Appendix A
Supplementary material on probability theory

A.1 Basics of probability theory

As the reader of this book very well knows, probability is an assignment of a number between zero and one to an event. This number tells how likely the event is — if it is nearly zero, the event is very unlikely, and if it is almost one, the event is very likely to occur. Mathematically probability is a *measure* defined on the collection of events. In this chapter we list the required results of *measure theory* and *probability theory*.

A.1.1 Measure theory

The basic concepts of measure theory that reader should be aware of are

- $(X, \mathscr{A})$ a measurable space: X is set, $\mathscr{A}$ its σ-algebra
- measurable function f, (positive) measure μ, integral $\int f \mathrm{d}\mu$
- Lebesgue measure on $\mathbb{R}^d$
- $L^p(\mu)$ space: Measurable f is in $L^p(\mu)$ if $\int |f|^p \mathrm{d}\mu < \infty$. Notation: $\|f\|_p = (\int |f|^p \mathrm{d}\mu)^{1/p}$.
- Product measures: If $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ are measure spaces, then their product space is $(X \times Y, \mathscr{A} \times \mathscr{B}, \mu \times \nu)$ where $X \times Y$ is Cartesian product, $\mathscr{A} \times \mathscr{B}$ the σ-algebra generated by $A \times B, A \in \mathscr{A}$ and $B \in \mathscr{B}$, and $\mu \times \nu$ the unique extension of $A \times B \mapsto \mu(A)\nu(B)$. (Here we have to assume that both measures μ and ν are σ-finite in the sense that X can be written as $X = \bigcup_{k=1}^{\infty} X_k$ where X_k a measurable sets with finite μ-measure and the same holds for Y and ν.)

Here is a summary of some results of measure theory. For the details and proof see any book on measure theory.

- Monotone convergence theorem: If f_n are measurable functions such that $0 \leq f_n \nearrow f$, then $\int f \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \mathrm{d}\mu$

- Dominated convergence theorem: If f_n are measurable functions and $f = \lim_{n\to\infty} f_n$ exists almost everywhere and $\exists g \geq 0$ such that $|f_n| \leq g$ for all n and $\int g \mathrm{d}\mu < \infty$, then $\int f \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \mathrm{d}\mu$.
- Fubini's theorem: Assume that μ and ν are σ-finite. Let $f \in \mathscr{A} \times \mathscr{B}$. If $f \geq 0$ or $\int |f| \mathrm{d}(\mu \times \nu) < \infty$ then $\int_X (\int_Y f \mathrm{d}\nu) \mathrm{d}\mu = \int_{X \times Y} f \mathrm{d}(\mu \times \nu) = \int_Y (\int_X f \mathrm{d}\mu) \mathrm{d}\nu$.
- Radon–Nikodym theorem: If ν is a σ-finite signed measure and μ is a σ-finite measure on $(X, \mathscr{A})$ and ν is absolutely continuous with respect to μ, then exist $g \in \mathscr{F}$ such that $\nu(A) = \int_A g \, \mathrm{d}\mu$. Here ν is absolutely continuous with respect to μ, if $\nu(A) = 0$ whenever $\mu(A) = 0$, $A \in \mathscr{F}$. A notation: $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ and it is called Radon–Nikodym derivative.

A notation which sometimes handy: $f \in \mathscr{A}$ where f is a function on X means that f is $\mathscr{A}$-measurable.

A.1.2 Probability theory

Probability theory is essentially measure theoretical formulation of probability. Therefore the basics of probability are easily accessible to anybody with background in measure theory. Here is a list of basic facts about probability:

- A probability space is a measure space $(\Omega, \mathscr{F}, \mathsf{P})$ such that P is a probability measure, i.e., $\mathsf{P}[\Omega] = 1$. Ω "outcomes", $\mathscr{F}$ "events"
- A random variable is a $\mathscr{F}$-measurable function $X : \Omega \to \mathbb{R}$. H-valued random variable is a measurable function $X : \Omega \to H$ (H is a measurable space).
- The expected value of X is $\mathsf{E}[X] = \int X \mathrm{d}\mathsf{P} \in [-\infty, \infty]$, which makes sense when $X \geq 0$ or when either $\int X^+ \mathrm{d}\mathsf{P} < \infty$ or $\int X^- \mathrm{d}\mathsf{P} < \infty$, where $X = X^+ - X^-$ is the decomposition of X into positive and negative part.
- $L^p(\mathsf{P})$ space: $\|X\|_p = (\mathsf{E}[|X|^p])^{1/p} < \infty$. By Hölder inequality, $\|X\|_p \leq \|X\|_q$ for $1 \leq p \leq q$ and hence $L^q(\mathsf{P}) \subset L^p(\mathsf{P})$ (a fact which isn't necessarily true for general measures).
- Independence: sub-σ-algebras $\mathscr{A}_1, \ldots, \mathscr{A}_n$ of $\mathscr{F}$ are *independent* if

