Lecture Notes For: Probability and Stochastic Process

Ali Fele Paranj

alifele@student.ubc.ca

January 30, 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 $I = \{1, \ldots, 10\}$ 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

$$\text{Atom}(\sigma(X)) = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}\},$$
$$\text{Atom}(\sigma(Y)) = \{\{a, b\}, \{c, d\}, \{e, f\}\},$$
$$\text{Atom}(\sigma(Z)) = \{\{a, b, c, d\}, \{e, f\}\}.$$

■

■ **Example 1.2 — Running example 2 (inspired from Nima Moshayedi's lecture notes).** $N \sim \text{Poisson}(\lambda)$. 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 $\mathbb{E}[X|Y]$, 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 $\mathbb{E}[X] \in \mathbb{R}$ is the unique number that minimizes
>
> $$\mathbb{E}\left[|X - n|^2\right]$$
>
> over all choices for $n \in \mathbb{R}$.

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

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

So

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

$\square$

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 $\mathbb{E}[X|Y] : \Omega \to \mathbb{R}$:
>
> (i) (Projective definition) is the unique $\sigma(Y)$ measurable random variable that minimizes $\mathbb{E}\left[|X' - X|^2\right]$ among all $\sigma(Y)$ measurable random variables X'. So $\mathbb{E}[X|Y]$ 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
>
> $$\mathbb{E}[X|Y](\omega) = \sum_{n \in I} n\mathbb{P}(X = n|Y = Y(\omega)).$$

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
>
> $$\mathbb{E}\left[\mathbb{E}[X|Y]|Z\right] = \mathbb{E}[X|Y].$$
>
> 2. Pulling out what is known: Let Z be $\sigma(Y)$ measurable. Then
>
> $$\mathbb{E}[XZ|Y] = Z\mathbb{E}[X|Y].$$
>
> .

■ **Example 1.4** In the running example in Example 1.1 calculate $\mathbb{E}[X|Y]$, $\mathbb{E}[X|Z]$, 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 $\mathbb{E}\left[X|Y\right]$: So we want to calculate $\mathbb{E}\left[X|Y\right]$. Let $\omega \in \Omega$. Then

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

When $\omega = a$ we have

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

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

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

With similar calculations we can see that

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

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

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

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

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

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

$$\text{Atom}(\sigma(X)) = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}\},$$
$$\text{Atom}(\sigma(Y)) = \{\{a, b\}, \{c, d\}, \{e, f\}\},$$
$$\text{Atom}(\sigma(Z)) = \{\{a, b, c, d\}, \{e, f\}\}.$$

So it is clear that $\mathbb{E}\left[X|Y\right]$ is $\sigma(Y)$-measurable, while $\mathbb{E}\left[X|Z\right]$ 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 $\tilde{X}_Y = \mathbb{E}\left[X|Y\right]$, and $\tilde{X}_Z = \mathbb{E}\left[X|Z\right]$. Then

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

and similarly we can compute

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

And similarly we can compute

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

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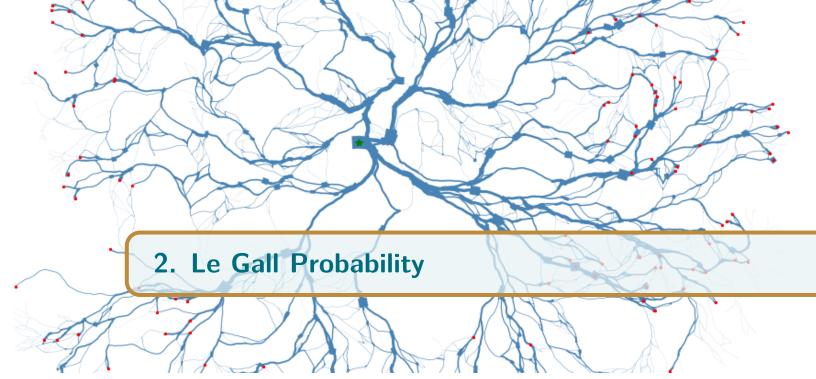
$$W = \begin{pmatrix} a & b & c & d & e & f \\ 5 & 5 & 6 & 6 & 7 & 7 \end{pmatrix}$$

Then

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

So we can compute

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

2. Le Gall Probability

In this chapter I will cover the proofs, problems, and etc that are left to the reader. I will also add some remarks as I read through the text. I will also add some theorems and their proofs that are hard to understand in the text but I have made some helpful remarks.

