Introduction and Techniques in Financial Mathematics
UiO-STK4510
Autumn 2015
Teacher: *S. Ortiz-Latorre*

Lecture 3

Probability Theory

---

1 Basic Definitions

Definition 1 *A probability space* $(\Omega, \mathcal{F}, P)$ *is a measure space such that* $P(\Omega) = 1$.

A probability space is a mathematical object that is used to model a random experiment. Each $\omega \in \Omega$ is a possible outcome of the random experiment and $P(B), B \in \mathcal{F}$ is the probability that the outcome belong to the event B.

Definition 2 *A random variable* $X : \Omega \to \mathbb{R}$ *is a* $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$*-measurable function.*

A random variable is the mechanism we use to observe the random experiment. That is, we do not observe ω directly but for all $B \in \mathcal{B}(\mathbb{R})$ we can decide if $X(\omega)$ belong to B or not.

Definition 3 *The law of a random variable* X*, denoted by* $\mathcal{L}(X)$*, is the image measure* P_X *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, that is,*

$$P_X(B) = P(X^{-1}B), \quad B \in \mathcal{B}(\mathbb{R}).$$

Note that $P_X(B) \leq 1$ for all $B \in \mathcal{B}(\mathbb{R})$ and, therefore, P_X is a Lebesgue-Stieltjes measure on $\mathcal{B}(\mathbb{R})$.

Definition 4 *The distribution function* $F_X(x)$ *of a r.v.* X *is the following distribution function associated to the Lebesgue-Stieltjes measure* P_X *:*

$$F_X(x) = P_X((-\infty, x]), \quad x \in \mathbb{R}.$$

Definition 5 *Let* $P_X \ll \lambda$*, then we define the density* f_X *of* X *to be the Radon-Nikodym derivative* $f_X = \frac{dP_X}{d\lambda}$ *and*

$$P_X(B) = \int_B dP_X = \int_B f_X d\lambda, \quad B \in \mathcal{B}(\mathbb{R}).$$

We have that $f_X(x) = \frac{d}{dx} F_X(x)$.

Definition 6 *Let* $g : \mathbb{R} \to \mathbb{R}$ *be a Borel measurable function. Then the expectation of* $g(X)$ *is defined to be*

$$\mathbb{E}\left[g(X)\right] = \int_\Omega g \circ X dP = \int_\mathbb{R} g dP_X.$$

If $P_X \ll \lambda$*, then*

$$\mathbb{E}\left[g(X)\right] = \int_\mathbb{R} g f_X d\lambda = \int_\mathbb{R} g(x) f_X(x) dx,$$

where the last integral is an improper Riemann integral.

Definition 7 *For any* $p \geq 1$ *we say that a r.v.* $X \in L^p := L^p(\Omega, \mathcal{F}, P)$ *iff* $\mathbb{E}[|X|^p] < \infty$.

Using Hölder inequality one can check that if $X \in L^p$ then $X \in L^q$ for all $q \leq p$.

Definition 8 *If $X \in L^1$ then $\mu = \mathbb{E}[X]$, the mean of X is well defined. If $X \in L^2$ then $\sigma^2 = \mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$, the variance of X is well defined.*

Definition 9 *We say that a random variable X is Gaussian or normal with mean μ and variance σ^2, and we denote it by $X \sim \mathcal{N}(\mu, \sigma^2)$, if $P_X \ll \lambda$ and*

$$f_X(x) = \frac{dP_X}{d\lambda}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right), \quad x \in \mathbb{R}.$$

We say that a Gaussian random variable with mean μ is degenerate if its law is given by the Dirac delta function on μ, that is, $\delta_\mu(B) = 1$ if $\mu \in B$ and zero otherwise, we denote it by $X \sim \mathcal{N}(\mu, 0)$. We say that X is a standard normal random variable if $X \sim \mathcal{N}(0, 1)$.

The same kind of definitions can be given for random vectors $X = (X_1, ..., X_d)'$ that is measurable transformations from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We shall use the column notation for vectors and the symbol $'$ is the transpose.

