Probabilité et Espérance conditionnelle

Dans ce chapitre, $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé et $\mathcal{G}$ sera une sous-tribu de $\mathcal{F}$. Nous donnerons pour simplifier la notion d'espérance conditionnelle en dimension 1 (variables aléatoires réelles). Bien entendu, tout ceci se généralise sans peine au cas multidimmensionnel (vecteurs aléatoires).

1 Introduction : variables aléatoires discrètes

On se donne deux variables aléatoires discrètes X et Z à valeurs dans $\mathbb{Z}$ à support fini. L'idée intuitive de probabilité conditionnelle nous amène à prendre pour valeur de la "probabilité de $\{X = i\}$ sachant $\{Z = j\}$" la quantité suivante

$$\mathbb{P}(X = i | Z = j) = \frac{\mathbb{P}(X = i,\ Z = j)}{\mathbb{P}(Z = j)}$$

pour j appartenant au support de Z. Si j n'appartient pas au support de Z, cette définition n'a pas de sens. Par suite, l'espérance de X sachant $\{Z = j\}$ sera définie comme

$$\mathbb{E}(X|Z = j) = \sum_{i \in \mathbb{Z}} i\mathbb{P}(X = i|Z = j)$$

Définissons maintenant l'espérance conditionnelle de X sachant Z comme la variable aléatoire $Y(\omega) = \mathbb{E}(X|Z)(\omega)$ donnée par

$$Y(\omega) = \sum_j \mathbb{1}_{Z^{-1}(\{j\})}(\omega)\mathbb{E}(X|Z = j)$$

En particulier, si ω n'appartient pas au support de Z, $Y(\omega) = 0$.
1) Remarquons que pour tout borélien $B \in \sigma(Z)$, c'est-à-dire $B \subset \mathcal{P}(Z^{-1}(\{z\})\ ;\ z \in \text{Support}(Z))$, on a

$$\mathbb{E}(Y \mathbb{1}_B) = \mathbb{E}(X \mathbb{1}_B)$$

2) Y est $\sigma(Z)$ mesurable.

La question qui se pose est de savoir comment généraliser cette notion d'espérance conditionnelle lorsque Z n'est pas à valeurs discrètes. En particulier, on remarqura que si Z est à densité, l'événement $\{Z = z\}$ est pour tout réel z de probabilité nulle si bien que la définition même de probabilité conditionnelle devient délicate. La première idéee consisterait à approximer l'événement $\{Z = z\}$ par $\lim_{\varepsilon \to 0}\{Z \in (z - \varepsilon; z + \varepsilon)\}$ et l'on peut effectivement procéder ainsi mais cela devient assez vite compliqué. On va procéder tout autrement.

2 Théorème d'existence et d'unicité

Théorème 2.1. *[Existence] Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ intégrable et $\mathbb{1}_G$ une sous-tribu de $\mathcal{F}$. Il existe une variable aléatoire Y telle que :*
1) Y est $\mathcal{G}$-mesurable.
2) $\mathbb{E}(|Y|) < +\infty$.
3) $\forall G \in \mathcal{G},\ \mathbb{E}(\mathbb{1}_G Y) = \mathbb{E}(\mathbb{1}_G X)$.

[Unicité] De plus, si $\tilde{Y}$ est une autre variable aléatoire vérifiant ces trois propriétés alors $Y = \tilde{Y}$ $\mathbb{P}$ p.s. On parlera alors de Y comme (d'une version) de l'espérance conditionnelle et on la notera $\mathbb{E}(X|\mathcal{G})$.

Avant d'entamer la preuve de l'existence de l'espérance conditionnelle, faisons quelques remarques.
Notations 1) Quand on considère au lieu de $\mathcal{G}$ la tribu engendrée par une variable aléatoire Z (resp. une famille de variables aléatoires $(Z_i)_{i\in I}$), il est d'usage de noter l'espérance conditionnelle de X sachant $\sigma(Z)$ par $\mathbb{E}(X|Z)$ (resp. $\mathbb{E}(X|Z_i,\ i \in I)$).
On remarquera que $\mathbb{E}(X|Z)$ est $\sigma(Z)$-mesurable donc elle s'écrit sous la forme $f(Z)$ où f est une certaine fonction borélienne.

Remarque 2.1. *Si l'on veut prouver* 3), *on peut se restreindre à montrer l'égalité pour des G qui appartiennent à un π-système contenant Ω et générant $\mathcal{G}$ (Théorème de la classe monotone/ machine standard : Un sous espace vectoriel des focntions mesurables bornées stable par limite croissante et contenant les indicatrices des éléments d'un π système générateur contient toutes les fonctions mesurables bornées). D'autre part, si on a* 3), *par le théorème de la classe monotone, on a aussi que*

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})U) = \mathbb{E}(XU)$$

pour toute variable aléatoire U qui est $\mathcal{G}$-mesurables bornée.

Démonstration. (du théorème (2.1)).Dans la démonstration, nous aurons besoin de considérer les espaces quotients $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ et $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. On a bien sûr l'inclusion triviale $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P}) \subset \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Cette inclusion passe au quotient puisque si $X, Y \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ définissent le même élément dans $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$, c'est que $\{Y \neq X\} \in \mathcal{G}$ est de $\mathbb{P}$ probabilité nulle donc que X et Y définissent aussi le même élément dans $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. D'autre part, $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ est fermé dans $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.
Existence : On commence par traiter le cas où X admet un moment d'ordre 2. $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sev complet de $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ pour la norme $\|\cdot\|_2$ dérivant du produit-scalaire noté $<\cdot,\cdot>$. On désigne alors par Y la projection orthogonale de X sur $\mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. Par définition de la projection orhogonale,

$$\forall Z \in \mathbb{L}^2(\Omega, \mathcal{G}, \mathbb{P}),\ \ < X - Y, Z >= 0$$

