example ① If $X \sim Bin(1,p) =$

$= Bernoulli(p)$

$$Var(X) = E\left(\underbrace{(X-p)^2}\right)$$

$$= (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p)$$

$$= p(1-p)\left[1-p+p\right] = p(1-p) \quad //$$

---

② $X \sim Unif\ [0,1]$

$$\mu = EX = \int_0^1 x \cdot 1\, dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$$

$$Var(X) = E\left((X-\mu)^2\right)$$

$$= \int_0^1 \left(x-\frac{1}{2}\right)^2 \cdot 1\, dx = \frac{1}{12}$$

or $\quad EX^2 = \int_0^1 x^2\, dx = \frac{1}{3}$

$$Var(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} \quad //$$

Def If X and Y are jointly distributed with means μ_X and μ_Y then their covariance is

$$cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$\mathrm{cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right].$$

<u>Note</u> $\quad \mathrm{cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$

$$= E\left[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y\right]$$

$$= E[XY] - \mu_X \underline{EY} - \mu_Y \underline{EX} + \underline{\mu_X\mu_Y}$$

$$= E[XY] - \mu_X\mu_Y$$

$$= E[XY] - EX\,EY. \quad //$$

<u>example</u> Suppose X, Y have joint pdf

$$f(x,y) = 2x + 2y - 4xy, \quad \begin{matrix} 0 \le x \le 1 \\ 0 \le y \le 1 \end{matrix}$$

Then

$$E(XY) = \int_0^1\int_0^1 (2x+2y-4xy)\cdot xy\,dx\,dy$$

$$= \dots = 2/9$$

$$E(X) = \int_0^1 x f_X(x)\,dx =$$

$$= \int_0^1 x \int_0^1 (2x+2y-4xy)\,dy\,dx$$

$$= \dots = \frac{1}{2}$$

$$E(Y) = \frac{1}{2}$$

$$\mathrm{cov}(X, Y) = 2/9 - \tfrac{1}{2} \cdot \tfrac{1}{2} = -\tfrac{1}{36}.$$

<u>Note</u>: $\mathrm{cov}(a + X, Y) = \mathrm{cov}(X, Y)$

$$\mathrm{cov}(aX, bY) = ab\,\mathrm{cov}(X, Y)$$

$$\mathrm{cov}(X, Y + Z) = \mathrm{cov}(X, Y) + \mathrm{cov}(X, Z)$$

Generally,

$$\mathrm{cov}(aX + bY + c,\ dU + eV + f) =$$

$$= ad\,\mathrm{cov}(X, U) + ae\,\mathrm{cov}(X, V) +$$

$$+ bd\,\mathrm{cov}(Y, U) + be\,\mathrm{cov}(Y, V)$$

$$\mathrm{cov}\left(a + \sum_{i=1}^{n} b_i X_i,\ c + \sum_{j=1}^{m} d_j Y_j\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j\,\mathrm{cov}(X_i, Y_j)$$

$$\mathrm{Var}(X) = E(X \cdot X) - (EX)(EX)$$

$$= \mathrm{cov}(X, X)$$

$$\mathrm{Var}(X + Y) = \mathrm{cov}(X + Y, X + Y)$$

$$= \mathrm{cov}(X, X) + 2\,\mathrm{cov}(X, Y) +$$

$$+ \mathrm{cov}(Y, Y)$$

$$= \mathrm{Var}(X) + 2\,\mathrm{cov}(X, Y) + \mathrm{Var}(Y).$$

$$Var\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \, cov(X_i, X_j)$$

If $a = 0, \; b_i = 1$

$$Var\left(\sum_{1}^{n} X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} Var(X_i) + 2 \underbrace{\sum\sum_{1 \le i < j \le n} cov(X_i, X_j)}_{\binom{n}{2} \text{ terms}}$$

If $X \perp\!\!\!\perp Y$ then

$$cov(X, Y) = E(XY) - EXEY$$

$$= EXEY - EXEY = 0$$

If $X \perp\!\!\!\perp Y \Rightarrow cov(X,Y) = 0$. (uncorrelated).

