Probabilités IV

STEP, MINES ParisTech*

1 décembre 2019 (`#dd190dd`)

Table des matières

Nous nous sommes intéressé jusqu'à présent à l'étude de variables aléatoires indépendantes. En pratique cependant, on rencontre souvent des variables dépendantes les unes des autres. Dans le cas de la météo, les variables température, vitesse du vent et pression en fournissent un exemple. Nous allons nous attacher dans ce chapitre à décrire les **lois conditionnelles** qui vont nous permettre de résumer l'information apportée par une variable (ou vecteur) sur une autre et nous intéresser en particulier à l'**espérance conditionnelle** qui nous indiquera le comportement moyen d'une variable conditionnellement à une autre.

*Ce document est un des produits du projet `boisgera/CDIS`, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS).

Lois conditionnelles dans un couple

Soient deux variables aléatoire X et Y définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Dans le cas où X et Y sont indépendantes, on a vu que pour tous boréliens de $\mathbb{R}$ B_1 et B_2\$, on a

$$\mathbb{P}(X \in \mathcal{B}_1, Y \in \mathcal{B}_2) = \mathbb{P}(X \in \mathcal{B}_1)\mathbb{P}(Y \in \mathcal{B}_2) = \mathbb{P}_X(B_1)\mathbb{P}_Y(B_2) = \int_{B_1} \mathbb{P}_Y(B_2)\mathbb{P}_X(dx),$$

où l'on a utilisé la représentation intégrale $\mathbb{P}_X(B_1) = \int_{B_1} \mathbb{P}_X(dx)$.

Du fait de l'indépendance, on a aussi $\mathbb{P}_Y(B_2) = \mathbb{P}(Y \in B_2) = \mathbb{P}(Y \in B_2 | X \in B_1) = \mathbb{P}_Y(B_2 | X \in B_1)$ ce qui exprime que pour tout borélien B_1, la loi conditionnelle de Y sachant $X \in B_1$ est identique à la loi de Y.

Dans le cas général, on va généraliser l'égalité ci-dessus en

$$\mathbb{P}(X \in \mathcal{B}_1, Y \in \mathcal{B}_2) = \mathbb{P}_X(B_1)\mathbb{P}_Y(B_2 | X \in B_1) = \int_{B_1} \mathbb{P}_Y(B_2 | X = x)\mathbb{P}_X(dx)$$

et s'intéresser à caractériser la loi conditionnelle de Y sachant $X = x$.

De même, on a pour toute application $g : \mathbb{R}^2 \to \mathbb{R}$ borélienne telle que $g(X, Y)$ admette une espérance, on pourra écrire :

$$\mathbb{E}(g(X, Y)) = \int_{\mathbb{R}} \mathbb{P}_X(dx) \int_{\mathbb{R}} g(x, y)\mathbb{P}_Y(dy | X = x)$$

La théorie générale

Cas discret

Densités conditionnelles

Nous avons étudié précédemment l'indépendance des variables aléatoires à densité. Nous allons nous intéresser ici à caractériser la dépendance entre variables aléatoires à densité et en particulier au problème du conditionnement.

On s'intéresse à caractériser la densité de la variable Y connaissant la valeur prise par la variable X, c'est la *densité conditionnelle* de Y sachant $\{X = x\}$:

Proposition

La formule suivante définit une densité sur $\mathbb{R}$, pour tout x tel que $f_X(x) > 0$:

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}.$$

Cette fonction s'appelle la *densité conditionnelle de Y sachant* $\{X = x\}$.

La preuve est immédiate puisque $f_{Y|X=x}$ est une fonction positive d'intégrale 1.

L'interprétation de cette définition est la suivante : la fonction $f(Y|X=x)$ est la densité de la "loi conditionnelle de Y sachant que $X=x$". Bien sûr, nous avons $\mathbb{P}(X=x)=0$ puisque X admet une densité, donc la phrase ci-dessus n'a pas réellement de sens, mais elle se justifie heuristiquement ainsi : dx et dy étant de "petits" accroissements des variables x et y et lorsque f est continue :

$$\begin{aligned} f_X(x)dx &\approx \mathbb{P}(X \in [x, x+dx]) \\ f_{X,Y}(x,y)dxdy &\approx \mathbb{P}(X \in [x, x+dx], Y \in [y, y+dy]) \end{aligned}$$

Par suite

$$\begin{aligned} f_{Y|X=x}(y)dy &\approx \frac{\mathbb{P}(X \in [x, x+dx], Y \in [y, y+dy])}{\mathbb{P}(X \in [x, x+dx])} \\ &\approx \mathbb{P}(Y \in [y, y+dy] | X \in [x, x+dx]) \end{aligned}$$

La proposition suivante nous donne une nouvelle caractérisation de l'indépendance de deux variables aléatoire à densité :

Proposition

Les variables aléatoires X et Y sont indépendantes si et seulement si la densité conditionnelle de Y sachant $\{X=x\}$ ne dépend pas de x.