$$\mathsf{P}[A_1 \cap A_2 \cap \ldots \cap A_n] = \mathsf{P}[A_1] \cdot \mathsf{P}[A_2] \cdot \ldots \cdot \mathsf{P}[A_n] \quad \text{for } A_k \in \mathscr{A}_k.$$

Random variables $X_1, X_2, \ldots, X_n$ are independent if the σ-algebras $\sigma(X_1)$, $\sigma(X_2), \ldots,$ $\sigma(X_n)$ are independent

$$\begin{aligned} \Leftrightarrow & \mathsf{P}[\{X_1 \in B_1\} \cap \{X_2 \in B_2\} \cap \ldots \cap \{X_n \in B_n\}] \\ & = \mathsf{P}[X_1 \in B_1] \cdot \mathsf{P}[X_2 \in B_2] \cdot \ldots \cdot \mathsf{P}[X_n \in B_n] \quad \text{for } B_k \in \mathscr{B}_{\mathbb{R}}. \end{aligned}$$

A couple of useful notations:

- $\mathsf{E}[X; E] = \int_E X \mathrm{d}\mathsf{P} = \int \mathbb{1}_E X \mathrm{d}\mathsf{P}$.
- A random variable X induces a measure on $\mathbb{R}$ by $\mu_X(B) = \mathsf{P}[X^{-1}(B)]$ where $B \in \mathscr{B}_{\mathbb{R}}$ and $\mathscr{B}_{\mathbb{R}}$ is the Borel σ-algebra on $\mathbb{R}$. The measure μ_X is called *distribution*

(or *law*) of X. When X and Y induce the same measure, we say that X and Y are equal in distribution and use the notation

$$X \stackrel{\mathrm{d}}{=} Y.$$

A.2 Conditional expected value

Definition A.1. Let X be a $L^1(\mathsf{P},\mathscr{F})$ random variable and let $\mathscr{G} \subset \mathscr{F}$ be a σ-algebra. The *conditional expected value of X given $\mathscr{G}$* is defined to be any random variable Y such that Y is (i) $\mathscr{G}$-measurable and (ii) for any $E \in \mathscr{G}$

$$\int_E X \mathrm{d}\mathsf{P} = \int_E Y \mathrm{d}\mathsf{P}.$$

We then use the notation $\mathsf{E}[X|\mathscr{G}]$ for the conditional expected value and any such Y is called a *version* of $\mathsf{E}[X|\mathscr{G}]$.

Proposition A.1. *The conditional expected value exists and is unique in the sense that if Y and Y' satisfy (i) and (ii) then $Y = Y'$ almost surely. Also the conditional expected value is integrable.*

Proof. Let $G = \{Y \geq 0\}$ which is $\mathscr{G}$-measurable. Then

$$\mathsf{E}[|Y|] = \int_G Y \mathrm{d}\mathsf{P} - \int_{G^c} Y \mathrm{d}\mathsf{P} = \int_G X \mathrm{d}\mathsf{P} - \int_{G^c} X \mathrm{d}\mathsf{P} \leq \|X\|_1,$$

where G^c is the complement $\Omega \setminus G$ of G. Therefore $\mathsf{E}[|Y|] < \infty$.

Existence follows from Radon–Nikodym theorem:

$$E \mapsto \int_E X \mathrm{d}\mathsf{P}$$

is a signed measure, which is absolutely continuous with respect to P. Then the Radon-Nikodym derivative Y of that measure satisfies the properties of the conditional expected value.

Uniqueness: If Y and Y' are version of $\mathsf{E}[X|\mathscr{G}]$, then let $E = \{Y > Y'\}$. Then if $\mathsf{P}[E] > 0$, $\int_E Y \mathrm{d}\mathsf{P} > \int_E Y' \mathrm{d}\mathsf{P}$ which is a contradiction. Hence $\mathsf{P}[\{Y = Y'\}] = 1$. $\square$

Intuitively $\mathsf{E}[X|\mathscr{G}]$ should be thought to the best guess of the value of X given the information contained in $\mathscr{G}$.

