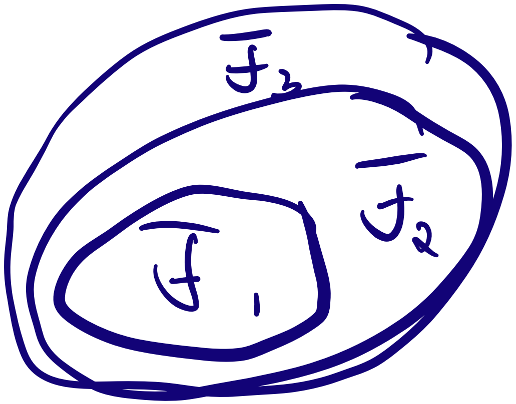
$\mathcal{F}_t \qquad s>t \quad \mathcal{F}_s \subseteq \mathcal{F}_t$

Lecture 4
Conditional Expectation



Peter Lukianchenko

$\mathcal{F}_t = \{\{2,4,6\}, \{1,3,5\}, \{1\}, \{3\}, \{5\}, \{2,3,4,5,6\}, \{1,2,4,5,6\}, \{1,2,3,4,6\}\}$

Karatzas

23 September 2023

$\mathcal{F}_0 = \{\phi, \Omega\}$

Ω

$2^{\Omega} =$

Lecture

- It all starts with the definition of conditional probability: $P(A|B) = P(AB)/P(B)$.
- If X and Y are jointly discrete random variables, we can use this to define a probability mass function for X *given* $Y = y$.
- That is, we write $p_{X|Y}(x|y) = P\{X = x|Y = y\} = \frac{p(x,y)}{p_Y(y)}$.
- In words: first restrict sample space to pairs (x, y) with given y value. Then divide the original mass function by $p_Y(y)$ to obtain a probability mass function on the restricted space.
- We do something similar when X and Y are continuous random variables. In that case we write $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$.
- Often useful to think of sampling (X, Y) as a two-stage process. First sample Y from its marginal distribution, obtain $Y = y$ for some particular y. Then sample X from its probability distribution *given* $Y = y$.
- Marginal law of X is weighted average of conditional laws.

$E[X|Y]$ - r.v. $\sim g(y)$!

$E[Z] = EX + EY = 2 \cdot EX = 2 \cdot \frac{1}{6} \cdot [1 + \dots + 6] = \frac{2 \cdot 21}{6} = \frac{21}{3} = 7$

1	2	3	4	5	6
$\frac{1}{6}$	..	.	.	..	$\frac{1}{6}$

Z	2	...	11	12

Lecture

$y = 5$

- Let X be value on one die roll, Y value on second die roll, and write $Z = X + Y$.
- What is the probability distribution for X given that $Y = 5$?
- Answer: uniform on $\{1, 2, 3, 4, 5, 6\}$.
- What is the probability distribution for Z given that $Y = 5$?
- Answer: uniform on $\{6, 7, 8, 9, 10, 11\}$ $\Rightarrow E[Z \mid Y=5] = \frac{6}{6} + \frac{7}{6} + \frac{8}{6} + \dots + \frac{11}{6} = \frac{51}{6}$
- What is the probability distribution for Y given that $Z = 5$?
- Answer: uniform on $\{1, 2, 3, 4\}$.

$E[Z \mid Y=1] = \frac{2+3+\dots+7}{6} = \frac{27}{6}$

$E[Z \mid Y=2] = \frac{3+\dots+8}{6} = \frac{33}{6}$

$E[Z \mid Y=3] = \dots = 39/6$

$E[Z \mid Y=4] = \dots = 45/6$

$E[Z] = 7$

$E[Z \mid Y=5] = \frac{51}{6}$

$E[Z \mid Y=6] = \frac{57}{6}$

$E_Y\left[E_Z[Z|Y]\right] = \left[\frac{27}{6} + \frac{33}{6} + \frac{39}{6} + \frac{45}{6} + \frac{51}{6} + \frac{57}{6}\right]\frac{1}{6} = \frac{252}{36} = 7 = E_Z[Z]$