Definition 10 *A random vector $X : \Omega \to \mathbb{R}^d, X = (X_1, ..., X_d)'$ is a $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$-measurable transformation.*

Definition 11 *The (joint) law of a random vector $X = (X_1, ..., X_d)'$, denoted by $\mathcal{L}(X)$, is the image measure P_X on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, that is,*

$$P_X(B) = P(X^{-1}B), \quad B \in \mathcal{B}(\mathbb{R}^d).$$

In this context, the law of X_i is called the marginal law of X_i. Note that P_X is a Lebesgue-Stieltjes measure on $\mathcal{B}(\mathbb{R}^d)$.

Definition 12 *The (joint) distribution function $F_X(x)$ of a random vector X is the following distribution function associated to the Lebesgue-Stieltjes measure P_X :*

$$\begin{aligned} F_X(x_1, ..., x_d) &= P_X((-\infty, x_1] \times \cdots \times (-\infty, x_d]) \\ &= P(\{X_1 \leq x_1, ..., X_d \leq x_d\}), \quad (x_1, ..., x_d) \in \mathbb{R}^d. \end{aligned}$$

We note that the joint law of X determines the marginal laws of X_i, but not the othere way around. Moreover,

$$F_{X_i}(x_i) = \lim_{x_j \to \infty} F_X(x_1, ..., x_d).$$

Definition 13 *Let $P_X \ll \lambda^d$, where $\lambda^d = \lambda \otimes \overset{d)}{\cdots} \otimes \lambda$. Then, we define the density f_X of X to be the Radon-Nikodym derivative $f_X = \frac{dP_X}{d\lambda^d}$ and*

$$P_X(B) = \int_B dP_X = \int_B f_X d\lambda^d, \quad B \in \mathcal{B}(\mathbb{R}^d).$$

We have that $f(x_1, ..., x_d) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F_X(x_1, ..., x_d)$.

Definition 14 *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a Borel measurable function. Then the expectation of $g(X)$ is defined to be*

$$\mathbb{E}[g(X)] = \int_\Omega g \circ X dP = \int_{\mathbb{R}^d} g dP_X.$$

If $P_X \ll \lambda^d$, then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g f_X d\lambda^d = \int_{\mathbb{R}^d} g(x_1, ..., x_d) f_X(x_1, .., x_d) dx_1 \cdots dx_d.$$

where the last integral is an improper Riemann integral.

Definition 15 *Let $X = (X_1, ..., X_d)'$. If $X_i \in L^1, i = 1, ..., d$ then $\mu = (\mu_1, ..., \mu_d)' = (\mathbb{E}[X_1], ..., \mathbb{E}[X_d])'$, the mean vector of X, is well defined. If $X_i \in L^2, i = 1, ..., d$ then*

$$\Sigma = \mathrm{Cov}(X) = (\mathrm{Cov}(X_i, X_j))_{i,j=1,\ldots,d} = \mathbb{E}[(X-\mu)(X-\mu)']$$
$$= \begin{pmatrix} \mathbb{E}[(X_1-\mu_1)^2] & \mathbb{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathbb{E}[(X_1-\mu_1)(X_d-\mu_d)] \\ \mathbb{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathbb{E}[(X_2-\mu_2)^2] & \cdots & \mathbb{E}[(X_2-\mu_2)(X_1-\mu_1)] \\ \vdots & \vdots & & \vdots \\ \mathbb{E}[(X_d-\mu_d)(X_1-\mu_1)] & \mathbb{E}[(X_d-\mu_d)(X_2-\mu_2)] & \cdots & \mathbb{E}[(X_d-\mu_d)^2] \end{pmatrix},$$

the covariance matrix of X, is well defined.

