

LOIS CONDITIONNELLES

PRÉPARATION À L'AGRÉGATION EXTERNE DE MATHÉMATIQUES DE L'UNIVERSITÉ RENNES 1¹
ANNÉE 2008/2009

1. LA NOTION DE LOI CONDITIONNELLE

Dans la suite, toute propriété de mesurabilité s'entend pour la tribu borélienne.

Il est facile de définir la loi conditionnelle de Y sachant X lorsque X est discrète, à valeurs dans \mathcal{D} . Supposons que Y est à valeurs dans \mathbb{R}^p . La loi conditionnelle de Y sachant X est la famille de lois $\{\mathbb{P}_{Y|X=x}, x \in \mathcal{D}\}$ sur \mathbb{R}^p , telle que si $\mathbb{P}(X = x) \neq 0$,

$$\mathbb{P}_{Y|X=x}(B) = \mathbb{P}(Y \in B | X = x) \quad \forall B \in \mathcal{B}(\mathbb{R}^p), \quad (*)$$

et, dans le cas contraire, $\mathbb{P}_{Y|X=x}$ est une probabilité quelconque sur \mathbb{R}^p . Ce procédé ne s'étend pas au cas général, en raison de l'impossibilité de diviser par 0. La notion de noyau de transition est introduite pour contourner cet écueil.

Définition 1.1 Une application $\nu : \mathbb{R}^p \times \mathbb{R}^d \rightarrow [0, 1]$ est appelée noyau de transition si elle satisfait les propriétés :

- (i) $\forall x \in \mathbb{R}^p, \nu(x, \cdot)$ est une probabilité sur \mathbb{R}^d ;
- (ii) $\forall A \in \mathcal{B}(\mathbb{R}^d)$, l'application $\nu(\cdot, A)$ est $\mathcal{B}(\mathbb{R}^p)$ -mesurable.

Si μ est une probabilité sur \mathbb{R}^p et ν un noyau de transition sur $\mathbb{R}^p \times \mathbb{R}^d$, on note $\mu \cdot \nu$ l'application définie sur $\mathcal{B}(\mathbb{R}^p) \times \mathcal{B}(\mathbb{R}^d)$ par :

$$\mu \cdot \nu(A \times B) = \int_A \nu(x, B) d\mu(x), \quad A \in \mathcal{B}(\mathbb{R}^p) \text{ et } B \in \mathcal{B}(\mathbb{R}^d).$$

Cette application, qui est σ -additive sur $\mathcal{B}(\mathbb{R}^p) \times \mathcal{B}(\mathbb{R}^d)$, se prolonge en une unique probabilité sur $\mathbb{R}^p \times \mathbb{R}^d$. Le prolongement est encore noté $\mu \cdot \nu$.

Définition 1.2 Soient X et Y des v.a. à valeurs dans \mathbb{R}^p et \mathbb{R}^d . On appelle loi conditionnelle de Y sachant X un noyau de transition ν sur $\mathbb{R}^p \times \mathbb{R}^d$ tel que $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \cdot \nu$.

Le théorème de Jirina, sur lequel nous ne nous attarderons pas, assure l'existence d'une telle loi conditionnelle, dans le cadre de la définition précédente.

Le plus souvent, le noyau de transition ν est noté $\mathbb{P}_{Y|X=\cdot}$. La loi conditionnelle est donc en réalité une famille de lois de probabilité. Dans le cas où X est à valeurs discrètes, on retrouve bien la loi conditionnelle définie par (*), car pour tout $B \in \mathcal{B}(\mathbb{R}^p)$ et $A \subset \mathcal{D}$, on a en notant $\mathcal{D}_0 = \{x \in \mathcal{D} : \mathbb{P}(X = x) \neq 0\}$:

$$\mathbb{P}_{(X,Y)}(A \times B) = \sum_{x \in \mathcal{D}_0 \cap A} \mathbb{P}_{Y|X=x}(B) \mathbb{P}_X(\{x\}) = \int_A \mathbb{P}_{Y|X=x}(B) \mathbb{P}_X(dx) = \mathbb{P}_X \cdot \mathbb{P}_{Y|X=\cdot}(A \times B).$$

Exercice 1.1 Soient X et Y des v.a.r. avec $X \sim \mathcal{U}[0, 1]$, et λ la mesure de Lebesgue sur \mathbb{R} .

