Lec 8 MATH 340/640

let $\vec{X}$ be a vector rv with k-dimension

$$\vec{\mu} := \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_k] \end{bmatrix}$$

e.g. if $\vec{X} \sim \text{Multinom}(n, \vec{p})$

$\Rightarrow X_i \sim \text{Bin}(n, p_i) \Rightarrow \vec{\mu} = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix} = n\vec{p}$

If M is a matrix of rv's of dim $m \times n$

$$E[M] := \begin{bmatrix} E[M_{1,1}] & \cdots & E[M_{1,n}] \\ \vdots & & \\ E[M_{m,1}] & \cdots & E[M_{m,n}] \end{bmatrix}$$

$\vec{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$

$$\text{Var}[\vec{X}] := E[\underbrace{\vec{X}\vec{X}^T}_{\text{outer product size } n \times n}] - E[\vec{X}]\, E[\vec{X}^T]$$

$$= E\begin{bmatrix} X_1X_1 & \cdots & X_1X_n \\ \vdots & & \\ X_nX_1 & \cdots & X_nX_n \end{bmatrix} - E\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} E[X_1 \ldots X_n]$$

$$= \begin{bmatrix} E[X_1X_1] - E[X_1]E[X_1] & E[X_1X_2] - E[X_1]E[X_2] & \cdots & E[X_1X_n] - E[X_1]E[X_n] \\ E[X_2X_1] - E[X_2]E[X_1] & E[X_2X_2] - E[X_2]E[X_2] & \cdots & E[X_2X_n] - E[X_2]E[X_n] \\ \vdots & & & \\ E[X_nX_1] - E[X_n]E[X_1] & E[X_nX_2] - E[X_n]E[X_2] & \cdots & E[X_nX_n] - E[X_n]E[X_n] \end{bmatrix}$$

$$= \{ Cov[X_i, X_j] \} = \begin{bmatrix} Var[X_1] & Cov[X_1,X_2] & \cdots & Cov[X_1,X_n] \\ Cov[X_1,X_2] & Var[X_2] & & \\ \vdots & & \ddots & \\ Cov[X_1,X_n] & & & Var[X_n] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & & \\ \vdots & & \ddots & \\ \sigma_{1n} & & & \sigma_n^2 \end{bmatrix}$$

if $X_1, \ldots X_n \overset{iid}{\sim}$

$$= \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix} = diag[\sigma_1^2, \ldots \sigma_n^2]$$

diagonal matrix

$\vec{X} \sim \text{Multi}(n, \vec{p})$, $\text{Cov}[X_i, X_j] = ?$

Back to main problem...

We know $X_i \sim \text{Bin}(n, p_i)$
$X_j \sim \text{Bin}(n, p_j)$

remember X_i, X_j are <u>not</u> independent

Binomials are equivalent to

$X_{i,1}, X_{i,2}, \ldots, X_{i,n} \overset{iid}{\sim} \text{Bern}(p_i)$

$X_{j,1}, X_{j,2}, \ldots, X_{j,n} \overset{iid}{\sim} \text{Bern}(p_j)$

Which are independent??

$X_{i,l}$ and $X_{i,m}$ if $l \neq m$

$X_{j,l}$ and $X_{j,m}$ if $l \neq m$

$X_{i,l}$ and $X_{j,m}$ if $l \neq m$

why? They are different draws from the bag with replacement!

$$\text{Cov}[X_i, X_j] = \text{Cov}[X_{i,1} + \ldots + X_{i,n}, X_{j,1} + \ldots + X_{j,n}]$$

$$= \sum_{l=1}^{n} \sum_{m=1}^{n} \text{Cov}[X_{i,l}, X_{j,m}]$$

generalization of $\text{Cov}[Y_1 + Y_2, Y_3] = \text{Cov}[Y_1, Y_3] + \text{Cov}[Y_2, Y_3]$

due to all the independences

$$= \sum_{l=1}^{n} \text{Cov}[X_{i,l}, X_{j,l}]$$

$$= \sum_{l=1}^{n} E[X_{i,l} X_{j,l}] - \underbrace{\mu_{i,l}}_{p_i} \underbrace{\mu_{j,l}}_{p_j}$$

$$E[X_{i,\ell}, X_{j,\ell}] = \sum_{x_{i,\ell} \in S_{X_{i,\ell}}} \sum_{x_{j,\ell} \in S_{X_{j,\ell}}} x_{i,\ell}\, x_{j,\ell}\, p_{X_{i,\ell}, X_{j,\ell}}(x_{i,\ell}, x_{j,\ell})$$

Bern(p_i) Bern(p_j)

$$= \sum_{x_{i,\ell} \in \{0,1\}} \sum_{x_{j,\ell} \in \{0,1\}} x_{i,\ell}\, x_{j,\ell}\, p_{X_{i,\ell}, X_{j,\ell}}(x_{i,\ell}, x_{j,\ell})$$

$$= p_{X_{i,\ell}, X_{j,\ell}}(1,1)$$

$= 0$ since you can't get a banana AND an apple simultaneously

$$\Rightarrow \mathrm{Cov}(X_i, X_j) = \sum_{\ell=1}^{n} -p_i p_j = -n p_i p_j$$

$X_i \uparrow \ominus X_j \downarrow$

$$\Rightarrow \mathrm{Var}[\vec{X}] = n \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_n \\ -p_1 p_2 & p_2(1-p_2) & & \\ \vdots & & \ddots & \\ -p_1 p_n & & & p_K(1-p_K) \end{bmatrix}$$

One more fact about the multinomial ....

