Recall

Conditional expectation, etc

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

How to compute? In general
- Find conditional pdf/pmf $f_{X|Y}(x|y)$ / $p_{X|Y}(x|y)$
- Mean of that is $\mathbb{E}(X|Y=y)$
- Variance of that is

$$Var\left(X - \mathbb{E}(X|Y) \middle| Y=y\right) = \begin{cases} \int_x \left(x - \mathbb{E}[X|Y=y]\right)^2 f_{X|Y}(x|y)\,dx \\ \text{OR} \\ \sum_x \left(x - \mathbb{E}[X|Y=y]\right)^2 p_{X|Y}(x|y) \end{cases}$$

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

Law of Total Variance: (a sometimes useful identity)

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var[\mathbb{E}(X|Y)]$$

Why does this hold? Recall

Uncorrelated so variances sum

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

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

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

(also because $X - \mathbb{E}(X|Y)$ has zero mean so

$$\mathrm{Var}\left(X - \mathbb{E}(X|Y)\right) = \mathrm{Var}\left[\left(X - \mathbb{E}(X|Y)\right)^2\right]$$
) — an aside

First Term $= \mathbb{E}\left(\mathrm{Var}(X|Y)\right)$

Example – See This In Action

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

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

Saw last time

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

$\mathrm{Var}(X|Y=y) =$ sum of variances of n-independent Bernoulli(y) trials

$= ny(1-y)$

Thus

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

uncorrelated ⇒ variances add

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

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

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

$Var(X|Y) = nY(1-Y) \Rightarrow \mathbb{E}(Var(X|Y)) = \mathbb{E}(nY) - \mathbb{E}(nY^2)$

$$= \frac{n}{2} - n\left(Var(Y) + (\mathbb{E}(Y))^2\right)$$

$$= \frac{n}{2} - n\left(\frac{1}{12} + \frac{1}{4}\right) = \frac{n}{6}$$

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

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

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

Next topic,

Moment Generating Function (MGF)

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

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

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

Some Examples

① $X \sim \text{Bernouilli}(p)$

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

② X = what comes up on a fair 6-sided die

$$e^{sX} = e^s \text{ when } X=1,\ e^{2s} \text{ when } X=2, \ldots e^{6s} \text{ when } X=6$$

$$M_X(s) = \mathbb{E}(sX) = \frac{1}{6}\left[e^s + e^{2s} + e^{3s} + e^{4s} + e^{5s} + e^{6s}\right]$$

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

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

$$M_X(s) = \mathbb{E}(e^{sX}) = \sum_{k=1}^{+\infty} p(1-p)e^{ks} = pe^s \sum_{k=1}^{\infty} \left((1-p)e^s\right)^{k-1} = \frac{pe^s}{1-(1-p)e^s}$$

caution: need $|(1-p)e^s| < 1$

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

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

Sort of Laplace transform-ish

NEED $s-\lambda<0$ here

$$= \frac{\lambda}{\lambda - s}$$

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

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

Complete square in exponent

$$M_Y(s) = \mathbb{E}(e^{sY}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{sy} e^{-\frac{y^2}{2}} dy = \frac{e^{s^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(s-y)^2}{2}} dy$$

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

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

Note

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

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^{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

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

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

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

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

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

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

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

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

Important Fact:

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

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