^{-\lambda x}$, qui ressemble quand même beaucoup à une solution d'équation différentielle.
Soit $x \in \mathbb{R}_+^*$. Alors :

$$\begin{aligned}
f_X(x) &= F_X'(x) \\
&= \frac{d}{dx} P(X \leq x) \\
&= \lim_{h \to 0} \frac{P(X \leq x+h) - P(X \leq x)}{h} \\
&= \lim_{h \to 0} \frac{1 - P(X > x+h) - 1 + P(X > x)}{h} \\
&= \lim_{h \to 0} \frac{P(X > x) - P(X > x+h)}{h} \\
&= \lim_{h \to 0} \frac{P(x < X < x+h)}{h} \\
&= \lim_{h \to 0} \frac{P(\{X > x\} \cap \{X < x+h\})}{h} \\
&= \lim_{h \to 0} \frac{P(X < x+h \mid X > x) P(X > x)}{h} && \text{par déf. de proba. conditionnelle} \\
&= \lim_{h \to 0} \frac{(1 - P(X > x+h \mid X > x)) P(X > x)}{h} && \text{par : } P(A^c|B) = 1 - P(A|B) \\
&= \lim_{h \to 0} \frac{(1 - P(X > h)) P(X > x)}{h} && \text{par la propriété sans mémoire} \\
&= P(X > x) \lim_{h \to 0} \frac{1 - P(X > h)}{h} \\
&= (1 - P(X \leq x)) \lim_{h \to 0} \frac{P(X > 0) - P(X > h)}{h} \\
&= (1 - F_X(x)) \lim_{h \to 0} \frac{1 - P(X \leq 0) - 1 + P(X \leq h)}{h} \\
&= (1 - F_X(x)) \lim_{h \to 0} \frac{P(X \leq h) - P(X \leq 0)}{h} \\
&= (1 - F_X(x)) \lim_{h \to 0} \frac{F_X(h) - F_X(0)}{h} \\
&= (1 - F_X(x)) F_X'(0) \\
&= \boxed{(1 - F_X(x)) f_X(0) = F_X'(x)}
\end{aligned}$$

Pour rendre le reste des calculs plus intelligible étudions alors $Q_X(x) = 1 - F_X(x) = 1 - P(X \leq x) = P(X > x)$. Notons que $Q_X'(x) = -F_X'(x)$, ainsi l'équation différentielle encadrée ci-dessus devient, pour $x \in \mathbb{R}_+^*$ et en notant $\lambda = f_X(0)$ :

$$\boxed{Q_X'(x) = -\lambda Q_X(x)}$$

Cette équation différentielle a pour solution $Q_X(x) = Ce^{-\lambda x}$ pour $C \in \mathbb{R}$ une constante à déterminer. $Q_X(x) = P(X > x)$ ainsi $Q_X(0) = P(X > 0) = 1$ car $X$ a pour support $D_X = \mathbb{R}_+^*$ donc $Ce^{-\lambda \times 0} = 1 \iff C = 1$. Ainsi, pour $x \in \mathbb{R}_+^*$ :

$$Q_X(x) = 1 - F_X(x) = e^{-\lambda x} \iff F_X(x) = 1 - e^{-\lambda x} \iff f_X(x) = \lambda e^{-\lambda x}$$

Nous avons donc prouvé que la seule variable aléatoire continue sans mémoire est l'exponentielle !

## 4.2 Ex 2

En appliquant la formule des transformations continues, avec $y = g(x) = e^x$ donc $x = g^{-1}(y) = \ln(y)$ :

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{f_X(g^{-1}(y))}{\exp(\ln(y))} = \frac{f_X(g^{-1}(y))}{y} = \frac{1}{y\sqrt{2\pi}}e^{-\frac{(\ln(y))^2}{2}}$$