$E_X[X]$ $E_X[X^2]$

- Now, what do we mean by $E[X|Y=y]$? This should just be the expectation of X in the conditional probability measure for X given that $Y=y$.
- Can write this as $E[X|Y=y] = \sum_x xP\{X=x|Y=y\} = \sum_x xp_{X|Y}(x|y)$.
- Can make sense of this in the continuum setting as well.
- In continuum setting we had $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$. So $E[X|Y=y] = \int_{-\infty}^{\infty} x\frac{f(x,y)}{f_Y(y)}dx$

Lecture

- Let X be value on one die roll, Y value on second die roll, and write $Z = X + Y$.
- What is $E[X|Y = 5]$?
- What is $E[Z|Y = 5]$?
- What is $E[Y|Z = 5]$?

Lecture

- Can think of $E[X|Y]$ as a function of the random variable Y. When $Y = y$ it takes the value $E[X|Y = y]$.
- So $E[X|Y]$ is itself a random variable. It happens to depend only on the value of Y.
- Thinking of $E[X|Y]$ as a random variable, we can ask what *its* expectation is. What is $E[E[X|Y]]$?
- **Very useful fact:** $E[E[X|Y]] = E[X]$.
- In words: what you expect to expect X to be *after learning* Y is same as what you *now* expect X to be.
- Proof in discrete case: $E[X|Y = y] = \sum_x xP\{X = x|Y = y\} = \sum_x x\frac{p(x,y)}{p_Y(y)}$.
- Recall that, in general, $E[g(Y)] = \sum_y p_Y(y)g(y)$.
- $E[E[X|Y = y]] = \sum_y p_Y(y) \sum_x x\frac{p(x,y)}{p_Y(y)} = \sum_x \sum_y p(x, y)x = E[X]$.

Lecture

- Definition:
  $\mathrm{Var}(X|Y) = E\big[(X - E[X|Y])^2|Y\big] = E\big[X^2 - E[X|Y]^2|Y\big]$.
- $\mathrm{Var}(X|Y)$ is a random variable that depends on Y. It is the variance of X in the conditional distribution for X given Y.
- Note $E[\mathrm{Var}(X|Y)] = E[E[X^2|Y]] - E[E[X|Y]^2|Y] = E[X^2] - E[E[X|Y]^2]$.
- If we subtract $E[X]^2$ from first term and add equivalent value $E[E[X|Y]]^2$ to the second, RHS becomes $\mathrm{Var}[X] - \mathrm{Var}[E[X|Y]]$, which implies following:
- **Useful fact:** $\mathrm{Var}(X) = \mathrm{Var}(E[X|Y]) + E[\mathrm{Var}(X|Y)]$.
- One can discover X in two stages: first sample Y from marginal and compute $E[X|Y]$, then sample X from distribution given Y value.
- Above fact breaks variance into two parts, corresponding to these two stages.

Lecture

- Let X be a random variable of variance σ_X^2 and Y an independent random variable of variance σ_Y^2 and write $Z = X + Y$. Assume $E[X] = E[Y] = 0$.
- What are the covariances $\mathrm{Cov}(X, Y)$ and $\mathrm{Cov}(X, Z)$?
- How about the correlation coefficients $\rho(X, Y)$ and $\rho(X, Z)$?
- What is $E[Z|X]$? And how about $\mathrm{Var}(Z|X)$?
- Both of these values are functions of X. Former is just X. Latter happens to be a constant-valued function of X, i.e., happens not to actually depend on X. We have $\mathrm{Var}(Z|X) = \sigma_Y^2$.
- Can we check the formula $\mathrm{Var}(Z) = \mathrm{Var}(E[Z|X]) + E[\mathrm{Var}(Z|X)]$ in this case?