Remark 16 *The covariance matrix of a random vector X is a symmetric nonnegative definite matrix, i.e. $\Sigma_{ij} = \Sigma_{j,i}, i, j = 1, ..., d$ and $y'\Sigma y \geq 0$ for all $y \neq 0, y \in \mathbb{R}^d$. If $\det \Sigma > 0$ then the probability distribution of X is a truly d-dimensional distribution in the sense that cannot be concentrated on a lower dimensional subspace of $\mathbb{R}^d$ and we say that the distribution is nondegenerate or regular. If $\det \Sigma = 0$ then the probability distribution of X is concentrated on such a lower dimensional subspace of $\mathbb{R}^d$ and we say that the distribution is degenerate or singular.*

Proposition 17 *Let X be a d-dimensional random vector with mean vector μ and covariance matrix Σ. Further, let B be an $m \times d$ matrix, let b be a constant m-dimensional vector, and set $Y = BX + b$. Then, Y is a m-dimensional random vector with mean vector and covariance matric given by*

$$\mathbb{E}[Y] = B\mu + b, \qquad \mathrm{Cov}(Y) = B\Sigma B'.$$

Theorem 18 *Let $S \in \mathcal{B}(\mathbb{R}^d)$ be partioned into disjoint measurable subsets $S_0, S_1, ..., S_m$ such that $S = \cup_{i=0}^m S_i$ and such that $\lambda^d(S_0) = 0$. Assume that for each $i = 1, ..., m$, the mapping $g : S_i \to \mathbb{R}^d$ is injective (one to one) and continuously differentiable with non-vanishing Jacobian. Let $Y = g(X)$, where $X : \Omega \to S$ is a random vector with density f_X with respect to λ^d. Then, Y has density (with respect to λ^d) given by*

$$f_Y(y) = \sum_{i=1}^m f_X(g^{-1}(y)) \left|\det J_{g_i^{-1}}(y)\right| \mathbf{1}_{g(S_i)}(y),$$

where g_i^{-1} denotes the inverse map $g_i^{-1} : g(S_i) \to S_i$ and $J_{g_i^{-1}}$ its corresponding Jacobian matrix.

Definition 19 *Let X be a d-dimensional random vector. The characteristic function φ_X of X is defined to be the Fourier transform of its law P_X, that is,*

$$\varphi_X(u) = \int_{\mathbb{R}^d} e^{iu'x} dP_X = \mathbb{E}[e^{iu'X}], \quad x \in \mathbb{R}^d.$$

Theorem 20 *X and Y have the the same characteristic function if and only if they have the same law,i.e., $P_X = P_Y$.*

Proposition 21 *Let X be a d-dimensional random vector with mean vector μ and covariance matrix Σ. Further, let B be an $m \times d$ matrix, let b be a constant m-dimensional vector, and set $Y = BX + b$. Then,*

$$\varphi_Y(u) = e^{iu'b}\varphi_X(B'u).$$

2 Independence

Definition 22 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A family of events $\{A_i\}_{i \in I} \subseteq \mathcal{F}$ are independent iff $\forall J \subset I, J$ finite, one has that*

$$P(\bigcap_{j\in J} A_j) = \prod_{j \in J} P(A_j).$$

Definition 23 *Let $\{\mathcal{E}_i\}_{i\in I}, \mathcal{E}_i \subset \mathcal{F}$, be a family of collections of events. We say that $\{\mathcal{E}_i\}_{i\in I}$ are independent iff every family $\{A_i\}_{i\in I}$, with $A_i \in \mathcal{E}_i$, is independent.*

Definition 24 *Let $\{X_i : \Omega \to \mathbb{R}\}_{i\in I}$ be a family of random variables. We say that $\{X_i\}_{i\in I}$ is independent iff the family of σ-algebras $\{\sigma(X_i)\}_{i\in I}$ is independent.*

Taking into account the previous definitions we have that $\{X_i\}_{i\in I}$ is independent iff for all finite subset $\{i_1, ..., i_n\} \subset I$, and for all $B_1, ...B_n \in \mathcal{B}(\mathbb{R})$, one has that

$$\begin{aligned} P(\{X_{i_1} \in B_1, ..., X_{i_n} \in B_n\}) &:= P(\{X_{i_1} \in B_1\} \cap ... \cap \{X_{i_n} \in B_n\}) \\ &= P(X_{i_1} \in B_1) \cdot \cdots \cdot P(X_{i_n} \in B_n). \end{aligned}$$

The next proposition characterizes the independence of a finite number or r.v.

