Advanced probability

These summaries aims to review the basic notions of probability theory in a more abstract setting. We will not prove any result here as most of them are from previous courses. Furthermore, we will skip some elementary definitions already defined in other summaries.

1. | Basics of measure theory and integration

Definition 1 (**σ-algebra**). Let E be a set. A *σ-algebra* $\mathcal{E}$ on E is a collection of subsets of E such that:

1. $\varnothing \in \mathcal{E}$.
2. $\forall A \in \mathcal{E}$, $A^c \in \mathcal{E}$.
3. $\forall (A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}$.

The pair $(E, \mathcal{E})$ is called a *measurable space*.

Definition 2. Let E be a set and $\mathcal{F}$ be a collection of subsets of E. The *σ-algebra generated by $\mathcal{F}$* is the smallest σ-algebra containing $\mathcal{F}$, i.e.:

$$\sigma(\mathcal{F}) := \bigcap_{\substack{\mathcal{E} \text{ is a } \sigma\text{-algebra} \\ \mathcal{F}\subseteq\mathcal{E}}} \mathcal{E}$$

Definition 3. Let $(E, \mathcal{E})$, $(F, \mathcal{F})$ be measurable spaces. A function $f : E \to F$ is said to be *measurable* if $\forall A \in \mathcal{F}$, $f^{-1}(A) \in \mathcal{E}$.

Definition 4 (**Measure**). Let $(E, \mathcal{E})$ be a measurable space. A function $\mu : \mathcal{E} \to [0, \infty]$ is said to be a *measure* if:

1. $\mu(\varnothing) = 0$.
2. μ is σ-additive, i.e. $\forall (A_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$ pairwise disjoint, we have:
$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

The triple $(E, \mathcal{E}, \mu)$ is called a *measure space*.

Definition 5. Let $(E, \mathcal{E}, \mu)$ be a measurable space and $f : E \to [0, \infty]$ be a measurable function. We define the *integral of f with respect to μ* as:

$$\int_E f \,\mathrm{d}\mu := \sup\left\{\int_E g \,\mathrm{d}\mu : g \leq f, g \text{ simple}\right\}$$

Definition 6. Let $(E, \mathcal{E}, \mu)$ be a measurable space and $f : E \to \mathbb{R}$ be a measurable function. Suppose that $\int_E |f| \,\mathrm{d}\mu < \infty$. Then, we define the *integral of f with respect to μ* as:

$$\int_E f \,\mathrm{d}\mu := \int_E f^+ \,\mathrm{d}\mu - \int_E f^- \,\mathrm{d}\mu$$

Theorem 7 (**Monotone convergence theorem**). Let $(E, \mathcal{E}, \mu)$ be a measurable space and $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions $f_n : E \to [0, \infty]$ such that $\forall n \in \mathbb{N}$, $f_n \leq f_{n+1}$. Then:

$$\int_E \lim_{n\to\infty} f_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int_E f_n \,\mathrm{d}\mu$$

Theorem 8 (**Fatou's lemma**). Let $(E, \mathcal{E}, \mu)$ be a measurable space and $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions $f_n : E \to [0, \infty]$. Then:

$$\int_E \liminf_{n\to\infty} f_n \,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_E f_n \,\mathrm{d}\mu$$

Theorem 9 (**Dominated convergence theorem**). Let $(E, \mathcal{E}, \mu)$ be a measurable space and $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions $f_n : E \to \mathbb{R}$ such that $\forall n \in \mathbb{N}$, $|f_n| \leq g$ for some $g : E \to [0, \infty]$ integrable. Then:

$$\int_E \lim_{n\to\infty} f_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int_E f_n \,\mathrm{d}\mu$$

Proposition 10. Let $(E, \mathcal{E}, \mu)$ be a measurable space and $f : E \times I \to \mathbb{R}$ be a measurable function, where $I \subseteq \mathbb{R}$ is an interval. Assume that $\forall \lambda \in I$, $f(\cdot, \lambda)$ is integrable and that for some $k \in \mathbb{N} \cup \{0\}$ and $\forall x \in E$ we have $f(x, \cdot) \in \mathcal{C}^k(I)$ and $\left|\partial_\lambda^k f(x, \lambda)\right| \leq g(x)$ for some $g : E \to [0, \infty]$ integrable. Then, the function $F : I \to \mathbb{R}$ defined by:

$$F(\lambda) := \int_E f(x, \lambda) \,\mathrm{d}\mu(x)$$

is in $\mathcal{C}^k(I)$ and $\forall j \in \{0, \ldots, k\}$ we have:

$$\partial_\lambda^j F(\lambda) = \int_E \partial_\lambda^j f(x, \lambda) \,\mathrm{d}\mu(x)$$

