Math 621 Lec 4 9/7/16

$$\vec{X} \sim \text{Multinom}(n, \vec{p}) := \binom{n}{x_1, x_2, \ldots x_k} p^{x_1} \cdots p^{x_k}$$

Sometimes you see a k to indicate $\dim[\vec{x}] = k$, $\dim[\vec{p}] = ?\ k$

No indicator functions necessary due to multichoose notation. If there needed to be...

$$\mathbb{1}_{\sum x_i = n} \prod_{i=1}^{k} \mathbb{1}_{x_i \in \mathbb{N}_0}$$ ugh!

Param Space

$\text{Supp}[\vec{X}] = \{\vec{x} : \vec{x} \in \mathbb{N}_0^k \text{ and } \vec{x} \cdot \vec{1} = n\}$ $\quad \mathcal{P} = \{\vec{p} : \vec{p} \in (0,1)^k \text{ and } \vec{1} \cdot \vec{p} = 1\}$

In a drawer of 10 fruits such that $p_1 = \frac{1}{4}$, $p_2 = \frac{1}{8}$, $p_3 = \frac{5}{8}$

$$P(\text{of getting 3 Apples, 2 bananas, 5 cantaloupes}) = \binom{10}{3,2,5} \left(\frac{1}{4}\right)^3 \left(\frac{1}{8}\right)^2 \left(\frac{5}{8}\right)^5$$

$$= P\left(\vec{X} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}\right) =$$

e.g. If $K=2$, $\vec{p} = \begin{bmatrix} p \\ ? \end{bmatrix} = \begin{bmatrix} p \\ 1-p \end{bmatrix}$ since $p_1 + p_2 \overset{\text{must}}{=} 1$

$\text{Multinom}\left(n, \begin{bmatrix} p_A \\ p_B \end{bmatrix}\right) = \text{Bin}(n, p_A)$

What is the support?

$\text{Supp}[X] = \mathbb{N}_0^{K}$ s.t. $\vec{X} \cdot \vec{1} = n$ i.e. $\sum x_k = n$ why?

By model assumption there's only n trials so there must be only n successes from all categories.

$$p(x_1, x_2) = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \text{Multinom}\left(n, \begin{bmatrix} p \\ 1-p \end{bmatrix}\right) = \binom{n}{x_1, x_2} p^{x_1} (1-p)^{x_2}$$

$\neq \text{Bin}(n,p)$ since it must be a $p(x_1,x_2)$!

joint mass function (JMF) tells you the prob of more than one r.v. value.

Is $X_1, X_2 \perp\!\!\!\perp$? What would this mean?

$$p(x_1, x_2) = p(x_1)\,p(x_2) \implies p(x_1 \mid x_2) = p(x_1) \text{ or } p(x_2 \mid x_1) = p(x_2)$$

conditional pmf

Why? Bayes Rule

If $x_2 \notin \text{Supp}(X_2)$... conditional PMF undefined...

$\forall x_1 \in \text{Supp}(X_1)$,
$\forall x_2 \in \text{Supp}(X_2)$

$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p(x_2)} \overset{\text{if ind}}{=} \frac{p(x_1)\cancel{p(x_2)}}{\cancel{p(x_2)}} = p(x_1)$$

$$p(x_2 \mid x_1) = \text{- - - - - - - - - - - -} = p(x_2)$$

Are X_1, X_2 ind? Interesting... no!! If you know x_2, then $X_1 = n - x_2$! They are the ultimate dependence...

Proof...

$$p(X_1 | x_2) = \frac{\overset{\nearrow \text{the PMF}}{p(X_1, X_2)}}{p(x_2)}$$

we need $p(x_2) \leftarrow$ the marginal distr.

$$p(x_2) = \sum_{x_1 \in Supp(X_1)} p(x_1, x_2) = \sum_{x_1 = 0}^{n} \frac{n!}{x_1! \, x_2!} p^{x_1} (1-p)^{x_2} \mathbb{1}_{x_1 + x_2 = n}$$

$\uparrow$

every possible way x_2 can be realized... you add up every comb of x_2, x_1 together

$$= \frac{n!}{x_2!} (1-p)^{x_2} \underbrace{\sum_{x_1 = 0}^{n} \frac{p^{x_1}}{x_1!} \mathbb{1}_{x_1 = n - x_2}}$$

all of these terms are zero except when $x_1 = n - x_2$

$$= \frac{n!}{x_2!} (1-p)^{x_2} \frac{p^{n - x_2}}{(n - x_2)!}$$

$$= \binom{n}{x_2} (1-p)^{x_2} p^{n - x_2}$$

Symmetric exercise!