## 4.3 Ex 3

En se rappelant que la formule de la surface d'un disque de rayon $r$ est $\pi r^2$, la surface du disque de l'exercice est alors $S = \pi R^2$, donc :

$$\begin{aligned}
E(S) &= E(\pi R^2) \\
&= \pi E(R^2) \\
&= \pi \int_0^1 r^2 f_R(r) dr \\
&= \pi \int_0^1 r^2 dr \\
&= \frac{\pi}{3}
\end{aligned}$$

En particulier, il est *faux* que $E(S) = \pi E(R)^2 = \frac{\pi}{2^2} = \frac{\pi}{4}$. On ne peut pas simplement remplacer des variables aléatoires par leurs espérances et cet exercice est une preuve de plus qu'en probas, appliquer calmement ses définitions et résultats mène à bien.

# ProbaStats - Compléments
# Variables aléatoires multiples

Salim Najib

Printemps 2022

## Introduction

C'est le chapitre assez fourre-tout du cours de probabilités et statistiques. Cette fiche risque d'être plus longue que les précédentes.

## Contents

# 1 The basics

Je m'en tiens, dans cette partie, à rappeler la formule de la marginale. Pour deux variables aléatoires continues $X$ et $Y$ jointes :

$$\forall x \in \mathbb{R} \ f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Ce n'est pas que les autres formules de base ne sont pas importantes, c'est simplement qu'on oublie souvent que la marginale peut s'obtenir ainsi.

# 2 Espérance conditionnelle (chaînée)

Rappelons d'abord l'espérance conditionnelle :

$$E\left(g(X,Y) \mid X = x\right) = \int_{-\infty}^{\infty} g(x,y) f_{Y|X}(y|x)dy$$

En particulier:

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$$

L'espérance conditionnelle est un outil pratique de calcul d'espérance avec une méthode dite "par étape". Commençons par adapter nos notations. Notons :

$$E_{X,Y}(g(X,Y)) = E(g(X,Y)) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \ g(x,y) f_{X,Y}(x,y)$$

L'idée est de noter en indice les variables aléatoires dont on prend l'espérance. De la même manière, nous pouvons noter :

$$E_X(g(X)) = E(g(X)) = \int_{-\infty}^{\infty} dx \ g(x) f_X(x)$$

Les quantités $E_X(g(X))$ et $E_{X,Y}(g(X,Y))$ sont des réels. Mais :

$$E_X(g(X,Y)) = \int_{-\infty}^{\infty} dx \ g(x,Y) f_X(x) \ \ \text{est une } \textbf{variable aléatoire} \text{ ! } (*)$$

Attention aux majuscules dans la formule ci-dessus.
Il est alors bénéfique de calculer $E(h(X,Y))$ ainsi parfois :

$$\begin{aligned}
E(h(X,Y)) &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \ h(x,y) f_{X,Y}(x,y) \\
&= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \ h(x,y) f_{Y|X}(y|x) f_X(x) \\
&= \int_{-\infty}^{\infty} dx \ f_X(x) \underbrace{\int_{-\infty}^{\infty} dy \ h(x,y) f_{Y|X}(y|x)}_{=E(h(X,Y) \mid X=x)} \\
&= \int_{-\infty}^{\infty} dx \ E(h(X,Y) \mid X = x) f_X(x) \\
&= E_X\left[E(h(X,Y) \mid X = x)\right]
\end{aligned}$$

La dernière ligne suit en appliquant la formule $(*)$ pour $g(x,Y) = E(h(X,Y)|X = x)$. Un exemple s'impose.
Soit la variable aléatoire $A(\Lambda) \sim \text{Pois}(\Lambda)$ où $\Lambda \sim \text{Pois}(\lambda^*)$ avec $\lambda^* \in \mathbb{R}_+^*$. Calculons $E(A(\Lambda))$ par étapes:

$$\begin{aligned}
E(A(\Lambda) \mid \Lambda = \lambda) &= \Lambda && \text{(espérance d'une Pois}(\lambda)\text{)} \\
E(A) &= E_\Lambda(E(A(\Lambda) \mid \Lambda = \lambda)) \\
&= E_\Lambda(\Lambda) \\
&= \lambda^*
\end{aligned}$$

A titre perso, je déteste la notation $\Lambda = \lambda$ car elle laisse à penser qu'on fixe $\Lambda = \lambda$, alors qu'au sein de $E_\Lambda$, cette écriture signifie juste qu'on calcule $E(A(\lambda))$ successivement pour toutes les valeurs de $\lambda$. Je préconise simplement $E_\Lambda(E(A(\Lambda) \mid \Lambda))$, dans le sens : l'espérance de $A$ lorsque nous fixons $\Lambda$.

