03/16

~~3.7~~

~~3.7. Computing probabilities by~~

3.7. Computing expectations by conditioning.

Recall: $\mathbb{E}[X \mid Y=y] =: \phi(y)$ is a function of y.

Now, $\phi(Y) =: \mathbb{E}[X \mid Y]$ is a random variable. (*)

<u>Ex</u> X = height, Y = age.

$\mathbb{E}[X \mid Y]$ = average height of ~~people of ~~ ~~(randomly~~ ~~(e.g. 37 y.old)~~ ~~of ra~~ people ~~a person~~ with <u>randomly chosen</u> age.

<u>Thm</u> (3.41) (<u>Law of Total Expectation</u>)

$$\boxed{\mathbb{E}[X] = \underset{\substack{\uparrow \\ \text{w.r.to } Y}}{\mathbb{E}}\Big[\underset{\substack{\uparrow \\ \text{w.r. to } X}}{\mathbb{E}}[X \mid Y]\Big]}$$

"Condition – uncondition"

(for discrete X, Y)

$$\text{RHS} = \sum_y \mathbb{E}[X \mid Y=y]\, \mathbb{P}\{Y=y\} \qquad \text{(by def (*))}$$

$$= \sum_y \sum_x x \underbrace{\mathbb{P}\{X=x \mid Y=y\} \cdot \mathbb{P}\{Y=y\}}_{= \mathbb{P}\{X=x,\ Y=y\}}$$

$$= \sum_x x \underbrace{\Big(\sum_y \mathbb{P}\{X=x, Y=y\}\Big)}_{= \mathbb{P}\{X=x\}} \qquad \text{(abs. convergence)}$$

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

See [Walsh] for proof for continuous distr's.

So,

$$\boxed{\begin{aligned} \mathbb{E}[X] &= \sum_y \mathbb{E}[X \mid Y=y]\ \mathbb{P}\{Y=y\} && \text{for discrete } Y \\ \mathbb{E}[X] &= \int_y \mathbb{E}[X \mid Y=y]\, f_Y(y)\, dy && \text{for continuous } Y. \end{aligned}}$$

just $E[N]$. Hence $E[N|Y=0] = 1 + E[N]$. Substituting Equation (3.6) into Equation (3.5) yields

$$E[N] = p + (1-p)(1+E[N])$$

or

$$E[N] = 1/p \quad \blacksquare$$

Because the random variable N is a geometric random variable with probability mass function $p(n) = p(1-p)^{n-1}$, its expectation could easily have been computed from $E[N] = \sum_1^\infty np(n)$ without recourse to conditional expectation. However, if you attempt to obtain the solution to our next example without using conditional expectation, you will quickly learn what a useful technique "conditioning" can be.

Example 3.12 A miner is trapped in a mine containing three doors. The first door leads to a tunnel that takes him to safety after two hours of travel. The second door leads to a tunnel that returns him to the mine after three hours of travel. The third door leads to a tunnel that returns him to his mine after five hours. Assuming that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until the miner reaches safety?

Solution: Let X denote the time until the miner reaches safety, and let Y denote the door he initially chooses. Now

$$\begin{aligned} E[X] &= E[X|Y=1]P\{Y=1\} + E[X|Y=2]P\{Y=2\} \\ &\quad + E[X|Y=3]P\{Y=3\} \\ &= \tfrac{1}{3}(E[X|Y=1] + E[X|Y=2] + E[X|Y=3]) \end{aligned}$$

However,

$$\begin{aligned} E[X|Y=1] &= 2, \\ E[X|Y=2] &= 3 + E[X], \\ E[X|Y=3] &= 5 + E[X], \end{aligned} \tag{3.7}$$

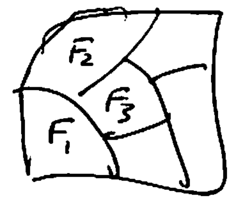
To understand why this is correct consider, for instance, $E[X|Y=2]$, and reason as follows. If the miner chooses the second door, then he spends three hours in the tunnel and then returns to the mine. But once he returns to the mine the problem is as before, and hence his expected additional time until safety is just $E[X]$. Hence $E[X|Y=2] = 3 + E[X]$. The argument behind the other equalities in Equation (3.7) is similar. Hence

$$E[X] = \tfrac{1}{3}(2 + 3 + E[X] + 5 + E[X]) \quad \text{or} \quad E[X] = 10 \quad \blacksquare$$

(Ross 3.12)

Remark (Law of Total Prob.) In the previous example, we ~~used~~ ~~the~~ could have used the LTP in the following ~~form~~ useful form:

$$\boxed{\mathbb{E}[X] = \sum_i \mathbb{E}[X \mid F_i]\, \mathbb{P}(F_i)}$$



where ~~...~~ events F_i form a partition of the sample space Ω (i.e. $\bigcup_i F_i = \Omega$; F_i disjoint)

and $\mathbb{E}[X \mid F_i]$ is defined in an obvious way $\left(\mathbb{E}(X \mid F_i) = \sum_x x\,\mathbb{P}\{X = x \mid F_i\} = \mathbb{E}[X \mid \mathbb{1}_{F_i}]\right)$

Ex (Ross 3.14) "Wait ~~...~~ for k ~~consecu.~~ successes in a row".

