2 Concepts of Probability Theory

2.1 Probability

The sample space Ω is a non-empty set that models the possible outcomes of a random experiment. Elements ω of Ω are called outcomes.

Definition 2.1. A family $\mathcal{F}$ of subsets of Ω is called a σ-field if

- $\Omega \in \mathcal{F}$,
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,
- $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Members of $\mathcal{F}$ are called measurable sets or events and the pair $(\Omega, \mathcal{F})$ is called a measurable space.

An event happens when we do the random experiment if the outcome ω of the random experiment is a member of the event.

Theorem 2.2.

1. $\emptyset \in \mathcal{F}$,
2. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$,
3. $A_1, \ldots, A_n \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{F}$,
4. $A_1, \ldots, A_n \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{n} A_i \in \mathcal{F}$,
5. $A, B \in \mathcal{F} \Rightarrow A \setminus B \in \mathcal{F}$.

Proof. Simple exercises. □

Intersections of σ-fields are σ-fields. Unions of σ-fields are usually not σ-fields but (using closedness under intersections) there is a smallest σ-field denoted $\bigvee_{\alpha \in A} \mathcal{F}_\alpha$ containing $\bigcup_{\alpha \in A} \mathcal{F}_\alpha$ called the σ-field generated by $\{\mathcal{F}_\alpha\}_{\alpha \in A}$. Likewise, for a family $\mathcal{G}$ of subsets of Ω there is a smallest σ-field $\sigma(\mathcal{G})$ containing the family called the σ-field generated by $\mathcal{G}$.[1]

Example 2.1.

1. $\{\emptyset, \Omega\}$ is a σ-field,
2. the family of all subsets of Ω is a σ-field,
3. $\{\emptyset, A, A^c, \Omega\}$ is a σ-field for any $A \subseteq \Omega$.

[1] Simple exercises.

Definition 2.3. The σ-field $\mathcal{B}$ of subsets of $\mathbb{R}$ generated by all intervals is called the Borel σ-field.

Definition 2.4. A probability measure $\mathbf{P}$ on a measureable space $(\Omega, \mathcal{F})$ is a function $\mathbf{P} : \mathcal{F} \to [0,1]$ such that

- $\mathbf{P}\{\Omega\} = 1,$
- $\mathbf{P}\{A^c\} = 1 - \mathbf{P}\{A\} \quad \text{for } \ A \in \mathcal{F},$
- $\mathbf{P}\{\bigcup_{n=1}^{\infty} A_n\} = \sum_{n=1}^{\infty} \mathbf{P}\{A_n\} \quad \text{for disjoint } \ A_1, A_2, \ldots \in \mathcal{F}.$

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Theorem 2.5. 1. $\mathbf{P}\{\emptyset\} = 0,$

2. $\mathbf{P}\{A \cup B\} = \mathbf{P}\{A\} + \mathbf{P}\{B\} - \mathbf{P}\{A \cap B\} \quad \text{for } \ A, B \in \mathcal{F},$
3. $\mathbf{P}\{A \setminus B\} = \mathbf{P}\{A\} - \mathbf{P}\{B\} \quad \text{for } \ \mathcal{F} \ni B \subseteq A \in \mathcal{F},$
4. $\mathbf{P}\{\bigcup_{n=1}^{N} A_n\} = \sum_{n=1}^{N} \mathbf{P}\{A_n\} \quad \text{for disjoint } \ A_1, \ldots, A_N \in \mathcal{F},$
5. $\mathbf{P}\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{n\to\infty} \mathbf{P}\{A_n\} \ \text{ for } \ A_1, A_2, \ldots \in \mathcal{F} \ \text{ with } \ A_1 \subseteq A_2 \subseteq \ \ldots \ ,$
6. $\mathbf{P}\{\bigcap_{n=1}^{\infty} A_n\} = \lim_{n\to\infty} \mathbf{P}\{A_n\} \ \text{ for } \ A_1, A_2, \ldots \in \mathcal{F} \ \text{ with } \ A_1 \supseteq A_2 \supseteq \ \ldots \ .$

Proof. 1-4. Simple exercises.