donc $\mathbb{E}(X\mathbb{1}_G) = \mathbb{E}(Y\mathbb{1}_G)$ pour tout $G \in \mathcal{G}$. Ceci prouve l'existence de l'espérance conditionnelle pour les variables aléatoires ayant des moments d'ordre 2. Passons maintenant au cas général.
Soit $X \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. En l'écrivant comme différence de sa partie positive et de sa partie négative : $X = X^+ - X^-$, il est clair que l'on peut se ramener au cas où X est positive. On suppose donc maintenant que X est positive. Soit pour tout entier n, $X_n = X \wedge n \in \mathbb{L}^2$. On peut donc par ce

qui précède prendre son espérance conditionnelle par rapport à $\mathcal{G}$, $Y_n = \mathbb{E}(X_n|\mathcal{G})$. D'autre part, X_n tend simplement en croissant vers X et on va voir que de même, $(Y_n)_n$ tend en croissant vers une certaine limite Y.

Lemme 2.1. $0 \leq Y_n \leq Y_{n+1}$, $\mathbb{P}$ *p.s.*

Démonstration. On prouve que si $U \geq 0$ admet un moment d'ordre 2 alors $W = \mathbb{E}(U|\mathcal{G}) \geq 0$. Si $\mathbb{P}(W < 0) > 0$ alors il existe n tel que $\mathbb{P}(W < -1/n) > 0$ et $\{W < -1/n\} \in \mathcal{G}$ donc

$$0 \leq \mathbb{E}(U\mathbb{1}_{W<-1/n}) = \mathbb{E}(W\mathbb{1}_{W<-1/n}) \leq -1/n$$

ce qui est exclu.
On en déduit immédiatement le lemme (par linéarité de l'espérance conditionnelle sur $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$).
□

On pose $Y = \limsup Y_n$ qui est $\mathcal{G}$-mesurable. Pour tout $G \in \mathcal{G}$,

$$\begin{aligned}
\mathbb{E}(Y\mathbb{1}_G) &= \lim_{n\to\infty} \mathbb{E}(Y_n\mathbb{1}_G) \text{ par convergence monotone} \\
&= \lim_{n\to\infty} \mathbb{E}(X_n\mathbb{1}_G) \\
&= \mathbb{E}(X\mathbb{1}_G) \text{ par convergence monotone}
\end{aligned}$$

En particulier, avec $G = \Omega$, $\mathbb{E}(Y) = \mathbb{E}(X) < +\infty$.

Unicité : Soit Y et $\tilde{Y}$ deux variables aléatoires vérifiant $1), 2)$ et $3)$. On veut montrer que $Y = \tilde{Y}$, $\mathbb{P}$ presque sûrement. On a donc :

$$\forall G \in \mathcal{G},\ \mathbb{E}(\mathbb{1}_G(Y - \tilde{Y})) = 0$$

Si $\mathbb{P}(Y = \tilde{Y}) < 1$ alors quitte à permuter Y et $\tilde{Y}$, on a

$$\mathbb{P}(Y > \tilde{Y}) > 0$$

Or $\lim_{n\to\infty} \mathbb{P}(Y - \tilde{Y} > 1/n) = \mathbb{P}(Y > \tilde{Y})$ et il existe donc n tel que $\mathbb{P}(Y - \tilde{Y} > 1/n) > 0$. Mais $G_n = \{Y - \tilde{Y} > 1/n\} \in \mathcal{G}$ par 1) donc

$$\mathbb{E}(\mathbb{1}_{G_n}(Y - \tilde{Y})) = 0 \Rightarrow \frac{1}{n}\mathbb{P}(Y - \tilde{Y} > 1/n) \leq 0 \quad : \text{exclu}$$

Ceci conclut la preuve de l'unicité et du théorème. □

Proposition 2.1. *L'espérance conditionnelle est une forme linéaire positive qui vérifie les propriétés suivantes*

(i) **Emboitement** *Si $\mathcal{H}$ est une sous-tribu de $\mathcal{G}$, alors*

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H}), \quad p.s.$$

(j) **"Sortir ce qui est connu"** *Si Z est $\mathcal{G}$-mesurable et bornée, alors*

$$\mathbb{E}(ZX|\mathcal{G}) = Z\mathbb{E}(X|\mathcal{G}), \ \ p.s.$$

(k) **Rôle de l'indépendance** *Si $\mathcal{H}$ est indépendant de $\sigma(\sigma(X),\mathcal{G})$, alors*

$$\mathbb{E}(X|\sigma(\mathcal{G},\mathcal{H})) = \mathbb{E}(X|\mathcal{G}), \ \ p.s.$$

En particulier, si X est indépendant de $\mathcal{H}$, alors $\mathbb{E}(X|\mathcal{H}) = E(X)$.

Remarque 2.2. *1) Quand on considère au lieu de $\mathcal{G}$ la tribu engendrée par une variable aléatoire Z (resp. une famille de variables aléatoires $(Z_i)_{i\in I}$), il est d'usage de noter l'espérance conditionnelle de X sachant $\sigma(Z)$ par $\mathbb{E}(X|Z)$ (resp. $\mathbb{E}(X|Z_i, \ i \in I)$).*
On remarquera que $\mathbb{E}(X|Z)$ est $\sigma(Z)$-mesurable donc elle s'écrit sous la forme $f(Z)$ où f est une certaine fonction borélienne.