Proposition 25 *Let $\{X_i\}_{i=1,...,d}$ be random variables with laws $P_{X_i}, i = 1, ..., d$, respectively, and let $X = (X_1, ..., X_d)$ be the random vector with law P_X. Then,*

1. *$\{X_i\}_{i=1,...,d}$ are independent iff $P_X = P_{X_1} \otimes \cdots \otimes P_{X_d}$.*
2. *$\{X_i\}_{i=1,...,d}$ are independent iff $F_X(x_1, ..., x_d) = F_{X_1}(x_1) \cdots F_{X_d}(x_d)$.*
3. *$\{X_i\}_{i=1,...,d}$ are independent iff $\varphi_X(u_1, ..., u_d) = \varphi_{X_1}(u_1) \cdots \varphi_{X_d}(u_d)$*
4. *Assume that $P_X \ll \lambda^d$. Then, $\{X_i\}_{i=1,...,d}$ are independent iff $f_X(x_1, ..., x_d) = f_{X_1}(x_1) \cdots f_{X_d}(x_d)$.*

The next proposition only gives necessary conditions for the independence.

Proposition 26

1. *Let $(E_i, \mathcal{E}_i)_{i\in I}$ and $(G_i, \mathcal{G}_i)$ be two families of measurable spaces. Let $\{X_i\}_{i\in I}$ be a family of random elements $X_i : \Omega \to E_i$, and $\{g_i\}_{i\in I}$ a family of measurable transformations $g_i : E_i \to G_i$. Then, $\{g_i \circ X_i\}_{i\in I}$ is a family of independent random elements $g_i \circ X_i : \Omega \to G_i$.*
2. *Let $\{X_i\}_{i=1,...,d}$ be independent random variables with finite expectation, i.e, $\mathbb{E}[|X_i|] < \infty, i = 1, .., d$, then $X_1 \cdots X_d$ has finite expectation and*

$$\mathbb{E}[X_1 \cdots X_d] = \mathbb{E}[|X_1|] \cdots \mathbb{E}[|X_d|].$$

3. *Let $\{X_i\}_{i=1,...,d}$ be independent r.v. with $\mathbb{E}[|X_i|^2] < \infty, i = 1, .., d$, then*

$$\mathrm{Var}[X_1 + \cdots + X_d] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_d].$$

3 The Multivariate Normal Distribution

Definition 27 *A d-dimensional random vector X is Gaussian (or multivariate normal) iff every linear combination $a'X = \sum_{i=1}^d a_i X_i$ has a (one dimensional) Gaussian distribution (possibly degenerate). The notation $X \sim \mathcal{N}(\mu, \Sigma)$ is used to denote that X has a multivariate normal distribution with mean vector μ and covariance matrix Σ. We say that X is a standard multivariate normal random vector if $X \sim \mathcal{N}(0, I_d)$ where I_d is the d-dimensional identity matrix.*

Theorem 28 *Suppose that $X \sim \mathcal{N}(\mu, \Sigma)$ and set $Y = BX + b$. Then $Y \sim \mathcal{N}(B\mu + b, B\Sigma B\prime)$.*

The previous theorem is useful to obtain a random vector with $X \sim \mathcal{N}(\mu, \Sigma)$ from an affine transform of a random vector $Z \sim \mathcal{N}(0, I_d)$. One has that $X = \Sigma^{1/2} Z + \mu$, where $\Sigma^{1/2}$ is a square root of the covariance matrix Σ, that is, $\Sigma^{1/2}$ is a matrix B such that $BB = \Sigma$.

Theorem 29 ***Theorem 30*** *Let $X \sim \mathcal{N}(\mu, \Sigma)$ and $\det \Sigma > 0$. Then $P_X \ll \lambda^d$ and*

$$f_X(x) = \frac{dP_X}{d\lambda^d}(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right), \qquad x \in \mathbb{R}^d.$$

Theorem 31 *Let $X \sim \mathcal{N}(\mu, \Sigma)$. Then $\{X_i\}_{i=1,...,d}$ are independent iff $\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j) = 0$. If, in addition, $\det \Sigma > 0$ then*

$$f_X(x) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi \Sigma_{i,i}}} \exp\left(-\frac{(x-\mu_i)^2}{\Sigma_{i,i}}\right).$$

4 Conditional Expectation

The concept of conditional expectation is crucial in many areas of modern Probability Theory. Important types of stochastic processes, like martingales and Markov processes, are defined using conditional expectation. Its definition stems from the following proposition.