Example A.1. (*Perfect information*) If X is $\mathscr{G}$ measurable then $\mathsf{E}[X|\mathscr{G}] = X$.

Example A.2. (*No information*) If X is independent of $\mathscr{G}$ then $\mathsf{E}[X|\mathscr{G}] = \mathsf{E}X$.

Example A.3. (*Relation to the usual conditional expected value*) Let $\Omega_1, \Omega_2, \ldots$ be a finite or countably infinite disjoint partition of Ω into $\mathscr{F}$-measurable sets, each of which has positive probability. If $\mathscr{G}$ is the σ-algebra generated by $\Omega_1, \Omega_2, \ldots$ then

$$\mathsf{E}[X|\mathscr{G}] = \frac{\mathsf{E}[X;\Omega_k]}{\mathsf{P}[\Omega_k]} \quad \text{on } \Omega_k.$$

Note that $\mathscr{G} = \{\bigcup_{k\in I}\Omega_k : I \subset \mathbb{N}\}$.

We list next some properties of conditional expected value, see Section 5.1 of [1] or Section 9.7 of [3].

Theorem A.1. *Let X,Y be $L^1(\mathsf{P},\mathscr{F})$ random variables and $a,b\in\mathbb{R}$ and $\mathscr{G},\mathscr{G}_1,\mathscr{G}_2 \subset \mathscr{F}$ be σ-algebras. Then*

1. $\mathsf{E}[aX+bY|\mathscr{G}] = a\mathsf{E}[X|\mathscr{G}] + b\mathsf{E}[Y|\mathscr{G}]$
2. $\mathsf{E}[\mathsf{E}[X|\mathscr{G}]] = \mathsf{E}[X]$
3. $\mathsf{E}[X\,Y|\mathscr{G}] = Y\,\mathsf{E}[X|\mathscr{G}]$ *if Y is $\mathscr{G}$-measurable and $X\cdot Y$ is $L^1(\mathsf{P},\mathscr{F})$*
4. (Tower property) $\mathsf{E}[\mathsf{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathsf{E}[X|\mathscr{G}_1]$ *if $\mathscr{G}_1 \subset \mathscr{G}_2$*
5. (Jensen's inequality) *If $\phi:\mathbb{R}\to\mathbb{R}$ is convex and $\mathsf{E}[|\phi(X)|]<\infty$ then $\phi(\mathsf{E}[X|\mathscr{G}]) \le \mathsf{E}[\phi(X)|\mathscr{G}]$.*
6. $|\mathsf{E}[X|\mathscr{G}]| \le \mathsf{E}[|X||\mathscr{G}]$ *and when $\mathsf{E}[|X|^2]<\infty$, $|\mathsf{E}[X|\mathscr{G}]|^2 \le \mathsf{E}[|X|^2|\mathscr{G}]$*
7. *If $X_n \to X$ in $L^p(\mathsf{P},\mathscr{F})$ then $\mathsf{E}[X_n|\mathscr{G}] \to \mathsf{E}[X|\mathscr{G}]$ in $L^p(\mathsf{P},\mathscr{F})$.*

The following notation is sometimes used: if X and Y are random variables and $\sigma(Y)$ is the σ-algebra generated by Y, then $\mathsf{E}[X|Y]$ means the same as $\mathsf{E}[X|\sigma(Y)]$.

A.3 Martingales

Definition A.2. A *filtration* on $(\Omega,\mathscr{F})$ is a collection $(\mathscr{F}_t)_{t\in\mathbb{R}_{\ge 0}}$ of sub-σ-algebras $\mathscr{F}_t \subset \mathscr{F}$ such that for each $0 \le s < t$, $\mathscr{F}_s \subset \mathscr{F}_t$.

Recall that a σ-algebra can be thought as information and thus the filtration should be thought as the information that we learn about as t increases. Hence $\mathscr{F}_t$ is the information available at time t. If $\mathscr{F}_t = \sigma(X_s, s\in[0,T])$ where $(X_t)_{t\in\mathbb{R}_{\ge 0}}$ is a stochastic process, then a random variable Y is $\mathscr{F}_t$ measurable if it is a function of the random variables $X_s, s\in[0,T]$.

The class of processes of the following definition is very important.

Definition A.3. A stochastic process $(M_t)_{t\in\mathbb{R}_+}$ is called a *(continuous-time) martingale* with respect to a filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ if

1. M_t is $\mathscr{F}_t$-measurable for each $t \ge 0$,
2. $\mathsf{E}[|M_t|] < \infty$ for each $t \ge 0$,
3. $\mathsf{E}[M_t|\mathscr{F}_s] = M_s$ for each $0 \le s < t$.