2.1 Variables à densité

Soit X et Z deux variables aléatoires réelles (pour simplifier les écritures) telles qu'elles aient une densité jointe $f_{X,Z}(x,z)$. On rappelle que X (resp. Z) a alors une densité notée $f_Z(z) = \int f_{X,Z}(x,z)dx$ (resp. $f_X(x) = \int f_{X,Z}(x,z)dz$). On pose alors

$$f_{X|Z}(x|z) = \mathbb{1}_{f_Z(z)\neq 0}\frac{f_{X,Z}(x,z)}{f_Z(z)}$$

Soit h une fonction borélienne (bornée pour simplifier) ; quelle est l'espérance conditionnelle de $h(X)$ sachant Z ?
Si g est une fonction borélienne bornée, par le théorème de Fubini, on a la suite d'égalités suivantes :

$$\begin{aligned}
\mathbb{E}(h(X)g(Z)) &= \int h(x)g(z)f_{X,Z}(x,z)dxdz \\
&= \int_{\mathbb{R}^2} g(z)f_Z(z)f_{X|Z}(x,z)h(x)dxdz \\
&= \int_{\mathbb{R}} g(z)f_Z(z)\left(\int_{\mathbb{R}} f_{X|Z}(x,z)h(x)dx\right) \\
&= \mathbb{E}(g(Z)\theta(Z))
\end{aligned}$$

où $\theta(z) = \int_{\mathbb{R}} f_{X|Z}(x,z)h(x)dx$. On a donc prouvé que

$$\mathbb{E}(h(X)|Z)(\omega) = \theta(Z(\omega)) = \int_{\mathbb{R}} f_{X|Z}(x,Z(\omega))h(x)dx$$

2.2 Variables discrètes

En se reportant à la section 1, on vérifiera sans peine que si X et Z sont deux variables aléatoires discrètes à valeurs dans $\mathbb{Z}$, on a :

$$\mathbb{E}(X|Z) = \sum_j j\left(\sum_i \mathbb{1}_{Z=i}\mathbb{P}(X=j|Z=i)\right)$$

Le terme

$$\sum_j j\mathbb{1}_{\{Z=i\}}\mathbb{P}(X=j|Z=i)$$

est souvent noté

$$\mathbb{E}(X|Z=i)$$

3 Noyaux et lois conditionnelles

Ce paragraphe est tiré de Ouvrard Tome II et donné sans la démonstration des propriétés. On s'y reportera pour de plus amples détails (mais ce qui est ici doit être su). La notion de loi conditionnelle est plus profonde que celle d'espérance conditionnelle qui n'est d'ailleurs que "l'espérance par rapport à la loi conditionnelle".
$(E,\mathcal{E})$ et $(F,\mathcal{F})$ sont deux espaces probabilisés (qui seront ici en fait des $(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$).

Définition 3.1. $\nu : E\times\mathcal{F}\to[0,1]$ *est un noyau (ou probabilité) de transition si il satisfait :*
1. $\forall x\in E,\ \nu(x,\cdot)$ *est une probabilité sur* $(F,\mathcal{F})$*.*
2. $\forall B\in\mathcal{F},\ \nu(\cdot,B)$ *est* $\mathcal{E}$*-mesurable.*

Définition 3.2. *Si* $X\in\mathbb{R}^d$ *et* $Y\in\mathbb{R}^k$ *sont deux variables aléatoires, on appelle loi conditionnelle de* Y *sachant* X *un noyau* ν *sur* $(\mathbb{R}^d\times\mathcal{B}(\mathbb{R}^k))$ *tel que :*

$$\mathbb{P}_{(X,Y)}=\mathbb{P}_X\cdot\nu$$

c'est-à-dire

$$\mathbb{P}(X\in A;Y\in B)=\int_{x\in A}\nu(x,B)\mathbb{P}_X(dx)$$

On notera, **et ce n'est qu'une notation**, un tel noyau ν de la manière suivante :

$$\nu(x,B)=\mathbb{P}(Y\in B|X=x)\ \text{ ou }\ \nu(x,B)=\mathbb{P}_Y(B|X=x)$$

Remarquer par exemple que si X est à densité, $\{X=x\}$ est un ensemble de probabilité nulle. Il est donc hors de question de la définir par :

$$\mathbb{P}(Y\in B|X=x)=\frac{\mathbb{P}(Y\in B;X=x)}{\mathbb{P}(X=x)}$$

Cependant cette formule est vraie dans le cas où $\mathbb{P}(X=x)\neq 0$.

Il reste à savoir quand on a l'existence de lois conditionnelles. La réponse à cette question est apportée par le théorème non trivial suivant.

Théorème 3.1. ***: théorème de Jirina***
Si X *et* Y *sont deux variables aléatoires à valeurs dans des* $(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))$ *alors il existe toujours une loi conditionnelle de* Y *sachant* X*.*

Le théorème suivant est une généralisation du théorème de Fubini dans le cadre conditionnel.

Théorème 3.2. *Soit* (X, Y) *une variable aléatoire à valeurs dans un espace probabilisable quelconque* $(E \times F, \mathcal{E} \otimes \mathcal{F})$ *telle qu'existe une loi conditionnelle* $\mathbb{P}(Y \in \cdot | X = \cdot)$ *de* Y *sachant* X. *Soit* f *une application mesurable de l'espace probabilisable* $(E \times F, \mathcal{E} \otimes \mathcal{F})$ *dans* $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$.
a) Si f *est positive, l'application* $x \to \int_F f(x, y) d\mathbb{P}_Y(dy|X = x)$ *est* $\mathcal{E}$*-mesurable et on a :*

$$\int_{E \times F} f(x, y)\mathbb{P}_{(X,Y)}(dxdy) = \int_E \left[\int_F f(x, y)\mathbb{P}_Y(dy|X = x) \right] \mathbb{P}_X(dx)$$

b) Si f *est de signe quelconque et* $\mathbb{P}_{(X,Y)}$*-intégrable, pour* $\mathbb{P}_X$ *presque tout* x, *l'application partielle* $f(x, \cdot)$ *est* $\mathbb{P}_Y(\cdot|X = x)$*-intégrable, et l'application définie* $\mathbb{P}_X$*-presque sûrement par* $x \to \int_F f(x, y) d\mathbb{P}_Y(dy|X = x)$ *est* $\mathbb{P}_X$*-intégrable et l'égalité précédente est encore vraie.*

Il est alors aisé d'établir le corrolaire suivant.