Démonstration Si X et Y sont indépendantes, $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$, d'où $f_{Y|X=x}(y) = f_Y(y)$.

Inversement, si $f_{Y|X=x}(y) = f_Y(y)$ alors $f_{(X,Y)}(x,y) = f_{Y|X=x}(y) f_X(x) = f_Y(y) f_X(x)$ et X et Y sont indépendantes. ■

Puisque $f_{Y|X=x}$ est une densité, on peut définir l'espérance qui lui est associée et introduire la notion d'espérance conditionnelle dans le cas où Y est intégrable.

Définition

Soit Y une variable aléatoire intégrable.

1. L'*espérance conditionnelle de Y sachant* $\{X=x\}$ est définie par

$$\mathbb{E}(Y|X=x) = \int_{\mathbb{R}} y f_{Y|X=x}(y) dy.$$

2. L'*espérance conditionnelle de Y sachant X* est la **variable aléatoire** définie par :

$$\mathbb{E}(Y|X) = \psi(X), \text{ avec } \psi(x) = \mathbb{E}(Y|X=x).$$

On peut étendre cette définition à toute variable de la forme $h(X,Y)$.

Définition

Soit Y une variable aléatoire et h une fonction mesurable positive ou bornée sur $\mathbb{R}^2$.

1. L'*espérance conditionnelle de* $h(X,Y)$ *sachant* $\{X=x\}$ est définie par
$$\mathbb{E}(h(X,Y)|X=x)=\int_{\mathbb{R}} h(x,y)f_{Y|X=x}(y)dy.$$
2. L'*espérance conditionnelle de* $h(X,Y)$ *sachant* X est la **variable aléatoire** définie par :
$$\mathbb{E}(h(X,Y)|X)=\psi(X),\ \text{avec}\ \psi(x)=\mathbb{E}(h(X,Y)|X=x).$$

Remarques

1. $\psi(x)$ n'est définie que pour $x\notin B=\{u,f_X(u)=0\}$, mais $\mathbb{P}(X\in B)=\int_B f_X(u)du=0$. Par conséquent, la définition définit bien l'espérance conditionnelle $\psi(X)=\mathbb{E}(Y|X)$ avec probabilité 1.
2. $\mathbb{E}(\mathbb{E}(|Y||X))=\int_{\mathbb{R}}\left(\int_R |y|\frac{f(x,y)}{f_X(x)}dy\right)f_X(x)dx=\mathbb{E}(|Y|)$ où nous avons utilisé le théorème de Fubini. L'espérance conditionnelle de Y sachant X est bien définie dès que Y est intégrable.
3. Les rôles de X et Y peuvent évidemment être inversés dans tous les résultats. On a en particulier pour tout x tel que $f_X(x)>0$ et tout y tel que $f_Y(y))>0$:
$$f_{X|Y=y}(x)=\frac{f(x,y)}{f_Y(y)}=\frac{f_{Y|X=x}(y)f_X(x)}{f_Y(y)}.$$
Ceci nous donne un équivalent de la *formule de Bayes pour les densités.*

Théorème

Si Y est intégrable, alors $\psi(X)=\mathbb{E}(Y|X)$ est intégrable, et
$$\mathbb{E}(\psi(X))=E(Y).$$

Démonstration On a

$$\begin{aligned}\mathbb{E}(\psi(X))&=\int_{\mathbb{R}}\psi(x)f_X(x)dx\\&=\int_{\mathbb{R}}\left(\int_R yf_{Y|X=x}dy\right)f_X(x)dx\\&=\int_{\mathbb{R}}\left(\int_R |y|\frac{f(x,y)}{f_X(x)}dy\right)f_X(x)dx\\&=\mathbb{E}(Y).\end{aligned}$$

où avons utilisé ici le théorème de Fubini dont l'application est justifiée par la remarque ci-dessus. ■

Ce résultat permet de calculer $\mathbb{E}(Y)$ en conditionnant par une variable auxiliaire X :

$$\mathbb{E}(Y) = \int_{\mathbb{R}} \mathbb{E}(Y|X = x) f_X(x) dx$$

Il généralise la formule des probabilités totales, qui correspond ici à $Y = 1_A$, et $B_x = \{X = x\}$ où les B_x forment cette fois une partition non dénombrable de $\mathbb{R}$. On l'écrit souvent sous forme

$$\mathbb{E}\left(\mathbb{E}(Y|X)\right) = \mathbb{E}(Y)$$

L'espérance conditionnelle étant définie comme l'espérance de la loi conditionnelle, elle hérite des propriétés usuelles de l'espérance :

1. si Y et Z sont intégrables, $\mathbb{E}(aY + bZ|X) = a\mathbb{E}(Y|X) + b\mathbb{E}(Z|X)$,
2. $\mathbb{E}(Y|X) \geq 0$ si $Y \geq 0$,
3. $\mathbb{E}(1|X) = 1$. De plus, si g est continue par morceaux positive ou bornée,

$$\mathbb{E}(Yg(X)|X) = g(X)\mathbb{E}(Y|X)$$

est une généralisation de l'égalité 1. ci-dessus, au cas où $a = g(X)$, qui doit être considéré “comme une constante” dans le calcul de l'espérance conditionnelle sachant X (X est fixée comme une donnée connue a priori).