# 3 Fonction génératrice de moments

Soit $X$ une variable aléatoire, on définit alors la fonction génératrice de moments de $X$ :

$$M_X(t) = E\left(e^{tX}\right) = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx & \text{si } X \text{ est continue} \\ \sum_{k=-\infty}^{\infty} e^{tk} f_X(k) & \text{si } X \text{ est discrète} \end{cases}$$

tant que $M_X(t) < +\infty$. C'est une sorte de transformée (inverse) de Laplace de $f_X$ - si vous ne prenez ni Analyse IV ni Signals and Systems, la transformée de Laplace ressemble beaucoup à la transformée de Fourier d'Analyse III et partage beaucoup de ses propriétés intéressantes, j'en reparle plus bas.

L'intérêt de la fonction génératrice de moments réside dans son développement en série de Taylor centrée en 0, tout comme la fonction génératrice de probabilités introduite dans la fiche sur les variables aléatoires discrètes. Rappel :

$$e^y = \sum_{k=0}^{\infty} \frac{y^k}{k!}$$

Donc :

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k) \\ &= 1 \quad + tE(X) \quad + \frac{t^2}{2} E(X^2) \quad + \frac{t^3}{6} E(X^3) \quad + \dots \\ &= M_X(0) \quad + tM'_X(0) \quad + \frac{t^2}{2} M''_X(0) \quad + \frac{t^3}{6} M'''_X(0) \quad + \dots \end{aligned}$$

On voit alors par unicité du développement en série de Taylor que

$$E(X^k) = M_X^{(k)}(0).$$

Parmi les autres propriétés intéressantes de la fonction génératrice de moments, si $X$ et $Y$ sont indépendantes :

$$M_{X+Y}(t) = E\left(e^{tX+tY}\right) = E\left(e^{tX}\right) E\left(e^{tY}\right) = M_X(t) M_Y(t)$$

Ainsi, si nous voyons $M_{X+Y}$ comme la transformée inverse de Laplace de $f_{X+Y}$, en utilisant la propriété de convolution de la transformée de Laplace $\mathcal{L}(g * h) = \mathcal{L}(g)\mathcal{L}(h)$ comme pour la transformée de Fourier, on obtient une preuve très rapide de :

$$X, Y \text{ indépendantes} \implies f_{X+Y} = f_X * f_Y$$

Aussi, et surtout,

une distribution est uniquement caractérisée par sa fonction génératrice de moment.

Ceci signifie que : si $X$ et $Y$ ont la même distribution, alors $M_X = M_Y$.
Je trouve que cela donne une intuition profonde sur les variables aléatoires : $M_X$ encode l'information de tous les moments de $X$, i.e tous les $E(X^k)$, et cette information à elle seule détermine complètement la distribution de $X$. Autrement dit, les distributions sont entièrement caractérisées par leurs moments, aussi bien que par leur densité de mesure ($f_X$) ou leur fonction de répartition ($F_X$) !

Enfin, une note sur la convergence en distribution. Soit la suite de variables aléatoires $\{X_n\}_{n \in \mathbb{N}}$ et la variable aléatoire $X$ telle que :

$$\exists \epsilon > 0 \; \forall t \in \, ]-\epsilon; \epsilon[ \; \lim_{n \to \infty} M_{X_n}(t) = M_X(t)$$

Alors :

$$X_n \xrightarrow{D} X, \text{ ie } \forall x \in \mathbb{R} \; \lim_{n \to \infty} F_{X_n}(x) = F_X(x).$$

En fait cette convergence est résultat du même théorème, dit de continuité de Lévy, que le résultat énoncé dans le paragraphe précédent.