■ **Problem 2.1** We define the discrete conditioning as

$$\mathbb{P}'(A) = \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

where $A, B \in \mathcal{F}$ are measurable sets. Furthermore, we define the conditional expectation of a non-negative, or L^1 random variable X as

$$\mathbb{E}\left[X|B\right] := \frac{\mathbb{E}\left[\mathbb{1}_B X\right]}{\mathbb{P}(X)}.$$

Show that $\mathbb{E}\left[X|B\right] = \mathbb{E}_{\mathbb{P}'}\left[X\right]$.

Solution We first show this for a random variable given as

$$X = \alpha \mathbb{1}_B$$

for some $A \in \mathcal{F}$. Observe that

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}'}\left[X\right] = \mathbb{E}_{\mathbb{P}'}\left[\alpha \mathbb{1}_A\right] &= \alpha \mathbb{E}_{\mathbb{P}'}\left[\mathbb{1}_A\right] \\
&= \alpha \mathbb{P}'(A) = \alpha \mathbb{P}(A|B) \\
&= \alpha \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \alpha \frac{\mathbb{E}\left[\mathbb{1}_A \mathbb{1}_B\right]}{\mathbb{P}(B)} \\
&= \frac{\mathbb{E}\left[\alpha \mathbb{1}_A \mathbb{1}_B\right]}{\mathbb{P}(B)} \\
&= \frac{\mathbb{E}\left[X \mathbb{1}_B\right]}{\mathbb{P}(B)} = \mathbb{E}\left[X|B\right].
\end{aligned}$$

Using the linearity of the expectation, the conclusion above also follows for a simple random variable $X = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{B_i}$ for $B_i \in \mathcal{F}$ for all $i = 1, \ldots, n$. Let X be a non-negative random variable. We can

construct a sequence of random variables $\{X_n\}$ such that $X_n \uparrow X$. From monotone convergence theorem we have $\mathbb{E}_{\mathbb{P}'}[X_n] \uparrow \mathbb{E}_{\mathbb{P}'}[X]$, as well as $\mathbb{E}[\mathbb{1}_B X_n] \uparrow \mathbb{E}[\mathbb{1}_B X]$ as $n \to \infty$. Since each X_n is a simple random variable we also have

$$\mathbb{E}_{\mathbb{P}'}[X_n] = \frac{\mathbb{E}[X_n \mathbb{1}_B]}{\mathbb{P}(B)}.$$

Letting $n \to \infty$ on both sides of the equation above we will have

$$\mathbb{E}_{\mathbb{P}'}[X] = \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}.$$

And finally for the case that $X \in L^1$ we can write $X = X^+ - X^-$ where X^+ and X^- are non-negative random variables and using the linearity of expectation it also follows that

$$\mathbb{E}_{\mathbb{P}'}[X] = \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}.$$

Definition 2.1 Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ be a random variable and $Y : (\Omega, \mathcal{A}) \to (E, 2^E)$ be a random variable that assumes its values in a countable set E. The conditional expectation of X knowing Y is defined as

$$\mathbb{E}[X|Y] = \varphi(Y),$$

where $\varphi : E \to \mathbb{R}$ is given as

$$\varphi(y) = \begin{cases} \mathbb{E}[X|Y=y] & y \in E' \\ 0 & y \notin E' \end{cases},$$

where $E' = \{y \in E : \mathbb{P}(Y=y) > 0\}$.