Proposition 3.1. *(Propriété de transfert conditionnel)*
Si $X \in \mathbb{R}^p$ *et* $Y \in \mathbb{R}^k$ *sont deux variables aléatoires et* $f : \mathbb{R}^{p+k} \to \mathbb{R}^q$ *borélienne, alors*

$$\mathbb{P}(f(X, Y) \in \cdot|X = x) = \mathbb{P}(f(x, Y) \in \cdot|X = x)$$

En particulier, si X *et* Y *sont indépendantes, on a*

$$\mathbb{P}(f(X, Y) \in \cdot|X = x) = \mathbb{P}(f(x, Y) \in \cdot)$$

Terminons par une définition avant de faire le lien avec l'espérance conditionnelle.

Définition 3.3. *Soit* Y *une variable aléatoire et* $(X_1, \dots, X_n)$ *une suite finie de variables aléatoires. On dira que relativement à* $Y = y$ *les variables aléatoires* $(X_1, \dots, X_n)$ *ont pour loi* μ_y *(probabilité sur* $\mathbb{R}^n$ *dépendant de* y*) ssi il existe une loi conditionnelle de* $(X_1, \dots, X_n)$ *relativement à* Y *telle que* $\mathbb{P}_Y$ *p.s.*

$$\mathbb{P}_{(X_1,\dots,X_n)}(\cdot|Y = y) = \mu_y(\cdot)$$

En particulier, on dira que $(X_1, \dots, X_n)$ *sont indépendantes conditionnellement à* $Y = y$ *ssi la loi conditionnelle de* $(X_1, \dots, X_n)$ *relativement à* Y *est une loi produit.*

L'espérance conditionnelle de la variable aléatoire X relativement à une autre variable aléatoire Y est

$$\mathbb{E}(X|Y)(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(dx|Y = Y(\omega))$$

Il suffit en effet de remarquer que la fonction définie par le membre de droite vérifie les propriétés caractéristiques de l'espérance conditionnelle. L'espérance conditionnelle apparaît donc comme "l'espérance relativement à la loi conditionnelle".

Exercice 3.1. Une poule pond N oeufs, N suivant une loi de Poisson de paramètre λ, c'est-à-dire $f_N(n) = \mathbb{P}(N = n) = e^{-\lambda}\frac{\lambda^n}{n!}$. Chaque oeuf éclôt avec probabilité p, indépendamment des autres oeufs. On note $q = 1-p$. Soit K le nombre de poussins. Notre but est de calculer, puis d'interpréter les valeurs de $\mathbb{E}(K|N)$, $\mathbb{E}(K)$ et $\mathbb{E}(N|K)$.

1) Montrer que $\mathbb{P}(K = k|N = n) = C_n^k p^k (1 - p)^{n-k}$ et que $\mathbb{E}(K|N) = pN$. En déduire que $\mathbb{E}(K) = p\lambda$. Interpréter.

2) Si vous ne l'avez jamais vue, montrer la règle de Bayes (disposer d'un moyen mnémotechnique pour s'en souvenir).

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

3) En utilisant la règle de Bayes, montrer que si $n \geq k$,

$$\mathbb{P}(N = n|K = k) = \frac{(q\lambda)^{n-k}}{(n-k)!}e^{-q\lambda}.$$

Conclure que $\mathbb{E}(N|K) = K + q\lambda$. Interpréter.

Solution de 1) Le loi du nombre de poussins K sachant le nombre d'oeufs N égale à n est la loi binomiale de paramètre (n, p).

Solution de 3).

$$\mathbb{P}(N = n \mid K = k) = \frac{\mathbb{P}(K = k \mid N = n)\mathbb{P}(N = n)}{\mathbb{P}(K = k)} =$$

$$\frac{C_n^k p^k (1-p)^{n-k} (\frac{\lambda}{n!}) e^{-\lambda}}{\sum_{m \geq k} C_m^k p^k (1-p)^{m-k} (\frac{\lambda^m}{m!}) e^{-\lambda}} = \frac{(q\lambda)^{n-k}}{(n-k)!}e^{-q\lambda}.$$

Donc

$$\mathbb{E}(N|K = k) = \sum_{n \geq k} n \frac{(q\lambda)^{n-k}}{(n-k)!}e^{-q\lambda} = k + q\lambda.$$

Donc $\mathbb{E}(N|K) = K + q\lambda$. On a $\mathbb{E}(N) = \lambda$, $\mathbb{E}K = p\lambda$. A titre de vérification des résultats, remarquer que

$$\lambda = \mathbb{E}N = \mathbb{E}(\mathbb{E}(N|K)) = \mathbb{E}(K + q\lambda) = p\lambda + q\lambda = \lambda.$$

L'espérance du nombre d'oeufs sachant le nombre d'oeufs éclos est égale à la somme du nombre, connu, d'oeufs éclos, K, plus l'espérance du nombre d'oeufs non éclos, soit $q\lambda$.

Exercice 3.2. Preuve de la formule générale caractérisant l'espérance conditionnelle dans le cas discret On considère deux v.a. discrètes X et Y et on définit

$$\phi(x) := \sum_y \mathbb{P}(Y = y|X = x), \quad Z := \phi(X).$$

Montrer que Z vérifie

$$\mathbb{E}(Zg(X)) = \mathbb{E}(Yg(X)) \tag{3.1}$$

pour toute fonction g telle que les deux espérances existent. Montrer qu'il y a une unique fonction ψ telle que $Z = \psi(X)$ vérifie (4.1).
Donc $Z = \phi(X) = \mathbb{E}(Y|X)$.

