

# Recall

Conditional expectation, etc

- $\text{IE}(X|Y)$  minimizes  $\text{IE}((X-h(Y))^2)$  over all functions  $h(Y)$ .
- Defined conditional variance of  $X$  given  $Y$  - it's a random variable
  - Notation:  $\text{Var}(X|Y)$
  - Definition:  $\text{Var}(X|Y) = \text{IE}\left[\underbrace{(X-\text{IE}(X|Y))^2}_{\text{another rv}} \mid Y\right]$

How to compute? In general

- Find conditional pdf/pmf  $f_{X|Y}(x|y) / p_{X|Y}(x|y)$
- Mean of that is  $\text{IE}(X|Y=y)$
- Variance of that is

$$\text{Var}(X-\text{IE}(X|Y) \mid Y=y) = \begin{cases} \int_x (X-\text{IE}[X|Y=y])^2 f_{X|Y}(x|y) dx \\ \quad \quad \quad \text{OR} \\ \sum_x (X-\text{IE}[X|Y=y])^2 p_{X|Y}(x|y) \end{cases}$$

- For each  $y$ , get a value - call that  $\gamma(y)$
- Plug in  $Y$  for  $y$ , get  $\text{Var}(X|Y) = \gamma(Y)$

**Law of Total Variance:** (a sometimes useful identity)

$$\text{Var}(X) = \text{IE}[\text{Var}(X|Y)] + \text{Var}[\text{IE}(X|Y)]$$

Why does this hold? Recall

$$X = (X - \mathbb{E}(X|Y)) + \mathbb{E}(X|Y)$$

Uncorrelated so  
variances sum

$$\text{Var}(X) = \text{Var}(\underline{X - \mathbb{E}(X|Y)}) + \text{Var}(\mathbb{E}(X|Y))$$

First term =  $\mathbb{E}[\mathbb{E}((X - \mathbb{E}(X|Y))^2 | Y)]$  by law of iterated expectations

(also because  $X - \mathbb{E}(X|Y)$  has zero mean so) - an aside  
 $\text{Var}(X - \mathbb{E}(X|Y)) = \text{Var}[(X - \mathbb{E}(X|Y))^2]$

$$\text{First Term} = \mathbb{E}(\text{Var}(X|Y))$$

### Example - See This In Action

$$Y \sim \text{Uniform}[0, 1]$$

$$X = \#\text{(heads)} \text{ in } n \text{ flips of coin w/ } P(\{H\}) = Y$$

Saw last time

$$\mathbb{E}(X|Y=y) = ny \Rightarrow \mathbb{E}(X|Y) = nY$$

$$\begin{aligned}\text{Var}(X|Y=y) &= \text{sum of variances of } n \text{-independent Bernoulli}(y) \text{ trials} \\ &= ny(1-y)\end{aligned}$$

Thus

$$\text{Var}(X|Y) = \gamma(Y) = nY(1-Y)$$

$$\text{Since } X = \underbrace{(X - \mathbb{E}(X|Y))}_{\text{uncorrelated}} + \underbrace{\mathbb{E}(X|Y)}_{\text{variance added}}$$

Find  $\text{Var}(X)$  by Law of Total Variance

$$\mathbb{E}(X|Y) = nY \Rightarrow \text{Var}(\mathbb{E}(X|Y)) = n^2/12$$

$$\begin{aligned}\text{Var}(X|Y) &= nY(1-Y) \Rightarrow \mathbb{E}(\text{Var}(X|Y)) = \mathbb{E}(nY) - \mathbb{E}(nY^2) \\ &= \frac{n}{2} - n\left(\text{Var}(Y) + (\mathbb{E}(Y))^2\right) \\ &= \frac{n}{2} - n\left(\frac{1}{12} + \frac{1}{4}\right) = \frac{n}{6}\end{aligned}$$

$$\text{Bottom Line: } \text{Var}(X) = \frac{n}{6} + \frac{n^2}{12}$$

Comment: In this context, could have figured  $\text{Var}(X)$  in other ways.

e.g.  $\mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}[\mathbb{E}((X - \mathbb{E}(X))^2 | Y)]$  (will be on HW)

Next topic,

### Moment Generating Function (MGF)

Given  $X$ , continuous or discrete, define MGF of  $X$  as

$$M_X(s) = \mathbb{E}(e^{sX}) \quad - s \text{ is "a variable"}$$

$M_X(s)$  is a function of  $s$  - need to be careful about its domain of definition

# Some Examples

①  $X \sim \text{Bernoulli}(p)$

$$e^{sx} = \begin{cases} 1 & , X=0 \\ e^s & , X=1 \end{cases} \Rightarrow M_x(s) = \mathbb{E}(e^{sx}) = pe^s + (1-p)$$

②  $X = \text{what comes up on a fair 6-sided die}$

$$e^{sx} = e^s \text{ when } X=1, e^{2s} \text{ when } X=2, \dots, e^{6s} \text{ when } X=6$$

$$M_x(s) = \mathbb{E}(e^{sx}) = \frac{1}{6} [e^s + e^{2s} + e^{3s} + e^{4s} + e^{5s} + e^{6s}]$$

