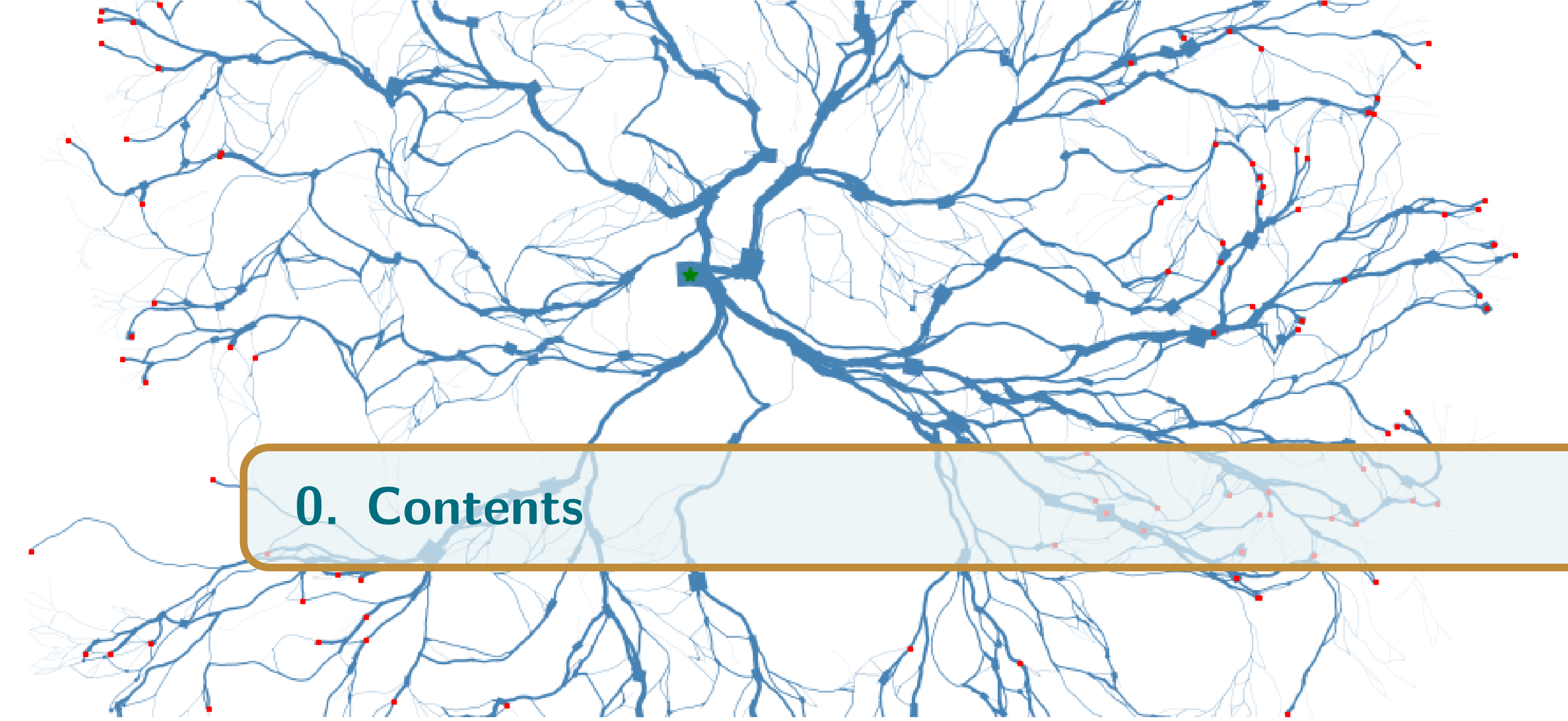
Lecture Notes For: Probability and Stochastic Process

Ali Fele Paranj
alifele@student.ubc.ca

January 15, 2025



0. Contents

1. Conditional Expectation

■ **Example 1.1 — Running example 1 (Inspired by Gordan Zitkovic lecture notes).** Throughout this chapter the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \{a, b, c, d, e, f\}$, $\mathbb{F} = \mathcal{P}(\Omega)$, and $\mathbb{P}$ uniform will be running example to demonstrate different notions in a tangible way. The following random variables defined as

$$X : (\Omega, \mathcal{F}) \to (I, \mathcal{I}), \qquad Y : (\Omega, \mathcal{F}) \to (I, \mathcal{I}), \qquad Z : (\Omega, \mathcal{F}) \to (I, \mathcal{I}),$$

where $I = \{1, \dots, 10\}$ and $\mathcal{I} = \mathcal{P}(I)$. will be in particular useful:

$$X = \begin{pmatrix} a & b & c & d & e & f \\ 1 & 3 & 3 & 5 & 5 & 7 \end{pmatrix}, \quad Y = \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 1 & 1 & 7 & 7 \end{pmatrix}, \quad Z = \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 2 & 2 \end{pmatrix}.$$