5. Taking $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$ we have $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$ with $B_1, B_2, \ldots \in \mathcal{F}$ disjoint so that, as $N \to \infty$,

$$\mathbf{P}\{\textstyle\bigcup_{n=1}^{\infty} A_n\} = \mathbf{P}\{\bigcup_{n=1}^{\infty} B_n\} = \sum_{n=1}^{\infty} \mathbf{P}\{B_n\} = \mathbf{P}\{A_N\} + \sum_{n=N+1}^{\infty} \mathbf{P}\{B_n\} \to \lim_{N\to\infty} \mathbf{P}\{A_N\}.$$

6. Using 5 we get

$$\mathbf{P}\{\textstyle\bigcap_{n=1}^{\infty} A_n\} = 1 - \mathbf{P}\{\bigcup_{n=1}^{\infty} A_n^c\} = 1 - \lim_{n\to\infty} \mathbf{P}\{A_n^c\} = \lim_{n\to\infty} \mathbf{P}\{A_n\}. \qquad \square$$

Definition 2.6. A random variable X is a function $X : \Omega \to \mathbb{R}$ such that

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B) \in \mathcal{F} \quad \text{for } \ B \in \mathcal{B}.$$

A random variable X is also called a measurable function from Ω to $\mathbb{R}$.

Theorem 2.7. A function $X : \Omega \to \mathbb{R}$ is a random variable if and only if

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}((-\infty, x]) \in \mathcal{F} \quad \text{for } \ x \in \mathbb{R}.$$

Hence the CDF $F_X(x) = \mathbf{P}\{X \leq x\}$ is always well-defined.

Proof. The family of sets $B \in \mathcal{B}$ such that $\{X \in B\} \in \mathcal{F}$ is a σ-field and $\sigma(\{X \leq x\}_{x\in\mathbb{R}}) = \mathcal{B}$. $\square$

Finite algebraic operations with random variables remain random variables. Point wise limits of random variables and the decomposition $g(X)$ of a $\mathcal{B}$-measurable function $g : \mathbb{R} \to \mathbb{R}$ with a random variable X are random variables. In particular the decomposition $g(X)$ of a continuous function $g : \mathbb{R} \to \mathbb{R}$ with a random variable X is a random variable.[2]

Definition 2.8. The σ-field $\sigma(X)$ generated by a random variable X is the sub-σ-field $X^{-1}(\mathcal{B})$ to $\mathcal{F}$.

$\sigma(X)$ is the smallest σ-field of subsets of Ω that X is measurable wrt.

σ-fields have an information interpretation: To know which of the events in $\mathcal{F}$ that happen when the random experiment is carried out means that we know the value of all random variables. To know which of the events in $\sigma(X)$ that happen for a random variable X when the random experiment is carried out means that we know the value of X.

Example 2.2. (Bernoulli random variable) The simplest non-trivial random variable is the indicator $\mathbf{1}_A$ of an $A \in \mathcal{F} \setminus \{\emptyset, \Omega\}$ given by $\mathbf{1}_A(\omega) = 1$ for $\omega \in A$ and $\mathbf{1}_A(\omega) = 0$ for $\omega \in A^c$. Clearly, $\sigma(I_A) = \{\emptyset, A, A^c, \Omega\}$.

2.2 Expectation

The Lebesgue integral of $f : [a, b] \to \mathbb{R}$ is done by dividing the range of f-values in small y-subintervals. The approximate value of the integral is the sum of the length of each x-interval for which $f(x)$ takes values in one of the y-subintervals times the lower endpoint of the corresponding y-subinterval. On paper:

$$\int_{[a,b]} f(x)\, dx \leftarrow \sum_{i=1}^{n} \operatorname{length}(\{x \in [a, b] : f(x) \in [y_{i-1}^n, y_i^n)\})\, y_{i-1}^n$$

where $\min(f) = y_0^n < y_1^n < \ldots < y_n^n = \max(f)$ and $\max_{1\leq i\leq n} y_i^n - y_{i-1}^n \downarrow 0$.

