Chapter 4 Moments

X a discrete RV

$$E[X] = \mu_X = \sum_{\text{All } x} x\, P[X=x] = \sum_{\text{All } x} x f_X(x)$$

X cts

$$E[X] = \mu_X = \int_{-\infty}^{+\infty} x f_X(x)\, dx$$

Ex

$X \sim \text{Poisson}(\lambda)$ $\quad P[X=k] = \dfrac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$

$$E[X] = \sum_{\text{All } x} x\, P[X=x] = \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

$$= \lambda e^{-\lambda} \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Ex $X \sim \text{Unif}[a,b]$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx = \int_a^b \frac{x}{b-a}\, dx = \frac{1}{b-a}\left(\frac{x^2}{2} \Big|_{x=a}^{x=b} \right)$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

$E[X]$ can be $+\infty$, $-\infty$, or not exist

Ex $\quad X \sim \text{Cauchy}(0,1)$

$$f_X(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < +\infty$$

Note that $\int_{-\infty}^{+\infty} f_X(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\pi}\left(\frac{1}{1+x^2}\right)dx$

$$= \frac{1}{\pi}\left[\tan^{-1}(x)\Big|_{x=-\infty}^{x=+\infty}\right] = \frac{1}{\pi}\left[\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)\right] = 1$$

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)dx = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x}{1+x^2}dx$$

$$= \frac{1}{\pi}\int_0^{\infty} \frac{x}{1+x^2}dx + \frac{1}{\pi}\int_{-\infty}^{0} \frac{x}{1+x^2}dx$$

$$= \frac{1}{\pi}\int_0^{\infty} \frac{x}{1+x^2}dx - \frac{1}{\pi}\int_0^{\infty} \frac{y}{1+y^2}dy \qquad y=-x$$

$= \infty - \infty$ Does not exist!

Since $\int_0^{\infty} \frac{x}{1+x^2}dx = \frac{1}{2}\ln(1+x^2)\Big|_{x=0}^{\infty} = \infty$

In $E[X] = \int_0^{\infty} x f_X(x)dx + \int_{-\infty}^{0} x f_X(x)dx \qquad y=-x$

$$= \int_0^{\infty} x f_X(x)dx + \int_{\infty}^{0} (-y) f_X(-y)(-dy)$$

$$= \int_0^{\infty} x f_X(x)dx - \int_0^{\infty} y f_X(-y)dy$$

$$= ① - ②$$

If $① = \infty$ & $② < \infty$, set $E[X] = +\infty$

If $① < \infty$ & $② = \infty$, set $E[X] = -\infty$

If $① = \infty$ & $② = \infty$, we say $E[X]$ Does not exist

<u>Ex</u> $X \in \{1, 2, 3, \dots\}$

$$P[X=k] = \frac{c}{k^2}, \quad k = 1, 2, \dots$$

a) What is c?

$$1 = \sum_{K=1}^{\infty} P(X=K) \Rightarrow 1 = \sum_{K=1}^{\infty} \frac{c}{K^2}$$

$$\Rightarrow 1 = c\left(\pi^2/6\right) \qquad c = \frac{6}{\pi^2}$$

b)
$$E[X] = \sum_{\text{All } x} x\, P[X=x]$$

$$= \sum_{K=1}^{\infty} K \frac{c}{K^2} = c \sum_{K=1}^{\infty} \frac{1}{K} = \infty$$

<u>Transformation Thm</u>

X a RV and g a function.

$$E[g(x)] = \sum_{\text{All } x} g(x)\, P[X=x] \qquad \text{discrete case}$$

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\, dx \qquad \text{cts case}$$

<u>proof</u>

Discrete case:

$$Y = g(x)$$

$$E[Y] = \sum_{\text{All } y} y\, P[Y=y]$$

$$= \sum_{\text{All } y} y\, P[g(x)=y]$$

$$= \sum_{\text{All } y} \sum_{\text{All } x:\, g(x)=y} y\, P[X=x]$$

$$= \sum_{\text{All } y} \sum_{\text{All } x:\, g(x)=y} g(x)\, P[X=x]$$

$$= \sum_{\text{All } x} g(x)\, P[X=x]$$

∎

Ex $X \sim \text{Unif}[0,1]$, $Y = X^2$

Thm gives $E[Y] = \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx$

$$= \int_0^1 x^2 \cdot 1\,dx = 1/3$$

To do this the "long way"

$$F_Y(y) = P[Y \leq y] = P[X^2 \leq y] = P[X \leq \sqrt{y}]$$

$$= \int_0^{\sqrt{y}} 1 \cdot dx = \sqrt{y}$$

$$f_Y(y) = \tfrac{1}{2} y^{-1/2}, \quad 0 \leq y \leq 1$$

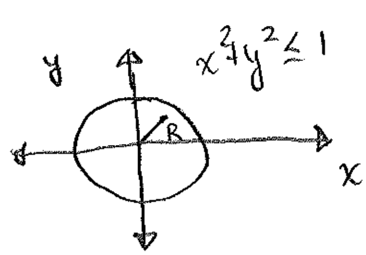
$$E[Y] = \int_0^1 y\, f_Y(y) = \tfrac{1}{2} \int_0^1 y^{1/2}\,dy = \tfrac{1}{2}\left(\frac{y^{3/2}}{3/2} \Bigg|_{y=0}^{y=1} \right)$$

$$= \frac{1}{2} \frac{1}{(3/2)} = 1/3$$

This result extends: If $Y = g(X_1, X_2, \ldots, X_n)$ is a function of the RVs $X_1, X_2, \ldots, X_n$

$$E[Y] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, x_2, \ldots, x_n)\, f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_n\, dx_{n-1} \cdots dx_1$$