Alternatively, let $B_n = Y^{-1}(n)$ for $n \in Y$. Then $\mathcal{B} = \{B_n\}_{n \in E}$ is a countable partition for Ω. Let $\mathcal{B} \supset \mathcal{B}' = \{B \in \mathcal{B} : \mathbb{P}(B) > 0\}$, i.e. those sets in the partition that has value zero. Then the conditional expectation of X knowing Y is given as

$$\mathbb{E}[X|Y] = \sum_{B \in \mathcal{B}'} \mathbb{E}[X|B] \mathbb{1}_B.$$

where $\mathbb{E}[X|B] = \mathbb{E}[X \mathbb{1}_B] / \mathbb{P}(B)$ is as defined in Le Gall, page 228, first section of the page. Observe that with this definition it automatically follows that $X = 0$ on any $B \notin \mathcal{B}'$.

■ **Remark**   Note that the value of $\varphi(y)$ when $y \notin E'$ can be any number and this will change $\mathbb{E}[X|Y]$ only on a set of measure zero. But it is convention to set it to be zero.

Theorem 2.1 Let $Y : (\Omega, \mathcal{A}) \to (E, 2^E)$ be a random variable where E is a countable set. Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Then we have $\mathbb{E}\left[\left|\mathbb{E}[X|Y]\right|\right] \leq \mathbb{E}[|X|]$, and thus

$$\mathbb{E}[X|Y] \in L^1(\Omega, \mathcal{A}, \mathbb{P}).$$

Furthermore, for every bounded $\sigma(Y)$-measurable random variable Z we have

$$\mathbb{E}[ZX] = \mathbb{E}\left[Z\mathbb{E}[X|Y]\right].$$

Proof. **Proof for** $\mathbb{E}\left[X|Y\right] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$**:** I will provide two proves for this. The first one is an elaborated version of Le Gall's proof, and the second one is my proof based on the stuff you can read in the appendix of this note.

- **Proof (1)** For Le Gall's proof we we will have

$$\mathbb{E}\left[\left|\mathbb{E}\left[X|Y\right]\right|\right] = \sum_{y \in E'} \mathbb{P}(Y = y) \left|\mathbb{E}\left[X|Y = y\right]\right|.$$

Note that for this step we used the definition of expectation for a random variable that takes countably many values (i.e. $\mathbb{E}\left[X|Y\right]$). Note that since Y takes countably many values, then by the definition of $\mathbb{E}\left[X|Y\right]$ (see Le Gall Definition 11.1), $\mathbb{E}\left[X|Y\right]$ also assumes countably many values. Furthermore, note that $E' = \{y \in E : \mathbb{P}(Y = y) > 0\}$. A different way to write this is to use the partition generated by Y. I.e. we can write

$$\mathbb{E}\left[\left|\mathbb{E}\left[X|Y\right]\right|\right] = \sum_{B \in \mathcal{B}'} \mathbb{P}(B) \left|\mathbb{E}\left[X|B\right]\right|$$

which is exactly the same as the first expression replacing $\{Y = y\}$ with B and changing the summation index accordingly. So we will have

$$\begin{aligned}
\mathbb{E}\left[\left|\mathbb{E}\left[X|Y\right]\right|\right] &= \sum_{y \in E'} \mathbb{P}(Y = y) \left|\mathbb{E}\left[X|Y = y\right]\right| \\
&= \sum_{y \in E'} \mathbb{P}(Y = y) \left|\frac{\mathbb{E}\left[X \mathbb{1}_{\{Y=y\}}\right]}{\mathbb{P}(Y = y)}\right| \\
&= \sum_{y \in E'} \left|\mathbb{E}\left[X \mathbb{1}_{\{Y=y\}}\right]\right| \\
&\leq \sum_{y \in E'} \mathbb{E}\left[|X| \, \mathbb{1}_{Y=y}\right] \\
&= \mathbb{E}\left[|X| \sum_{y \in E'} \mathbb{1}_{\{Y=y\}}\right] \\
&= \mathbb{E}\left[|X|\right].
\end{aligned}$$

- **Proof (2)** For this proof we will use the fact that since $\mathbb{E}\left[X|Y\right]$ is $\sigma(Y)$-measurable, then we can write it as

$$\mathbb{E}\left[X|Y\right] = \sum_{y \in E'} \mathbb{E}\left[X|Y = y\right] \mathbb{1}_{Y=y}$$

or equivalently

$$\mathbb{E}\left[X|Y\right] = \sum_{B \in \mathbb{B}'} \mathbb{E}\left[X|B\right] \mathbb{1}_B,$$

where E' and $\mathcal{B}'$ are as in Definition 2.1. For simplicity for the notation we will write $\alpha_B = \mathbb{E}\left[X|B\right]$.