Flip a biased coin with $\mathbb{P}(\text{Heads}) = p$ until k consecutive heads.

$\mathbb{E}[\#\text{flips}] = ?$
$\qquad \| $
$\quad N_k$

~~... # flips to get k consecutive heads ...~~

Condition on N_{k-1}, # flips to get $k-1$ consequtive heads.

$$\overbrace{\text{HTHHTHTHHH}}^{}_{N_3 = 10}$$

$$N_k \mid N_{k-1} = \begin{cases} N_{k-1} + 1, & \text{if } k\text{'th flip} = H \text{ (done)} \\ N_{k-1} + 1 + \#(\text{flips add'l to get } k \text{ conseq. heads}), & \text{if } k\text{'th flip} = T \text{ ("reset")} \end{cases}$$

$$\Rightarrow \mathbb{E}[N_k \mid N_{k-1}] = (N_{k-1} + 1)\,p + \left(N_{k-1} + 1 + \mathbb{E}[N_k]\right)(1-p)$$

$$= N_{k-1} + 1 + (1-p)\,\mathbb{E}[N_k].$$

law of Total Exp $\Rightarrow$ $E_k := \mathbb{E}[N_k]$ satisfies

$$E_k = E_{k-1} + 1 + (1-p)E_k$$

$$\Rightarrow E_k = \frac{1}{p} + \frac{E_{k-1}}{p}$$

<u>Induction</u>: N_1 = # flips to get the first head $\sim \text{Geom}(p)$

$$E_1 = \mathbb{E}[N_1] = \frac{1}{p}.$$

$$E_2 = \frac{1}{p} + \frac{1}{p^2}; \quad E_3 = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3}, \quad \ldots$$

$$E_k = \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k} = \boxed{\frac{1/p^k - 1}{1-p}}.$$

<u>Remark</u> (Conditional probabilities) ~~One~~ Similarly, we can define

$$\mathbb{P}\{X \in A \mid Y = y\}$$

even for continuous r.v. Y. ~~Indeed,~~
Just use the conditional distr. of X for that. Similarly,

$$\mathbb{P}\{X \in A \mid Y\},$$

<u>Thm</u> (Computing probabilities by conditioning) $\forall A \subset \mathbb{R}^2$ Borel,

$$\mathbb{P}\{(X,Y) \in A\} = \underset{\substack{\uparrow \\ \text{wrt } Y}}{\mathbb{E}}\left[\mathbb{P}\{(X,Y) \in A \mid Y\}\right]$$

<u>Ex</u>: prove for discrete Y in a similar way to ~~Thm p.92~~ LTP.

We can also derive Thm directly from LTP:

$$\mathbb{P}\{(X,Y)\in A\} = \mathbb{E}\left[\mathbb{1}_{\{(X,Y)\in A\}}\right] = \mathbb{E}\Big[\underbrace{\mathbb{E}\left[\mathbb{1}_{\{(X,Y)\in A\}} \mid Y\right]}_{\mathbb{P}\{(X,Y)\in A \mid Y\}}\Big].$$

Thus, for continuous:

$$\mathbb{P}\{(X,Y)\in A\} = \int_{-\infty}^{\infty} \mathbb{P}\{(X,Y)\in A \mid Y=y\}\, f_Y(y)\, dy.$$

Ex 1 $X \sim Exp(\lambda)$, $Y \sim Exp(\mu)$ } independent

$P\{X > Y\} = ?$ (wait in one line less than in the other)

~~Condition on Y: $P\{X<Y\} = E[P\{X<Y \mid Y\}]$~~

Condition on Y:

$$P\{X>Y\} = \int_0^\infty \underbrace{P\{X>Y \mid Y=y\}}_{} f_Y(y)\, dy$$

$$P\{X>y \mid Y=y\} = P\{X>y\} \text{ (by independence)}$$

$$= e^{-\lambda y} \quad (\text{since } X \sim Exp(\lambda)).$$

$$= \int_0^\infty e^{-\lambda y} \cdot \mu e^{-\mu y}\, dy$$

$$= \mu \cdot \int_0^\infty e^{-(\lambda+\mu) y}\, dy = \boxed{\frac{\mu}{\lambda+\mu}}.$$

Ex 2 A processor consists of n independent cores, each having lifetime $\sim Exp(\lambda_i)$.

(1) What is the expected lifetime of the processor? (it fails if one core fails)

$T = \min(T_i)$, $T_i \sim Exp(\lambda_i)$ indep.

$$P\{T > x\} = P\{T_i > x \ \forall i\} = \prod_i P\{T_i > x\} \quad (\text{indep.})$$

$$= \prod_i e^{-\lambda_i x} = \exp\left(-\left(\sum_1^n \lambda_i\right) x\right)$$

$$\Rightarrow \boxed{T \sim Exp\left(\sum_1^n \lambda_i\right)}$$

$$\Rightarrow E[T] = \frac{1}{\sum_1^n \lambda_i}.$$