<u>example</u> (uncorrelated $\not\Rightarrow$ indepence)

$$X : \left.\begin{matrix} 3 \\ 1 \\ -1 \\ -3 \end{matrix}\right\} \text{ w/ prob. } 1/4$$

So $EX = 0$ and $EX^3 = 0$

Let $Y = X^2$

$$cov(X, Y) = cov(X, X^2) = E(X \cdot X^2) - $$

$$- EXEX^2 = 0.$$

But $X \not\perp\!\!\!\perp Y$ //

But $X \not\perp Y$. //

<u>example</u> $X \sim Bin(n, p)$

$Var(X) = ?$

We could $Var(X) = EX^2 - (np)^2 \ldots$

Instead, $X_i = \begin{cases} 1, & \text{if trial } \underline{i} \text{ is a success} \\ 0, & \text{o.w.} \end{cases}$

$$X = \sum_{i=1}^{n} X_i$$

$$Cov(X_i, X_j) = 0 \quad \text{for } i \neq j$$

$$Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

$X_i \sim Bernoulli(p)$, $Var(X_i) = p(1-p)$

$$\Rightarrow Var(X) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$ //

$X \sim Bin(n, p) \Rightarrow EX = np$, $Var(X) = np(1-p)$.

Def The <u>correlation coefficient of X and Y</u> is

$$\rho_{XY} = \frac{Cov(X, Y)}{SD(X)\, SD(Y)} = corr(X, Y)$$

(μ, σ, ρ)
mean, SD, correlation, often.

<u>Remark</u> Covariance can be <u>any</u> real number, the correlation coefficient is always $[-1, 1]$.

Let $\sigma_X = SD(X)$, $\sigma_Y = SD(Y)$

$$0 \le Var\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y} \right)$$

$$= Var\left(\frac{X}{\sigma_X}\right) + Var\left(\frac{Y}{\sigma_Y}\right) \pm 2\,cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$

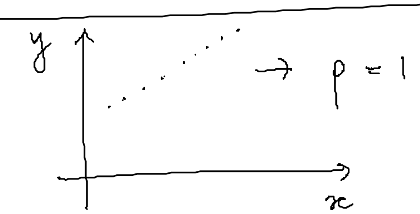
$$= \frac{1}{\sigma_X^2} Var(X) + \frac{1}{\sigma_Y^2} Var(Y) \pm \frac{2}{\sigma_X \sigma_Y} cov(X, Y)$$

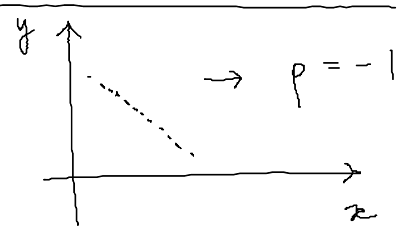
$$= 1 + 1 \pm \frac{2\,cov(X,Y)}{\sigma_X \sigma_Y} = 2 \pm 2\rho_{XY}$$

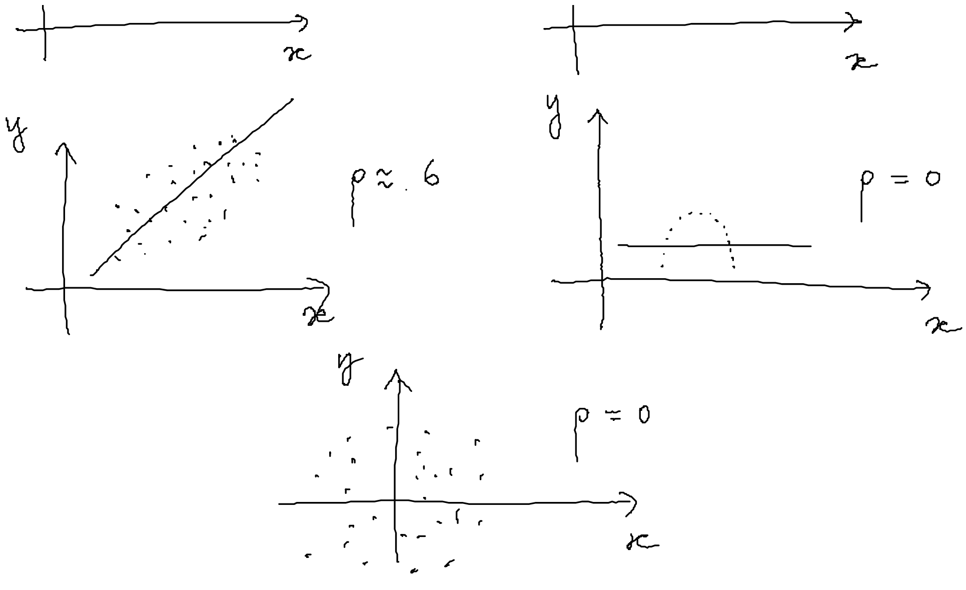
$$\Rightarrow \quad 2 + 2\rho_{XY} \geq 0 \quad \Rightarrow \quad \rho_{XY} \geq -1$$

$$2 - 2\rho_{XY} \geq 0 \quad \Rightarrow \quad \rho_{XY} \le 1.$$

$$|\rho_{XY}| \le 1.$$







//

Moments of Distribution.