Lecture

- Sometimes think of the expectation $E[Y]$ as a "best guess" or "best predictor" of the value of Y.
- It is best in the sense that at among all constants m, the expectation $E[(Y-m)^2]$ is minimized when $m = E[Y]$.
- But what if we allow non-constant predictors? What if the predictor is allowed to depend on the value of a random variable X that we can observe directly?
- Let $g(x)$ be such a function. Then $E[(y-g(X))^2]$ is minimized when $g(X) = E[Y|X]$.

Lecture

- Toss 100 coins. What's the conditional expectation of the number of heads given the number of heads among the first fifty tosses?
- What's the conditional expectation of the number of aces in a five-card poker hand given that the first two cards in the hand are aces?

Lecture

> *Conditional expectation, $\mathbb{E}(X \mid Y)$, is a random variable with randomness inherited from Y, not X.*

$$\frac{567}{1024} = \frac{87}{128} + \frac{567}{1024} = \frac{1263}{1024} \approx 1.2$$

$$\mathrm{Var}\,X = \underbrace{E_Y(\mathrm{Var}(X|Y))}_{} + \mathrm{Var}_Y(E(X|Y))$$

$$\mathrm{Var}_Y E[X|Y] = \mathrm{Var}_Y\left(\frac{9}{4}y\right) = \frac{81}{16}\mathrm{Var}\,y = \frac{81}{16}\left[E(y^2) - E^2 y\right] = \frac{81}{16}\left[\frac{1}{8} + \frac{4\cdot 7}{8} - \left(\frac{15}{8}\right)^2\right] = \frac{81}{16}\left[\frac{29}{8} - \frac{225}{64}\right] = \frac{567}{1024}$$

$$= E_y\left(\frac{3}{16}y^2\right) = \frac{3}{16}E_y y^2 = \frac{3}{16}\cdot\frac{29}{8} = \frac{87}{128}$$

Example: Suppose $Y = \begin{cases} 1 & \text{with probability } 1/8\,, \\ 2 & \text{with probability } 7/8\,, \end{cases}$

and $\quad X\,|\,Y = \begin{cases} 2Y & \text{with probability } 3/4\,, \\ 3Y & \text{with probability } 1/4\,. \end{cases}$

$$E_X[X|Y] = 2y\cdot\frac{3}{4} + 3\cdot y\cdot\frac{1}{4} = \frac{9}{4}\cdot y$$

$$E(X) = E_y\left[E_x[X|Y]\right] = E_y\left[\frac{9}{4}Y\right] = \frac{9}{4}\left[1\cdot\frac{1}{8} + 2\cdot\frac{7}{8}\right] = \frac{9\cdot 15}{4\cdot 8} = \frac{135}{32} \approx 4$$

$$\mathrm{Var}_X(X|Y) = E_X[X^2|Y] - E_X^2[X|Y] = E[X^2|Y] - \left(\frac{9}{4}y\right)^2 =$$

$$= \frac{4y^2 \cdot 3}{4} + \frac{9y^2 \cdot 1}{4} - \frac{81}{16}y^2 = 3y^2 + \frac{9}{4}y^2 - \frac{81}{16}y^2 = \frac{3}{16}y^2$$

Lecture

Conditional variance

The conditional variance is similar to the conditional expectation.

- $\text{Var}(X \mid Y = y)$ is the variance of X, when Y is fixed at the value $Y = y$.
- $\text{Var}(X \mid Y)$ is a random variable, giving the variance of X when Y is fixed at a value to be selected randomly.

Definition: Let X and Y be random variables. The **conditional variance of X, given Y,** is given by

$E[x^2|Y]$ $E[E(X|Y)] = [\quad \ldots \quad]$

$$\boxed{\text{Var}(X \mid Y) = \mathbb{E}(X^2 \mid Y) - \left\{\mathbb{E}(X \mid Y)\right\}^2 = \mathbb{E}\left\{(X - \mu_{X \mid Y})^2 \mid Y\right\}}$$

$\text{Var}\, X = E(X^2) - (E(X))^2$

$E[P_t] = \underset{\mathcal{F}_t}{E}\big[E[P_t \mid \mathcal{F}_t]\big]$!!

