Feuille de TP n°5
Espérance conditionnelle en modélisation

1 Espérance conditionnelle.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Pour $p \geqslant 1$, on note $\mathbf{L}^p(\mathcal{A})$ l'espace des variables aléatoires X définies sur $(\Omega, \mathcal{A}, \mathbf{P})$ et telles que $\|X\|_p^p = \mathbb{E}(|X|^p)$ soit finie. $\mathbf{L}^2(\mathcal{A})$ est un espace de Hilbert pour le produit scalaire $\langle X, Y\rangle = \mathbb{E}(XY)$. Si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$, $\mathbf{L}^2(\mathcal{B})$ est un sous-espace de Hilbert de $\mathbf{L}^2(\mathcal{A})$. Pour $X \in \mathbf{L}^2(\mathcal{A})$, l'espérance conditionnelle de X sachant $\mathcal{B}$ est la projection orthogonale de X sur $\mathbf{L}^2(\mathcal{B})$. Elle correspond à la v.a. de $\mathbf{L}^2(\mathcal{B})$ la plus proche de X au sens des moindres carrés, que l'on note $\mathbb{E}(X \mid \mathcal{B})$. On a $\mathbb{E}((X - \mathbb{E}(X \mid \mathcal{B}))^2) = \inf \left\{\mathbb{E}((X - Y)^2) \text{ avec } Y \in \mathbf{L}^2(\mathcal{B})\right\}$. L'espérance conditionnelle est une v.a. définie p.s. et toutes les relations où elle figure sont des relations entre classes d'équivalence de variables aléatoires. Toutefois, on ne précisera plus dans la suite le terme p.s.
Propriété 1. Pour tout $Y \in \mathbf{L}^2(\mathcal{B})$, on a $\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{B})Y)$ et $\mathbb{E}(\mathbb{E}(X \mid \mathcal{B}))^2) \leqslant \mathbb{E}(X^2)$.
Propriété 2. L'espérance conditionnelle est un opérateur linéaire croissant de $\mathbf{L}^2(\mathcal{A})$ dans $\mathbf{L}^2(\mathcal{B})$. En particulier, si $X \in \mathbf{L}^2(\mathcal{A})$ avec $X \geqslant 0$, alors $\mathbb{E}(X \mid \mathcal{B}) \geqslant 0$.
Propriété 3. Si X est $\mathcal{B}$-mesurable, $\mathbb{E}(X \mid \mathcal{B}) = X$. Si X est indépendante de $\mathcal{B}$, $\mathbb{E}(X \mid \mathcal{B}) = \mathbb{E}(X)$. Si X et $XY \in \mathbf{L}^1(\mathcal{A})$ et si Y est $\mathcal{B}$-mesurable, on a la factorisation $\mathbb{E}(XY \mid \mathcal{B}) = Y\mathbb{E}(X \mid \mathcal{B})$.
Propriété 4. Si $\mathcal{C}$ est une sous-tribu de $\mathcal{B}$, on a $\mathbb{E}(\mathbb{E}(X \mid \mathcal{B}) \mid \mathcal{C}) = \mathbb{E}(X \mid \mathcal{C})$. En particulier, $\mathbb{E}(\mathbb{E}(X \mid \mathcal{B})) = \mathbb{E}(X)$.

Soit X une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $Y \in \mathbf{L}^2(\mathcal{A})$. L'espérance conditionnelle de Y sachant X, notée $\mathbb{E}(Y \mid X)$, correspond à l'espérance conditionnelle de Y sachant $\sigma(X)$. Soit f la fonction mesurable définie par $f(X) = \mathbf{E}(Y \mid X)$. Pour toute fonction borélienne bornée g, on a

$$\mathbf{E}(Yg(X)) = \mathbf{E}(f(X)g(X)) \text{ et } \mathbf{E}(f(X)^2) \leqslant \mathbf{E}(Y^2).$$

2 Utilisation en modélisation.

Exercice 2.1 (Régression linéaire). Soit $X = (X_1, \cdots, X_p)$ un vecteur aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ de matrice de covariance inversible Γ et soit $Y \in \mathbf{L}^2(\mathcal{A})$. Montrer que la meilleure approximation de Y par une fonction affine de X au sens des moindres carrés est donnée par $f(X) = a + b^t X$ avec $a = \mathbb{E}(Y) - b^t\mathbb{E}(X)$ et $b = \Gamma^{-1}\mathbf{Cov}(X, Y)$. On a toujours $\mathbb{E}((Y - \mathbb{E}(Y|X))^2) \leqslant \mathbb{E}((Y - f(X))^2)$ avec égalité dans le cadre gaussien. Vérifier que si (X, Y) est un vecteur gaussien, alors $f(X)$ correspond à $\mathbb{E}(Y|X)$, $Y - f(X)$ est indépendante de X et

$$\mathbb{E}(Y|X) = a + b_1X_1 + \cdots + b_pX_p = \mathbb{E}(Y) + \langle b, X - \mathbb{E}(X)\rangle.$$