Definition 11 (**Product measure**). Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two measurable spaces. We define the *product measure* $\mu \otimes \nu$ on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ as:

$$\forall A \in \mathcal{E}, B \in \mathcal{F}, \quad \mu \otimes \nu(A \times B) := \mu(A)\nu(B)$$

Definition 12. Let $(E, \mathcal{E}, \mu)$ be a measurable space. We say that μ is *σ-finite* if there exists a sequence $(E_n)_{n\in\mathbb{N}} \subseteq \mathcal{E}$ such that $\forall n \in \mathbb{N}$, $\mu(E_n) < \infty$ and $\bigcup_{n\in\mathbb{N}} E_n = E$.

Theorem 13 (**Fubini**). Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two σ-finite measurable spaces and $f : E \times F \to \mathbb{R}$ be a measurable function. Then, the following are equivalent:

1. f is integrable with respect to $\mu \otimes \nu$.
2. $\displaystyle\int_E \left(\int_F |f(x, y)| \,\mathrm{d}\nu(y)\right) \mathrm{d}\mu(x) < \infty$.
3. $\displaystyle\int_F \left(\int_E |f(x, y)| \,\mathrm{d}\mu(x)\right) \mathrm{d}\nu(y) < \infty$.

And if any of the above holds, then:

$$\int_{E\times F} f \,\mathrm{d}(\mu \otimes \nu) = \int_E \left(\int_F f(x,y)\,\mathrm{d}\nu(y) \right) \mathrm{d}\mu(x)$$
$$= \int_F \left(\int_E f(x,y)\,\mathrm{d}\mu(x) \right) \mathrm{d}\nu(y)$$

Definition 14. Let $(E, \mathcal{E}, \mu)$ be a measurable space and $f : E \to [0, \infty]$ be a measurable function. We define the measure ν on $(E, \mathcal{E})$ as $\forall A \in \mathcal{E}$:

$$\nu(A) := \int_A f \,\mathrm{d}\mu$$

In that case, we say that f is the *density* of ν with respect to μ, also denoted by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = f$.

Definition 15. Let $(E, \mathcal{E}, \mu)$ be a measurable space. A measure ν on $(E, \mathcal{E})$ is said to be *absolutely continuous* with respect to μ if $\forall A \in \mathcal{E}$ such that $\mu(A) = 0$, we have $\nu(A) = 0$.

Theorem 16 (Radon-Nikodym). Let μ, ν be two σ-finite on a measurable space $(E, \mathcal{E})$ such that ν is absolutely continuous with respect to μ. Then, ν admits a density f with respect to μ.

2. | Probability spaces and random variables

Definition 17. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a measure on $(\Omega, \mathcal{F})$ such that $\mathbb{P}(\Omega) = 1$. In this context, the elements if $\mathcal{F}$ are called *events*.

Definition 18 (Random variable). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ be a measurable space. An *E-valued random variable* is a measurable function from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$[1].

Definition 19 (Expectation). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable. We define the *expectation of X* as:

$$\mathbb{E}(X) := \int_\Omega X \,\mathrm{d}\mathbb{P}$$

Definition 20. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(E, \mathcal{E})$ be a measurable space and X be a E-valued random variable. We define the *law of X* as the measure image on E, defined for all $A \in \mathcal{E}$ as:

$$\mathcal{L}^X(A) := \mathbb{P} \circ X^{-1}(A) = \mathbb{P}(X \in A)$$

Proposition 21. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X be a random variable and $h : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $h(X)$ is integrable. Then:

$$\mathbb{E}(h(X)) = \int_\mathbb{R} h(x) \,\mathrm{d}\mathcal{L}^X(x)$$

[1] When E is not specified, we will assume that $E = \mathbb{R}$.