Now f could have an infinite range of values. And we want the approximating sum (unlike approximating Riemann sums) to always converges. Also, we want to be able to integrate over abstract spaces $(\Omega, \mathcal{F})$, not just $(\mathbb{R}, \mathcal{B})$. And we are interested in random variables X and their expectation $\mathbf{E}\{X\}$ rather than math functions f and their integral $\int_\Omega f\, dP$.

[2] Useful exercises.

We first define $\mathbf{E}\{X\}$ for $X \geq 0$ and then $\mathbf{E}\{X\} = \mathbf{E}\{X^+\} - \mathbf{E}\{X^-\}$ in the general case whenever at least one of $\mathbf{E}\{X^+\}$ and $\mathbf{E}\{X^-\}$ are finite. The definition for $X \geq 0$ is

$$\mathbf{E}\{X\} = \lim_{n\to\infty} \sum_{k=0}^{2^n n-1} \tfrac{k}{2^n}\, \mathbf{P}\big\{X \in [\tfrac{k}{2^n}, \tfrac{k+1}{2^n})\big\} = \lim_{n\to\infty} \sum_{k=0}^{2^n n-1} \tfrac{k}{2^n}\, \mathbf{P}\big\{\omega \in \Omega : X(\omega) \in [\tfrac{k}{2^n}, \tfrac{k+1}{2^n})\big\}. \qquad (\star)$$

The limit exists (although possibly is infinite) as the sum increases with n.

Theorem 2.9. 1. $\mathbf{E}\{Y\} \geq \mathbf{E}\{X\}$ if $Y \geq X$ (positivity),

2. $\mathbf{E}\{\alpha X + \beta Y\} = \alpha\, \mathbf{E}\{X\} + \beta\, \mathbf{E}\{Y\}$ (linearity),

3. $\mathbf{E}\{g(X)\} \geq g(\mathbf{E}\{X\})$ for $g : \mathbb{R} \to \mathbb{R}$ convex (Jensen's inequality),

4. if $X_n \geq 0$ and $X_n \uparrow X$ (pointwise) then $\lim_{n\to\infty} \mathbf{E}\{X_n\} = \mathbf{E}\{X\}$ (monotone convergence),

5. $\mathbf{E}\{\liminf_{n\to\infty} X_n\} \leq \liminf_{n\to\infty} \mathbf{E}\{X_n\}$ if $X_n \geq 0$ (Fatou's lemma),

6. if $X_n \to X$ (pointwise) and $\mathbf{E}\{\sup_{n\geq 1} |X_n|\} < \infty$ then $\lim_{n\to\infty} \mathbf{E}\{X_n\}$ $= \mathbf{E}\{X\}$ (dominated convergence).

Proof. 1-2. By inspection of $(\star)$.

3. By convexity $g(x) \geq a + b\,x$ for $x \in \mathbb{R}$ with equality for $x = \mathbf{E}\{X\}$, for some $a, b \in \mathbb{R}$, so

$$\mathbf{E}\{g(X)\} \geq \mathbf{E}\{a + b\,X\} = a + b\,\mathbf{E}\{X\} = g(\mathbf{E}\{X\}). \qquad :)$$

4. For $\mathbf{E}\{X\} = 0$ the claim is trivial.

For $\mathbf{E}\{X\} \in (0, \infty)$ take $\varepsilon \geq 0$ and set $A_n = \{\omega \in \Omega : X - X_n \geq \varepsilon\}$. Then

$$\mathbf{E}\{X\} = \mathbf{E}\{X_n\} + \mathbf{E}\{\mathbf{1}_{A_n}(X - X_n)\} + \mathbf{E}\{\mathbf{1}_{A_n^c}(X - X_n)\} \leq \mathbf{E}\{X_n\} + \mathbf{E}\{\mathbf{1}_{A_n} X\} + \varepsilon,$$

where $\mathbf{Q}\{A_n\} = \mathbf{E}\{\mathbf{1}_{A_n} X\}/\mathbf{E}\{X\} \downarrow 0$ as $n \to \infty$ by 6 in Theorem 2.5 as $A_1 \supseteq A_2 \supseteq \ldots$ with $\cap_{n=1}^{\infty} A_n = \emptyset$. Hence $\liminf_{n\to\infty} \mathbf{E}\{X_n\} \geq \mathbf{E}\{X\} - \varepsilon$ for each $\varepsilon \geq 0$.