Ex Dart board toss



Let $R = \sqrt{X^2 + Y^2}$
= distance to origin

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$E[R] = \iint_{\text{Unit Circle}} \sqrt{x^2+y^2}\; f_{X,Y}(x,y)\,dy\,dx$$

$$= \iint_{\text{Unit Circle}} \sqrt{x^2+y^2}\; \frac{1}{\pi}\,dy\,dx$$

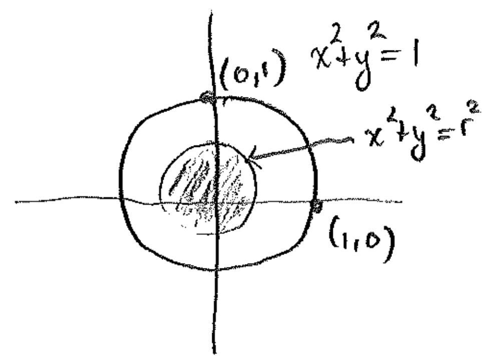
Polar Swap

$$= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 r \quad r\,dr\,d\theta$$

$$= \frac{1}{\pi}\int_0^{2\pi} \frac{1}{3}\,d\theta = 2/3$$

The long way: $0 \le r \le 1$

$$F_R(r) = P[R \le r] = P\left[\sqrt{x^2+y^2} \le r\right]$$



$$= \iint_{\text{Shaded Region}} f_{X,Y}(x,y)\,dy\,dx$$

$$= \frac{1}{\pi}\pi r^2 = r^2, \quad 0 \le r \le 1$$

$$f_R(r) = \frac{d}{dr}(r^2) = 2r, \quad 0 \le r \le 1$$

$$E[R] = \int_{-\infty}^{+\infty} r f_R(r)\,dr = \int_0^1 r \cdot 2r\,dr$$

$$= 2/3\, r^3 \Big|_{r=0}^{r=1}$$

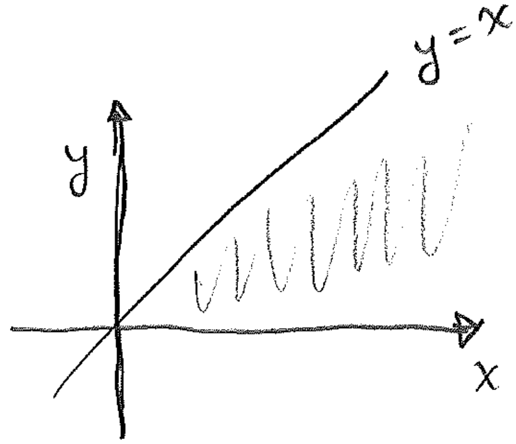
$$= 2/3$$

Tail Integration

Suppose $X \geq 0$ is a cts RV.

Then $E[X] = \int_0^\infty x f_X(x)\,dx$

$$= \int_0^\infty \int_0^x dy \; f_X(x)\,dx$$

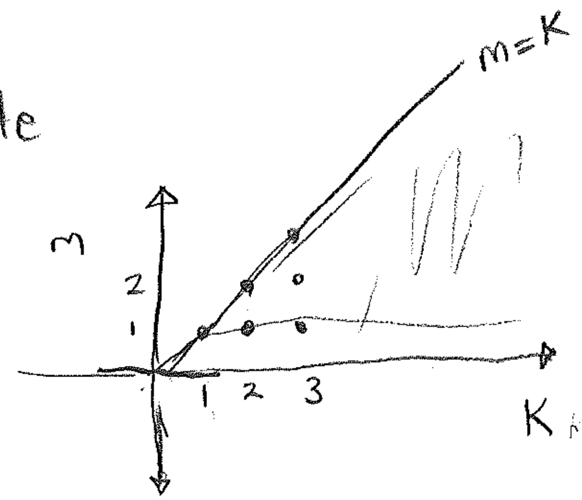


$$= \int_0^\infty \int_0^x f_X(x)\,dy\,dx$$

$$= \int_0^\infty \int_y^\infty f_X(x)\,dx\,dy$$

$$= \int_0^\infty P(X>y)\,dy$$

If $X \in \{0, 1, 2, \dots\}$ is discrete



$$E[X] = \sum_{K=1}^{\infty} K\, P[X=K]$$

$$= \sum_{K=1}^{\infty} \sum_{m=1}^{K} P[X=K]$$

$$= \sum_{m=1}^{\infty} \sum_{K=m}^{\infty} P[X=K]$$

$$= \sum_{m=1}^{\infty} P[X \geq m] = \sum_{m=0}^{\infty} P[X>m]$$

Ex $\quad X \sim \text{Geometric}(p)$

$$P[X=K] = p\,q^{K-1}, \qquad K = 1, 2, \dots$$

$$q = 1-p$$

X counts the # of IID trials until a success is achieved. Each trial has success probability p.