Def For any r.v. X, the expectation $E(X^k)$, $k \in \mathbb{N}$, is called the k^{th} moment of X (or its distribution).

EX is just 1^{st} moment of X.

Def. For any r.v. X, the expectation

$$E\left((X-\mu)^k \right), \quad k \in \mathbb{N}, \quad \mu = EX$$

is called the k^{th} central moment of X (or its dist'n).

$Var(X)$ – 2^{nd} central moment.

$Var(X)$

Theorem If $E|X|^k < \infty$ for $k \in \mathbb{N}$ then $E|X|^j < \infty$ for all $j \le k$.

(try to find densities s.t. EX, EX^2, EX^3 exist but not EX^4, say).

Moment Generating Function.

Def For $t \in \mathbb{R}$, let $\psi(t) = E(e^{tX})$, $\psi(t)$ is the moment generating function (mgf) of X.

Suppose that $\psi(t)$ exists. Then

$$\psi'(0) = \frac{d}{dt}\left[Ee^{tX}\right]\bigg|_{t=0}$$

$$= E\left[\frac{d}{dt} e^{tX}\right]\bigg|_{t=0}$$

$$= E\left[Xe^{tX}\right]\bigg|_{t=0} = E[X]$$

In fact, (try this!)

$$\psi^{(n)}(0) = EX^n$$

This explains the name "mgf".

Just like cdf, pdf, pmf, the mgf uniquely discribes distributions.

<u>example</u> 1) $X \sim \text{Bin}(n, p)$

$$\psi(t) = E(e^{tX})$$

$$= \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} (\underbrace{pe^t}_{a})^x (\underbrace{1-p}_{b})^{n-x} \cdot \binom{n}{x}$$

$$= (pe^t + 1 - p)^n, \quad t \in \mathbb{R}.$$

$$\psi'(t)\Big|_{t=0} = n(pe^t + 1-p)^{n-1} \cdot pe^t \Big|_{t=0}$$

$$= np$$

$$\psi''(t)\Big|_{t=0} = n(n-1)(pe^t + 1-p)^{n-2}(pe^t)^2$$

$$+ n(pe^t + 1-p)^{n-1} \cdot pe^t \Big|_{t=0}$$

$$= n(n-1)p^2 + np = EX^2$$

$$\text{Var}(X) = n(n-1)p^2 + np - (np)^2 =$$

$$= np - np^2 = np(1-p) \quad //$$

Sum of independent random variables

Let $X_1, X_2, \ldots, X_n$ be independent r.v.'s

and $Y = \sum_{i=1}^{n} X_i$ then

mgf
$$\psi_Y(t) = E\left(e^{tY}\right)$$
$$= E\left(e^{t\sum_1^n X_i}\right)$$
$$= E\left(\prod_1^n e^{tX_i}\right)$$
$$\overset{\text{indep.}}{=} \prod_1^n Ee^{tX_i} = \prod_{i=1}^n \psi_{X_i}(t).$$

If X_i's are iid (same marginals)

then $\psi_Y(t) = \left(\psi_X(t)\right)^n.$

$$EY = \psi_Y'(t)\Big|_{t=0} = \left[\left(\psi_X(t)\right)^n\right]'\Big|_{t=0}$$
$$= n\,\underbrace{\psi_X^{n-1}(t)}\cdot \psi_X'(t)\Big|_{t=0}$$
$$= nEX$$

$$EY^2 = \psi_Y''(t)\Big|_{t=0} = n(n-1)\cancelto{1}{\psi_X^{n-2}(t)}\left(\psi_X'(t)\right)^2$$
$$+\; n\,\cancelto{1}{\psi_X^{n-1}(t)}\cdot\psi_X''(t)\Big|_{t=0}$$

$$= n(n-1)\left(EX\right)^2 + nE(X^2) = EY^2$$

$$Var(Y) = n(n-1)(EX)^2 + \underbrace{n\,E(X^2)} - (n\,EX)^2$$

$$= n\,E(X^2) - n(EX)^2 = n\,Var\,X$$

as expected.