③  $X \sim \text{Geometric}(p) \rightsquigarrow P_x(k) = \begin{cases} p(1-p)^{k-1} & , k > 0 \\ 0 & , k \leq 0 \end{cases}$

$$e^{sx} = e^{ks} \text{ when } X=k.$$

$$M_x(s) = \mathbb{E}(e^{sx}) = \sum_{k=1}^{\infty} p(1-p)e^{ks} = pe^s \sum_{k=1}^{\infty} ((1-p)e^s)^{k-1} = \frac{pe^s}{1 - (1-p)e^s}$$

↑  
Caution: need  $|1-p|e^s < 1$

$$(4) X \sim \text{Exponential}(\lambda) \rightarrow f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{else} \end{cases}$$

$$M_X(s) = \mathbb{E}(e^{sx}) = \int_{-\infty}^{+\infty} e^{sx} f_X(x) dx = \int_0^{\infty} \lambda e^{(s-\lambda)x} dx = \left. \frac{\lambda}{s-\lambda} e^{(s-\lambda)x} \right|_{x=0}^{x=\infty}$$

sort of Laplace transform-ish

$$\text{NEED } s-\lambda < 0 \text{ here} \Rightarrow = \frac{\lambda}{\lambda-s}$$

Side observation:  $M_X(0)$  always well defined - in fact,  $M_X(0) = 1$  for ANY  $\sim X$ .

$$(5) X \sim \text{Gaussian}(\mu, \sigma^2), Y \sim \text{Gaussian}(\mu=0, \sigma^2=1)$$

$$M_Y(s) = \mathbb{E}(e^{sy}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{sy} e^{-\frac{y^2}{2}} dy \stackrel{\substack{\text{Complete square in} \\ \text{exponent}}}{=} \frac{e^{s^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(s-y)^2}{2}} dy$$

by Normalization property  
of a Gaussian with  
mean  $\mu=s$

$$= e^{s^2/2}$$

Note

$$X = \sigma Y + \mu \implies M_X(s) = e^{\mu s} M_Y(\sigma s) = e^{\mu s} e^{\frac{\sigma^2 s^2}{2}}$$

General rule: Moment generating function of  $X = aY + b$  is

$$M_X(s) = e^{bs} M_Y(as)$$

To see this

$$M_X(s) = \mathbb{E}(e^{sx}) = \mathbb{E}(e^{s(aY+b)}) = e^{sb} \mathbb{E}(e^{asY}) = e^{sb} M_Y(as)$$

Why is it called a moment generating function?

Recall that the  $k^{\text{th}}$  moment of  $X = \mathbb{E}(X^k)$ ,  $k > 0$ .

Fact:  $\forall k > 0, \mathbb{E}(X^k) = \left. \frac{d^k}{ds^k} M_k(s) \right|_{s=0}$

Why? Say for simplicity  $X$  is continuous, so

$$M_X(s) = \int_{-\infty}^{+\infty} e^{sx} f_x(x) dx$$

$$\Rightarrow \frac{d}{ds} M_X(s) = \int_{-\infty}^{+\infty} x e^{sx} f_x(x) dx$$

$$\Rightarrow \frac{d}{ds} \left[ \frac{d}{ds} M_X(s) \right] = \int_{-\infty}^{+\infty} x^2 e^{sx} f_x(x) dx ; \text{ etc, etc, etc}$$

Evaluate at  $s=0$  get  $\int_{-\infty}^{+\infty} x f_x(x) = \mathbb{E}(X)$

Comment: this requires interchangeability of  $\frac{d}{ds}$  and  $\int dx$  (or  $\frac{d}{ds}$  and  $\sum_x$  in discrete world) - turns out to be legal in this context.

## Example - Applying the moment-generating property

$$\textcircled{1} \quad X \sim \text{Exponential}(\lambda) \Rightarrow M_X(s) = \frac{1}{\lambda-s}$$

$$E(X) = \left. \frac{d}{ds} M_X(s) \right|_{s=0} = \left. \frac{\lambda}{(\lambda-s)^2} \right|_{s=0} = \frac{1}{\lambda}$$

$$E(X^2) = \left. \frac{d^2}{ds^2} M_X(s) \right|_{s=0} = \left. \frac{2\lambda}{(\lambda-s)^3} \right|_{s=0} = \frac{2}{\lambda^2}$$

$$E(X^3) = \left. \frac{d^3}{ds^3} M_X(s) \right|_{s=0} = \left. \frac{6\lambda}{(\lambda-s)^4} \right|_{s=0} = \frac{6}{\lambda^3}$$

**Important Fact:**

$$\left. \begin{array}{l} f_X(x) \\ P_X(x) \end{array} \right\} \xleftrightarrow[\text{one-to-one correspondence}]{\quad} M_X(s)$$

i.e.  $M_X(s)$  determines pmf or pdf of  $X$  completely.