Il existe aussi une fonction génératrice de moments pour vecteurs aléatoires. Si $X = (X_1, \ldots, X_p)$ est un vecteur aléatoire sur $\mathbb{R}^p$, alors nous définissons la fonction génératrice de moments de $X$ par :

$$M_X(t) = E(e^{t^T X}) = E\left[\exp\left(\sum_{i=1}^{p} t_i X_i\right)\right]$$

pour tout $t \in \mathbb{R}^p$ tel que la quantité ci-dessus est finie. Aussi, comme dans le cas univarié :

$$E(X) = \nabla M_X(0),\ E(XX^T) = \text{Hess}_{M_X}(0), \ldots$$

# 4 Distribution normale

Il existe nombre de fausses implications en probabilités. Par exemple :

$$\text{Il est } \textbf{vrai} \text{ que : } X, Y \text{ indépendantes } \implies \text{cov}(X,Y) = E(XY^T) - E(X)E(Y)^T = 0$$

Mais :

$$\text{Il est } \textbf{faux} \text{ que : } \text{cov}(X,Y) = 0 \implies X, Y \text{ indépendantes, } \textbf{SAUF...}$$

**sauf** quand $X$ et $Y$ sont normales. N'oubliez jamais que :

$$X, Y \text{ gaussiennes } \textbf{jointes} \implies \left(X, Y \text{ indépendantes } \iff \text{cov}(X,Y) = 0\right)$$

C'était un premier point, important en pratique car il y a assez systématiquement un exercice utilisant ce résultat en examen. Le résultat est faux pour $X, Y$ des gaussiennes non nécéssairement jointes : prenons $X \sim \mathcal{N}(0,1)$ et $Y = ZX$ où $Z = \pm 1$ équiprobablement. On peut calculer que $\text{cov}(X,Y) = 0$ et que $Y \sim \mathcal{N}(0,1)$ mais $(X,Y)$ n'est pas une gaussienne multivariée, i.e $X, Y$ ne sont pas des gaussiennes jointes, et en l'occurence $Y$ et $X$ ne sont pas indépendantes : si l'on connaît $X$, on connaît $Y$ au signe près.

Un second concerne la normale multivariée et ses propriétés, dont je rappelle ici quelques-unes :

- Toute combinaison linéaire (et même affine) de normales est normale.
- Toute marginale de normale multivariée est normale : si $X = (X_1, \ldots, X_d) \sim \mathcal{N}_d(\mu, \Omega)$ alors $X_{\mathcal{I}} = (X_{i_1}, \ldots, X_{i_p}) \sim \mathcal{N}_p(\mu_{\mathcal{I}}, \Omega_{\mathcal{I}})$ avec $\mathcal{I} = \{i_1, \ldots, i_p\} \subseteq [\![1..d]\!]$.

# 5 Exercices

## 5.1 Ex 1 : Calcul d'espérance

Soit $N$ une variable aléatoire discrète dénotant une taille de fichier aléatoire, telle que $E(N) = \lambda > 0$. Soient les $N$ bits de ce fichier : $X_1, \ldots, X_N$ des variables aléatoires de Bernouilli de paramètre $p$. Notons $Y$ le nombre de bits $X_i$ égaux à 1.
Calculer $E(Y)$.

# 6 Solutions

## 6.1 Ex 1

On calcule avec l'espérance en chaîne :

$$\begin{aligned}
E(Y) &= E_N(E(Y \mid N = n)) \\
&= \sum_{n=0}^{\infty} E(Y \mid N = n) f_N(n) \\
&= \sum_{n=0}^{\infty} np f_N(n) \qquad \text{car, en fixant } N = n,\ Y \sim \text{Binomial}(n, p) \\
&= p \sum_{n=0}^{\infty} n f_N(n) \\
&= pE(N) \\
&= p\lambda
\end{aligned}$$