<u>example</u> $X \sim Bin(n,p)$

$$X = \sum_{i=1}^{n} X_i, \quad X_i = \begin{cases} 1, & \text{if } i^{th} \text{ is success} \\ 0, & \text{o.w.} \end{cases}$$

$$\psi_{X_i}(t) = E e^{tX_i} = e^0(1-p) + e^t \cdot p$$

$$= pe^t + 1 - p, \quad t \in \mathbb{R}$$

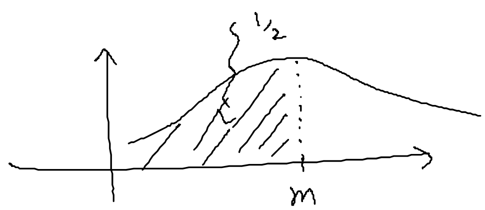
$$\psi_X(t) = \left(\psi_{X_i}(t)\right)^n = \left(1-p+pe^t\right)^n, \quad t \in \mathbb{R}$$

same as before. //

One more important measure of "center" of a distribution

<u>Def</u> <u>Median</u> of distribution of X is a number m s.t.

$$P(X \le m) \ge \frac{1}{2} \quad \underline{\text{and}} \quad P(X \ge m) \ge \frac{1}{2}.$$



<u>example</u> (a) $X \sim Unif\,[a,b]$

example (a) $X \sim \text{Unif}\,[a, b]$

$$\frac{1}{2} = \int_a^m \frac{1}{b-a}\,dx = \frac{m-a}{b-a}$$

$$m = a + \frac{1}{2}(b-a) = \frac{a+b}{2}$$ — midpoint of $[a, b]$.

$$EX = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2} = m$$

Here, $\mu = m$, but not always!

(b) X has density

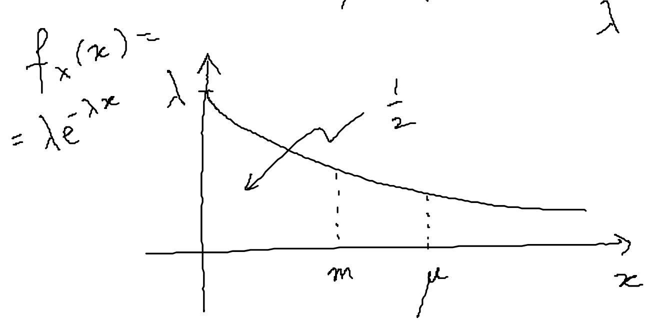
$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

Before: $\mu = EX = \frac{1}{\lambda}$.

$$\frac{1}{2} = \int_0^m \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^m$$

$$= 1 - e^{-\lambda m}$$

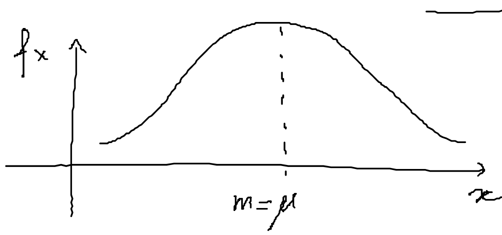
$$\Rightarrow m = \frac{\ln 2}{\lambda} < \mu = \frac{\ln e}{\lambda}$$