Solution :

$$\mathbb{E}(\mathbb{E}(Y|X)g(X)) = \sum_x yg(x)f_{Y|X}(y|x)f_X(x) = \sum_{x,y} yg(x)f_{X,Y}(x,y) = \mathbb{E}(Yg(X)).$$

Unicité : soit $Z = \psi(X)$ vérifiant (4.1). On choisit $g(X) = \mathbb{1}_{X=x_0}$. Cela donne

$$\psi(x_0)\mathbb{P}(X = x_0) = \sum_{x,y} y\mathbb{1}_{x=x_0}\mathbb{P}(X = x; Y = y) = \sum_y y\mathbb{P}(X = x_0; Y = x).$$

Donc

$$\psi(x_0) = \sum_y y\mathbb{P}(Y = y|X = x_0),$$

ce qui est bien la définition.

Exercice 3.3. (Grimmett-Stirzaker, exercice 3.7.7 p. 69). Une usine produit n lampes. Chacune est défectueuse avec probabilité ϕ. On teste chaque lampe pour détecter un éventuel défaut. La probabilité de détecter le défaut quand il est présent est δ. On note X le nombre d'ampoules defectueuses et Y le nombre d'ampoules détectées comme defectueuses. Montrer que

$$\mathbb{E}(X|Y) = \frac{n\phi(1-\delta) + (1-\phi)Y}{1-\phi\delta}.$$

Solution : Une lampe non detectée défecteuse est en réalité défectueuse avec probabilité $\pi = \frac{\phi(1-\delta)}{1-\phi\delta}$. *Donc le nombre de lampes défectueuses sachant* Y *suit une binomiale* $\mathbb{B}(n-Y,\pi)$ *et a pour moyenne* $\frac{(n-Y)(phi(1-\delta))}{1-\phi\delta}$. *Donc*

$$\mathbb{E}(X|Y) = Y + \frac{(n-Y)\phi(1-\delta)}{1-\phi\delta}.$$

Exercice 3.4. Les propriétés fondamentales : les lire (Williams, Probability with martingales p. 88 et la dernière page) Soit $(\Omega,\ \mathcal{F},\ \mathbb{P})$ un espace de probabilité. On suppose que X, $\mathcal{F}$-mesurable, vérifie $\mathbb{E}|X| < \infty$. $\mathcal{G}$ et $\mathcal{H}$ sont des sous-tribus de $\mathcal{F}$. On rappelle qu'il existe une unique v.a. $\mathcal{G}$-mesurable et sommable, notée $\mathbb{E}(X|G)$ telle que

$$\forall G \in \mathcal{G},\ \int_G \mathbb{E}(X|\mathcal{G})d\mathbb{P} = \int_G Xd\mathbb{P}. \tag{3.2}$$

(L'unicité est vraie modulo un ensemble de mesure nulle). De manière plus condensée :

$$\forall G \in \mathcal{G},\ \mathbb{E}(\mathbb{E}(X|\mathcal{G}); G) = E(X;\mathcal{G}),$$

où $E(X;G) = \int_G Xd\mathbb{P}$.

Montrer les propriétés suivantes :

(a) On pose $Y = \mathbb{E}(X|\mathcal{G})$, alors $\mathbb{E}(Y) = \mathbb{E}(X)$.

(b) Si X est $\mathcal{G}$-mesurable, alors $\mathbb{E}(X|\mathcal{G}) = X$ p.p.

(c) **Linéarité** $\mathbb{E}(aX_1 + a_2X_2|\mathcal{G}) = a_1\mathbb{E}(X_1|\mathcal{G}) + a_2\mathbb{E}(X_2|\mathcal{G})$, p.p.

(d) **Positivité** Si $X \geq 0$, alors $\mathbb{E}(X|\mathcal{G}) \geq 0$, p.p.

(e) **C-Convergence monotone** Si $0 \leq X_n \uparrow X$, alors $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$, p.p.

(f) **C-Fatou** Si $X_n \geq 0$, alors $\mathbb{E}(\liminf X_n|\mathcal{G}) \leq \liminf \mathbb{E}(X_n|\mathcal{G})$, p.p.

(g) **C-Convergence dominée** Si $\forall n\ |X_n(\omega)| \leq V(\omega)$, $\mathbb{E}V < \infty$ et $X_n \to X$ p.p., alors

$$\mathbb{E}(|X_n - X|\ |\mathcal{G}) \to 0,\ p.p.$$

(Par Jensen juste après, on déduit que $\mathbb{E}(X_n\ |\mathcal{G}) \to \mathbb{E}(X\ |\mathcal{G}),\ p.p.$

(h) **C-Jensen** Si $c : \mathbb{R} \to \mathbb{R}$ est convexe et $\mathbb{E}|c(X)| < \infty$, alors

$$\mathbb{E}(c(X)|\mathcal{G}) \geq c(\mathbb{E}(X|\mathcal{G})), \quad \text{p.p.}$$

Corollaire : $||\mathbb{E}(X|\mathcal{G})||_p \leq ||X||_p$.

(i) **Emboitement** Si $\mathcal{H}$ est une sous-tribu de $\mathcal{G}$, alors

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|H), \text{ p.p.}$$

(j) **"Sortir ce qui est connu"** Si Z est $\mathcal{G}$-mesurable et bornée, alors

$$\mathbb{E}(ZX|\mathcal{G}) = Z\mathbb{E}(X|\mathcal{G}), \text{ p.p. .}$$

Généralisations : encore vrai si $X \in L^p$ et $Z \in L^{p'}$, ou si $X \in (m\mathcal{F})^+$, $Z \in (m\mathcal{G})^+$, $\mathbb{E}(X) < \infty$ et $\mathbb{E}(ZX) < \infty$.