So we will have

$$
\mathbb{E}\left[\left|\mathbb{E}\left[X|Y\right]\right|\right] = \mathbb{E}\left[\left|\sum_{B\in\mathcal{B}'}\alpha_B\mathbb{1}_B\right|\right]
$$

$$
\leq \mathbb{E}\left[\sum_{B\in\mathcal{B}'}|\alpha_B|\,\mathbb{1}_B\right]
$$

$$
= \sum_{B\in\mathcal{B}'}\mathbb{E}\left[|\alpha_B|\,\mathbb{1}_B\right]
$$

$$
= \sum_{B\in\mathcal{B}'}|\alpha_B|\,\mathbb{P}(B)
$$

$$
= \sum_{B\in\mathcal{B}'}\frac{\left|\mathbb{E}\left[X\mathbb{1}_B\right]\right|}{\mathbb{P}(B)}\mathbb{P}(B)
$$

$$
\leq \sum_{B\in\mathcal{B}'}\mathbb{E}\left[|X|\,\mathbb{1}_B\right]
$$

$$
= \mathbb{E}\left[|X|\sum_{B\in\mathcal{B}'}\mathbb{1}_B\right]
$$

$$
= \mathbb{E}\left[|X|\right].
$$

Proof for $\mathbb{E}\left[ZX\right] = \mathbb{E}\left[Z\mathbb{E}\left[X|Y\right]\right]$. Since Z is $\sigma(Y)$-measurable, by Proposition 8.9 Le Gall, we can find a bounded function φ such that $Z = \varphi(Y)$. So we can write

$$
\mathbb{E}\left[Z\mathbb{E}\left[X|Y\right]\right] = \mathbb{E}\left[\varphi(Y)\mathbb{E}\left[X|Y\right]\right]
$$

$$
= \sum_{y\in E}\varphi(y)\mathbb{E}\left[X|Y=y\right]\mathbb{P}(Y=y)
$$

$$
= \sum_{y\in E}\varphi(y)\mathbb{E}\left[X\mathbb{1}_{\{Y=y\}}\right]
$$

$$
= \mathbb{E}\left[X\sum_{y\in E}\varphi(y)\mathbb{1}_{\{Y=y\}}\right]
$$

$$
= \mathbb{E}\left[XZ\right],
$$

where we have used the Fubini's theorem to interchange $\mathbb{E}$ with sum, and also we used the fact that $Z = \sum_{y\in E}\varphi(y)\mathbb{1}_{Y=y}$ (that follows from $Z = \varphi(Y)$).

$\square$

■ **Problem 2.2** In the properties of conditional expectation of random variables in Le Gall (page 234-245), right after part (d) he concludes that using the property (d) for any non-negative random variable $(Y_n)_{n\in\mathbb{N}}$ we have

$$
\mathbb{E}\left[\sum_n Y_n\Big|\mathcal{B}\right] = \sum_n \mathbb{E}\left[Y_n|\mathcal{B}\right].
$$

Prove this!

Solution   Since $(Y_n)_{n\in\mathbb{N}}$ is non-negative, then the sequence of partial sums form an increasing sequence $\sum_{n=1}^{m} Y_n = X_m \uparrow \sum_n Y_n$. Using the linearity of conditional expectation we can write

$$\mathbb{E}\left[\sum_{n=1}^{m} Y_n \middle| \mathcal{B}\right] = \sum_{n=1}^{m} \mathbb{E}\left[Y_n \middle| \mathcal{B}\right].$$

Taking the limit of both sides and using the property (d) we will have

$$\mathbb{E}\left[\sum_{n} Y_n \middle| \mathcal{B}\right] = \sum_{n} \mathbb{E}\left[Y_n | \mathcal{B}\right].$$