Lecture

If all the expectations below are finite, then for ANY random variables X and Y, we have:

i) $\boxed{\mathbb{E}(X) = \mathbb{E}_Y\big(\mathbb{E}(X \mid Y)\big)}$ *Law of Total Expectation.*

Note that we can pick any r.v. Y, to make the expectation as easy as we can.

ii) $\mathbb{E}(g(X)) = \mathbb{E}_Y\Big(\mathbb{E}(g(X) \mid Y)\Big)$ *for any function g.*

iii) $\boxed{\mathrm{Var}(X) = \overbrace{\mathbb{E}_Y\Big(\mathrm{Var}(X \mid Y)\Big)}^{\mathrm{I}} + \overbrace{\mathrm{Var}_Y\Big(\mathbb{E}(X \mid Y)\Big)}^{\mathrm{II}}}$

Law of Total Variance.

Lecture

1. Swimming with dolphins

Fraser runs a dolphin-watch business. Every day, he is unable to run the trip due to bad weather with probability p, independently of all other days. Fraser works every day except the bad-weather days, which he takes as holiday.

Let Y be the number of consecutive days Fraser has to work between bad-weather days. Let X be the total number of customers who go on Fraser's trip in this period of Y days. Conditional on Y, the distribution of X is

$$(X \mid Y) \sim \text{Poisson}(\mu Y).$$

(a) Name the distribution of Y, and state $\mathbb{E}(Y)$ and $\text{Var}(Y)$.

(b) Find the expectation and the variance of the number of customers Fraser sees between bad-weather days, $\mathbb{E}(X)$ and $\text{Var}(X)$.

Lecture

(a) Let 'success' be 'bad-weather day' and 'failure' be 'work-day'.
Then $\mathbb{P}(\text{success}) = \mathbb{P}(\text{bad-weather}) = p$.
Y is the number of failures before the first success.
So

$$Y \sim \text{Geometric}(p).$$

Thus

$$\begin{aligned} \mathbb{E}(Y) &= \frac{1-p}{p}, \\ \text{Var}(Y) &= \frac{1-p}{p^2}. \end{aligned}$$

Lecture

(b) *We know* $(X \mid Y) \sim$ *Poisson*(μY)*: so*

$$\mathbb{E}(X \mid Y) = \text{Var}(X \mid Y) = \mu Y.$$

By the Law of Total Expectation:

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}_Y\Big\{\mathbb{E}(X \mid Y)\Big\} \\ &= \mathbb{E}_Y(\mu Y) \\ &= \mu \mathbb{E}_Y(Y) \end{aligned}$$

$$\therefore \quad \mathbb{E}(X) = \frac{\mu(1-p)}{p}.$$

By the Law of Total Variance:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}_Y\Big(\text{Var}(X \mid Y)\Big) + \text{Var}_Y\Big(\mathbb{E}(X \mid Y)\Big) \\ &= \mathbb{E}_Y\Big(\mu Y\Big) + \text{Var}_Y\Big(\mu Y\Big) \\ &= \mu \mathbb{E}_Y(Y) + \mu^2 \text{Var}_Y(Y) \\ &= \mu\left(\frac{1-p}{p}\right) + \mu^2\left(\frac{1-p}{p^2}\right) \\ &= \frac{\mu(1-p)(p+\mu)}{p^2}. \end{aligned}$$

Lecture

2. Randomly stopped sum

This model arises very commonly in stochastic processes. A random number N of events occur, and each event i has associated with it some cost, penalty, or reward X_i. The question is to find the mean and variance of the total cost / reward: $T_N = X_1 + X_2 + \ldots + X_N$.
The difficulty is that the number N of terms in the sum is itself random.