$$E[X] = \sum_{K=1}^{\infty} K\,P[X=K] = \sum_{K=1}^{\infty} K p q^{K-1}$$

$$= p \sum_{K=1}^{\infty} K q^{K-1}$$

We need $\sum_{K=1}^{\infty} K x^{K-1} = \sum_{K=1}^{\infty} \frac{d}{dx}(x^K)$ $\quad 0<x<1$

$$= \frac{d}{dx}\left(\sum_{K=1}^{\infty} x^K \right) = \frac{d}{dx}\left(\frac{x}{1-x} \right)$$

$$= \frac{(1-x)(1) - x(-1)}{(1-x)^2} = \frac{1}{(1-x)^2}$$

So $E[X] = p\left(\frac{1}{(1-q)^2} \right) = \frac{p}{p^2} = 1/p$

If you didn't see "that trick",

$P(X>m) = P(T_1 T_2 T_3 \ldots T_m) = q^m, \quad m = 0, 1, 2, \ldots$

$\sum_{m=0}^{\infty} P(X>m) = \sum_{m=0}^{\infty} q^m = \frac{1}{1-q} = \frac{1}{p}$

If $g(x) = ax+b$,

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\,dx \qquad \text{cts case}$$

$$= \int_{-\infty}^{+\infty} (ax+b) f_X(x)\,dx$$

$$= a\int_{-\infty}^{+\infty} x f_X(x)\,dx + b\int_{-\infty}^{+\infty} f_X(x)\,dx$$

$= aE[X] + b$

Because of this, we call expectation a "linear operator".

Note that if g_1 & g_2 are two functions of $X_1, X_2, \dots, X_n$

$$E[g_1(X_1, X_2, \dots, X_n) + g_2(X_1, \dots, X_n)] =$$

$$\int_{-\infty}^{-\infty} \cdots \int_{-\infty}^{+\infty} [g_1(x_1, \dots, x_n) + g_2(x_1, \dots, x_n)] f_{X_1, \dots, X_n}(x_1, \dots, x_n)\, dx_n\, dx_{n-1} \dots dx_1$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g_1(x_1, \dots, x_n) f_{X_1, \dots X_n}(x_1, \dots x_n)\, dx_n\, dx_{n-1} \dots dx_1 \quad +$$

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g_2(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n)\, dx_n \dots dx_1$$

$$= E[g_1(X_1, \dots, X_n)] + E[g_2(X_1, \dots, X_n)]$$

Ex $X \sim Bin(n,p)$.

Write $X = X_1 + X_2 + \dots + X_n$

$$X_i = \begin{cases} 1, & \text{if trial } i \text{ is a success} \\ 0, & \text{otherwise} \end{cases}$$

$$E[X] = E[X_1] + \dots + E[X_n]$$

$$E[X_i] = 1\, P(X_i = 1) + 0\, P(X_i = 0) = 1 \cdot p + 0(1-p) = p$$

So $E[X] = p + p + \dots + p = np$ Try to sum $\sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k}$

Variance $\qquad X$ is a RV

$$Var(X) := E[(X-\mu_X)^2] \qquad \mu_X = E[X]$$

$$Var(X) =: \sigma_X^2 \geq 0$$

σ_X^2 measures spread in the distribution.

$$Var(X) = E[(X-\mu_X)^2] = E[X^2 - 2\mu_X X + \mu_X^2]$$
$$= E[X^2] - 2\mu_X E[X] + \mu_X^2$$
$$= E[X^2] - \mu_X^2$$
$$= E[X^2] - E[X]^2$$

Shortcut Formula

Ex $\quad X \sim Unif(a,b)$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} \qquad E[X] = \frac{a+b}{2}$$

$$Var(X) = E[X^2] - E[X]^2$$
$$= \int_a^b x^2 \frac{1}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2 =$$
$$= \frac{1}{b-a}\left(\frac{x^3}{3} \Big|_{x=a}^{x=b} \right) - \left(\frac{a+b}{2}\right)^2$$
$$= \frac{b^3-a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} - \frac{a^2+2ab+b^2}{4}$$
$$= \text{For You} = \frac{(b-a)^2}{12}$$

$Var(ax+b) = ?$

Let $Y = ax+b$ $\qquad E[Y] = a\mu_x + b$

$$Var(Y) = E[(Y-\mu_Y)^2]$$

$$= E[(ax+b-(a\mu_x+b))^2] =$$

$$E[(a(X-\mu_x))^2] =$$

$$E[a^2(X-\mu_x)^2] = a^2 E[(X-\mu_x)^2]$$

$$= a^2 Var(X)$$

b does not enter into things.

Adding constants does not affect variabilities....

