Stat 567
2-28-17
(1)

Law of Iterated Expectations said $E[E[X|Y]] = E[X]$

Example: Suppose N = # accidents per week and $N \sim \text{Poisson}(\lambda)$

Let X_i be the # workers injured in the i^{th} accident

Assume $X_1, \ldots, X_N \sim \text{iid } (\mu, \sigma^2)$

Let $X = \sum_{i=1}^{N} X_i$ Find $E[X]$.

(2)

$$E[X] = E[\,E[X|N]\,]$$

$$E[X|N=n] = E\left[\sum_{i=1}^{n} X_i \,\middle|\, N=n\right]$$

$$= E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = n\mu$$

So $E[X|N] = N\mu$

Then $E[X] = E[E[X|N]] = E[N\mu] = \mu E[N]$
$= \mu\lambda$

3

Conditional Variance

Defn: $\text{Var}[X \mid Y=y] = E[X^2 \mid Y=y] - \left(E[X \mid Y=y]\right)^2$

and $\text{Var}[X \mid Y] = E[X^2 \mid Y] - \left(E[X \mid Y]\right)^2$

$\text{Var}[X \mid Y]$ is a r.v. which is a function of Y.

Find $E[\text{Var}[X \mid Y]]$

$$= E\left[E[X^2 \mid Y] - \left(E[X \mid Y]\right)^2\right]$$

4

$$= E\left[E[X^2 \mid Y]\right] - E\left(E[X \mid Y]\right)^2$$

$$= E[X^2] - E\left(E[X \mid Y]\right)^2$$

$$= \underbrace{E[X^2] - (E[X])^2}_{\text{Var}[X]} + (E[X])^2 - E\left(E[X \mid Y]\right)^2$$

So $\text{Var}[X] = E[\text{Var}[X \mid Y]] + E\left(E[X \mid Y]\right)^2 - (\underbrace{E[X]}_{= E(E[X \mid Y])})^2$

⑤

last 2 terms:

$$E\left(E[X|Y]\right)^2 - \left(E[E[X|Y]]\right)^2$$

Let $W = E[X|Y]$

$$E[W^2] - (E[W])^2 = \mathrm{Var}[W]$$

$$\therefore \mathrm{Var}[X] = E[\mathrm{Var}[X|Y]] + \mathrm{Var}[E[X|Y]]$$

⑥

Back to the injured workers example.

$$\mathrm{Var}[X] = E[\mathrm{Var}[X|N]] + \underline{\mathrm{Var}[E[X|N]]}$$

We saw that $E[X|N] = N\mu$

$$\mathrm{Var}[E[X|N]] = \mathrm{Var}[N\mu]$$

$$= \mu^2 \mathrm{Var}[N] = \mu^2\lambda$$

$$\mathrm{Var}[X|N=n] = \mathrm{Var}\left[\sum_{i=1}^{N} X_i \middle| N=n\right]$$

$$= \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = n\sigma^2$$

(7)

So $Var[X|N] = N\sigma^2$

$$E[Var[X|N]] = E[N\sigma^2] = \sigma^2 E[N]$$
$$= \sigma^2 \lambda$$

$\therefore Var[X] = \sigma^2\lambda + \mu^2\lambda = \lambda(\mu^2+\sigma^2)$

Example: Best prize problem

There are n prizes, appearing in a random order.
Once you reject a prize, you cannot choose it again.

(8)

Our strategy: Fix a number k.

Reject the 1st k prizes, no matter how good they look.

Then accept the 1st prize that is better than the k rejected ones.

Goals: Find the value of k that maximizes the prob. of selecting the best prize,
and find that probability.

(9)

Let $P_k(\text{best})$ be Prob[best prize is selected with the strategy of k rejections]

Let X be the position of the best prize.