It is also important to describe $\sigma(X), \sigma(Y)$, and $\sigma(Z)$ explicitly. The atoms of $\sigma(X)$ will be the $X^{-1}(1) = \{a\}, X^{-1}(2) = \{b\}, X^{-1}(3) = \{c\}, X^{-1}(5) = \{d\}, X^{-1}(7) = \{e\}, X^{-1}(11) = \{f\}$. Thus $\sigma(X) = \mathcal{P}(\Omega)$. With a similar argument the atoms of $\sigma(Y)$ is $Y^{-1}(4) = \{a, b\}, Y^{-1}(4) = \{c, d\}, Y^{-1}(6) = \{e, f\}$. And finally, the atoms of Z will be $Z^{-1}(8) = \{a, b, c, d\}$, and $Z^{-1}(9) = \{e, f\}$. In summary

$$\begin{aligned} \mathrm{Atom}(\sigma(X)) &= \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}\}, \\ \mathrm{Atom}(\sigma(Y)) &= \{\{a, b\}, \{c, d\}, \{e, f\}\}, \\ \mathrm{Atom}(\sigma(Z)) &= \{\{a, b, c, d\}, \{e, f\}\}. \end{aligned}$$

■

■ **Example 1.2 — Running example 2 (inspired from Nima Moshayedi's lecture notes).** $N \sim \mathrm{Poisson}(\lambda)$. Consider a game, where we say that when $N = n$ we do n independent tossing of a coin where each time one obtains 1 with probability $p \in [0, 1]$ and 0 with probability $1 - p$. Define also S to be the random variable giving the total number of 1 obtained in the game. Therefore, if $N = n$ is given, we get that S has binomial distribution with parameters (p, n). ■

■ **Example 1.3 — Running example 3.** I will add some suitable random variable with density function $f_{X,Y}(x, y)$. The goal is to later calculate $f_{X|Y}(x, y)$ and $\mathbb{E}\left[X|Y\right]$, etc. TODO: Will be designed and added later. ■

We will be using the simple lemma below to demonstrate the main ideas of the conditional expectation.

Lemma 1.1 — Projection Lemma. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a non-negative random variable. Then $\mathbb{E}\Big[X\Big] \in \mathbb{R}$ is the unique number that minimizes

$$\mathbb{E}\Big[|X - n|^2\Big]$$

over all choices for $n \in \mathbb{R}$.

Proof. By differentiating and setting equal to zero we will have

$$0 = \frac{d}{dn}\mathbb{E}\Big[\frac{d}{dn}|X - n|^2\Big] = \mathbb{E}\Big[|X|\Big] - \mathbb{E}\Big[|n|\Big].$$

So

$$\mathbb{E}\Big[X\Big] = n.$$

□

1.1 Conditional Probability (Discrete Case)

Summary 1.1 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X, Y : \Omega \to I$ discrete random variables. Then the conditional expectation $\mathbb{E}\Big[X|Y\Big] : \Omega \to \mathbb{R}$:

(i) (Projective definition) is the unique $\sigma(Y)$ measurable random variable that minimizes $\mathbb{E}\Big[|X' - X|^2\Big]$ among all $\sigma(Y)$ measurable random variables X'. So $\mathbb{E}\Big[X|Y\Big]$ can be thought as the orthogonal projection of $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ to the subspace $L^2(\Omega, \sigma(Y), \mathbb{P})$.

(ii) (Alternative definition) is a random variable whose values are given as

$$\mathbb{E}\Big[X|Y\Big](\omega) = \sum_{n \in I} n\mathbb{P}(X = n | Y = Y(\omega)).$$

The nice thing about considering the conditional probability in the discrete case is the ability to do some explicit calculations.

Proposition 1.1 — Properties of conditional expectation.
1. Tower property: If $\sigma(Z) \subseteq^{\sigma} \sigma(Y)$ then
$$\mathbb{E}\Big[\mathbb{E}\Big[X|Y\Big]|Z\Big] = \mathbb{E}\Big[X|Y\Big].$$
2. Pulling out what is known: Let Z be $\sigma(Y)$ measurable. Then
$$\mathbb{E}\Big[XZ|Y\Big] = Z\mathbb{E}\Big[X|Y\Big].$$

.

■ **Example 1.4** In the running example in Example 1.1 calculate $\mathbb{E}\Big[X|Y\Big]$, $\mathbb{E}\Big[X|Z\Big]$, and explicitly verify if these random variables are $\sigma(Y)$ and $\sigma(Z)$ measurable (respectively). Then check the properties in Proposition 1.1. ■

Solution Recall that we had

$$X = \begin{pmatrix} a & b & c & d & e & f \\ 1 & 3 & 3 & 5 & 5 & 7 \end{pmatrix}, \quad Y = \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 1 & 1 & 7 & 7 \end{pmatrix}, \quad Z = \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 2 & 2 \end{pmatrix}.$$

Calculating $\mathbb{E}\Big[X|Y\Big]$: So we want to calculate $\mathbb{E}\Big[X|Y\Big]$. Let $\omega \in \Omega$. Then

$$\mathbb{E}\Big[X|Y\Big](\omega) = \sum_{n=0}^{\infty} n\mathbb{P}(X = n|Y = Y(\omega)).$$