If X & Y are independent,

$$E[XY] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f_X(x) f_Y(y)\, dy\, dx$$

$$= \int_{-\infty}^{+\infty} x f_X(x)\, dx \int_{-\infty}^{+\infty} y f_Y(y)\, dy$$

$$= E[X]\,E[Y]$$

Indeed, if X & Y are indep and $g_1(\cdot)$ and $g_2(\cdot)$ are functions,

$$E[g_1(X)g_2(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x)g_2(y) f_{X,Y}(x,y)\,dy\,dx$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x)g_2(y) f_X(x) f_Y(y)\,dy\,dx$$

$$= \int_{-\infty}^{+\infty} g_1(x) f_X(x)\,dx \int_{-\infty}^{+\infty} g_2(y) f_2(y)\,dy$$

$$= E[g_1(X)]\,E[g_2(Y)]$$

If $X_1, X_2, \ldots, X_n$ are independent,

$$E[X_1 X_2 \ldots X_n] = \prod_{i=1}^{n} E[X_i] = E[X_1]\,E[X_2] \ldots E[X_n]$$

<u>Thm</u> If X & Y are independent, $Var(X+Y) = Var(X) + Var(Y)$.

$$Var(X+Y) = E[(X+Y-(\mu_X+\mu_Y))^2]$$

$$= E[\{(X-\mu_X)+(Y-\mu_Y)\}^2] =$$

$$E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] + 2E[(X-\mu_X)(Y-\mu_Y)]$$

Cross Term: $E[(X-\mu_X)(Y-\mu_Y)] =$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y)\,dy\,dx =$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f_X(x) f_Y(y)\,dy\,dx$$

$$= \left[\int_{-\infty}^{+\infty} (x-\mu_x) f_X(x)\,dx\right]\left[\int_{-\infty}^{+\infty} (y-\mu_Y) f_Y(y)\,dy\right]$$

$$= 0 \cdot 0 = 0$$

So $Var(X+Y) = Var(X) + Var(Y)$

Ex. $X \sim Bin(n,p)$

$X = X_1 + X_2 + \dots + X_n$.

$Var(X) = Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$

since the trials are independent.

$Var(X_1) = Var(X_2) = \dots = Var(X_n)$ since the trials are identically distributed

$Var(X_i) = E[X_i^2] - E[X_i]^2$

$E[X_i] = p$ $\qquad$ $E[X_i^2] = \sum_{\text{All } x} x^2 P[X_i = x]$

$= 0^2 P[X_i = 0] + 1^2 P[X_i = 1] =$

$P[X_i = 1] = p$

$Var(X_i) = p - p^2 = p(1-p)$

So $Var(X) = Var(X_1 + \dots + X_n) = np(1-p)$

Chebyshev's ≠ :

If $X \geq 0$, $E[X] = \int_0^\infty x f_X(x)\,dx$

$$\geq \int_c^\infty x f_X(x)\,dx \quad \text{for any } c > 0$$

$$\geq c \int_c^\infty f_X(x)\,dx = c\,P[X \geq c]$$

$$P[X \geq c] \leq \frac{E[X]}{c} \quad (\text{Markov's } \neq)$$

Now use this as follows:

$$P[\,|X - \mu_X| \geq K\sigma_X\,] = P[\,(X-\mu_X)^2 \geq K^2\sigma_X^2\,]$$

$$\leq E[\,(X-\mu_X)^2\,] / K^2\sigma_X^2$$

$$= \frac{\sigma_X^2}{\sigma_X^2}\,\frac{1}{K^2} = \frac{1}{K^2}$$

At least $1 - \frac{1}{K^2}$ of a distribution's probability lies between $\mu_X - K\sigma_X$ & $\mu_X + K\sigma_X$

<u>Fact</u> Chebyshev's ≠ cannot be improved upon: there exists distributions where

$$P[\,\mu_X - K\sigma_X \leq X \leq \mu_X + K\sigma_X\,] = 1 - \frac{1}{K^2}$$

Moments

The α^{th} moment of X is $E[X^\alpha]$, $\alpha \geq 0$

The α^{th} central moment of X is $E[(X-\mu_X)^\alpha]$, $\alpha \geq 0$

Fact If $E[|X|^\beta] < \infty$, then $E[|X|^\alpha] < \infty$ for $0 \leq \alpha \leq \beta$.

Why Take $X \geq 0$ WOLOG to avoid writing absolute values. Cts case

$$E[X^\alpha] = \int_0^\infty x^\alpha f_X(x)\,dx = \int_0^1 x^\alpha f_X(x)\,dx + \int_1^\infty x^\alpha f_X(x)\,dx$$

$$\leq \int_0^1 f_X(x)\,dx + \int_1^\infty x^\beta f_X(x)\,dx$$

$$= \underbrace{\int_0^\infty x^\beta f_X(x)\,dx}_{<\infty} + \underbrace{P[X \in (0,1]]}_{\leq 1} - \underbrace{\int_0^1 x^\beta f_X(x)\,dx}_{\leq 1}$$

$< \infty$ ∎

Skewness is defined as $E[(X-\mu_X)^3] / \sigma_X^3$

Moment Generating Functions

If X is a RV, its MGF is defined as

$$\psi_X(t) = E[e^{tX}]$$

This may be infinite

Ex $X \sim \mathrm{Exp}(\beta)$

$$f_X(x) = \beta e^{-\beta x}, \quad x \geq 0$$

$$E\left[e^{tX}\right] = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\,dx$$

$$= \int_0^{\infty} e^{tx} \beta e^{-\beta x}\,dx$$

$$= \beta \int_0^{\infty} e^{-(\beta-t)x}\,dx$$

finite if $\beta - t > 0 \Leftrightarrow t < \beta$

$$= \beta \left. \frac{e^{-(\beta-t)x}}{-(\beta-t)} \right|_{x=0}^{x=\infty}$$

$$= \frac{\beta}{\beta - t}$$

$$\psi_X(t) = \begin{cases} \beta/\beta-t, & t < \beta \\ \infty, & t \geq \beta \end{cases}$$