For $\mathbf{E}\{X\} = \infty$ take $N > 0$ and chose an $n \in \mathbb{N}$ such that the sum in $(\star)$ is greater than N. Then 5 in Theorem 2.6 shows that $\mathbf{P}\big\{X_m \in [\tfrac{k}{2^n}, \tfrac{k+1}{2^n})\big\} \uparrow \mathbf{P}\big\{X \in [\tfrac{k}{2^n}, \tfrac{k+1}{2^n})\big\}$ as $m \to \infty$ for $k = 0, \ldots, 2^n n^n - 1$ so that also the sum in $(\star)$ with X replaced by X_m is greater than N for m large enough. Hence $\liminf_{m\to\infty} \mathbf{E}\{X_m\} > N$ for each $N > 0$.

5. As $\inf_{k\geq n} X_k(\omega) \uparrow \liminf_{k\to\infty} X_k(\omega)$ as $n \to \infty$ 4 shows that

$$\mathbf{E}\{\liminf_{k\to\infty} X_k\} = \lim_{n\to\infty} \mathbf{E}\{\inf_{k\geq n} X_k\} = \liminf_{n\to\infty} \mathbf{E}\{\inf_{k\geq n} X_k\} \leq \liminf_{n\to\infty} \mathbf{E}\{X_n\}.$$

6. Write $Y = \sup_{n\geq 1} |X_n|$ and note that by 2 and 3

$$\limsup_{n\to\infty} |\mathbf{E}\{X_n\} - \mathbf{E}\{X\}| \le \limsup_{n\to\infty} \mathbf{E}\{|X_n - X|\} = \mathbf{E}\{2Y\} - \liminf_{n\to\infty} \mathbf{E}\{2Y - |X_n - X|\}.$$

Here the right hand side is at most $\mathbf{E}\{2Y\} - \mathbf{E}\{\liminf_{n\to\infty}(2Y - |X_n - X|)\} = 0$ by 5. $\square$[3]

The expectation is also denoted $\int_{\omega\in\Omega} X(\omega)\, dP(\omega)$ indicating that we move over different outcomes ω in Ω, check the value $X(\omega)$ of the random variable, weigh with how likely $d\mathbf{P}(\omega)$ the outcome ω is, and sum up to get the expectation (average value).

Our presentation carries over to any measurable function $f : \Omega \to \mathbb{R}$ from a measurable space $(\Omega, \mathcal{F})$ with a (not necessarily total mass 1) measure $\mathbf{P}$: We build our expectation

$$\mathbf{E}\{X\} = \int_\Omega X\, d\mathbf{P} = \int_{\omega\in\Omega} X(\omega)\, d\mathbf{P}(\omega)$$

in the same way as the math integral $\int_\Omega f\, d\mathbf{P}$ when f and $\mathbf{P}$ do not come from probability.

A random variable is called simple if a linear combination of Bernoulli random variables

$$X(\omega) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(\omega)$$

where $A_1, \ldots, A_n \in \mathcal{F}$ are disjoint and $a_1, \ldots, a_n \in \mathbb{R}$. This readily gives

$$E(X) = \sum_{i=1}^{n} a_i\, \mathbf{P}\{X \in A_i\}.$$

Any $X \ge 0$ can be approximated by simple $X_n \ge 0$ with $X_n(\omega) \uparrow X(\omega)$ as $n \to \infty$: E.g.,

$$X_n(\omega) = \sum_{k=0}^{2^n n-1} \frac{k}{2^n}\, \mathbf{1}_{[\frac{k}{2^n}, \frac{k+1}{2^n})}(X(\omega)).$$

By monotone convergence $\mathbf{E}\{X\} = \lim_{n\to\infty} \mathbf{E}\{X_n\}$ for any simple sequence $X_n \ge 0$ with $X_n \uparrow X \ge 0$. So we can define the expectation for simple random variables and then extend it to positive random variable by approximating from below with positive simple ones.