Exercice 2.2 (Modèle linéaire gaussien). On considère la régression linéaire gaussienne simple défini, pour $n > 2$, par

$$Y_i = a + bx_i + \varepsilon_i;\ i = 1, \ldots, n,$$

où $(x_i)_i$ est une suite de nombres réels connus non tous égaux et où $(\varepsilon_i)_i$ est une suite i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. Déterminer les estimateurs de moindres carrés $\widehat{\theta} = (\widehat{a}, \widehat{b})$ et $\widehat{\sigma}^2$ de $\theta = (a, b)$ et σ^2. Montrer que, si $X = (x_1, \ldots, x_n)^T$, alors $\widehat{\theta} \sim \mathcal{N}(\theta, \sigma^2(XX^T)^{-1})$ donc

$$\widehat{a} \sim \mathcal{N}\left(a, \frac{\sigma^2\overline{x^2}}{n\mathbf{Var}(x)}\right) \quad \text{et} \quad \widehat{b} \sim \mathcal{N}\left(b, \frac{\sigma^2}{n\mathbf{Var}(x)}\right).$$

Montrer que les estimateurs $\widehat{\theta}$ et $\widehat{\sigma^2}$ sont indépendants et $(n-2)\widehat{\sigma^2} \sim \sigma^2\chi^2(n-2)$. En déduire que

$$\sqrt{\frac{n\mathrm{Var}(x)}{\overline{x^2}}}\left(\frac{\widehat{a}-a}{\widehat{\sigma}}\right) \sim t(n-2) \quad \text{et} \quad \sqrt{n\mathbf{Var}(x)}\left(\frac{\widehat{b}-b}{\widehat{\sigma}}\right) \sim t(n-2).$$

On peut ainsi effectuer des tests sur les valeurs a et b et obtenir des intervalles de confiance pour a et b. Montrer que, si $a=0$, $\sum_{i=1}^n(\widehat{a}+\widehat{b}x_i-\tilde{b}x_i)^2/\widehat{\sigma}^2 \sim F(1,n-2)$ avec $\tilde{b} := \overline{xY}/\overline{x^2}$ et, si $b=0$, $\sum_{i=1}^n(\widehat{a}+\widehat{b}x_i-\overline{Y})^2/\widehat{\sigma}^2 \sim F(1,n-2)$. On peut ainsi tester H_0 : « $a=0$ » contre H_1 : « $a\neq 0$ » ou bien encore H_0 : « $b=0$ » contre H_1 : « $b\neq 0$ ».

Exercice 2.3 (Régression linéaire). Créer un code Matlab permettant de générer une régression linéaire gaussienne simple où les valeurs n, a, b et σ^2 sont affectées par l'utilisateur et où $(x_i)_i$ est une réalisation d'un n-échantillon de loi uniforme sur $[0,1]$. Calculer les estimateurs des moindres carrés $\widehat{\theta}=(\widehat{a},\widehat{b})$ et $\widehat{\sigma}^2$. Représenter graphiquement le nuage de points formé par les couples $(x_i,y_i)_i$ et tracer la droite des moindres carrés $y=\widehat{a}+\widehat{b}x$. Donner pour chaque paramètre a, b et σ^2 un intervalle de confiance de risque $\alpha=5\%$. Reprendre cet exercice en faisant varier n, a, b et σ^2 ainsi que la loi associée à $(x_i)_i$.

Exercice 2.4 (Régression linéaire). Un particulier cherche à acquérir un appartement aux alentours immédiats de la place du Capitole à Toulouse. Il a sélectionné les 24 offres de vente suivantes où x représente la surface en mètres carrés et Y correspond au prix en milliers d'Euros :

x	28	50	196	55	190	110	60	48	90	35	86	65
Y	130	280	800	268	790	500	320	250	378	250	350	300
x	32	52	40	70	28	30	105	52	80	60	20	100
Y	155	245	200	325	85	78	375	200	270	295	85	495

Etudier la régression linéaire simple du prix en fonction de la surface grâce à `linreg` de Matlab.

Exercice 2.5 (Processus de Galton-Watson sur la taille d'une population). On considère le processus de Galton-Watson

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}$$

avec $X_0=1$. La v.a. X_n représente le nombre d'individus à la n^e génération et pour $1\leqslant k\leqslant X_n$, $Y_{n,k}$ correspond au nombre de descendants du k^e individu de la n^e génération. On suppose que $(Y_{n,k})$ est i.i.d. à valeurs dans $\mathbb{N}$, de moyenne m et de variance σ^2. Montrer que ce processus peut s'ecrire sous la forme $X_{n+1}=mX_n+\varepsilon_{n+1}$ où (ε_n) satisfait $\mathbb{E}(\varepsilon_{n+1}|\mathcal{F}_n]=0$ et $\mathbb{E}(\varepsilon_{n+1}^2|\mathcal{F}_n)=\sigma^2X_n$. On note q la probabilité d'extinction de la population. On a la trichotomie suivante :

1. Cas *sous-critique* : $m<1$, $q=1$, $(X_n)_n$ tend vers 0 p.s.
2. Cas *critique* : $m=1$, $q=1$, $(X_n)_n$ tend vers 0 p.s. et $n\mathbb{P}(X_n>0)\to 2/\sigma^2$.
3. Cas *sur-critique* : $m>1$, q est l'unique point fixe de la fonction génératrice associée à la loi de $(Y_{n,k})$ et (X_n/m^n) converge p.s. vers une v.a. finie L avec $\mathbb{E}(L)=1$, $\mathbf{Var}(L)=\sigma^2/(m^2-m)$ et $\mathbb{P}(L=0)=q$. Vérifier par simulation ces résultats lorsque $(Y_{n,k})$ est associée à la loi Binomiale $\mathcal{B}(a,1/2)$ avec $a=1,2,4$. Essayer d'autres lois.