Markov Decision Processes **January 16, 2024**

Module 1.A: Foundations of Probability

Lecturer: Avik Kar

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

1 Preliminaries

Notation: We denote the set of first N positive integers by $[N] \triangleq \{1,2,\ldots,N\}$, the set of first N positive integers and 0 by $[N]_0 \triangleq \{0,1,2,\ldots,N\}$, the set of positive integers (natural numbers) by $\mathbb{N} \triangleq \{1,2,\ldots\}$, the set of integers by $\mathbb{Z} \triangleq \{\ldots,-1,0,1,\ldots\}$, the set of non-negative integers by $\mathbb{Z}_+ \triangleq \{0,1,\ldots\}$, the set of rational numbers by $\mathbb{Q}$, the set of reals by $\mathbb{R}$, and the set of non-negative reals by $\mathbb{R}_+$. Power set of a set A is the collection of all subsets of A and is denoted by $2^A \triangleq \{B : B \subseteq A\}$.

Definition 1.1 (Function)**.** For sets A, B, we denote a function from set B to set A by $f : B \to A$, where $(b, f(b)) \in B \times A$ and for each $b \in B$ there is only one value $f(b) \in A$. That is,

$$\{(b, f(b)) : b \in B\} \subseteq B \times A.$$

The set B and A are called the domain and co-domain of function f, and the set $f(B) = \{f(b) \in A : b \in B\}$ is called the range of function f.

The collection of all A-valued functions with domain B is denoted by A^B.

Definition 1.2 (Inverse map)**.** Let $f \in A^B$. The inverse map of f, denoted by $f^{-1} : 2^A \to 2^B$, is defined as

$$f^{-1}(A) = \{x \in B : f(x) \in A\},\ A \subseteq A.$$

Definition 1.3 (Sequence)**.** A B-valued sequence is an element of $B^{\mathbb{N}}$.

We will denote the cardinality of set A by $|A|$. Two sets, A and B are called equi-cardinal if there exists a bijection between them.

Definition 1.4 (Countable and finite sets)**.** A is called a countable set if A is equi-cardinal with $\mathbb{N}$. If $|A| < \infty$ then A is said to be a finite set.

Exercise 1.5. Show that $\mathbb{Z}$ is a countable set.

2 Probability space

Two terms are not formally defined in probability theory; these are 'random experiment' and 'outcome.'

Definition 2.1 (Sample space and sample point)**.** 'Sample space', denoted by Ω, is the collection of all possible 'outcomes' of a 'random experiment.' Elements of sample space, $\omega \in \Omega$ are called sample points.

Definition 2.2 (Event (Informal))**.** 'Interesting' subsets of sample space on which we assign probabilities are called events. Collection of all events corresponding to a 'random experiment' is called the 'event space.'

Example 2.3. Consider the random experiment of tossing a coin once. Head and Tail are the only possible outcomes. So, we write the abstract set, sample space, as $\Omega = \{H, T\}$; H, T are the sample points; $\{H\}$ is an event.

Example 2.4. Consider the random experiment of tossing a coin thrice. define the sample space as

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Let us denote $\mathbb{P}(A)$ as the probability of event A informally for the following discussion. Consider a sample space, Ω; let $A, B \subset \Omega$ are events such that $A \cap B = \phi$. Then, from our intuitive understanding of probability, we expect

1. $\mathbb{P}(\phi) = 0$.
2. $\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A)$.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Underlying this, we are also demanding that

1. ϕ is an event.
2. If A is event, then so is $\Omega \setminus A$, i.e., event space is closed under complements.
3. If A, B are events then $A \cup B$ is also an event, i.e., event space is closed under finite union.

It is seen that for large enough sample spaces, countable additivity of probability and closure under countable union for event space is required. The finite counterpart follows from there. This reasons next two definitions.

Definition 2.5 (σ-algebra)**.** A collection of subsets of Ω, $\mathcal{F}$ is called a σ-algebra if the following properties are satisfied.

- **(s1)** $\phi \in \mathcal{F}$.
- **(s2)** If $S \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$.
- **(s3)** If $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{F}$, then $\cup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

A σ-algebra $\mathcal{F}$ on Ω serves as the event space. We call $(\Omega, \mathcal{F})$ a measurable space.

Exercise 2.6. If $\mathcal{F}$ is a σ-algebra on Ω, then show the followings:

1. $\Omega \in \mathcal{F}$.
2. (Closure under finite union). If $\{A_n\}_{n \in [N]} \subset \mathcal{F}$, then $\cup_{n \in [N]} A_n \in \mathcal{F}$.
3. (Closure under finite and countable intersection). Let $\mathcal{I}$ be an at most countable index set. If $\{A_n\}_{n \in \mathcal{I}} \subset \mathcal{F}$, then $\cap_{n \in \mathcal{I}} A_n \in \mathcal{F}$.