A random variable X generates a so called Stieltjes probability measure dF_X on $(\mathbb{R}, \mathcal{B})$ by $dF_X(B) = \mathbf{P}\{X \in B\}$ for $B \in \mathcal{B}$. Note that $dF_X((-\infty, x]) = F_X(x)$.

Theorem 2.10. For a random variable X and a measurable $g : \mathbb{R} \to \mathbb{R}$ we have $\mathbf{E}\{g(X)\} = \int_{-\infty}^{\infty} g(x)\, dF_X(x)$ whenever the expectation is well-defined.

Note that Theorem 1.8 is for ALL random variables – continuous, discrete and others.

Proof.[4] As $\mathbf{E}\{\mathbf{1}_{[-N,N]}(X) g(X)\} \to \mathbf{E}\{g(X)\}$ as $N \to \infty$ by 6 in Theorem 2.9 we need only do the proof for $X \in [-N, N]$. Taking $s_i^n, t_i^n \in [\frac{i-1}{2^n}, \frac{i}{2^n}]$ such that $g(s_i^n) = \min_{x\in[\frac{i-1}{2^n}, \frac{i}{2^n}]} g(x)$

[3] The proof illustrates that 4-6 are more less the same and are close to the third axiom in Defintion 2.4.

and $g(t_i^n) = \max_{x\in[\frac{i-1}{2^n},\frac{i}{2^n}]} g(x)$ for $i \in \{-N2^n+1,\ldots,N2^n\}$ and $n \in \mathbb{N}$ we have

$$\mathbf{E}\{g(X)\} \begin{cases} \leq \sum_{i=-N2^n+1}^{N2^n} g(t_i^n)\,\mathbf{P}\big\{X\in(\frac{i-1}{2^n},\frac{i}{2^n}]\big\} = \sum_{i=-N2^n+1}^{N2^n} g(t_i^n)\,\big(F_X(\frac{i}{2^n})-F_X(\frac{i-1}{2^n})\big) \\ \geq \sum_{i=-N2^n+1}^{N2^n} g(s_i^n)\,\mathbf{P}\big\{X\in(\frac{i-1}{2^n},\frac{i}{2^n}]\big\} = \sum_{i=-N2^n+1}^{N2^n} g(s_i^n)\,\big(F_X(\frac{i}{2^n})-F_X(\frac{i-1}{2^n})\big) \end{cases}.$$

Here the upper and lower bounds differ by at most $\sup_{i\in\{-N2^n+1,\ldots,N2^n\}} \sup_{x,y\in[\frac{i-1}{2^n},\frac{i}{2^n}]} |g(y) - g(x)| \to 0$ as $n \to \infty$ by uniform continuity so both sums converge to $\mathbf{E}\{g(X)\} = \int_{-\infty}^{\infty} g\,dF_X$.□

2.3 Conditional Expectation

Definition 2.11. Let $\mathbf{P}$ and $\mathbf{Q}$ be probability measures on a measurable space $(\Omega,\mathcal{F})$. We say that $\underline{\mathbf{Q} \text{ is absolutely continuous with respect to } \mathbf{P}}$ ($\mathbf{Q} << \mathbf{P}$) if $\mathbf{P}\{A\} = 0 \Rightarrow \mathbf{Q}\{A\} = 0$ for $A \in \mathcal{F}$.

Theorem 2.12. (Radon-Nikodym[5]) If $\mathbf{Q} << \mathbf{P}$ then there exists a unique (except for the values on an event of probability zero) random variable $\Lambda \geq 0$ with $\mathbf{E}\{\Lambda\} = 1$ such that

$$\mathbf{Q}\{A\} = \mathbf{E_P}\{\mathbf{1}_A\Lambda\} = \int_A \Lambda\,d\mathbf{P} \quad \text{for } A \in \mathcal{F}.$$