Definition 32 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and $X \in L^1(\Omega, \mathcal{F}, P)$. The conditional expectation of X given the σ-algebra $\mathcal{G}$, denoted by $\mathbb{E}[X|\mathcal{G}]$, is the ($P$-a.s unique) $\mathcal{G}$-measurable random variable satisfying*

$$\mathbb{E}[\mathbf{1}_B X] = \int_B X dP = \int_B \mathbb{E}[X|\mathcal{G}] dP = \mathbb{E}[\mathbf{1}_B \mathbb{E}[X|\mathcal{G}]], \quad \forall B \in \mathcal{G}.$$

Proposition 33 $\mathbb{E}[X|\mathcal{G}]$ *exists.*
Proof. *Assume first that $X \geq 0$. Then, define the measure μ by*

$$\mu(B) = \int_B X dP, \quad A \in \mathcal{F}.$$

The measure $\mu \ll P$. This property also holds when we restrict both measures to $\mathcal{G}$. By the Radon-Nikodym Theorem, there exists a density f measurable with respect to $\mathcal{G}$ such that

$$\mu(B) = \int_B f dP, \quad A \in \mathcal{G}.$$

If we define $\mathbb{E}[X|\mathcal{G}] := f$ the statement of the proposition is satisfied. The general case follows by considering the decomposition $X = X^+ - X^-$ and using the hypothesis (included in the definition) that $X \in L^1(\Omega, \mathcal{F}, P)$. ■

The intuitive meaning of the conditional expectation is as follows. A random experiment has been performed. The only information available to you regarding which sample point ω has been chosen is the set of values $Z(\omega)$ for every $\mathcal{G}$-measurable random variable Z. Then $\mathbb{E}[X|\mathcal{G}](\omega)$ is the expected value of $X(\omega)$ given this information.

Lemma 34 (Properties of the conditional expectation) *Assume that all the r.v. appearing below are integrable.*

1. *Conservation of the expectation:* $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.
2. *Linearity:* $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$, *for all* $a, b \in \mathbb{R}$.
3. *Positivity:* $X \geq 0 \Rightarrow \mathbb{E}[X|\mathcal{G}] \geq 0$.
4. *Conditional monotone convergence: If $\{X_n\}_{n\geq 1}$ is an increasing sequence of positive random variables such that $X_n \uparrow X$, a.s., then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$, a.s..*
5. *Conditional Fatou lemma: If $X_n \geq 0, \forall n \geq 1$, then $\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]$.*
6. *Conditional dominated convergence: $|X_n| \leq Y \in L^1(\Omega, \mathcal{F}, P), \forall n \geq 1$ and $X_n \to X$, a.s., then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$, a.s..*
7. *Conditional Jensen's inequality: Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, $\varphi(X) \in L^1(\Omega, \mathcal{F}, P)$. Then, $\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}]$.*
8. *Tower property: Let $\mathcal{H} \subseteq \mathcal{G}$ be a sub-σ-algebra of $\mathcal{G}$. Then,*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}].$$

9. *"What is measurable goes out": Assume that XY and X are integrable and that Y is $\mathcal{G}$-measurable. Then, $\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$. In particular, if X is $\mathcal{G}$-measurable then $\mathbb{E}[X|\mathcal{G}] = X$.*
10. *"What is independent vanishes": If $\mathcal{H}$ is a σ-algebra independent of $\sigma(\mathcal{G}, \sigma(X))$, then $\mathbb{E}[X|\sigma(\mathcal{G}, \sigma(X))] = \mathbb{E}[X|\mathcal{G}]$. In particular, if X is independent of $\mathcal{G}$ then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.*

11. *Substitution property: Let X and Y be two random variables and $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a measurable function. If X is $\mathcal{G}$-measurable and $h(X,Y)$ is integrable, then* $\mathbb{E}[h(X,Y)|\mathcal{G}] = \mathbb{E}[h(a,Y)|\mathcal{G}]|_{a=X}$.