In particular, if the law of X admits a density f with respect to the Lebesgue measure, then:

$$\mathbb{E}(h(X)) = \int_\mathbb{R} h(x) f(x) \,\mathrm{d}x$$

Definition 22. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable. We define the *σ-algebra generated by X* as the smallest σ-algebra containing X, i.e.:

$$\sigma(X) := \sigma(X^{-1}(A) : A \in \mathcal{E})$$

Proposition 23. Let X be a $(E, \mathcal{E})$-valued random variable and Y be a $\sigma(X)$-measurable random variable. Then, there exists a measurable function $f : E \to \mathbb{R}$ such that $Y = f(X)$.

Proposition 24 (Jensen's inequality). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X be a random variable and $h : \mathbb{R} \to \mathbb{R}$ be a convex function. Then:

$$h(\mathbb{E}(X)) \le \mathbb{E}(h(X))$$

as long as the expectations are well-defined.

Proposition 25. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X be a random variable and $h : \mathbb{R} \to \mathbb{R}$ be a non-decreasing positive function. Then:

$$\mathbb{P}(X \ge t) \le \frac{\mathbb{E}(h(X))}{h(t)}$$

3. | Conditional expectation

Proposition 26. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Then, for any integrable random variable X, there exists a unique (up to a.s.) random variable Y such that:

1. Y is $\mathcal{G}$-measurable.
2. For any Z $\mathcal{G}$-measurable such that XZ is integrable, we have that $\mathbb{E}(XZ) = \mathbb{E}(YZ)$.

We denote $Y = \mathbb{E}(X \mid \mathcal{G})$ and call it the *conditional expectation of X given $\mathcal{G}$*.

Remark. If the variable X is not integrable but it is non-negative, then the above holds for any Z non-negative as well.

Remark. The conditional expectation, when restricted to $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, is the orthogonal projection of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Proposition 27. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra and X, Y be random variables. Then, assuming that all the expectations below are well-defined, we have:

1. If Y, Z are $\mathcal{G}$-measurable random variables, then $\mathbb{E}(XY + Z \mid \mathcal{G}) = Y\mathbb{E}(X \mid \mathcal{G}) + Z$.
2. If $X \overset{\text{a.s.}}{\le} Y$, then $\mathbb{E}(X \mid \mathcal{G}) \overset{\text{a.s.}}{\le} \mathbb{E}(Y \mid \mathcal{G})$.

3. $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$.

4. $\mathbb{E}(|\mathbb{E}(X \mid \mathcal{G})|) \leq \mathbb{E}(|X|)$.

5. *Tower property*: if $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ are σ-algebras, then $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H})$.

6. If X is independent of $\mathcal{G}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$.

7. If X is independent of $\mathcal{G}$ and Y is $\mathcal{G}$-measurable, then for any measurable function f, we have that $\mathbb{E}(f(X,Y) \mid \mathcal{G}) = g(Y)$, where $g(y) = \mathbb{E}(f(X,y))$. This is often written as:

$$\mathbb{E}(f(X,Y) \mid \mathcal{G}) = \mathbb{E}(f(X,y))|_{y=Y}$$

Definition 28. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X, Y be random variables. We define the *conditional expectation of X given Y* as:

$$\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \sigma(Y))$$

Remark. It can be seen that this definition coincides with the one given by:

$$\mathbb{E}(X \mid Y) = \sum_{y \in \operatorname{supp}(Y)} \mathbb{E}(X \mid Y = y)\mathbf{1}_{Y=y}$$

Proposition 29. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X, Y be random variables. Assume that (X,Y) has a law which admits a density $f = f(x,y)$ (which for simplicity we may think it with respect to $\mathrm{d}x\,\mathrm{d}y$). Then, for any function h such that $\mathbb{E}(h(X))$ makes sense:

$$\mathbb{E}(h(X) \mid Y) \overset{\text{a.s}}{=} \frac{\int_{\mathbb{R}} h(x)f(x,Y)\,\mathrm{d}x}{\int_{\mathbb{R}} f(x,Y)\,\mathrm{d}x}$$

Definition 30. A *probability kernel* on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a function $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$ such that:

1. $\forall y \in \mathbb{R}$, $K(y, \cdot)$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

2. $\forall A \in \mathcal{B}(\mathbb{R})$, $K(\cdot, A)$ is measurable.