- (i) On suppose que pour tout $x \in [0, 1/2[$, $\mathbb{P}_{Y|X=x} = \mathbf{1}_{[1/2, 1]} \lambda + (1/2) \delta_x$, et pour tout $x \in [1/2, 1]$, $\mathbb{P}_{Y|X=x} = \mathcal{U}[0, 1]$. Calculer la loi de Y .
- (ii) Calculer la loi conditionnelle de Y sachant X , lorsque pour tous boréliens A de $[0, 1]$ et B de \mathbb{R} :

$$\mathbb{P}_{(X,Y)}(A \times B) = \lambda(A \cap [0, 1/2[) \lambda(B \cap [1/2, 1]) + \frac{1}{2} \lambda(A \cap B \cap [0, 1/2]) + \lambda(A \cap [1/2, 1]) \lambda(B \cap [0, 1]).$$

2. PROPRIÉTÉS DES LOIS CONDITIONNELLES

Dorénavant, X et Y désignent des v.a. à valeurs dans \mathbb{R}^p et \mathbb{R}^d respectivement.

Le théorème de type Fubini qui est énoncé ci-dessous, et sur lequel on serait tenté de jeter un coup d'oeil distrait, est d'utilité constante dans la manipulation des lois conditionnelles.

¹Benoît Cadre - ENS Cachan Bretagne

Théorème 1.1 [FUBINI] Soit $\varphi : \mathbb{R}^p \times \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable.

(i) Si φ est positive, l'application $x \mapsto \int \varphi(x, y) \mathbb{P}_{Y|X=x}(dy)$ est $\mathcal{B}(\mathbb{R}^p)$ -mesurable, et

$$\int_{\mathbb{R}^p \times \mathbb{R}^d} \varphi d\mathbb{P}_{(X,Y)} = \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}^d} \varphi(x, y) \mathbb{P}_{Y|X=x}(dy) \right] \mathbb{P}_X(dx).$$

(ii) Si φ est $\mathbb{P}_{(X,Y)}$ -intégrable, alors pour \mathbb{P}_X -p.t. x , l'application $\varphi(x, \cdot)$ est $\mathbb{P}_{Y|X=x}$ -intégrable, l'application définie pour \mathbb{P}_X -p.t. x par $x \mapsto \int \varphi(x, y) \mathbb{P}_{Y|X=x}(dy)$ est \mathbb{P}_X -intégrable et l'égalité de (i) est encore vraie.

APPLICATION : ESPÉRANCE CONDITIONNELLE. Fixons $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $\psi(Y) \in L^1$. Une simple application du théorème de Fubini montre que pour tout $A \in \mathcal{B}(\mathbb{R}^p)$,

$$\int_{\{X \in A\}} \psi(Y) d\mathbb{P} = \int_A \left[\int_{\mathbb{R}^d} \psi(y) \mathbb{P}_{Y|X=x}(dy) \right] \mathbb{P}_X(dx),$$

et donc que

$$\mathbb{E}[\psi(Y)|X = x] = \int_{\mathbb{R}^d} \psi(y) \mathbb{P}_{Y|X=x}(dy).$$

Par ailleurs, on peut aussi calculer $\mathbb{E}\psi(Y)$ en utilisant l'espérance conditionnelle de $\psi(Y)$ sachant X :

$$\mathbb{E}\psi(Y) = \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}^d} \psi(y) \mathbb{P}_{Y|X=x}(dy) \right] \mathbb{P}_X(dx) = \int_{\mathbb{R}^p} \mathbb{E}[\psi(Y)|X = x] \mathbb{P}_X(dx).$$

Exercice 2.1 Dans le contexte de l'exercice 1.1, calculer $\mathbb{E}XY$.

On peut aussi calculer une loi conditionnelle en se ramenant à un calcul d'espérance conditionnelle (cf complément de cours : Espérance conditionnelle et introduction aux martingales). Reprenons le contexte de l'application au calcul d'espérances conditionnelles. On calcule dans un premier temps $\mathbb{P}(\psi(Y) \leq y | X = x)$, puis on utilise le théorème de Dynkin pour en déduire $\mathbb{P}(\psi(Y) \in \cdot | X = x)$. La définition 1.2 et la définition d'une espérance conditionnelle nous montrent alors que $\mathbb{P}(\psi(Y) \in \cdot | X = \cdot)$ est le noyau de transition recherché.

Exercice 2.2 Soient X_1, \dots, X_n i.i.d. de loi $\mathcal{U}[0, 1]$ et $T = \sup(X_1, \dots, X_n)$. Le vecteur aléatoire (X_1, T) possède-t-il une densité? Calculer la loi conditionnelle de X_1 sachant T .