Exemple

Soient X et Y de densité jointe $f_{X,Y}(x, y) = \frac{1}{x} 1_T(x, y)$ où T est le triangle $T = \{0 < y < x < 1\}$

La densité marginale de X est donnée par $f_X(x) = \int f_{X,Y}(x, y)dy = 1_{]0,1[}(x)$ et pour $x \in]0, 1[$,

$$f_{Y|X=x}(y) = \frac{1}{x} 1_{]0,x[}(y)$$

Ainsi X est uniformément distribué sur $]0, 1[$, et la loi de Y sachant $X = x$ est uniforme sur $]0, x[$ pour $(0 < x < 1)$. Pour un tel x, l'espérance conditionnelle $\mathbb{E}(Y|X = x)$ vaut ainsi $x/2$ et nous obtenons $\mathbb{E}(Y|X) = \frac{X}{2}$.

note bofbof cet exemple. . .

Cas général

Vecteurs Gaussiens

Dans cette partie, on s'intéresse à l'exemple canonique des vecteurs gaussiens. Les vecteurs gaussiens (et leur extension que sont les fonctions aléatoires gaussiennes, qui, une fois discrétisées, se ramènent à des vecteurs) se rencontrent dans un grand nombre d'applications. L'ubiquité de la loi gaussienne se justifie également du fait du théorème central limite que nous verrons au prochain chapitre.

Régression

La régression est un ensemble de méthodes d'apprentissage statistique très utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres. Ces méthodes visent notamment à décrire les liens de dépendance entre variable mais aussi de prédire au mieux la valeur d'une quantité non observée en fonction d'une ou plusieurs autres variables. On va en décrire ici le principe du point de vue probabiliste dans le cas particulier des variables de carré intégrable. On verra dans ce cadre, que l'on rencontre très fréquemment en pratique, une interprétation géométrique très éclairante de l'espérance conditionnelle.

Régression linéaire

On considère deux variables aléatoires de carré intégrable dont on suppose connues les variances et la covariance. Nous souhaitons trouver la meilleure approximation de Y par une fonction affine de X de la forme $aX + b$, au sens des moindres carrés, c'est à dire qui minimise la quantité $\mathbb{E}((Y - (aX + b))^2)$. Il s'agit de déterminer les constantes a et b telles que $\mathbb{E}((Y - (aX + b))^2)$ soit minimale. Or, par linéarité,

$$\mathbb{E}((Y - (aX + b))^2) = \mathbb{E}(Y^2) - 2a\mathbb{E}(XY) + a^2\mathbb{E}(X^2) + 2ab\mathbb{E}(X) + b^2$$

L'annulation des dérivées partielles par rapport à a et b entraîne que les solutions sont

$$\begin{aligned} a &= \frac{\mathrm{Cov}(X, Y)}{\mathbb{V}(X)} = \rho(X, Y)\frac{\sigma_Y}{\sigma_X} \\ b &= \mathbb{E}(Y) - a\mathbb{E}(X) \end{aligned}$$

On vérifie aisément que ces valeurs donnent bien un minimum pour $\mathbb{E}((Y - (aX + b))^2)$, et déterminent ainsi la meilleure approximation linéaire de Y basée sur X au sens de l'erreur quadratique moyenne.

Cette approximation linéaire vaut

$$\mathbb{E}(Y) + \rho(X, Y)\frac{\sigma_Y}{\sigma_X}(X - \mathbb{E}(X))$$

et l'erreur quadratique moyenne vaut alors

$$\begin{aligned} \mathbb{E}\left(\left(Y - \mathbb{E}(Y) - \rho(X, Y)\frac{\sigma_Y}{\sigma_X}(X - \mathbb{E}(X))\right)^2\right) &= \sigma_Y^2 + \rho^2(X, Y)\sigma_Y^2 - 2\rho^2(X, Y)\sigma_Y^2 \\ &= \sigma_Y^2(1 - \rho^2(X, Y)). \end{aligned}$$

On voit ainsi que cette erreur est proche de 0 lorsque $|\rho(X,Y)| \approx 1$ tandis qu'elle est proche de $\mathbb{V}(Y) = \sigma_Y^2$ lorsque $\rho(X,Y) \approx 0$.

Dans ce paragraphe, on s'est intéressé à caractériser la relation linéaire entre deux variables aléatoires de carré intégrable. On va montrer dans ce paragraphe que la meilleure approximation, au sens des l'erreur quadratique moyenne, de Y par une fonction de X est précisément donnée par $\mathbb{E}(Y|X)$.

Espace de Hilbert des variables aléatoires de carré intégrable