Theorem 2.13. Let X be a random variable on a probability space $(\Omega,\mathcal{F},\mathbf{P})$ with $\mathbf{E}(|X|) < \infty$. For a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$ there exists a unique $\mathcal{G}$-measurable random variable Y with $\mathbf{E}\{|Y|\} < \infty$ such that

$$\mathbf{E}\{\mathbf{1}_A Y\} = \mathbf{E}\{\mathbf{1}_A X\} \quad \text{for } A \in \mathcal{G}.$$

Proof. As $X = X^+ - X^-$ it is enough to prove the theorem for non-negative X and set $\mathbf{E}\{X|\mathcal{G}\} = \mathbf{E}\{X^+|\mathcal{G}\} - \mathbf{E}\{X^-|\mathcal{G}\}$ afterwards. Now, for $X \geq 0$ with $\mathbf{E}\{X\} = 0$ we see that $Y = 0$ works. So assume $\mathbf{E}\{X\} > 0$ and define a new probability measure

$$\mathbf{Q}\{A\} = \mathbf{E}\{\mathbf{1}_A X\}/\mathbf{E}\{X\} \quad \text{for } A \in \mathcal{G}.$$

Then $\mathbf{P}\{A\} = 0 \Rightarrow \mathbf{Q}\{A\} = 0$ for $A \in \mathcal{G}$ so that Radon-Nikodym gives

$$\mathbf{Q}\{A\} = \mathbf{E_P}\{\mathbf{1}_A\Lambda\} \quad \text{for } A \in \mathcal{G},$$

for some $\mathcal{G}$-measurable Λ. And so we may take $Y = \Lambda\,\mathbf{E}\{X\}$ as

$$\mathbf{E}\{\mathbf{1}_A X\} = \mathbf{Q}\{A\}\,\mathbf{E}\{X\} = \mathbf{E_P}\{\mathbf{1}_A\Lambda\}\,\mathbf{E}\{X\} \quad \text{for } A \in \mathcal{G}. \qquad \square$$

[4]We do the proof for g continuous only to avoid unwanted measure theoretical complications.

[5]A theorem where functional analysis and integration meet. Cannot be proved at home from scratch.

Definition 2.14. The random variable Y in Theorem 2.12 is denoted $\mathbf{E}\{X|\mathcal{G}\}$ and called the conditional expectation of X with respect to $\mathcal{G}$.

Conditional probability is defined by $\mathbf{P}\{A|\mathcal{G}\} = \mathbf{E}\{I_A|\mathcal{G}\}$ for $A \in \mathcal{F}$.

Unlike naive conditional expectation, the abstract $\mathbf{E}\{X|\mathcal{G}\}$ is a random variable.

Definition 2.15. Two σ-fields $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$ are independent if

$$\mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\}\,\mathbf{P}\{Y \in B\} \quad \text{for } A \in \mathcal{G}_1 \text{ and } B \in \mathcal{G}_2.$$

X is independent of a σ-field $\mathcal{G} \subseteq \mathcal{F}$ if $\sigma(X)$ is independent of $\mathcal{G}$.

Theorem 2.16.