long right tail $\Rightarrow$

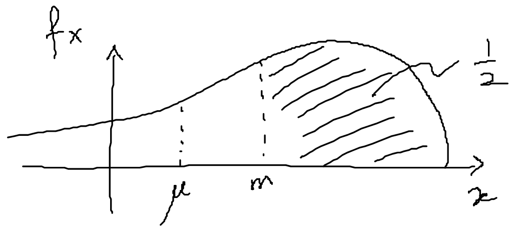
$\Rightarrow$ skewed to the right

$\Rightarrow m < \mu.$



Symmetric distribution

$m = \mu$



long left tail ⟹

⟹ skewed to the left

⟹ $\mu < m$. //

In continuous dist'ns, find median by solving

$$\int_{-\infty}^{m} f(x) = \frac{1}{2} = \int_{m}^{\infty} f(x)\,dx.$$

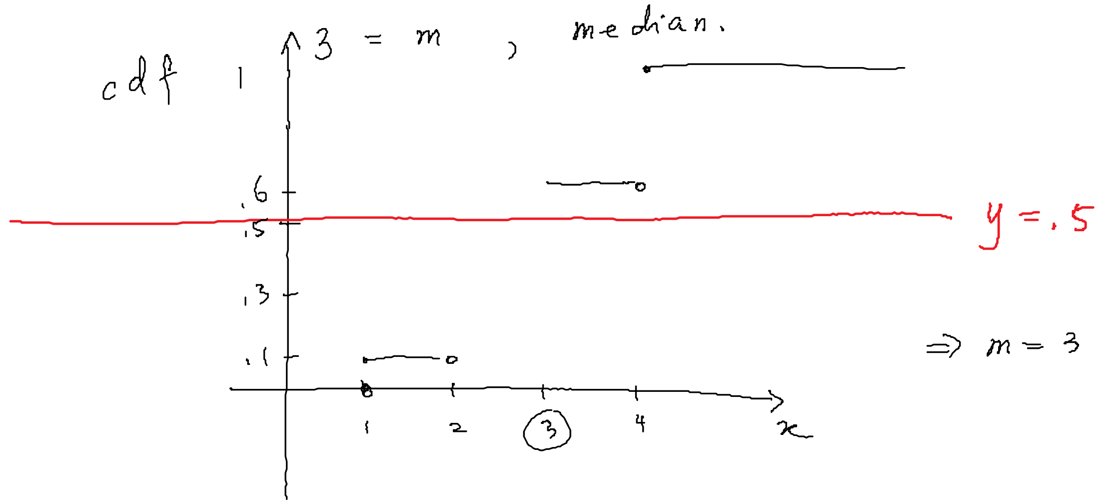
example (Median of a discrete dist'n).

x	1	2	3	4
$p_X(x)$	.1	.2	.3	.4

$$P(X \le m) \ge \frac{1}{2} \quad \& \quad P(X \ge m) \ge \frac{1}{2}$$

$$P(X \le 3) = .6 > \frac{1}{2} \quad \& \quad P(X \ge 3) = .7 > \frac{1}{2}$$

3 = m, median.



⟹ m = 3

Consider a slight modification

Consider a slight modification

x	1	2	3	4
$p_X(x)$	.2	.3	.4	.1

$$P(X \le 3) = .9 > \tfrac{1}{2} \quad \& \quad P(X \ge 3) = .5 \ge \tfrac{1}{2}$$

but also

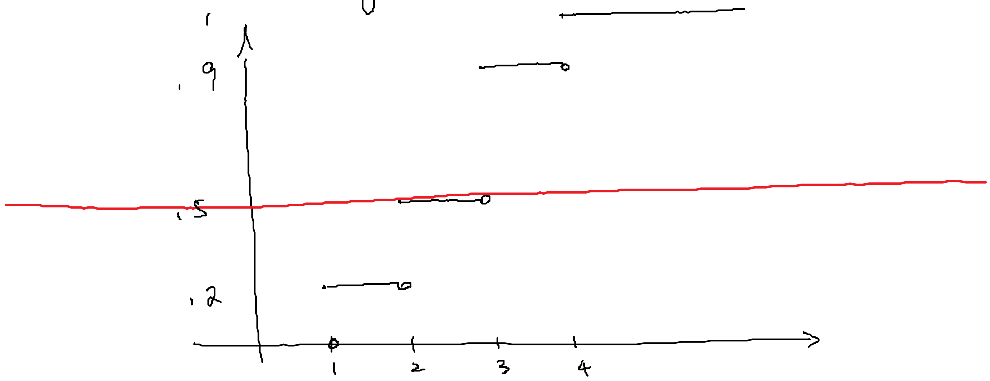
$$P(X \le 2) = .5 \ge \tfrac{1}{2} \quad \& \quad P(X \ge 2) = .8 > \tfrac{1}{2}$$

$\Rightarrow$ both 2 & 3 can be medians, in fact

for $a \in (2, 3)$

$$P(X \le a) = \tfrac{1}{2} = P(X \ge a)$$

$\Rightarrow$ any $a \in \underline{[2, 3]}$ is a median.



median is not unique in some discrete dist'ns.

If you have an interval satisfying definition for median, take the midpoint,

$$m = \frac{2+3}{2} = 2.5.$$

//

Suppose X is s.t. $\mu = EX$, $\sigma^2 = \mathrm{Var}\, X$

Let's find a scalar d s.t.