$\vec{X} \sim \text{Multinom}(n, \vec{p})$

What if I tell you $X_j = 5$. What is the JMF of all the other $K-1$ rv's

known value

$$P(\vec{X}_{-j} = \vec{x}_{-j} \mid X_j = c) \leftarrow \text{Conditional JMF}$$

$$= \frac{P(\vec{X} = \vec{x})}{P(X_j = c)} = \frac{\text{Multinom}(n, \vec{p})}{\text{Binom}(n, p_j)} = \frac{\binom{n}{x_1, \ldots, x_K} p_1^{x_1} \cdots p_K^{x_K}}{\binom{n}{x_j} p_j^{x_j} (1-p_j)^{n-x_j}}$$

$$= \frac{\frac{\cancel{n!}}{x_1! \cdots x_{j-1}! \cancel{x_j!} x_{j+1}! \cdots x_K!} p_1^{x_1} \cdots p_{j-1}^{x_{j-1}} \cancel{p_j^{x_j}} \cdots p_K^{x_K} \mathbb{1}_{n \in \mathbb{N}_0} \mathbb{1}_{x_1 + \ldots + x_K = n} \prod_{\ell=1}^{K} \mathbb{1}_{x_\ell \in \{0, \ldots, n\}}}{\frac{\cancel{n!}}{\cancel{x_j!} (n-x_j)!} \cancel{p_j^{x_j}} (1-p_j)^{n-x_j} \mathbb{1}_{n \in \mathbb{N}_0} \mathbb{1}_{x_j \in \{0, \ldots, n\}}}$$

Define $U_a := \frac{\mathbb{1}_a}{\mathbb{1}_a}$ which is 1 if $\mathbb{1}_a = 1$ otherwise undefined

$$= \frac{\frac{(n-x_j)!}{x_1! \cdots x_{j-1}! x_{j+1}! \cdots x_K!} p_1^{x_1} \cdots p_{j-1}^{x_{j-1}} p_{j+1}^{x_{j+1}} \cdots p_K^{x_K} \mathbb{1}_{x_1 + \ldots + x_{j-1} + x_{j+1} + \ldots + x_K = n - x_j} \prod_{\ell \neq j} \mathbb{1}_{x_\ell \in \{0, \ldots, n\}} U_{n \in \mathbb{N}_0} U_{x_j \in \{0, \ldots, n\}}}{(1-p_j)^{n-x_j}}$$

Note: $x_1 + \ldots + x_{j-1} + x_{j+1} + \ldots + x_K = n - x_j \Rightarrow (1-p_j)^{n-x_j} = (1-p_j)^{x_1} \cdot \ldots \cdot (1-p_j)^{x_{j-1}} (1-p_j)^{x_{j+1}} \cdot \ldots \cdot (1-p_j)^{x_K}$

$$= \binom{n - x_j}{x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_K} \left(\frac{p_1}{1-p_j}\right)^{x_1} \cdot \ldots \cdot \left(\frac{p_{j-1}}{1-p_j}\right)^{x_{j-1}} \left(\frac{p_{j+1}}{1-p_j}\right)^{x_{j+1}} \cdot \ldots \cdot \left(\frac{p_K}{1-p_j}\right)^{x_K} U_{n \in \mathbb{N}_0} U_{x_j \in \{0, \ldots, n\}}$$

$$= \text{Multinom}(n', \vec{p}') \, U_{n \in \mathbb{N}_0} U_{x_j \in \{0, \ldots, n\}}$$

usually completely ignored

($K-1$ dimensional)

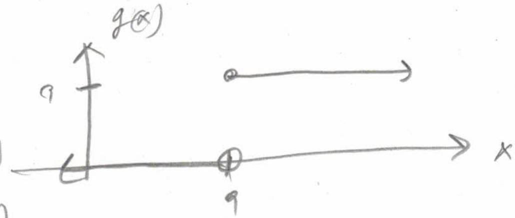
where $n' := n - x_j$, $\vec{p}' = \frac{1}{1-p_j} \begin{bmatrix} p_1 \\ p_{j-1} \\ p_{j+1} \\ \vdots \\ p_K \end{bmatrix}$ Does this make sense?