1. $\mathbf{E}\{Y|\mathcal{G}\} \geq \mathbf{E}\{X|\mathcal{G}\}$ when $Y \geq X$ (positivity),
2. $\mathbf{E}\{\alpha X + \beta Y|\mathcal{G}\} = \alpha\,\mathbf{E}\{X|\mathcal{G}\} + \beta\,\mathbf{E}\{Y|\mathcal{G}\}$ (linearity),
3. $g(\mathbf{E}\{X|\mathcal{G}\}) \leq \mathbf{E}\{g(X)|\mathcal{G}\}$ when g is convex (Jensen's inequality),
4. $\lim_{n\to\infty} \mathbf{E}\{X_n|\mathcal{G}\} = \mathbf{E}\{X|\mathcal{G}\}$ when $0 \leq X_n \uparrow X$ (monotone convergence),
5. $\mathbf{E}\{\liminf_{n\to\infty} X_n|\mathcal{G}\} \leq \liminf_{n\to\infty} \mathbf{E}\{X_n|\mathcal{G}\}$ when $X_n \geq 0$ (Fatou's lemma),
6. $\lim_{n\to\infty} \mathbf{E}\{X_n|\mathcal{G}\} = \mathbf{E}\{X|\mathcal{G}\}$ when $X_n \to X$ and $\mathbf{E}\{\sup_{n\geq 1}|X_n|\} < \infty$ (dominated convergence),
7. $\mathbf{E}\{X|\{\emptyset, \Omega\}\} = \mathbf{E}\{X\}$,
8. $\mathbf{E}\{\mathbf{E}\{X|\mathcal{G}_2\}|\mathcal{G}_1\} = \mathbf{E}\{X|\mathcal{G}_1\}$ when $\mathcal{G}_1 \subseteq \mathcal{G}_2$ (towering),
9. $\mathbf{E}\{X|\mathcal{G}\} = \mathbf{E}\{X\}$ when $\mathcal{G}$ and $\sigma(X)$ are independent,
10. $\mathbf{E}\{\mathbf{E}\{X|\mathcal{G}\}\} = \mathbf{E}\{X\}$,
11. $\mathbf{E}\{X|\mathcal{F}\} = X$ when X is $\mathcal{F}$-measurable,
12. $\mathbf{E}\{XY|\mathcal{G}\} = X\,\mathbf{E}\{Y|\mathcal{G}\}$ when X is $\mathcal{G}$-measurable,
13. $\mathbf{E}\{X|\mathcal{G}_1 \vee \mathcal{G}_2\} = \mathbf{E}\{X|\mathcal{G}_1\}$ when $\mathcal{G}_2$ is independent of $\sigma(X)$ and $\mathcal{G}_1$,
14. $\mathbf{E}\{g(X,Y)|\mathcal{G}_2\} = \mathbf{E}\{g(X,Y)|\mathcal{G}_1\}$ when $\mathcal{G}_1 \subseteq \mathcal{G}_2$, X is $\mathcal{G}_1$-measurable and Y is independent of $\mathcal{G}_2$,
15. $\mathbf{E}\{g(X,Y)|\mathcal{G}\} = \mathbf{E}\{g(x,Y)\}\,\big|_{x=X}$ when X is $\mathcal{G}$-measurable and Y is independent of $\mathcal{G}$.

Proof. 1-6. Proved in the same way as the corresponding properties for usual expectations.

7-9. Proved in the exercises.

10. Take $\mathcal{G}_1 = \{\emptyset, \Omega\}$ in 8.

11. By inspection of the definition.

12. Using $X = X^+ - X^-$ and $Y = Y^+ - Y^-$ as in the proof of Theorem 2.12 it is enough to do the proof for $X, Y \geq 0$. Now, for $X, Y \geq 0$ we may approximate X with $0 \leq X_n \uparrow X$ as in the definition of the integral with $X_n(\omega) = \sum_{i=1}^n a_i \mathbf{1}_{A_i}(\omega)$ and $A_i \in \mathcal{G}$. For a $B \in \mathcal{G}$ we may now use monotone convergence for expectations and conditional expectations together with 2, 10 and the definition of conditional expectation for Y to obtain

$$\mathbf{E}\{\mathbf{1}_B X\, \mathbf{E}\{Y|\mathcal{G}\}\} = \lim_{n\to\infty} \mathbf{E}\Big\{\mathbf{1}_B \sum_{i=1}^n a_i \mathbf{1}_{A_i} \mathbf{E}\{Y|\mathcal{G}\}\Big\} = \lim_{n\to\infty} \sum_{i=1}^n a_i \mathbf{E}\{\mathbf{1}_B \mathbf{1}_{A_i} Y\} = \mathbf{E}\{\mathbf{1}_B XY\}.$$