$\underbrace{E\left((X-d)^2\right)}$ is minimised.

MSE = mean-squared error of d as an estimator of X.

$$f(d) = E\left((X-d)^2\right) = EX^2 - 2dEX + d^2$$

$$f'(d) = -2\mu + 2d \Rightarrow d = \mu$$

$\Rightarrow$ $d = \mu = EX$ is the minimizer of the MSE.

$$\underbrace{MSE(d)} = E\left((X-d)^2\right) =$$

$$= E\left((X-\mu+\mu-d)^2\right)$$

$$= \underbrace{E\left((X-\mu)^2\right)} + 2(\mu-d)\overbrace{E\left(\cancel{X-\mu}\right)}^{0} +$$

$$+ (\mu-d)^2$$

$$= \underbrace{MSE(\mu)} + (\mu-d)^2 =$$

$$= \underbrace{\mathrm{Var}\, X} + \underbrace{(\mu-d)^2}_{(\text{bias})^2}$$

(bias = μ − expected value of your estimator) //

Median minimizes the absolute value error

$E|X-d|$, so

d that minimizes this is the median.

(proof: textbook).

Conditional Expection.

Suppose X, Y are jointly continuously dist'd and $g(y|x)$ is the conditional density of $Y | X = x$. Then

$E(Y|X=x)$ is the mean of this conditional distribution.

non-random

$$\varphi(x) = E(Y/X=x) = \int_{-\infty}^{\infty} y\, g(y|x)\, dy$$

- function of x!

Then we define $E(Y|X) = \varphi(X)$

- a random variable! We will it the conditional expectation of Y given X.

example $P \sim$ Unif $[0,1]$

<u>example</u> $P \sim Unif\,[0,1]$

$X \mid P = p \sim Bin\,(n, p)$

Then

$$\varphi(p) = E(X \mid P = p) = np$$

So $E(X \mid P) = \varphi(P) = nP$ – random!

$$E(X \mid P) \sim Unif\,[0, n].$$

$$E\left[E(X \mid P)\right] = \frac{n}{2}.$$

Recall, $X \sim Unif\,\{0, 1, \ldots, n\}$

$$EX = \sum_{x=0}^{n} x \cdot \frac{1}{n+1} = \frac{1}{n+1} \cdot \frac{n(n+1)}{2} = \frac{n}{2} =$$

$$= E\big(E(X \mid P)\big)$$

– this is not a coincidence!

<u>Theorem</u> Suppose that Y is s.t. EY exists then

$$E\big(E(Y \mid X)\big) = EY$$

In continuous case, $\varphi(x) = E(Y \mid X = x)$

$$E\big(E(Y \mid X)\big) = E\,\varphi(X) =$$

$$= \int_{\mathbb{R}} \underbrace{\varphi(x)}_{} f_X(x)\,dx =$$

$\varphi(x)$

$$= \int_{\mathbb{R}} \varphi(x) f_X(x)\,dx$$

$$= \int_{\mathbb{R}} \Big[\overbrace{\int_{\mathbb{R}} y\, f_{Y|X}(y|x)\,dy}^{\varphi(x)} \Big] f_X(x)\,dx$$

$$= \iint_{\mathbb{R}\,\mathbb{R}} y\, f_{Y|X}(y|x)\, f_X(x)\,dx\,dy$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} y\, f(x,y)\,dx\,dy$$

$$= \int_{\mathbb{R}} y\, f_Y(y)\,dy = EY. \quad //$$

In the same manner, we define the conditional variance

$$\mathrm{Var}(Y|X) = \eta(X), \text{ where}$$

$$\eta(x) = \mathrm{Var}(Y|X=x)$$

$$\boxed{\begin{aligned} &E(E(Y/X)) = EY \\ &\underbrace{\mathrm{Var}(E(Y|X)) + E(\mathrm{Var}(Y|X))}_{\text{analysis of variance}} = \mathrm{Var}\,Y \end{aligned}}$$

<u>example</u> $P \sim \mathrm{Unif}\,[0,1]$

$X|P=p \sim \mathrm{Bin}(n,p)$

$$\begin{aligned}
Var(X) &= Var(E(X \mid P)) + E(Var(X \mid P)) \\
&= Var(nP) + E(nP(1-P)) \\
&= n^2 \cdot Var P + n EP - n EP^2 \quad \ominus
\end{aligned}$$

$P \sim Unif\,[0,1]$, $EP = \frac{1}{2}$, $Var P = \frac{1}{12}$

$$\begin{aligned}
\ominus\ & n^2 \cdot \frac{1}{12} + n \cdot \frac{1}{2} - n\left(\frac{1}{12} + \left(\frac{1}{2}\right)^2\right) \\
&= \frac{n^2+2n}{12} = Var(X). \qquad //
\end{aligned}$$

Markov & Chebyshev Inequalities.