(k) **Rôle de l'indépendance** Si $\mathcal{H}$ est indépendant de $\sigma(\sigma(X), \mathcal{G})$, alors

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X|\mathcal{G}), \text{ p.p.}$$

En particulier, si X est indépendant de $\mathcal{H}$, alors $\mathbb{E}(X|\mathcal{H}) = E(X)$.

Indications :

(a) Provient de la relation fondamentale (4.2) appliquée à $G = \Omega$.

(b) Découle de l'unicité de la fonction $\mathcal{G}$-mesurable satisfaisant (4.2)*, i.e. de l'unicité de l'espérance conditionnelle.*

(c) De même (expliquer !).

(d) Considérer l'ensemble $G = \{\mathbb{E}(X|\mathcal{G}) < -\frac{1}{n}\}$ et appliquer la relation fondamentale.

(e) Poser $Y = \limsup \mathbb{E}(X_n|\mathcal{G})$, montrer qu'elle est $\mathcal{G}$-mesurable, qu'elle est la limite presque partout des $\mathbb{E}(X_n|\mathcal{G})$. Montrer que $\lim_n \mathbb{E}(\mathbb{E}(X_n|\mathcal{G}); G) = \mathbb{E}(Y; G)$ et conclure.

(f) Reprendre la démonstration qui dérive le lemme de Fatou du théorème de converge monotone pour déduire **C-Fatou** *de* **C-convergence monotone**. *(Poser $Y_n = \mathbb{E}(X_n, \mathcal{G})$, puis $Z_n = \inf_{k \geq n} Y_n$ et lui appliquer* **C-monotone**.*)*

(g) Reprendre la démonstration qui déduit le théorème de convergence monotone du lemme de Fatou pour déduire **C-convergence dominée** *de* **C-Fatou**.

(h)

(h1) Soit c : $\mathbb{R} \to \mathbb{R}$ convexe. Montrer qu'il existe une suite dénombrable $(a_n, b_n)_n \in \mathbb{R}^2$ telle que

$$c(x) = \sup_n (a_n x + b_n), \; x \in \mathbb{R}.$$

(Remarquer que c est continue, d'épigraphe donc fermé, et utiliser par exemple Hahn-Banach : Il y a un hyperplan séparant tout point n'appartenant pas à un convexe fermé de ce convexe).

(h2) Déduire que

$$\mathbb{E}(c(X)|\mathcal{G}) \geq a_n \mathbb{E}(X|\mathcal{G}) + b_n$$

et conclure. Pour montrer l'extension de (h), prendre $c(x) = |x|^p$.

(i) Remaquer que pour tout $H \in \mathcal{H}$, $\mathbb{E}(\mathbb{1}_H X) = \mathbb{E}(\mathbb{1}_H \mathbb{E}(X \mid \mathcal{G}))$ puisque $H \in \mathcal{G}$. Donc par unicité, $0 = \mathbb{E}(X - \mathbb{E}(\mathbb{1}_H \mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}))$.

(j) La propriété est linéaire : Appliquer la "machine standard", c'est-à-dire montrer que la propriété est vraie quand Z est une fonction indicatrice d'élément G de $\mathcal{G}$, puis une fonction en escalier positive, puis une fonction $\mathcal{G}$-mesurable positive quelconque, puis une fonction $\mathcal{G}$-mesurable quelconque telle que $\mathbb{E}(|ZX|) < \infty$. Vérifier que la preuve s'applique aux généralisations indiquées.

(k) Supposer s.p.d.g. que $X \geq 0$ et $\mathbb{E}(X) < \infty$.

(k1) Montrer que si $G \in \mathcal{G}$ et $H \in \mathcal{H}$, alors

$$\mathbb{E}(X; G \cap H) = E(X \mathbb{1}_G)\mathbb{P}(H).$$

(k2) Poser $Y = \mathbb{E}(X|\mathcal{G})$. Montrer que

$$\mathbb{E}((Y \mathbb{1}_G)\mathbb{1}_H) = \mathbb{E}(Y \mathbb{1}_G)\mathbb{P}(H).$$

(k3) Déduire que les mesures $F \to \mathbb{E}(X; F)$ et $F \to \mathbb{E}(Y; F)$ définies sur $\sigma(\mathcal{G}, \mathcal{H})$ coïncident sur le π-système des ensembles de la forme $G \cap H$ ($G \in \mathcal{H}$), $H \in \mathcal{H}$) et donc sur $\sigma(\mathcal{G}, \mathcal{H})$. Conclure.

Exercice 3.5. Comprendre de quoi on parle
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $A_1, \dots,\ A_n$ une partition de Ω en ensembles mesurables et $\mathcal{G}$ la sous-tribu engendrée par ces ensembles. Décrire cette tribu et calculer $E(X|\mathcal{G})$, pour X $\mathcal{F}$-mesurable.

Exercice 3.6. Lois de Bernoulli et uniforme (Ouvrard Tome II Ex. 11.2 p.171)
Soient n variables aléatoires $X_1, X_2, \dots, X_n$ indépendantes de même loi de Bernoulli de paramètre $p \in (0,1)$. On note X le vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ à valeurs dans $\{0,1\}^n$ et $S_n = \sum_{i=1}^n X_i$. Montrer que conditionnellement à $S_n = s$, le vecteur X est réparti selon une loi uniforme sur l'hyperplan $\{(k_1, k_2, \dots, k_n) | \sum_{i=1}^n k_i = s\}$ pour $0 \leq s \leq n$.