If the last property is replaced by $\mathsf{E}[M_t|\mathscr{F}_s] \ge M_s$, the process is called *submartingale*, and if the last property is replaced by $\mathsf{E}[M_t|\mathscr{F}_s] \le M_s$, the process is called *supermartingale*.

Quite many results for martingales are proved using discrete-time martingales.

Definition A.4. A *discrete-time filtration* on $(\Omega, \mathscr{F})$ is a collection $(\mathscr{F}_t)_{t\in\mathbb{Z}_+}$ of sub-σ-algebras $\mathscr{F}_t \subset \mathscr{F}$ such that for each $t \in \mathbb{Z}_+$, $\mathscr{F}_t \subset \mathscr{F}_{t+1}$.

A stochastic process $(M_t)_{t\in\mathbb{Z}_+}$ is called a *(discrete-time) martingale* with respect to a filtration $(\mathscr{F}_t)_{t\in\mathbb{Z}_+}$ if

1. M_t is $\mathscr{F}_t$-measurable for each $t \in \mathbb{Z}_+$,
2. $\mathsf{E}[|M_t|] < \infty$ for each $t \in \mathbb{Z}_+$,
3. $\mathsf{E}[M_{t+1}|\mathscr{F}_t] = M_t$ for each $t \in \mathbb{Z}_+$.

If the last property is replaced by $\mathsf{E}(M_{t+1}|\mathscr{F}_t) \geq M_t$, the process is called *submartingale*, and if the last property is replaced by $\mathsf{E}(M_{t+1}|\mathscr{F}_t) \leq M_t$, the process is called *supermartingale*.

Example A.4. Let $X \in L^1(\mathsf{P}, \mathscr{F})$ and let $(\mathscr{F}_t)_{t\geq 0}$ be a filtration. Then $M_t = \mathsf{E}(X|\mathscr{F}_t)$ is a martingale: 1 holds by the definition of conditional expected value, 2 holds by items 2 and 6 of Theorem A.1 and 3 holds by item 4 of Theorem A.1.

Example A.5. Let $X_0, X_1, X_2, \ldots$ be a sequence of independent integrable random variables such that $\mathsf{E}(X_k) = 0$ for each k and let $\mathscr{F}_n = \sigma(X_0, X_1, X_2, \ldots, X_n)$. Then $(M_n)_{n\in\mathbb{Z}_+}$ defined by

$$M_n = \sum_{k=0}^{n} X_k$$

is a martingale with respect to $(\mathscr{F}_n)_{n\in\mathbb{Z}_+}$.

Example A.6 (The name martingale*).* There is a *gambling strategy* called martingale. Consider a gambler that is playing roulette, where the outcome is either red or black with probability $1/2$ each. After a loss the gambler always doubles his bet and keeps playing until the first time when he wins. After that he stops playing. If the first bet is x, then the gambler is sure to win x by this strategy! Do you see any problem with the martingale strategy? This is related to the previous example when we consider $X_0, X_1, \ldots$ such that $X_0 = 0$, $X_1 = \hat{X}_1$ and

$$X_k = \hat{X}_k \, \mathbb{1}_{\{\text{no wins during rounds } 1,2,\ldots,k-1\}}$$

for $k \geq 2$, where $\hat{X}_k$ are independent and $\mathsf{P}[\hat{X}_k = \pm x 2^k] = 1/2$. Then M_n is the wealth of the gambler after n rounds relative to the wealth at time zero.

Example A.7. Let $X_0, X_1, X_2, \ldots$ be a sequence of independent integrable random variables such that $\mathsf{E}(X_k) = 1$ for each k and let $\mathscr{F}_n = \sigma(X_0, X_1, X_2, \ldots, X_n)$. Then $(M_n)_{n\in\mathbb{Z}_+}$ defined by

$$M_n = \prod_{k=0}^{n} X_k$$

is a martingale with respect to $(\mathscr{F}_n)_{n\in\mathbb{Z}_+}$.

Example A.8. There are many martingales related to Brownian motions. In the main text, we will check the following formulas

$$\mathsf{E}[B_t \mid \mathscr{F}_s] = B_s$$
$$\mathsf{E}[B_t^2 - t \mid \mathscr{F}_s] = B_s^2 - s$$
$$\mathsf{E}\left[\exp\left(\theta B_t - \frac{\theta^2}{2}t\right) \,\middle|\, \mathscr{F}_s\right] = \exp\left(\theta B_s - \frac{\theta^2}{2}s\right).$$

A result that we need about martingales is the next inequality. Its proof is given in the exercises.