Back to theory.... Let's do two famous inequalities

Let X be a non-neg r.v. i.e. $S_X \geq 0$

that has $\mu < \infty$. Let $a > 0$, a constant. Consider

$$g(x) = a\,\mathbb{1}_{x \geq a} \sim \begin{cases} a & \text{wp } P(X \geq a) \\ 0 & \text{wp } P(X < a) \end{cases}$$



Is $a\,\mathbb{1}_{X \geq a} \leq X$? Strange question...

"tail probability" or prob of "extreme events"

Only two scenarios..

if $X \geq a \Rightarrow a\,\mathbb{1}_{X \geq a} = a \leq X$ by assumption

if $X < a \Rightarrow a\,\mathbb{1}_{X \geq a} = 0 \leq X$ since $S_X \geq 0$

Let's take expectation of both sides:

Markov's Inequality

HW 1

$$E\left[a\,\mathbb{1}_{X \geq a}\right] \leq E(X) \Rightarrow a\,E\left[\mathbb{1}_{X \geq a}\right] \leq \mu \Rightarrow a P(X \geq a) \leq \mu \Rightarrow P(X \geq a) \leq \frac{\mu}{a}$$

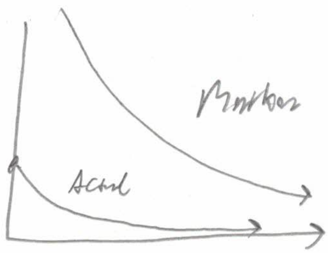
This bound is crude which means it's usually too big to be interesting. For example

$X \sim Exp(1) := e^{-x}$, $\mu = 1$, $F(x) = 1 - e^{-x}$

$P(X \geq 4) = 1 - F(4) = e^{-4} = .018$

Markov Bound $P(X \geq 4) \leq \frac{1}{4} = .25$. This is what "crude" means

How



Corollaries

• Let X be any r.v. $|X|$ is non-neg since its support ≥ 0

$$\Rightarrow P(|X| \geq a) \leq \frac{E[|X|]}{a}$$

Interesting since it bounds both the left and right tail

How to calculate $E[|X|]$

Note: $|X| = X \mathbb{1}_{X \geq 0} - X \mathbb{1}_{X<0} \Rightarrow E[|X|] = E[X \mathbb{1}_{X \geq 0}] - E[X \mathbb{1}_{X<0}]$

$$= \int_{\mathbb{R}} x \mathbb{1}_{x \geq 0} f(x)\,dx - \int_{\mathbb{R}} x \mathbb{1}_{x<0} f(x)\,dx$$

$$= \int_0^\infty x f(x)\,dx - \int_{-\infty}^0 x f(x)\,dx$$

Still sometimes difficult...

• Let h be any strictly monotonically increasing function i.e. $|\cdot|$

$$P(h(X) \geq h(a)) \leq \frac{E[h(X)]}{h(a)}$$

$$\Rightarrow P(X \geq a) \leq \frac{E[h(X)]}{h(a)}$$ Sometimes this can give better bounds.

$X \sim Exp(1)$. Let $h(X) = X^2 \Rightarrow E[h(X)] = E[X^2] = Var(X) + E[X]^2 = 1 + 1^2 = 2$

$P(X \geq 4) = \frac{2}{4^2} = \frac{1}{8}$ which is better than the $\frac{1}{4}$ from the "vanilla" bound

Let X be any rv with finite μ, σ^2

Let $Y = (X-\mu)^2$ $\quad S_Y \geq 0$! Use Markov on a^2

$$P(Y \geq a^2) \leq \frac{E[(X-\mu)^2]}{a^2}$$

$$\Rightarrow P((X-\mu)^2 \geq a^2) \leq \frac{\sigma^2}{a^2}$$

$$\Rightarrow P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Chebyshev's Inequality

This is a much ~~less~~ crude bound than Markov but requires more information about X, its variance

both tails for the centered r.v.

Can we get this to look like Markov's one tail style?

Yes... assume $S_X \geq 0$

disjoint events

$$P(|X-\mu| \geq a) = P(X-\mu \geq a) + P(-(X-\mu) \geq a) = P(X-\mu \geq a) + P(X-\mu \leq -a)$$

$$= P(X \geq \mu+a) + P(X \leq \mu-a)$$

Assume $a \geq \mu$ $\;\Rightarrow\;$ $P(X \geq \mu+a) \leq \frac{\sigma^2}{a^2}$

let $b = \mu+a$ $\Rightarrow a = b-\mu$ $\;\Rightarrow\;$ $P(X \geq b) \leq \frac{\sigma^2}{(b-\mu)^2}$ $\quad$ Valid for non-neg rv's and $b \geq 2\mu$

Note $b \geq 2\mu$

$X \sim Exp(1)$, $\mu = 1$, $\sigma^2 = 1$

$$P(X \geq 4) \leq \frac{1}{(4-1)^2} = \frac{1}{9}$$

better than the bound that needed $E(X^2)$ which is the same as knowing variance

Note $4 \geq 2\mu = 2$