Lorsque la v.a. X apparaît aussi dans Y , le calcul de la loi conditionnelle de Y sachant X peut être facilité :

Théorème 2.1 [TRANSFERT CONDITIONNEL] Soit φ une fonction mesurable définie sur $\mathbb{R}^p \times \mathbb{R}^d$. Alors, pour tout $x \in \mathbb{R}^p$, $\mathbb{P}_{\varphi(X,Y)|X=x} = \mathbb{P}_{\varphi(x,Y)|X=x}$. En particulier, si X et Y sont indépendantes, on a pour tout $x \in \mathbb{R}^p$: $\mathbb{P}_{\varphi(X,Y)|X=x} = \mathbb{P}_{\varphi(x,Y)}$.

Sous les conditions d'intégrabilité adéquates, le calcul de quantités du type $\mathbb{E}[\varphi(X, Y)|X]$ se ramène à un calcul de la loi conditionnelle de Y sachant X , car

$$\mathbb{E}[\varphi(X, Y)|X = x] = \mathbb{E}[\varphi(x, Y)|X = x] = \int_{\mathbb{R}^d} \varphi(x, y) \mathbb{P}_{Y|X=x}(dy).$$

Si, de plus, X et Y sont indépendantes, on a :

$$\mathbb{E}[\varphi(X, Y)|X = x] = \int_{\mathbb{R}^d} \varphi(x, y) \mathbb{P}_Y(dy).$$

Lorsque le couple (X, Y) possède une densité, on a là encore une situation assez confortable. Précisons auparavant la notion de *densité conditionnelle* : si, pour \mathbb{P}_X -p.t. $x \in \mathbb{R}^p$, $\mathbb{P}_{Y|X=x}$ possède une densité, celle-ci est appelée *densité conditionnelle* de Y sachant $X = x$.

Théorème 2.2 Supposons que (X, Y) possède une densité f . Alors, pour tout x tel que $f_X(x) > 0$, la loi de Y sachant $X = x$ admet une densité, qui vaut :

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

Preuve Soit ρ une probabilité quelconque sur \mathbb{R}^d . Pour tout $x \in \mathbb{R}^p$ et $B \in \mathcal{B}(\mathbb{R}^d)$, on note

$$\nu(x, B) = \int_B \frac{f(x, y)}{f_X(x)} dy,$$

si $f_X(x) > 0$, et $\nu(x, B) = \rho(B)$ si $f_X(x) = 0$. Une loi conditionnelle de Y sachant X est ν , car ν est un noyau de transition sur $\mathbb{R}^p \times \mathbb{R}^d$ qui vérifie $\mathbb{P}_{(X, Y)} = \mathbb{P}_X \cdot \nu$. Par ailleurs, si $f_X(x) > 0$, la mesure $\nu(x, \cdot)$ est à densité $f_{(X, Y)}(x, \cdot) / f_X(x)$ par rapport à la mesure de Lebesgue. •

Exercice 2.3 Soient $X \perp\!\!\!\perp Y$, avec $X \sim \mathcal{E}(\lambda)$ et $Y \sim \mathcal{E}(\lambda)$. Calculer la loi conditionnelle de $X + Y$ sachant X .

Comme d'habitude, les vecteurs gaussiens ont un comportement très particulier.

Théorème 2.3 [CONDITIONNEMENT GAUSSIEN] *Supposons que (X, Y) est un vecteur gaussien, que Y est une v.a.r., et que X possède une matrice de variance inversible. On note $a^t = (\text{cov}(Y, X_1), \dots, \text{cov}(Y, X_p)) \mathbb{W}(X)^{-1}$. Alors, la loi conditionnelle de Y sachant X est gaussienne, de moyenne $\mathbb{E}(Y|X) = \langle a, X - \mathbb{E}X \rangle + \mathbb{E}(Y)$, et de variance indépendante de X .*

Preuve La valeur de $\varphi(X) := \mathbb{E}(Y|X)$ a été calculée dans le complément de cours "Vecteurs gaussiens". A cette occasion, on a aussi remarqué que $Y - \varphi(X) \perp\!\!\!\perp X$. On en déduit du théorème de transfert conditionnel que

$$\mathbb{P}_{Y|X=x} = \mathbb{P}_{Y-\varphi(X)+\varphi(X)|X=x} = \mathbb{P}_{Y-\varphi(X)+\varphi(x)|X=x} = \mathbb{P}_{Y-\varphi(X)+\varphi(x)}.$$

Comme φ est une fonction affine, $Y - \varphi(X)$, qui s'exprime comme une combinaison linéaire des v.a. X, Y , est donc une v.a.r. gaussienne. Enfin, on déduit de l'égalité ci-dessus que

$$\text{var}(Y|X = x) = \text{var}(Y - \varphi(X) + \varphi(x)) = \text{var}(Y - \varphi(X)),$$

et donc que le variance de la loi conditionnelle de Y sachant X est indépendante de X . •