Exercise 2.7. Show that 2^{Ω} is a σ-algebra on Ω.

Definition 2.8 (Probability measure)**.** Let $(\Omega, \mathcal{F})$ be a measurable space, then $\mathbb{P} : \mathcal{F} \to [0,1]$ is called a probability measure on $(\Omega, \mathcal{F})$ if the following properties are satisfied.

- **(p1)** (Certainty) $\mathbb{P}(\Omega) = 1$.
- **(p2)** (Non-negativity) $\mathbb{P}(A) \geq 0 \; \forall A \in \mathcal{F}$.

(p3) (Countable additivity). Let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ be a pairwise disjoint collection of events. Then

$$\mathbb{P}\left(\cup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

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

Exercise 2.9. If $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Show the followings:

1. For any $A \in \mathcal{F}$, $\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A)$.
2. (Monotonicity). If $A, B \in \mathcal{F}$ such that $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
3. (Finite additivity). If $\{A_n\}_{n\in[N]} \subset \mathcal{F}$ is a collection of pairwise disjoint sets, then $\mathbb{P}(\cup_{n\in[N]} A_n) = \sum_{n\in[N]} \mathbb{P}(A_n)$.
4. (Countable sub-additivity). If $\{A_n\}_{n\in[N]} \subset \mathcal{F}$, then $\mathbb{P}(\cup_{n\in[N]} A_n) \leq \sum_{n\in[N]} \mathbb{P}(A_n)$.
5. (Continuity). If $\{A_n\}_{n\in[N]} \subset \mathcal{F}$ and $\lim_{n\to\infty} A_n = A$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Definition 2.10 (σ-algebra generated by a collection)**.** let $\mathcal{C}$ be a collection of subsets of Ω. The σ-algebra generated by $\mathcal{C}$, $\sigma(\mathcal{C})$ is defined as

$$\sigma(\mathcal{C}) \triangleq \cap \{\mathcal{H} : \mathcal{C} \subseteq \mathcal{H} \textit{ and } \mathcal{H} \textit{ is a } \sigma\textit{-algebra}\}.$$

Exercise 2.11. Show that $\sigma(\mathcal{C})$ is a σ-algebra where $\mathcal{C}$ is a non-empty collection of subsets of Ω.

Now we will introduce a popular non-trivial σ-algebra on $\mathbb{R}$.

Definition 2.12 (Borel σ-algebra on $\mathbb{R}$)**.** The Borel σ-algebra on $\mathbb{R}$ is the σ-algebra generated by sets of the form $(-\infty, x], x \in \mathbb{R}$, i.e.,

$$\mathcal{B}(\mathbb{R}) \triangleq \sigma\left(\{(-\infty, x] : x \in \mathbb{R}\}\right).$$

Exercise 2.13. Show that $\{x\}, (x, y), [x, y] \in \mathcal{B}(\mathbb{R})\ \forall x, y \in \mathbb{R}$.

3 Independence of events and conditional probability

Theorem 3.1 (Law of total probability)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Consider a sequence of events $\{B_n\}_{n\in\mathbb{N}}$ that partitions the sample space Ω, i.e., $B_m \cap B_n = \phi\ \forall m, n \in \mathbb{N}$ with $m \neq n$ and $\cup_{n\in\mathbb{N}} B_n = \Omega$. Then, for any $A \in \mathcal{F}$, we have*

$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}} \mathbb{P}(A \cap B_n).$$

Exercise 3.2. Prove the law of total probability.

Definition 3.3 (Independence of events)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A family of events $\{A_n\}_{n\in I}$ is said to be independent if for any finite set $F \subset I$, we have

$$\mathbb{P}(\cup_{n\in F} A_n) = \prod_{n\in F} \mathbb{P}(A_n).$$

Definition 3.4 (Conditional probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For any event $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$, the conditional probability, $\mathbb{P}(\cdot \mid B) : \mathcal{F} \to [0,1]$ of any event $A \in \mathcal{F}$ conditioned on the event B is defined as

$$\mathbb{P}(A \mid B) \triangleq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Exercise 3.5. For any event $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$, show that $\mathbb{P}(\cdot \mid B)$ is a probability measure on $(\Omega, \mathcal{F})$.