13. Since the family of events $C \in \mathcal{G}_1 \vee \mathcal{G}_2$ such that $\mathbf{E}\{\mathbf{1}_C \mathbf{E}\{X|\mathcal{G}\}\} = \mathbf{E}\{\mathbf{1}_C X\}$ is a σ-field and $\sigma\{A \cap B : A \in \mathcal{G}_1,\, B \in \mathcal{G}_2\} = \mathcal{G}_1 \vee \mathcal{G}_2$, it is enough to check that $\mathbf{E}\{\mathbf{1}_A \mathbf{1}_B \mathbf{E}\{X|\mathcal{G}\}\} = \mathbf{E}\{\mathbf{1}_A \mathbf{1}_B X\}$ for $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$. But by 10 and 12 together with the assumed independence

$$\begin{aligned}&\mathbf{E}\{\mathbf{1}_A \mathbf{1}_B \mathbf{E}\{X|\mathcal{G}_1\}\}\\ &\quad= \mathbf{E}\{\mathbf{1}_B \mathbf{E}\{\mathbf{1}_A X|\mathcal{G}_1\}\} = \mathbf{E}\{\mathbf{1}_B\}\, \mathbf{E}\{\mathbf{E}\{\mathbf{1}_A X|\mathcal{G}_1\}\} = \mathbf{E}\{\mathbf{1}_B\}\, \mathbf{E}\{\mathbf{1}_A X\} = \mathbf{E}\{\mathbf{1}_B \mathbf{1}_A X\}.\end{aligned}$$

14. The proof requires a bit more technical details about measures than we want to go into.

15. We get $\mathbf{E}\{g(X,Y)|\mathcal{G}\} = \mathbf{E}\{g(X,Y)|\sigma(X)\}$ by 14. Now use that an $A \in \sigma(X)$ satisfies $\mathbf{1}_A(\omega) = \mathbf{1}_B(X(\omega))$ for some $B \in \mathcal{B}$ to obtain

$$\begin{aligned}\mathbf{E}\big\{\mathbf{1}_A \mathbf{E}\{g(x,Y)\}\big|_{x=X}\big\} &= \textstyle\int_{-\infty}^{\infty} \mathbf{1}_B(x) \mathbf{E}\{g(x,Y)\}\, dF_X(x)\\ &= \textstyle\int_{-\infty}^{\infty} \mathbf{1}_B(x) \Big(\int_{-\infty}^{\infty} g(x,y)\, dF_Y(y)\Big) dF_X(x) = \mathbf{E}\{\mathbf{1}_B(X) g(X,Y)\}. \quad \square\end{aligned}$$

Theorem 2.17. For X and Y continuous random variables with $\mathbf{E}\{|X|\} < \infty$ and a well-defined conditional density function $f_{X|Y}(x|y)$, the relation between the naive conditional expectation $\mathbf{E}\{X|Y = y\}$ and the abstract $\mathbf{E}\{X|\sigma(Y)\}$ is that $\mathbf{E}\{X|\sigma(Y)\} = g(Y)$ where $g(y) = \mathbf{E}\{X|Y = y\}$.

Proof. We check that $g(Y)$ works as $\mathbf{E}\{X|\sigma(Y)\}$ in Definition 2.13: As an $A \in \sigma(Y)$ satisfies $A = Y^{-1}(B)$, i.e., $\mathbf{1}_A(\omega) = \mathbf{1}_B(Y(\omega))$ for $\omega \in \Omega$, for an $B \in \mathcal{B}$, we have

$$\begin{aligned}\mathbf{E}\{\mathbf{1}_A\, g(Y)\} &= \mathbf{E}\Big\{\mathbf{1}_B(Y) \textstyle\int_{-\infty}^{\infty} x\, \frac{f_{X,Y}(x,Y)}{f_Y(Y)}\, dx\Big\}\\ &= \textstyle\int_{-\infty}^{\infty} \mathbf{1}_B(y) \big(\int_{-\infty}^{\infty} x\, \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dx\big)\, f_Y(y)\, dy = \mathbf{E}\{\mathbf{1}_B(Y) X\} = \mathbf{E}\{\mathbf{1}_A X\}.\end{aligned}$$

Here $g(Y) = g \circ Y$ is $\sigma(Y)$-measurable as g (being an integral) is Borel-measurable so that $\{g(Y) \in B\} = \{\omega \in \Omega : g(Y(\omega)) \in B\} = Y^{-1}(g^{-1}(B)) \in \sigma(Y)$ for $B \subseteq \mathbb{R}$ Borel-measurable. $\square$