X takes on the values $1, \ldots, n$, each with probability $\frac{1}{n}$

$$P_k(\text{best}) = \sum_{x=1}^{n} P_k(\text{best} \cap X = x)$$

$$= \sum_{x=1}^{n} P_k(\text{best} \mid X = x)\, P(X = x)$$

(10)

$$= \frac{1}{n} \sum_{x=1}^{n} P_k(\text{best} \mid X = x)$$

<u>Case 1</u>: $x \leq k$ Then $P_k(\text{best} \mid X = x) = 0$

<u>Case 2</u>: $x > k$ Then $P_k(\text{best} \mid X = x) =$

$P_k(\text{prizes } k+1, k+2, \ldots, x-1 \text{ are not selected})$

$= P_k(\text{best of the } 1^{\text{st}}\ x-1 \text{ prizes occurs within the } 1^{\text{st}}\ k \text{ that were rejected})$

$$= k \cdot \frac{1}{x-1}$$

(11)

$$P_k(\text{best}) = \frac{1}{n} \sum_{x=1}^{n} P_k(\text{best} \mid X = x)$$

$$= \frac{1}{n} \sum_{x=k+1}^{n} \frac{k}{x-1} = \frac{k}{n} \sum_{x=k+1}^{n} \frac{1}{x-1}$$

$$= \frac{k}{n} \sum_{x=k}^{n-1} \frac{1}{x}$$

$$\approx \frac{k}{n} \int_{k}^{n-1} \frac{1}{x}\, dx = \frac{k}{n} \ln x \Big|_{k}^{n-1}$$

(12)

$$P_k(\text{best}) \approx \frac{k}{n}\left[\ln(n-1) - \ln k\right]$$

Let $g(k) = \frac{k}{n}\left[\ln(n-1) - \ln k\right]$

$$g'(k) = \frac{1}{n}\ln(n-1) - \frac{1}{n}\left[k \cdot \frac{1}{k} + \ln k\right]$$

$$= \frac{\ln(n-1)}{n} - \frac{1}{n} - \frac{1}{n}\ln k \overset{\text{set}}{=} 0$$

$$\ln k = \ln(n-1) - 1$$

$$k = e^{\ln(n-1)} e^{-1} = \frac{n-1}{e}$$

If $k = \frac{n-1}{e}$, then (13)

$$P_k(\text{best}) = \frac{n-1}{e}\frac{1}{n}\left[\ln(n-1) - \ln\left(\frac{n-1}{e}\right)\right]$$

$$= \frac{n-1}{en}\left[1\right] \approx \frac{1}{e} \doteq .37$$

17. Let Y be a gamma random variable with parameters (s, α). That is, its density is

$$f_Y(y) = Ce^{-\alpha y}y^{s-1}, \quad y > 0$$

where C is a constant that does not depend on y. Suppose also that the conditional distribution of X given that $Y = y$ is Poisson with mean y. That is,

$$P\{X = i | Y = y\} = e^{-y}y^i/i!, \quad i \geqslant 0$$

Show that the conditional distribution of Y given that $X = i$ is the gamma distribution with parameters $(s + i, \alpha + 1)$.

31. Each element in a sequence of binary data is either 1 with probability p or 0 with probability $1 - p$. A maximal subsequence of consecutive values having identical outcomes is called a run. For instance, if the outcome sequence is $1, 1, 0, 1, 1, 1, 0$, the first run is of length 2, the second is of length 1, and the third is of length 3.
(a) Find the expected length of the first run.
(b) Find the expected length of the second run.

37. A manuscript is sent to a typing firm consisting of typists A, B, and C. If it is typed by A, then the number of errors made is a Poisson random variable with mean 2.6; if typed by B, then the number of errors is a Poisson random variable with mean 3; and if typed by C, then it is a Poisson random variable with mean 3.4. Let X denote the number of errors in the typed manuscript. Assume that each typist is equally likely to do the work.
(a) Find $E[X]$.
(b) Find $\mathrm{Var}(X)$.

41. A rat is trapped in a maze. Initially it has to choose one of two directions. If it goes to the right, then it will wander around in the maze for three minutes and will then return to its initial position. If it goes to the left, then with probability $\frac{1}{3}$ it will depart the maze after two minutes of traveling, and with probability $\frac{2}{3}$ it will return to its initial position after five minutes of traveling. Assuming that the rat is at all times equally likely to go to the left or the right, what is the expected number of minutes that it will be trapped in the maze?

57. The number of storms in the upcoming rainy season is Poisson distributed but with a parameter value that is uniformly distributed over $(0, 5)$. That is, Λ is uniformly distributed over $(0, 5)$, and given that $\Lambda = \lambda$, the number of storms is Poisson with mean λ. Find the probability there are at least three storms this season.