Lemma 3.6 (Bayes rule)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$,*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{1}$$

4 Random variables

Definition 4.1 (Measurable function)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $X : \Omega \to \mathbb{R}$ is called $\mathcal{F}$-measurable if for every $B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) \in \mathcal{F}$.

Now we are in a good position to introduce random variables. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition 4.2 (Random variable)**.** An $\mathbb{R}$-valued function $X : \Omega \to \mathbb{R}$ is called a random variable if it is $\mathcal{F}$-measurable.

So, each event in the probability space is mapped to a set of real numbers via the random variable. Correspondingly the probability measure of events induces to a measure on $\mathbb{R}$.

Definition 4.3 (Probability law)**.** The probability law of random variable X, denoted by $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \to [0,1]$, is defined as

$$\mathbb{P}_X(A) \triangleq \mathbb{P}(X^{-1}(A)), \; A \in \mathcal{B}(\mathbb{R}).$$

Definition 4.4 (Distribution function)**.** The distribution function of random variable X, denoted by $F_X : \mathbb{R} \to [0,1]$, is defined as

$$F_X(x) \triangleq \mathbb{P}(X^{-1}((-\infty, x]), \; x \in \mathbb{R}.$$

Exercise 4.5. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable $X : \Omega \to \mathbb{R}$, show that $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ is a probability space.

Example 4.6 (Indicator random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space and $A \in \mathcal{F}$. The indicator random variable of event A, $\mathbb{1}_A$ is defined as

$$\mathbb{1}_A = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise.} \end{cases}$$

The probability law, $\mathbb{P}_{\mathbb{1}_A}$ is given by

$$\mathbb{P}_{\mathbb{1}_A}(B) = \mathbb{1}_{\{0 \in B\}}(1 - \mathbb{P}(A)) + \mathbb{1}_{\{1 \in B\}}\mathbb{P}(A),$$

and the distribution function, $F_{\mathbb{1}_A}$ is given by

$$F_{\mathbb{1}_A}(x) = \mathbb{1}_{\{x \geq 0\}}(1 - \mathbb{P}(A)) + \mathbb{1}_{\{x \geq 1\}}\mathbb{P}(A).$$

Exercise 4.7. Let $f:\mathbb{R}\to\mathbb{R}$ be a $\mathcal{B}(\mathbb{R})$-measurable function, i.e., $f^{-1}(B)\in\mathcal{B}(\mathbb{R})\ \forall B\in\mathcal{B}(\mathbb{R})$. If X is a random variable on probability space $(\Omega,\mathcal{F},\mathbb{P})$, then show that $f(X)$ is a random variable on $(\Omega,\mathcal{F},\mathbb{P})$.

Definition 4.8. Let $X;\Omega\to\mathbb{R}$ be a random variable defined on probability space $(\Omega,\mathcal{F},\mathbb{P})$. The σ-algebra generated by X, denoted by $\sigma(X)$, is defined as

$$\sigma(X)\triangleq\sigma\left(\left\{X^{-1}((-\infty,x];x\in\mathbb{R}\right\}\right).$$

Exercise 4.9. Let $(\Omega,\mathcal{F},\mathbb{P})$ be the probability space and $A\in\mathcal{F}$. Show that $\sigma(\mathbb{1}_A)=\{\phi,A,\Omega\setminus A,\Omega\}$.

Definition 4.10 (Independence of σ-algebras and random variables)**.** Let $(\Omega,\mathcal{F},\mathbb{P})$ be the probability space.

1. Let $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras of $\mathcal{F}$. Then $\mathcal{H}$ and $\mathcal{G}$ are independent if for any two events $H\in\mathcal{H}$, $G\in\mathcal{G}$, we have
$$\mathbb{P}(H\cap G)=\mathbb{P}(H)\mathbb{P}(G).$$
2. Let X,Y are two random variables defined on $(\Omega,\mathcal{F},\mathbb{P})$. X,Y are said to be independent if $\sigma(X)$ and $\sigma(Y)$ are independent.

Proposition 4.11. *Let X,Y be two random variables on the probability space $(\Omega,\mathcal{F},\mathbb{P})$. Show that the followings are equivalent.*

1. *X,Y are independent.*
2. *Their joint distribution function factorizes, i.e.,*

$$\begin{aligned}F_{XY}(x,y)&\triangleq\mathbb{P}(\{\omega\in\Omega:X(\omega)\le x,Y(\omega)\le y\})\\&=\mathbb{P}(\{\omega\in\Omega:X(\omega)\le x\})\mathbb{P}(\{\omega\in\Omega:Y(\omega)\le y\})\\&=F_X(x)F_Y(y).\end{aligned}$$

5 Expectation

Let $(\Omega,\mathcal{F},\mathbb{P})$ be the probability space.