The MGF Generates moments.

$$\psi_X(t) = E\left[e^{tX}\right] = E\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots\right]$$

$$= 1 + t\,E[X] + \frac{t^2}{2} E[X^2] + \frac{t^3}{6} E[X^3] + \cdots$$

$$\psi_X'(t) = E[X] + t\,E[X^2] + \frac{t^2}{2} E[X^3] + \cdots$$

So $\psi_X'(0) = E[X]$

$$\psi_X''(t) = E[X^2] + t\,E[X^3] + \cdots$$

$$\psi_X''(0) = E[X^2]$$

In general $\quad \psi_X^{(k)}(0) = E[X^k]$.

Suppose that $X_1, X_2, \ldots, X_n$ are independent

$$\psi_{X_1+X_2+\ldots+X_n}(t) = E\left[e^{t(X_1+X_2+\ldots+X_n)}\right]$$
$$= E\left[e^{tX_1} e^{tX_2} \cdots e^{tX_n}\right]$$
$$= E\left[e^{tX_1}\right] E\left[e^{tX_2}\right] \cdots E\left[e^{tX_n}\right]$$
$$= \psi_{X_1}(t)\,\psi_{X_2}(t) \cdots \psi_{X_n}(t)$$

__Thm__ The MGF is unique: If $\psi_X(t) = \psi_Y(t)$ "for all t", then $F_X(t) = F_Y(t) \quad \forall t$.

No proof

This is very useful:

Suppose $X \sim \text{Poisson}(\lambda_X)$ & $Y \sim \text{Poisson}(\lambda_Y)$ & X & Y are indep.

$$\psi_X(t) = E\left[e^{tX}\right] = \sum_{k=0}^{\infty} e^{tk} P(X=k)$$
$$= \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda_X} \lambda_X^k}{k!} = e^{-\lambda_X} \sum_{k=0}^{\infty} \frac{(\lambda_X e^t)^k}{k!}$$

$$= e^{-\lambda_X} e^{\lambda_X e^t} = e^{\lambda_X(e^t-1)}$$

Always finite

So $\psi_{X+Y}(t) = \psi_X(t)\,\psi_Y(t)$

$$= e^{\lambda_x(e^t-1)}\, e^{\lambda_Y(e^t-1)}$$

$$= e^{(\lambda_x+\lambda_Y)(e^t-1)}$$

But this is the MGF of a Poisson$(\lambda_x+\lambda_Y)$ distribution

$X+Y \sim \text{Poisson}(\lambda_x+\lambda_Y)$

Compare to the grind it out way:

$$P[X+Y=K] = P(X=0)P(Y=K) + P(X=1)P(Y=K-1) + \dots + P(X=K)P(Y=0)$$

$$= \frac{e^{-\lambda_x}\lambda_x^0}{0!}\frac{e^{-\lambda_Y}\lambda_Y^K}{K!} + \frac{e^{-\lambda_x}\lambda_x}{1!}\frac{e^{-\lambda_Y}(\lambda_Y)^{K-1}}{(K-1)!} + \dots + \frac{e^{-\lambda_x}\lambda_x^K}{K!}\frac{e^{-\lambda_Y}\lambda_Y^0}{0!}$$

$$= \sum_{m=0}^{K} \frac{e^{-\lambda_x}\lambda_x^m}{m!}\frac{e^{-\lambda_Y}(\lambda_Y)^{K-m}}{(K-m)!}$$

$$= \frac{e^{-(\lambda_x+\lambda_Y)}}{K!}\sum_{M=0}^{K}\frac{K!}{M!(K-M)!}\lambda_x^M\lambda_Y^{K-M}$$

$$= \frac{e^{-(\lambda_x+\lambda_Y)}}{K!}\sum_{M=0}^{K}\binom{K}{m}\lambda_x^m\lambda_Y^{K-m}$$

Binomial Thm

$$= \frac{e^{-(\lambda_x+\lambda_Y)}(\lambda_x+\lambda_Y)^K}{K!}$$

While here, note that

$$\psi_X(t) = e^{\lambda_X(e^t-1)} \implies$$

$$\psi_X'(t) = e^{\lambda_X(e^t-1)} \lambda_X e^t \implies E[X] = \psi_X'(0) = \lambda_X$$

$$\psi_X''(t) = e^{\lambda_X(e^t-1)} \lambda_X e^t + \lambda_X e^t \, e^{\lambda_X(e^t-1)} \lambda_X e^t$$

So $E[X^2] = \psi_X''(0) = \lambda_X + \lambda_X^2$

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \lambda_X + \lambda_X^2 - \lambda_X^2 = \lambda_X$$