Theorem 35 (Factorization Theorem) *Let Ω be a set. Let $(E, \mathcal{E})$ a measurable space. Let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to E$ be two mappings. Then, X is $\sigma(Y)$-measurable iff there exists a function $\varphi : E \to \mathbb{R}, (\mathcal{E}, \mathcal{B}(\mathbb{R}))$-measurable such that*

$$X = \varphi(Y).$$

Corollary 36 *If $Y : \Omega \to \mathbb{R}^d$, then $X : \Omega \to \mathbb{R}$ is $\sigma(Y)$-measurable iff there exists a measurable function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that*

$$X = \varphi(Y).$$

5 Conditional Probability*

Definition 37 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$ and $A \in \mathcal{F}$. We call the conditional probability of A given $\mathcal{G}$, the r.v.*

$$P(A|\mathcal{G}) = \mathbb{E}[\mathbf{1}_A|\mathcal{G}], \quad P\text{-a.s.}$$

Using the properties of the conditional expectation one can see that $P(\cdot|\mathcal{G})$ has the same properties as those of a probability measure, with the exeception that they only hold a.s.. It is natural to ask wether $P(A|\mathcal{G})(\omega)$ as a function of $A \in \mathcal{F}$ is a probability measure, at least P-a.s. $\omega \in \Omega$. Unfortunately the answer is negative in general because to find a version of such object we need to consider an uncountable number of exceptional sets. However, when we add some topological structure to Ω we have a positive result.

Definition 38 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. A regular conditional probability of P given $\mathcal{G}$ is a mapping*

$$\begin{array}{rccc} p & : \Omega \times \mathcal{F} & \to & [0,1] \\ & (\omega, A) & \to & p(\omega, A) \end{array},$$

such that

1. *For each fixed $A \in \mathcal{F}$,*
$$\begin{array}{rccc} p(\cdot, A) & : \Omega & \to & [0,1] \\ & \omega & \to & p(\omega, A) \end{array},$$
is a version of the conditional probability of A given $\mathcal{G}$,i.e.,
   (a) *$p(\cdot, A)$ is $\mathcal{G}$-measurable.*
   (b) *$\forall G \in \mathcal{G}, P(A \cap G) = \int_G \mathbf{1}_A dP = \int_G p(\omega, A) P(d\omega)$.*
2. *P-a.s. $\omega \in \Omega$,*
$$\begin{array}{rccc} p(\omega, \cdot) & : \mathcal{F} & \to & [0,1] \\ & A & \to & p(\omega, A) \end{array},$$
is a probability in $(\Omega, \mathcal{F})$.

In the previous definition we can change the "P-a.s. $\omega \in \Omega$" in condition 2. by "$\forall \omega \in \Omega$".
The following is general positive result on the existence of conditional probabilities.

Theorem 39 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $Y : \Omega \to E$ a measurable function from $(\Omega, \mathcal{F})$ to a measurable space $(E, \mathcal{E})$, where E is a Polish space (i.e. metric, separable and complete) and $\mathcal{E} = \mathcal{B}(E)$. Let $\mathcal{G}$ be sub-σ-algebra of $\mathcal{F}$. Then, there exists a regular conditional probability of Y given $\mathcal{G}$. That is: There exists a mapping*

$$\begin{array}{rccc} p & : \Omega \times \mathcal{E} & \to & [0,1] \\ & (\omega, A) & \to & p(\omega, A) \end{array},$$

such that

1. $\forall A \in \mathcal{E}$,

$$\begin{array}{rccc} p(\cdot, A) & : \Omega & \rightarrow & [0,1] \\ & \omega & \rightarrow & p(\omega, A) \end{array},$$

satisfies

(a) $p(\cdot, A)$ *is* $\mathcal{G}$*-measurable.*

(b) $\forall G \in \mathcal{G}$

$$P\left(\{Y \in A\} \cap G\right) = \int_G \mathbf{1}_{\{Y \in A\}} dP = \int_G p(\cdot, \{Y \in A\})(\omega) P(d\omega).$$

2. P*-a.s.* $\omega \in \Omega$,

$$\begin{array}{rccc} p(\omega, \cdot) & : \mathcal{F} & \rightarrow & [0,1] \\ & A & \rightarrow & p(\omega, A) \end{array},$$

is a probability in $(E, \mathcal{E})$.