Definition 5.1. $X:\Omega\to\mathbb{R}$ is called a simple random variable if

1. X is $\mathcal{F}$-measurable,
2. $|X(\Omega)|<\infty$.

Definition 5.2 (Expectation of simple non-negative random variable)**.** Let $X:\Omega\to\mathcal{X}\subset\mathbb{R}_+$ be a simple non-negative random variable, and let $\mathcal{X}$ be finite. Then, expectation of X, $\mathbb{E}[X]$ is defined as

$$\begin{aligned}\mathbb{E}[X]&\triangleq\sum_{x\in\mathcal{X}}x\mathbb{P}\left(X^{-1}(\{x\})\right)\\&=\sum_{x\in\mathcal{X}}x\mathbb{P}_X(x).\end{aligned}$$

We need the following theorem to extend the definition of $\mathbb{E}[X]$ to all non-negative random variables.

Theorem 5.3 (Monotone approximation)**.** *For every non-negative random variable, there exists an increasing sequence of simple non-negative random variables, $\{X_n\}_{n\in\mathbb{N}}$, which converges to X pointwise, i.e.,*

$$\lim_{n\to\infty}X_n(\omega)=X(\omega)\ \textit{for every}\ \omega\in\Omega.$$

Definition 5.4 (Expectation of non-negative random variable)**.** Let $X : \Omega \to \mathbb{R}_+$ be a non-negative random variable. Then, expectation of X, $\mathbb{E}[X]$ is defined as

$$\mathbb{E}[X] \triangleq \lim_{n\to\infty} \mathbb{E}[X_n],$$

where $\{X_n\}_{n\in\mathbb{N}}$ is an increasing sequence of simple non-negative random variables that converges to X pointwise.

Now we will extend the definition of $\mathbb{E}[X]$ to all random variables.

Exercise 5.5. Let $X : \Omega \to \mathbb{R}$ be a random variable. Define

$$X_+ \triangleq \max\{X, 0\} \quad \text{and} \quad X_- \triangleq -\min\{X, 0\}.$$

Show that X_+, X_- are random variables.

Definition 5.6. Let $X : \Omega \to \mathbb{R}_+$ be a random variable. Then, expectation of X, $\mathbb{E}[X]$ is defined as

$$\mathbb{E}[X] \triangleq \mathbb{E}[X_+] + \mathbb{E}[X_-]$$

if at least one of $\mathbb{E}[X_+]$ and $\mathbb{E}[X_-]$ is finite.

5.1 Properties of expectation

1. **Linearity:** Let $a, b \in \mathbb{R}$, and X, Y are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathbb{E}[X], \mathbb{E}[Y]$ and $\mathbb{E}[aX + bY]$ are well-defined, then
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$
2. **Monotonicity:** If $\mathbb{P}\{X \geq Y\} = 1$ and both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are well-defined with $\mathbb{E}[Y] > -\infty$, then $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

Exercise 5.7. Prove linearity and monotonicity of expectation.

6 Conditional expectation

The 'conditional expectation' allows us to deal with the expectation of a random variable given the value of another random variable, or more generally, given some σ-algebra.

Example 6.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space that models the random experiment of throwing an unbiased dice, i.e., $\Omega = [6], \mathcal{F} = 2^{\Omega}$; for any $A \in \mathcal{F}$, $\mathbb{P}(A) = \frac{1}{6}$. Let X and Y, two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, are defined as, $X(\omega) = \omega$ and $Y = \mathbb{1}_{\{\omega \geq 3\}}$.

Let us compute $\mathbb{E}[X|Y = 1]$ and $\mathbb{E}[X|Y = 0]$.

$$\mathbb{E}[X \mid Y = 1] = \sum_{x\in[6]} x \frac{\mathbb{P}(\{X = x\} \cap \{X > 3\})}{\mathbb{P}(\{X > 3\})} = \frac{1}{3}(4 + 5 + 6) = 5$$

and

$$\mathbb{E}[X \mid Y = 0] = \sum_{x\in[6]} x \frac{\mathbb{P}(\{X = x\} \cap \{X \leq 3\})}{\mathbb{P}(\{X \leq 3\})} = \frac{1}{3}(1 + 2 + 3) = 2.$$

Precisely, $\mathbb{E}[X \mid Y] = 5Y + 2(1 - Y) = 2 + 3Y$.

Note that conditional expectation is a random variable.