Theorem 31. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X, Y be random variables. Then, there exists a probability kernel $\mathcal{L}^{X|Y}$, called the *conditional law of X given Y*, such that for any bounded measurable function f we have:

$$\mathbb{E}(f(X) \mid Y) = \int_{\mathbb{R}} f(x)\,\mathrm{d}\mathcal{L}^{X|Y}(Y,x)$$

4. | Martingales

Definition 32. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is a sequence of sub-σ-algebras $(F_n)_{n\in\mathbb{N}}$ such that $\forall n \in \mathbb{N}$, $F_n \subseteq F_{n+1}$. The tuple $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ is called a *filtered probability space*.

Definition 33. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space. A stochastic process $(X_n)_{n\in\mathbb{N}}$ is *adapted* to $(F_n)_{n\in\mathbb{N}}$ if $\forall n \in \mathbb{N}$, X_n is F_n-measurable.

Definition 34. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space and $(M_n)_{n\in\mathbb{N}}$ be an adapted stochastic process. We say that $(M_n)_{n\in\mathbb{N}}$ is a *martingale* if $\forall n \in \mathbb{N}$, $\mathbb{E}(|M_n|) < \infty$ and $\mathbb{E}(M_{n+1} \mid F_n) = M_n$.

Remark. A *submartingale* and *supermartingale* are defined similarly, but with $\mathbb{E}(M_{n+1} \mid F_n) \geq M_n$ and $\mathbb{E}(M_{n+1} \mid F_n) \leq M_n$ respectively.

Definition 35. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space. A *stopping time* is a random variable $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ such that $\forall n \in \mathbb{N}$, $\{\tau \leq n\} \in F_n$.

Definition 36. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space, τ be a stopping time and $M := (M_n)_{n\in\mathbb{N}}$ be a process. We define the *stopped process* $M^\tau := (M_{\tau\wedge n})_{n\in\mathbb{N}}$.

Proposition 37. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space, τ be a stopping time and $M := (M_n)_{n\in\mathbb{N}}$ be a martingale. Then, $M^\tau := (M_{\tau\wedge n})_{n\in\mathbb{N}}$ is a martingale.

Corollary 38. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space, τ be a bounded stopping time and $M := (M_n)_{n\in\mathbb{N}}$ be a martingale. Then, $\mathbb{E}(M_\tau) = \mathbb{E}(M_0)$.

Definition 39. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space, τ be a stopping time. We define:

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : \forall n \in \mathbb{N}, A \cap \{\tau = n\} \in F_n\}$$

Remark. It can be seen that in the above definition, $\mathcal{F}_\tau$ is a σ-algebra.

Proposition 40. Let $(\Omega, \mathcal{F}, \mathbb{P}, (F_n)_{n\in\mathbb{N}})$ be a filtered probability space, $\rho \leq \tau$ be two bounded stopping times, and M be a martingale. Then, $\mathbb{E}(M_\tau \mid \mathcal{F}_\rho) = M_\rho$.

Theorem 41. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(M_n)_{n\in\mathbb{N}}$ be a martingale such that $\sup_{n\in\mathbb{N}} \mathbb{E}(|M_n|) < \infty$. Then, there exists a random variable M_∞ such that $M_n \overset{\text{a.s.}}{\to} M_\infty$.

Definition 42. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A family of random variables $(X_t)_{t\in I}$ is said to be *uniformly integrable* if:

$$\lim_{a\to\infty} \sup_{t\in I} \mathbb{E}(|X_t|\mathbf{1}_{|X_t|>a}) = 0$$

Proposition 43. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n\in\mathbb{N}}$ be a family of uniformly integrable random variables such that $X_n \overset{\text{a.s.}}{\to} X$. Then, X is integrable and $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Theorem 44. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(M_n)_{n\in\mathbb{N}}$ be a martingale bounded in L^p, $1 < p < \infty$. Then, there exists a random variable M_∞ such that $M_n \overset{L^p}{\to} M_\infty$ and $M_n \overset{\text{a.s.}}{\to} M_\infty$.