When $\omega = a$ we have

$$\mathbb{E}\Big[X|Y\Big](a) = 1 \cdot \frac{\mathbb{P}(X = 1, Y = 2)}{\mathbb{P}(Y = 2)} + 3 \cdot \frac{\mathbb{P}(X = 3, Y = 2)}{\mathbb{P}(Y = 2)} = (1 + 3)/2 = 2.$$

With a similar argument we can calculate $\mathbb{E}\Big[X|Y\Big](b) = 2$. Let $\omega = c$. Then

$$\mathbb{E}\Big[X|Y\Big](c) = 3 \cdot \frac{\mathbb{P}(X = 3, Y = 1)}{\mathbb{P}(Y = 1)} + 5 \cdot \frac{\mathbb{P}(X = 5, Y = 1)}{\mathbb{P}(Y = 1)} = (3 + 5)/2 = 8.$$

With similar calculations we can see that

$$\mathbb{E}\Big[X|Y\Big] = \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 4 & 4 & 6 & 6 \end{pmatrix}.$$

Calculating $\mathbb{E}\Big[X|Z\Big]$: Similar to above, let $\omega = a$. Then

$$\mathbb{E}\Big[X|Z\Big](a) = 1\cdot\frac{\mathbb{P}(X = 1, Z = 3)}{\mathbb{P}(Z = 3)} + 3\cdot\frac{\mathbb{P}(X = 3, Z = 3)}{\mathbb{P}(Z = 3)} + 3\cdot\frac{\mathbb{P}(X = 3, Z = 3)}{\mathbb{P}(Z = 3)} + 5\cdot\frac{\mathbb{P}(X = 5, Z = 3)}{\mathbb{P}(Z = 3)} = (1+3+3+5)/4 = 3.$$

And with a similar computation we will get $\mathbb{E}\Big[X|Y\Big](e) = 6$. So we can write

$$\mathbb{E}\Big[X|Z\Big] = \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 6 & 6 \end{pmatrix}.$$

Measurability of $\mathbb{E}\Big[X|Y\Big], \mathbb{E}\Big[X|Z\Big]$: Recall the atoms of the σ-algebra generated by X, Y, Z as

$$\begin{aligned} \mathrm{Atom}(\sigma(X)) &= \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}\}, \\ \mathrm{Atom}(\sigma(Y)) &= \{\{a, b\}, \{c, d\}, \{e, f\}\}, \\ \mathrm{Atom}(\sigma(Z)) &= \{\{a, b, c, d\}, \{e, f\}\}. \end{aligned}$$

So it is clear that $\mathbb{E}\Big[X|Y\Big]$ is $\sigma(Y)$-measurable, while $\mathbb{E}\Big[X|Z\Big]$ is $\sigma(Z)$-measurable.
Verification of the projection interpretation. Recall Lemma 1.1. Then it immediately follows that the only function that assumes constant values on the atoms of $\sigma(Y)$ (or $\sigma(Z)$), hence $\sigma(Y)$-measurable (or $\sigma(Z)$-measurable) is the function that assumes the average value of Y (or Z) on the atoms of $\sigma(Y)$ (or $\sigma(Z)$) with the law $\mathcal{L}(\dot{|}A)$ (or $\mathcal{L}(\dot{|}B)$) where A is an atom of $\sigma(X)$ (or where B is an atom of $\sigma(Y)$).
Checking the Tower property. For an easier notation we will write $\tilde{X}_Y = \mathbb{E}\Big[X|Y\Big]$, and $\tilde{X}_Z = \mathbb{E}\Big[X|Z\Big]$. Then

$$\mathbb{E}\Big[\mathbb{E}\Big[X|Y\Big]|Z\Big](a) = \mathbb{E}\Big[\tilde{X}_Y|Z\Big] = (2 + 2 + 4 + 4)/4 = 3,$$

and similarly we can compute

$$\mathbb{E}\Big[\mathbb{E}\Big[X|Y\Big]|Z\Big] = \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 6 & 6 \end{pmatrix}.$$

And similarly we can compute

$$\mathbb{E}\Big[X|Z\Big] = \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 6 & 6 \end{pmatrix}.$$

In this case that the sample space and the random variables are finite, this property exactly translates to the fact that in order to compute the average of say n numbers, it is the same if we compute the average for the some disjoint sub-collections and then average those average values.

Checking the pull out property. Let W be a random variable that is $\sigma(Y)$ measurable, say given as

$$W = \begin{pmatrix} a & b & c & d & e & f \\ 5 & 5 & 6 & 6 & 7 & 7 \end{pmatrix}$$

Then

$$WX = \begin{pmatrix} a & b & c & d & e & f \\ 5 & 15 & 18 & 30 & 35 & 49 \end{pmatrix}$$

So we can compute

$$\mathbb{E}\Big[WX|Y\Big] = \begin{pmatrix} a & b & c & d & e & f \\ 10 & 10 & 24 & 24 & 42 & 42 \end{pmatrix}.$$