Exercice 3.7. Application importante
Soient $X_1,\ X_2, \dots X_n$ des variables aléatoires i.i.d. sommables et $S_n = X_1 + \dots\ + X_n$. On va montrer que $\mathbb{E}(X_1|S_n) = \dots = \mathbb{E}(X_n|S_n) = \dfrac{S_n}{n}$. On pose

$$\mathcal{G}_n = \sigma(S_n,\ S_{n+1}, \dots) = \sigma(S_n, X_{n+1}, X_{n+2}, \dots).$$

1) Montrer en utilisant (k) que

$$\mathbb{E}(X_1|\mathcal{G}_n) = \mathbb{E}(X_1|S_n).$$

2) Montrer que $\mathbb{E}(X_1|S_n) = \dots = \mathbb{E}(X_n|S_n)$.
(On montrera que pour tout borélien B, $\mathbb{E}(X_1; S_n \in B) = \dots = \mathbb{E}(X_n; S_n \in B)$.)

3) En déduire que

$$\mathbb{E}(X_1|S_n) = \dots = \mathbb{E}(X_n|S_n) = \frac{1}{n}\mathbb{E}(X_1 + \dots + X_n|S_n) = \frac{1}{n}S_n.$$

Exercice 3.8. Densités conditionnelles 1

Théorème 3.3. rappel (Williams, Probability with martinagles, p. 87) et application : Brémaud P., *Introduction aux probabilités : Modélisation des phénomènes aléatoires* p. 212.

Soient X et Z deux v.a. qui ont une densité de probabilité jointe $f_{X,Z}(x,z)$. Alors $f_Z(z) = \int_{\mathbb{R}} f_{X,Z}(x,z)dx$ est une densité de probabilité pour Z. On définit la densité conditionnelle $f_{X|Z}$ de X sachant Z par

$$f_{X|Z}(x|z) := \frac{f_{X,Z}(x,z)}{f_Z(z)} \text{ si } f_Z(z) \neq 0, = 0 \text{ sinon.}$$

On pose alors

$$g(z) := \int_{\mathbb{R}} h(x) f_{X|Z}(x|z)dx.$$

Alors l'espérance conditionnelle de $h(X)$ sachant $\sigma(Z)$ est égale à $g(Z)$:

$$\mathbb{E}(h(X)|\sigma(Z)) = g(Z) = \int_{\mathbb{R}} h(x) f_{X|Z}(x|Z)dx.$$

Soient X, Y, Z trois variables aléatoires réelles telles que
1) X est uniformément répartie sur $[0, 1]$;
2) Sachant que $X = x$, Y admet une densité conditionnelle $f_{Y|X}(y|x)$ donnée par

$$f_{Y|X}(y|x) = \begin{cases} (y-x)e^{-(y-x)} & \text{si } \ y > x \\ 0 \ \text{ sinon} \end{cases}$$

3) Sachant que $X = x$ et $Y = y$, Z admet une densité conditionnelle $f_{Z|(X,Y)}(z|(x, y))$ définie par

$$f_{Z|(X,Y)}(z|(x,y)) = \begin{cases} (y-x)e^{-z(y-x)} & \text{si } \ z > 0 \\ 0 \ \text{ sinon} \end{cases}$$

pour $(x, y) \in A = \big\{(x, y) \in \mathbb{R}^2 | x \in [0, 1] \ \text{ et } \ y > x\big\}$.

1) Montrer que (X, Y, Z) admet une densité sur $\mathbb{R}^3$ donnée par

$$f_{(X,Y,Z)}(x,y,z) = (y-x)^2 e^{-(y-x)(z+1)} \mathbb{1}_B(x,y,z)$$

où $B = \big\{(x, y, z) \in \mathbb{R}^3 | x \in [0, 1], \ y > x, \ z > 0\big\}$.

2) Montrer que la loi de Z est donnée par la densité

$$f_Z(z) = \frac{2}{(z+1)^3} \mathbb{1}_{z>0}$$

3) Montrer que la loi conditionnelle de (X, Y) sachant $Z = z$ est donnée par la densité conditionnelle

$$f_{(X,Y)|Z}((x,y)|z) = \frac{1}{2}(z+1)^3 (y-x)^2 e^{-(y-x)(z+1)} \mathbb{1}_A(x,y)$$

4) Calculer $\mathbb{E}(\sqrt{Y-X}|Z=z)$, puis $\mathbb{E}(\sqrt{Y-X})$.

5) On pose $U = Y - X$ et $V = Z(Y-X)$. Montrer que (X,U,V) admet pour densité

$$f_{(X,U,V)}(x,u,v) = \mathbb{1}_{[0,1]}(x) u e^{-u} \mathbb{1}_{u>0,v>0}$$

6) X, U, V sont-elles indépendantes ?

Exercice 3.9. Densités conditionnelles 2

Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de densité commune $f(x)$ et de fonction de répartition $F(x) = \int_{-\infty}^{x} f(u)du$. On pose

$$Y = \max(X_1, \ldots, X_n) \text{ et } Z = \min(X_1, \ldots, X_n).$$

1) En remarquant que $F_{Y,Z}(y,z) = \mathbb{P}(Y \leq y) - \mathbb{P}(Y \leq y, Z > z)$, montrer que

$$F_{Y,Z}(y,z) = F(y)^n - (F(y) - F(z))^n \text{ si } y \geq z,$$

$$\text{et } F_{Y,Z}(y,z) = F(y)^n \text{ si } y < z.$$

2) En déduire que

$$f_{Y,Z}(y,z) = n(n-1)(F(y) - F(z))^{n-2} f(y) f(z) \mathbb{1} y \geq z.$$

3) En déduire une formule intégrale pour $\mathbb{E}(Z|Y)$.

4) Si les X_n sont équiréparties sur $[0,1]$, en déduire que

$$f_{Z|Y}(z,y) = (n-1)\frac{(y-z)^{n-2}}{y^n} \mathbb{1}_{0 \leq z \leq y \leq 1}.$$

Exercice 3.10. Processus de Poisson (Ouvrard, exercice 11.3, pp. 171-177)

Soit $(W_n)_{n \in \mathbb{N}^*}$ une suite croissante de variables aléatoires positives telles que $W_0 = 0$. Soit, pour $n \in \mathbb{N}^*$, la variable aléatoire $T_n = W_n - W_{n-1}$. On suppose que les variables aléatoires T_n forment une famille de variables aléatoires indépendantes de même loi exponentielle $\exp(\lambda)$ où λ est un réel strictement positif. On pose $X_0 = 0$ et pour tout $t > 0$,

$$X_t = \sum_{n \in \mathbb{N}^*} 1_{\{W_n \leq t\}}$$

La famille de variables aléatoires $(X_t)_{t \in \mathbb{R}^+}$ est appelée processus de Poisson d'intensité λ.