Theorem A.2 (Doob's martingale inequality). *Suppose that $(M_t)_{t\in\mathbb{R}_+}$ is a martingale, which has a continuous path almost surely. Then for each $p \geq 1$, $T > 0$, $\lambda > 0$,*

$$\mathsf{P}\left[\sup_{0\leq s\leq t} |M_s| \geq \lambda\right] \leq \frac{1}{\lambda^p}\mathsf{E}[|M_T|^p].$$

This result follows from the following auxiliary results.

Lemma A.1. *If $(M_t)_{t\in\mathbb{R}_+}$ is a martingale, $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function and $\mathsf{E}[|\phi(M_t)|] < \infty$ for all $t \in \mathbb{R}$, then $(\phi(M_t))_{t\in\mathbb{R}_+}$ is a submartingale.*

Proof. Follows from the conditional version of Jensen's inequality which is presented in Theorem A.1. □

Theorem A.3 (Doob's submartingale inequality). *Suppose that $(M_t)_{t\in\mathbb{R}_+}$ is a non-negative submartingale, which has a continuous path almost surely. Then for each $\lambda > 0$,*

$$\mathsf{P}\left[\sup_{0\leq s\leq t} M_s \geq \lambda\right] \leq \frac{1}{\lambda}\mathsf{E}[M_T].$$

The proof is left as an exercise.

A.4 Stopping times and optional stopping

Optional stopping is a concept that extends the martingale property to random times.

Definition A.5. If τ is a stopping time with respect to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$, define the *stopping time σ-algebra* as

$$\mathscr{F}_\tau = \{A \in \mathscr{F} : A \cap \{\tau \leq t\} \in \mathscr{F}_t \text{ for all } t \in \mathbb{R}_+\}$$

In the same way, as $\mathscr{F}_t$ can be thought as the information available at time t, a stopping time σ-algebra $\mathscr{F}_\tau$ can be thought as the information available at a random time τ. The following set of results extends the martingale property to random times.

Theorem A.4. *Let $(M_t)_{t\in\mathbb{R}_+}$ be a continuous martingale and τ and σ stopping times with respect to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$. Then for each $t \in \mathbb{R}_+$*

$$\mathsf{E}[M_{t\wedge\tau}|\mathscr{F}_\sigma] = M_{t\wedge\sigma\wedge\tau}$$

Remark A.1. As seen below, we have to care about the integrability of quantities such as M_τ. Here the non-random number t in $M_{t\wedge\tau}$ guarantees that $\mathsf{E}|M_{t\wedge\tau}| < \infty$.

Corollary A.1. *Let $(M_t)_{t\in\mathbb{R}_+}$ be a continuous martingale and τ be a stopping time with respect to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$. Then the process $(M_t^\tau)_{t\in\mathbb{R}_+}$ defined by*

$$M_t^\tau = M_{t\wedge\tau}$$

is a continuous martingale with respect to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$.

Remark A.2. Stopped local martingales are local martingales by the same argument.

Definition A.6. A collection $\mathscr{C}$ of random variables is said to be *uniformly integrable* if

$$\lim_{m\to\infty} \sup_{X\in\mathscr{C}} \mathsf{E}[|X|\,;\,|X| \geq m] = 0.$$

Corollary A.2. *Let $(M_t)_{t\in\mathbb{R}_+}$ be a continuous martingale and τ and σ almost surely finite stopping times with respect to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$. Assume that $\sigma \leq \tau$. Then*

$$\mathsf{E}[M_\tau|\mathscr{F}_\sigma] = M_\sigma$$

under any of the following conditions:

- *For some constant $C > 0$, $\sigma \leq \tau \leq C$*
- *For some constant $C > 0$, $|M_t| \leq C$ for all t.*
- *The collection of random variables M_t, $t \in \mathbb{R}_+$, is uniformly integrable.*

Remark A.3. In a sense, the first two cases are special cases of the last case.

Remark A.4. In the last case, $M_\sigma = \mathsf{E}[M_\tau|\mathscr{F}_\sigma] = \mathsf{E}[M_\infty|\mathscr{F}_\sigma]$ for some random variable M_∞ and $M_t \to M_\infty$ in L^1.

References

1. Durrett, R.: Probability: theory and examples, fourth edn. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2010)
2. Kallenberg, O.: Foundations of modern probability. Springer Verlag (2002)
3. Williams, D.: Probability with martingales. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, Cambridge (1991)