Definition 6.2 (Conditional expectation (Informal))**.** If $X:\Omega\to\mathcal{X}$ and $Y:\Omega\to\mathcal{Y}$ with $\mathcal{X},\mathcal{Y}\subset\mathbb{R}$ and $|\mathcal{X}|,|\mathcal{Y}|<\infty$, then $\mathbb{E}[X\mid Y]:\Omega\to\mathbb{R}$ is the random variable given by $\mathbb{E}[X\mid Y](\omega)\triangleq\mathbb{E}[X\mid Y=Y(\omega)]$ where

$$\begin{aligned}\mathbb{E}[X\mid Y=y] &= \sum_{x\in\mathcal{X}} x\mathbb{P}(\{X=x\}\mid\{Y=y\})\\ &= \sum_{x\in\mathcal{X}} x\frac{\mathbb{P}(\{X=x\}\cap\{Y=y\})}{\mathbb{P}(\{Y=y\})}.\end{aligned}$$

If we go by this definition, $\mathbb{E}[X\mid Y]$ is undefined on $\{\omega\in\Omega:\mathbb{P}(\{Y=Y(\omega)\})=0\}$ and if Y is a continuous random variable, $\{\omega\in\Omega:\mathbb{P}(\{Y=Y(\omega)\})=0\}=\Omega$. So, we need a generalized definition of conditional expectation, which would be in coherence with the above definition. To do that, let us look into some properties of $\mathbb{E}[X\mid Y]$ from the above informal definition.

1. **Measurability:** $\mathbb{E}[X\mid Y](\omega)$ should only depend on $Y(\omega)$ and so it should be $\sigma(Y)$-measurable.
2. **Orthogonality:** For a measurable $A\subset\mathcal{Y}$,

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{Y^{-1}(A)}\mathbb{E}[X\mid Y]] &= \sum_{y\in A}\mathbb{P}(\{Y=y\})\mathbb{E}[X\mid Y=y]\\ &= \sum_{y\in A}\sum_{x\in\mathcal{X}} x\mathbb{P}(\{X=x\}\cap\{Y=y\})\\ &= \sum_{x\in\mathcal{X}} x\mathbb{P}(\{X=x\}\cap\{Y\in A\})\\ &= \mathbb{E}[\mathbb{1}_{Y^{-1}(A)}X].\end{aligned}$$

We are ready to define 'conditional expectation' using the above properties.

Definition 6.3 (Conditional expectation)**.** The conditional expectation of random variable X given σ-algebra $\mathcal{G}$ is a random variable $\mathbb{E}[X\mid\mathcal{G}];\Omega\to\mathbb{R}$ defined on the same probability space such that

1. $Z\triangleq\mathbb{E}[X\mid\mathcal{G}]$ is $\mathcal{G}$-measurable.
2. We have $\mathbb{E}[X\mathbb{1}_G]=\mathbb{E}[Z\mathbb{1}_G]$ for all $G\in\mathcal{G}$.

We define the conditional expectation of random variable X given another random variable Y as $\mathbb{E}[X\mid Y]\triangleq\mathbb{E}[X\mid\sigma(Y)]$.

Lemma 6.4. *The conditional expectation of random variable X given $\mathcal{G}$ is a unique random variable with probability 1, i.e., if $Z_1=\mathbb{E}[X\mid\mathcal{G}]$ and $Z_2=\mathbb{E}[X\mid\mathcal{G}]$, then $\mathbb{P}(Z_1=Z_2)=1$.*

Exercise 6.5. Let X be a random variable defined on the probability space $(\Omega,\mathcal{F},\mathbb{P})$. Show that

1. $\mathbb{E}[X\mid\mathcal{F}]=X$ with probability 1.
2. $\mathbb{E}[X\mid\{\phi,\Omega\}]=\mathbb{E}[X]$ with probability 1.
3. If X is independent of $\mathcal{G}$, i.e., $\sigma(X)$ and $\mathcal{G}$ are two independent σ-algebras, then $\mathbb{E}[X\mid\mathcal{G}]=\mathbb{E}[X]$ with probability 1.

6.1 Properties of conditional expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space. $a, b \in \mathbb{R}$, X, Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G}, \mathcal{H}$ be sub σ-algebras of $\mathcal{F}$.

1. **Linearity:** $\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$.
2. **Monotonicity:** If $X \leq Y$ with probability 1, then $\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]$ with probability 1.
3. **Pulling out what's known:** If Y is $\mathcal{G}$-measurable, then with probability 1 $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$.
4. **Tower property:** If $\mathcal{H} \subseteq \mathcal{G}$, then with probability 1 $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$.