In particular if we choose $(\Omega, \mathcal{F}) = (E, \mathcal{E})$ and Y being the identity we get that if $(\Omega, \mathcal{F})$ is a Polish space with its Borel σ-algebra then there exists a conditional probability of P given $\mathcal{G}$.

Theorem 40 *Let* $X \in L^1(\Omega, \mathcal{F}, P)$*. Let* p *be a regular conditional probability of* P *given* $\mathcal{G}$ *a sub-*σ*-algebra of* $\mathcal{F}$*. Then,* P*-a.s.* $\omega \in \Omega$*,* X *is integrable with respect to* $p(\omega, \cdot)$ *and*

$$\int_\Omega X(\omega') p(\omega, d\omega') = \mathbb{E}[X|\mathcal{G}](\omega).$$

Definition 41 *We say that a* σ*-algebra* $\mathcal{G}$ *is countably generated (or separable) if there exists a sequence* $\{A_n\}_{n \geq 1} \subset \mathcal{G}$ *such that* $\sigma\left(\{A_n\}_{n\geq 1}\right) = \mathcal{G}$*.*

Theorem 42 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space,* $\mathcal{G}$ *a countably generated sub-*σ*-algebra of* $\mathcal{F}$*. If a regular conditional probability of* P *given* $\mathcal{G}$ *exists, then we can choose a version of it such that*

$$p(\omega, A) = \mathbf{1}_A(\omega), \quad \forall \omega \in \Omega, \forall A \in \mathcal{G}. \tag{1}$$

Some authors require property (1) to hold when defining a regular conditional probability.

As a consequence of the previous arguments we have that there always exists a regular version of the law of a random vector conditioned to other random vector.

6 Conditional Expectation and Random Variables

Definition 43 *Let* $X, Y : \Omega \rightarrow \mathbb{R}$ *be two random variables. We call law of* Y *conditioned by* X *any transition probability measure* p *from* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *to* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *satisfying*

$$P(X \in A, Y \in B) = \int_A p\,(x, B)\, P_X(dx), \quad \forall A, B \in \mathcal{B}(\mathbb{R}).$$

Sometimes one writes $P(Y \in B | X = x) := p\,(x, B)$, but one has to keep in mind that this function is only defined P_X-a.s. as the following proposition shows.

Proposition 44 *Given two random variables* X, Y *there always exists a law of* Y *conditioned to* X*. If* p *and* $p\prime$ *are two of them then for all* $B \in \mathcal{B}(\mathbb{R}), p(\cdot, B) = p'(\cdot, B), P_X$*-a.s.*

Example 45 *Let* (X, Y) *be a random vector with* $P_{(X,Y)} \ll \lambda^2$*. Then, we can take*

$$p(x, B) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & \text{if} \quad f_X(x) \neq 0 \\ Q(B) & \text{if} \quad f_X(x) = 0 \end{cases},$$

where Q *is any probability measure. Let's check it:*

$$P(\{X \in A, Y \in B\}) = \int_{A \times B} f_{X,Y}(x, y) dx dy$$

$$
\begin{aligned}
&= \int_A \left(\int_B f_{X,Y}(x,y) \mathbf{1}_{\{f_X(x)\neq 0\}} dy \right) dx \\
&+ \int_A \left(\int_B f_{X,Y}(x,y) \mathbf{1}_{\{f_X(x)= 0\}} dy \right) dx \\
&= \int_A \left(\int_B f_X(x) \frac{f_{X,Y}(x,y)}{f_X(x)} \mathbf{1}_{\{f_X(x)\neq 0\}} dy \right) dx \\
&= \int_A p(x,B) f_X(x) dx = \int_A p(x,B) P_X(dx).
\end{aligned}
$$

Where we have used the image measure theorem,then Fubini's Theorem and finally we take into account that

$$
\begin{aligned}
\int_A \left(\int_B f_{X,Y}(x,y) \mathbf{1}_{\{f_X(x)=0\}} dy \right) dx &\leq \int_A \mathbf{1}_{\{f_X(x)=0\}} \left(\int_{\mathbb{R}} f_{X,Y}(x,y) dy \right) dx \\
&= \int_A \mathbf{1}_{\{f_X(x)=0\}} f_X(x) dx = 0
\end{aligned}
$$