T_N is called a ***randomly stopped sum: it is a sum of X_i's, randomly stopped at the random number of N terms.***

Example: Think of a cash machine, which has to be loaded with enough money to cover the day's business. The number of customers per day is a random number N. Customer i withdraws a random amount X_i. The total amount withdrawn during the day is a randomly stopped sum: $T_N = X_1 + \ldots + X_N$.

Cash machine example

$E[T_N] = E_N[E[T_N|N]] = E_N(N100]$
100λ.

The citizens of Remuera withdraw money from a cash machine according to the following probability function (X):

Amount, x (\$)	50	100	200
$\mathbb{P}(X = x)$	0.3	0.5	0.2

The number of customers per day has the distribution $N \sim \text{Poisson}(\lambda)$.

Let $T_N = X_1 + X_2 + \ldots + X_N$ be the total amount of money withdrawn in a day, where each X_i has the probability function above, and $X_1, X_2, \ldots$ are independent of each other and of N.
T_N is a randomly stopped sum, stopped by the random number of N customers.

(a) Show that $\mathbb{E}(X) = 105$, and $\text{Var}(X) = 2725$.

(b) Find $\mathbb{E}(T_N)$ and $\text{Var}(T_N)$: the mean and variance of the amount of money withdrawn each day.

$E[T_N|N] = NE[X_i|N] = N \cdot 100$

Lecture

Similarly,

$$
\begin{aligned}
\text{Var}(T_N \mid N) &= \text{Var}(X_1 + X_2 + \ldots + X_N \mid N) \\
&= \text{Var}(X_1 + X_2 + \ldots + X_N) \\
&\qquad\qquad \textit{where } N \textit{ is now considered constant;} \\
&\qquad\qquad \textit{(because all } X_i\textit{'s are independent of } N\textit{)} \\
&= \text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_N) \\
&\qquad \textit{(we DO need independence of } X_i\textit{'s for this)} \\
&= N \times \text{Var}(X) \quad \textit{(because all } X_i\textit{'s have same variance, } \text{Var}(X)\textit{)} \\
&= 2725N.
\end{aligned}
$$

So

$$\begin{aligned}\mathbb{E}(T_N) &= \mathbb{E}_N\left\{\mathbb{E}(T_N \mid N)\right\} \\ &= \mathbb{E}_N(105N) \\ &= 105\mathbb{E}_N(N) \\ &= 105\lambda,\end{aligned}$$

because $N \sim$ *Poisson*(λ) *so* $\mathbb{E}(N) = \lambda$.

Similarly,

$$\begin{aligned}\textit{Var}(T_N) &= \mathbb{E}_N\left\{\textit{Var}(T_N \mid N)\right\} + \textit{Var}_N\left\{\mathbb{E}(T_N \mid N)\right\} \\ &= \mathbb{E}_N\{2725N\} + \textit{Var}_N\{105N\} \\ &= 2725\mathbb{E}_N(N) + 105^2\,\textit{Var}_N(N) \\ &= 2725\lambda + 11025\lambda \\ &= 13750\lambda,\end{aligned}$$

because $N \sim$ *Poisson*(λ) *so* $\mathbb{E}(N) = \textit{Var}(N) = \lambda$.

General result for randomly stopped sums:

Suppose $X_1, X_2, \ldots$ each have the same mean μ and variance σ^2, and $X_1, X_2, \ldots,$ and N are mutually independent. Let $T_N = X_1 + \ldots + X_N$ be the randomly stopped sum. By following similar working to that above:

$$\boxed{\begin{aligned} \mathbb{E}(T_N) = \mathbb{E}\left\{\sum_{i=1}^{N} X_i\right\} &= \mu\,\mathbb{E}(N) \\ \mathrm{Var}(T_N) = \mathrm{Var}\left\{\sum_{i=1}^{N} X_i\right\} &= \sigma^2\,\mathbb{E}(N) + \mu^2\,\mathrm{Var}(N). \end{aligned}}$$

Thank you for your attention!

See next week!