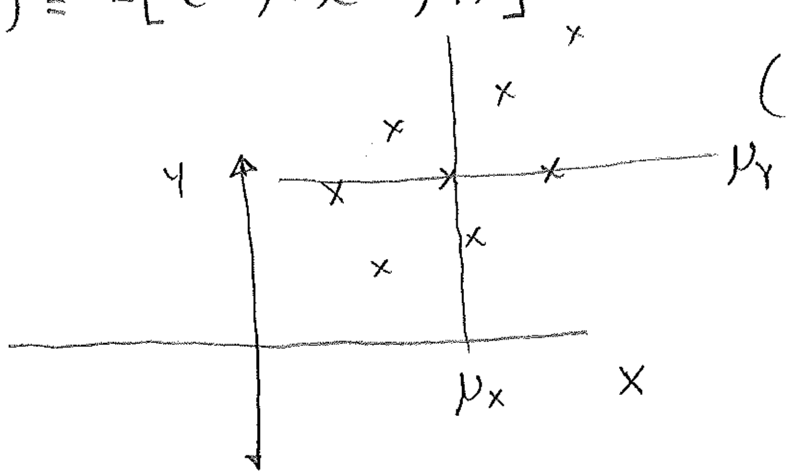
A Poisson variable has unit dispersion D:

$$D = \frac{\sigma_X^2}{\mu_X} = \frac{\lambda_X}{\lambda_X} = 1$$

Covariance & Correlation

Def: $\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$

Linear Association



$(X_i, Y_i)_{i=1}^n$ IID pairs

For this scatterplot, if $X > \mu_X$, Y tends to be $> \mu_Y$
if $X < \mu_X$, Y tends to be $< \mu_Y$

In this case, $E\left[\overset{+}{\underset{-}{(X-\mu_X)}}\overset{+}{\underset{-}{(Y-\mu_Y)}}\right] > 0$

$$
\begin{aligned}
Cov(X,Y) &= E\left[(X-\mu_X)(Y-\mu_Y)\right] \\
&= E\left[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y\right] \\
&= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X\mu_Y \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

Note that if X & Y are independent, $E[XY] = E[X]E[Y]$

$\Rightarrow$ $Cov(X,Y) = 0$

The converse is not true!

Ex

x	-1	0	$+1$
$P[X=x]$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$
\begin{aligned}
E[X] &= \tfrac{1}{3}(-1) + \tfrac{1}{3}(0) + \tfrac{1}{3}(+1) \\
&= 0
\end{aligned}
$$

$Y = X^2$

y	0	1
$P[Y=y]$	$1/3$	$2/3$

$$
\begin{aligned}
E[X^3] &= \tfrac{1}{3}(-1)^3 + \tfrac{1}{3}(0)^3 + \tfrac{1}{3}(+1)^3 \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
Cov(X,Y) &= E[XY] - E[X]E[Y] \\
&= \cancelto{0}{E[X^3]} - \cancelto{0}{E[X]}\,E[Y] = 0
\end{aligned}
$$

So X & Y are uncorrelated

X & Y are dependent

But $P[X=1 \cap Y=1] = P[X=1] = 1/3$

$\overset{?}{=} P[X=1]P[Y=1] = 1/3 \cdot 2/3$ ✗

Def $\quad \mathrm{Corr}(X,Y) = \rho_{X,Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$

Fact: $\quad -1 \le \rho_{X,Y} \le +1$

Why

$$0 \le \mathrm{Var}\left(\frac{X-\mu_X}{\sigma_X} - \left(\frac{Y-\mu_Y}{\sigma_Y} \right) \right)$$

$$= \mathrm{Var}\left(\frac{X-\mu_X}{\sigma_X} \right) + \mathrm{Var}\left(\frac{Y-\mu_Y}{\sigma_Y} \right) - 2\,\mathrm{Cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_X} \right)$$

$$= 1 + 1 - 2\,E\left[\frac{(X-\mu_X)(Y-\mu_Y)}{\sigma_X \sigma_Y} \right] \quad \Rightarrow$$

$$0 \le 2 - 2\rho_{X,Y} \Rightarrow \quad \rho_{X,Y} \le 1$$

Also

$$0 \le \mathrm{Var}\left(\frac{X-\mu_X}{\sigma_X} + \frac{Y-\mu_Y}{\sigma_Y} \right) = \mathrm{Var}\left(\frac{X-\mu_X}{\sigma_X} \right) + \mathrm{Var}\left(\frac{Y-\mu_Y}{\sigma_Y} \right) + 2\,\mathrm{Cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y} \right)$$

$$\Rightarrow \quad 0 \le 2 + 2\rho_{X,Y} \Rightarrow \quad \rho_{X,Y} \ge -1. \quad \blacksquare$$

Properties of Covariance

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
2. $\mathrm{Cov}(X, aY+b) = a\,\mathrm{Cov}(X,Y)$
3. $\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$
4. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
5. $\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i + b_i, \sum_{j=1}^{m} c_j Y_j + d_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i c_j \,\mathrm{Cov}(X_i, Y_j)$

Why (2)

$$\mathrm{Cov}(X, aY+b) = E[X(aY+b)] - E[X]E[aY+b]$$

$$= aE[XY] + bE[X] - E[X](aE[Y] + b)$$

$$= a\,E[XY] - a\,E[X]E[Y] = a\,\mathrm{Cov}(X,Y)$$

③ $$\mathrm{Cov}(X, Y+Z) = E[X(Y+Z)] - E[X]E[Y+Z]$$
$$= E[XY + XZ] - E[X](E[Y] + E[Z])$$
$$= E[XY] - E[X]E[Y] + E[XZ] - E[X]E[Z]$$
$$= \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$$