Definition 46 *Let $X, Y : \Omega \to \mathbb{R}$ be two random variables. Let p be a law of Y conditioned to X. Let $g : \mathbb{R} \to \mathbb{R}$ a Borel measurable function such that the integral of g with respect to $p(x, \cdot)$. One defines*

$$
\mathbb{E}\left[g(Y)|X = x\right] := \int_{\mathbb{R}} g(y) p(x, dy),
$$

which is called the conditional expectation of $g(Y)$ by $\{X = x\}$.

The previous definition is not precise because it depends on the particular version that we use of the law of Y conditioned to X. That is, as p and p' are only defined P_X-a.s., it does not make sense to talk about the value of such functions for a particular x. For that reason, it is better to leave x as a variable and require that exists the integral of g with respect to $p(x, \cdot)$, P_X-a.e. x and think about the function

$$
x \longmapsto \mathbb{E}\left[g(Y)|X = x\right].
$$

Definition 47 *Given a r.v. X, the conditional expectation of a random variable Y with respect to X is the conditional expectation of Y given $\sigma(X)$.*

The connection of the latter definition with the previous one is as follows. By the Factorization Theorem we know that exists $\varphi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathbb{E}[Y|\sigma(X)] = \varphi(X)$. Therefore, we may define $\mathbb{E}[Y|X = x] = \varphi(x)$. However, the function φ is only defined outside sets of P_X-measure zero. Hence, we have to understand $\mathbb{E}[Y|X = x]$ as the function that satisfies

$$
\begin{aligned}
\int_{X^{-1}(B)} Y(\omega) P(d\omega) &= \int_{X^{-1}(B)} \mathbb{E}[Y|X](\omega) P(d\omega) = \int_{X^{-1}(B)} (\varphi \circ X)(\omega) P(d\omega) \\
&= \int_B \varphi(x) P_X(dy) = \int_B \mathbb{E}[Y|X = x] P_X(dx), \quad B \in \mathcal{B}(\mathbb{R}),
\end{aligned}
$$

by the Image Measure Theorem.

Theorem 48 *Assume that $\mathbb{E}\left[g(Y)\right]$ exists. Then, $\mathbb{E}\left[g(Y)|X = x\right]$ exists P_X-a.s. $x \in \mathbb{R}$, and*

$$
\mathbb{E}\left[g(Y)\right] = \int_{\mathbb{R}} \mathbb{E}\left[g(Y)|X = x\right] P_X(dx).
$$

In particular, if $g(Y)$ is integrable with respect to P, then $\mathbb{E}\left[g(Y)|X = x\right]$ is finite P_X-a.s.

Proof. We have that

$$
\begin{aligned}
\mathbb{E}\left[g(Y)\right] &= \int_{\mathbb{R}^2} g(y) P_{X,Y}(d(x,y)) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(y) p(x, dy) \right) P_X(dx)
\end{aligned}
$$

$$= \int_{\mathbb{R}} \mathbb{E}\left[g(Y)|X = x\right] P_X(dx),$$

where in the first equality we have used the image measure theorem and in the second equality we have used the product measure theorem. Note that the law $P_{X,Y}$ is the product measure constructed from the probability P_X and the transition probability p. ■

References

[1] A. Alabert. *Mesura i Probabilitat.* Servei de Publicacions de la UAB. 1997. (In Catalan)

[2] A. Alabert. *Difusions I.* Unpublished. You cand find it at *http://gent.uab.cat/aureli_ alabert/.* (In Catalan)

[3] R. B. Ash. *Real Analysis and Probabilty.* Academic Press, New York. 1972.

[4] J. Jacod and P. Protter. *Probability Essentials.* Second Edition. Springer, New York. 2003.

[5] R. Leadbetter, S. Cambanis and V. Pipiras. *A Basic Course in Measure and Probability. Theory for Applications.* Cambridge University Press, New York. 2014.