Markov Suppose X is a r.v. s.t.

$EX < \infty$ and $P(X \geq 0) = 1$.

Then for $t > 0$

$$P(X \geq t) \leq \frac{EX}{t}.$$

$\triangleright$

$$\begin{aligned}
EX &= \int_{\mathbb{R}} x f(x)\,dx = \int_0^\infty x f(x)\,dx \\
&\geq \int_t^\infty x f(x)\,dx \geq t \int_t^\infty f(x)\,dx \\
&= t P(X \geq t). \qquad \square
\end{aligned}$$

Let $X = (Y - EY)^2 \geq 0$, assume
$Var(Y) < \infty$. $EX = Var(Y)$ & $P(X \geq 0) = 1$.

Fix $t = s^2 > 0$ then by Markov Ineq.

$$P(X \geq t) \leq \frac{EX}{t} = \frac{Var(Y)}{s^2}$$
$$\parallel$$
$$P((Y-EY)^2 \geq t) = P(|Y-EY| \geq s),$$

So we get Chebyshev Ineq.

$$P(|Y-EY| \geq s) \leq \frac{Var\, Y}{s^2}$$
$(Var\, Y < \infty,\ s > 0)$.

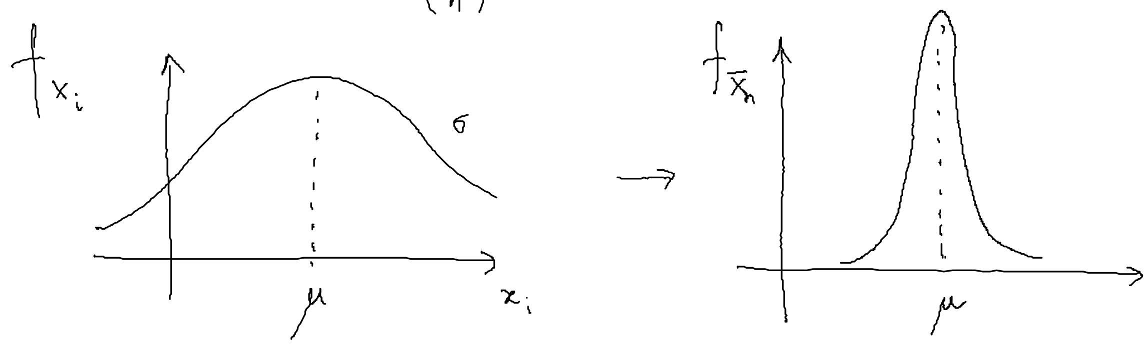
<u>example</u> Suppose that $X_1, \dots, X_n$ are iid with mean μ, variance σ^2.

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$ – sample average = sample mean.

$$E(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu,$$ same as individual X_i's.

$$Var(\bar{X}_n) = \frac{1}{n^2} \cdot Var\left(\sum_1^n X_i\right) =$$
$$= \left(\frac{1}{n}\right)^2 \cdot n\sigma^2 = \sigma^2/n.$$

$$= \left(\frac{1}{n}\right)^2 \cdot n\sigma^2 = \sigma^2/n .$$



Fix $s > 0$, then

$$0 \le P(|\bar{X}_n - \mu| \ge s) \le \frac{Var\,\bar{X}_n}{s^2} =$$

$$= \frac{\sigma^2}{ns^2} \longrightarrow 0$$

$$\Rightarrow \quad \lim_{n\to\infty} P(|\bar{X}_n - \mu| \ge s) = 0.$$

Convergence in Probability

Def $Z_1, Z_2, Z_3 \ldots$ converges to b in probability if for every $\varepsilon > 0$

$$\lim_{n\to\infty} P(|Z_n - b| > \varepsilon) = 0$$

we write $Z_n \xrightarrow[p]{} b$ as $n \to \infty$.

example $\bar{X}_n \xrightarrow[p]{} \mu$ as $n \to \infty$
$(\mu, \sigma^2 < \infty)$.

Weak Law of Large Numbers.

0