④ Induction with ① & ③ provide
$$\mathrm{Cov}\left(X, \sum_{j=1}^{m} c_j Y_j + d_j\right) = \sum_{j=1}^{m} c_j \mathrm{Cov}(X, Y_j)$$

Now use this with X replaced by $\sum_{i=1}^{n} a_i X_i + b$ to get
$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i + b, \sum_{j=1}^{m} c_j Y_j + d_j\right) = \sum_{j=1}^{m} c_j \mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i + b, Y_j\right)$$
$$= \sum_{j=1}^{m} c_j \sum_{i=1}^{n} a_i \mathrm{Cov}(X_i, Y_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i c_j \mathrm{Cov}(X_i, Y_j)$$

A classic formula:

$X_1, X_2, \dots, X_n$ RVs $\qquad \bar{X} = \dfrac{X_1 + \dots + X_n}{n}$

$$\mathrm{Var}(\bar{X}) = \mathrm{Cov}(\bar{X}, \bar{X})$$
$$= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{j=1}^{m} X_j\right)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) + \frac{1}{n^2}\sum\sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) + \frac{2}{n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1} Cov(X_i, X_j)$$

$$\begin{array}{ccccc} 1,1 & 1,2 & 1,3 & \cdots & 1,n \\ 2,1 & 2,2 & 2,3 & \cdots & 2,n \\ 3,1 & 3,2 & 3,3 & \cdots & 3,n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n,1 & n,2 & n,3 & \cdots & n,n \end{array}$$

(i,j) pairs from 1 to n

If $X_1, X_2, \ldots, X_n$ are IID with $E[X_i] \equiv \mu_X$ & $Var(X_i) \equiv \sigma_X^2$, $Cov(X_i, X_j) = 0$ and

$$Var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma_X^2}{n}$$

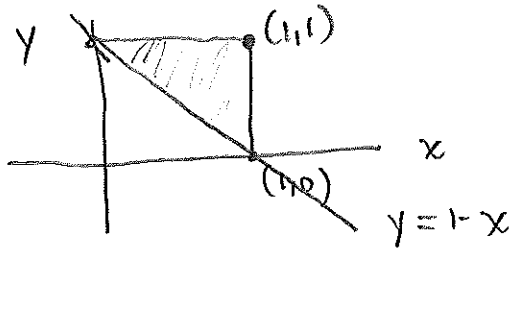
Conditional Expectation

Suppose X & Y are jointly cts.

$$E[X \mid Y=y] \stackrel{def}{=} \int_{-\infty}^{+\infty} x\, f_{X|Y=y}(x)\, dx$$

$$E[Y \mid X=x] \stackrel{def}{=} \int_{-\infty}^{+\infty} y\, f_{Y|X=x}(y)\, dy$$

Ex $\quad f_{X,Y}(x,y) = \frac{3}{2}(x+y) \qquad 0 \le x, y \le 1$, $y > 1-x$



Let's get $E[Y \mid X=x]$

We need $f_{Y|X=x}(y) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \int_{1-x}^{1} \tfrac{3}{2}(x+y)\,dy \qquad 0<x<1$$

$$= \tfrac{3}{2}xy + \tfrac{3}{4}y^2 \Big|_{y=1-x}^{y=1}$$

$$= \tfrac{3}{2}x + \tfrac{3}{4} - \tfrac{3}{2}x(1-x) - \tfrac{3}{4}(1-x)^2$$

$$= \tfrac{3}{4} + \tfrac{3}{2}x^2 - \tfrac{3}{4}(1-2x+x^2)$$

$$= \tfrac{3}{4}x^2 + \tfrac{3}{2}x \quad , \qquad 0<x<1$$

So $f_{Y|X=x}(y) = f_{X,Y}(x,y) / f_X(x)$

$$= \tfrac{3}{2}(x+y) \Big/ \tfrac{3}{2}x\left(\tfrac{x}{2}+1\right) = \frac{x+y}{x\left(\frac{x}{2}+1\right)}, \quad 1-x \le y \le 1$$

Thus, $$E[Y|X=x] = \int_{1-x}^{1} \frac{y(x+y)}{x\left(\frac{x}{2}+1\right)}\,dy$$

$$= \frac{\left(x\,y^2/2 + y^3/3 \Big|_{y=1-x}^{y=1}\right)}{x\left(x/2+1\right)}$$

$$= \frac{\left(x+\frac{1}{3}\right) - \left(\frac{x(1-x)^2}{2} + \frac{(1-x)^3}{3}\right)}{x\left(x/2+1\right)}$$ For You to Simplify

If $x=1$ $\quad E[Y|X=1] = \dfrac{4/3}{3/2} = 8/9$

Note that $E[Y|X=x]$ is a function of x only.