1) Montrer que $\mathbb{P}$ presque sûrement $t \to X_t$ est une fonction càdlàg (continue à droite avec des limites à gauches) croissante et que les points de discontinuités sont les W_n.

2) Soit s, t tels que $0 \leq s < t$. Calculer l'intégrale définie pour tout $n \in \mathbb{N}^*$ par

$$I_n(s,t) = \int_{\mathbb{R}}^{n} 1_{(s \leq x_1 \leq \ldots \leq x_n \leq t)} dx_1 \ldots dx_n$$

3) a) Calculer pour tout $n \in \mathbb{N}^*$ et toute famille $(f_j)_{j\in\{1,\ldots,n}$ de fonctions positives mesurables bornées sur $\mathbb{R}$ la quantité

$$\mathbb{E}\left[1_{(X_t=n)}\prod_{j=1}^n f_j(W_j)\right]$$

En déduire la loi de X_t et une loi conditionnelle de $(W_1,\ldots,W_n)$ sachant $\{X_t = n\}$.
b) En utilisant la loi des T_i et leur indépendance, montrer que

$$\begin{aligned}\mathbb{E}\left[1_{(X_t=n)}\prod_{j=1}^n f_j(W_j)\right] &= \int_{(\mathbb{R}_+)^{n+1}}\left[1_{(\sum_{i=1}^n t_i\le t)\cap(\sum_{i=1}^{n+1} t_i>t)}\prod_{j=1}^n f_j\left(\sum_{i=1}^j t_i\right)\right]\\ &\times\ \lambda^{n+1}\exp\left(-\lambda\sum_{i=1}^{n+1} t_i\right)dt_1\ldots dt_{n+1}\end{aligned}$$

Cette formule démontre alors deux choses. D'une part l'indépendance des variables aléatoires $(X_{t_{i+1}} - X_{t_i})_{i=0,\ldots,k-1}$ et d'autre part que la loi de $X_t - X_s$ est une loi de Poisson de paramètre $\lambda(t-s)$.

4) Soient $\alpha_1,\alpha_2,\ldots,\alpha_k$ des entiers positifs quelconques. Notons n leur somme. On pose $l_j = \alpha_1+\ldots+\alpha_j$. En utilisant la question précédente, montrer

$$\mathbb{P}\left[\cap_{j=1}^k(X_{t_j} - X_{t_{j-1}} = \alpha_j))\right] = \lambda^n\exp(-\lambda t)\psi(t_1,\ldots,t_k)$$

où

$$\Psi(t_1,\ldots,t_k) = \int_{\mathbb{R}^n}\prod_{j=1}^{k-1}1_{(w_{l_j}\le t_j)\cap(w_{l_{j+1}}>t_j)}\times 1_{(0\le w_1\le w_2\ldots\le w_n\le t)}dw_1\ldots dw_n$$

En déduire

$$\mathbb{P}\left[\cap_{j=1}^k(X_{t_j} - X_{t_{j-1}} = \alpha_j)\right] = \prod_{j=1}^k \exp(-\lambda(t_j-t_{j-1}))\frac{(\lambda(t_j-t_{j-1}))^{\alpha_j}}{\alpha_j!}$$

Solutions :
2) On pourra soit procéder par récurrence comme Ouvrard, soit remarquer que si $\mathcal{S}_n$ désigne le groupe des permutations sur $\{1,\ldots,n\}$, on a

$$\forall\sigma\in\mathcal{S}_n,\ I_n(s,t) = \int_{\mathbb{R}^n}1_{(s\le x_{\sigma(1)}\le\ldots\le x_{\sigma(n)}\le t)}dx_1\ldots dx_n$$

et

$$\sum_{\sigma\in\mathcal{S}_n}1_{(s\le x_{\sigma(1)}\le\ldots\le x_{\sigma(n)}\le t)} = 1_{(x_1,\ldots,x_n)\in[s,t]^n}.$$

3) a) Commencer par remarquer que $\{X_t = n\} = \{W_n \le t\}\cap\{W_{n+1} > t\}$ et remplacer dans la quantité à calculer les W_k par $\sum_{i=1}^k T_i$. b) Faire le changement de variables $w_1 = t_1, w_2 =$

$t_1 + t_2, \ldots, w_{n+1} = t_1 + \ldots + t_{n+1}$ *et en déduire*

$$\begin{aligned}\mathbb{E}\left[1_{(X_t=n)} \prod_{j=1}^{n} f_j(W_j)\right] &= \lambda^n \exp(-\lambda t) \\ &\times \left[\int_{\mathbb{R}^n} \prod_{j=1}^{n} f_j(w_j) 1_{(0 \leq w_1 \leq w_2 \ldots \leq w_n \leq t)} dw_1 \ldots dw_n\right]\end{aligned}$$

En faisant dans cette formule, $f_j = 1$, montrer que X_t suit une loi de Poisson de paramètre λt. En déduire que $(W_1, \ldots, W_n)$ admet une densité conditionnelle sachant $(X_t = n)$ donnée par

$$f_{(W_1,\ldots,W_n)}(w_1, \ldots, w_n | X_t = n) = \frac{n!}{t^n} 1_{(0 \leq w_1 \leq w_2 \ldots \leq w_n \leq t)} \text{ \textit{(loi de Dirichlet)}}$$