Suppose $E[X|Y=y] = h(y)$ for some function h

Define $E[X|Y] = h(Y)$ This is a RV

Law of Total Expectation $E[X] = E[\,E[X|Y]\,]$

Proof

Cts case

Let $h(y) = E[X|Y=y]$

$$
\begin{aligned}
E[\,E[X|Y]\,] &= E[\,h(Y)\,] \\
&= \int_{-\infty}^{+\infty} h(y) f_Y(y)\,dy \\
&= \int_{-\infty}^{+\infty} E[X|Y=y]\, f_Y(y)\,dy \\
&= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} \frac{f_{X,Y}(x,y)}{f_Y(y)}\,dx \right] f_Y(y)\,dy \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy = E[X]
\end{aligned}
$$

Ex Toss a fair L-sided die and then flip a fair coin the # of times shown on the die. Let X = # heads

D = die toss $\in \{1, 2, \ldots, L\}$
Discrete Uniform

What is $E[X]$?

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

$X \mid D=d \sim \text{Bin}(d, p=\tfrac{1}{2})$

So $E[X \mid D=d] = d/2$

$E[X \mid D] = D/2$

$$E[X] = E\big[E[X \mid D]\big] = E[D/2] = \tfrac{1}{2}E[D]$$

$$E[D] = \sum_{l=1}^{L} l\, P[D=l] = \sum_{l=1}^{L} l \frac{1}{L} = \frac{1}{L}\frac{L(L+1)}{2}$$

$$= \frac{L+1}{2}$$

So $E[X] = \dfrac{L+1}{4}$

Suppose we have a joint RV pair (X,Y) and we observe X.

We want to predict Y from X - Say $\hat{Y} = h(X)$

What function h should we use?

We will minimize the mean squared error $E[(Y-\hat{Y})^2]$

$$E[(Y-\hat{Y})^2] = E\big[(Y - E[Y|X] + E[Y|X] - \hat{Y})^2\big]$$

$$= E\big[(Y-E[Y|X])^2\big] + E\big[(E[Y|X]-\hat{Y})^2\big]$$

$$+ 2E\big[(Y-E[Y|X])(E[Y|X]-\hat{Y})\big]$$

$$= ① + ② + ③$$

Now 1 does not depend on choice of h
2 $= 0$ if we choose $\hat{Y} = E[Y|X]$

Claim: 3 $= 0$

Why

For $\hat{Y}$ = any function of X,

Now $E\Big[(Y - E[Y|X])(E[Y|X] - \hat{Y}) \Big] =$

$E\Big[E\big[(Y - E[Y|X])(E[Y|X] - \hat{Y}) \mid X \big] \Big] =$

$E\Big[(E[Y|X] - \hat{Y})\, E\big[(Y - E[Y|X]) \mid X \big] \Big] =$

$E\Big[(E[Y|X] - \hat{Y}) \cdot 0 \Big] = 0$

We have used $E[\, g(x)\, Z \mid X] = g(x)\, E[Z|X]$

So we get the best squared error by choosing $\hat{Y} = E[Y|X]$

∎

Conditional expectations obey all the laws of regular expectation

For example, $E[X+Y|Z] = E[X|Z] + E[Y|Z]$

Many times, conditional expectations are hard to compute

Often, folks will examine a linear predictor:

$$\hat{Y} = aX + b$$

How to choose a & b? Minimize $E\left[(Y-\hat{Y})^2\right] = E\left[(Y-(aX+b))^2\right]$

$$= E\left[Y^2 - 2Y(aX+b) + (aX+b)^2\right]$$

$$= E\left[Y^2\right] - 2a\,E[XY] - 2b\,E[Y] + a^2 E\left[X^2\right] + 2ab\,E[X] + b^2$$

$$= g(a,b)$$

$$\frac{\partial g(a,b)}{\partial a} \overset{\text{set}}{=} 0 \implies -2E[XY] + 2aE\left[X^2\right] + 2bE[X] = 0$$

$$\frac{\partial g(a,b)}{\partial b} \overset{\text{set}}{=} 0 \implies -2E[Y] + 2a\,E[X] + 2b = 0$$

$$a\,E\left[X^2\right] + b\,E[X] = E[XY] \qquad (1)$$

$$a\,E[X] + b = E[Y] \qquad (2)$$

$$b = E[Y] - aE[X] \implies a\,E\left[X^2\right] + \left(E[Y] - a\,E[X]\right)E[X] = E[XY]$$

$$\implies a\left(E\left[X^2\right] - E[X]^2\right) = E[XY] - E[X]E[Y]$$

$$\implies a = \mathrm{Cov}(X,Y) / \mathrm{Var}(X)$$

$$b = E[Y] - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} E[X]$$

$$\hat{Y} = \frac{Cov(X,Y)}{Var(X)} X + E[Y] - \frac{Cov(X,Y)}{Var(X)} E[X]$$

$$= E[Y] + \frac{Cov(X,Y)}{Var(X)} \left(X - E[X] \right)$$

Linear Predictors may be much worse than conditional expectation predictions

Ex $\quad X \sim N(0,1) \qquad Y = X^2 \qquad$ Predict Y from X

Conditional mean $\quad \hat{Y} = E[Y|X] = E[X^2 | X] = X^2 = Y$

Perfect prediction — no error.

Linear Prediction

$$Cov(X,Y) = E[XY] - \cancelto{0}{E[X]} E[Y]$$

$$= E[X^3] = \int_{-\infty}^{+\infty} x^3 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 0$$

So $\hat{Y} = E[Y]$

There is no information for Y from X in a linear prediction