$\Rightarrow \quad X_2 \sim \text{Binom}(n, 1-p) \quad$ ALSO... $X_1 \sim \text{Binom}(n, p)$

the marginal distr. in a multinomial is a binomial... makes sense...

$$P(X_1 | X_2) = \frac{\frac{\cancel{n!}}{X_1!\,\cancel{X_2!}}\, p^{X_1}\, \cancel{(1-p)^{X_2}}\, \mathbb{1}_{X_1+X_2=n}}{\frac{\cancel{n!}}{\cancel{X_2!}\,(n-X_2)!}\, \cancel{(1-p)^{X_2}}\, p^{n-X_2}}$$

$$= \frac{(n-X_2)!}{X_1!}\, p^{X_1+X_2-n}\, \underset{\substack{\uparrow \\ 0 \text{ unless } X_1 = n - X_2}}{\mathbb{1}_{X_1+X_2=n}}$$

$$\forall x_2 \quad P(X_1 = n - X_2 | X_2) = \frac{\cancel{(n-X_2)!}}{\cancel{(n-X_2)!}}\, \cancel{p^{0}}^{\to 1} = 1$$

Why? $\neq \text{Bin}(n,p)$

$\Rightarrow$ Not independent!

Let's do general case...

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

$$P(X_{-j} | X_j) = \frac{P(X_1, \ldots, X_K)}{P(X_j)} = \frac{\text{Multin}(n, \vec{p})}{\text{Binom}(n, p_j)}$$

all elements except the j^{th}

$$= \frac{\frac{n!}{x_1! \cdots x_K!}\, p_1^{x_1} \cdots p_K^{x_K}}{\frac{n!}{x_j!\,(n-x_j)!}\, p_j^{x_j} (1-p_j)^{n-x_j}}$$

$$= \frac{(n-x_j)!}{x_1! \cdots x_{j-1}!\, x_{j+1}! \cdots x_K!}\; \frac{p_1^{x_1} \cdots p_{j-1}^{x_{j-1}}\, p_{j+1}^{x_{j+1}} \cdots p_K^{x_K}}{(1-p_j)^{n-x_j}}$$

let $n' := n - x_j$

Note: $\sum_{j=1}^{K} x_j = n \Rightarrow x_1 + \ldots + x_{j-1} + x_j + x_{j+1} + \ldots + x_K = n \Rightarrow$

$x_1 + \ldots + x_{j-1} + x_{j+1} + \ldots + x_K = n - x_j = n'$

$$= \binom{n'}{x_1,\dots,x_{j-1},x_{j+1},\dots,x_K} \frac{p_1^{x_1}\cdots p_{j-1}^{x_{j-1}}\cdot p_{j+1}^{x_{j+1}}\cdots p_K^{x_K}}{(1-p_j)^{n'}}$$

let $p_1' = \frac{p_1}{1-p_j}$, $p_2' = \frac{p_2}{1-p_j}$, $\dots$, $p_K' = \frac{p_K}{1-p_K} \Rightarrow p' := \begin{bmatrix} p_1' \\ \vdots \\ p_K' \end{bmatrix}$

$$= \binom{n'}{\cdots} \frac{(p_1'(1-p_j))^{x_1}\cdots(p_{j-1}'(1-p_{j-1}))^{x_{j-1}}(p_{j+1}'(1-p_{j+1}))^{x_{j+1}}\cdots(p_K'(1-p_K))^{x_K}}{(1-p_j)^{n'}}$$

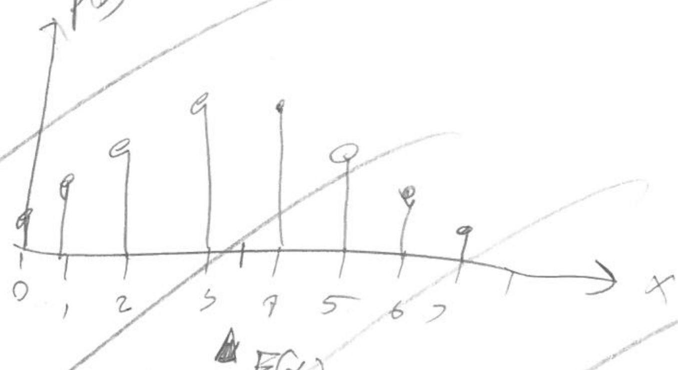
$$= \binom{n'}{\cdots} \frac{(p_1')^{x_j}\cdots(p_{j-1}')^{x_{j-1}}(p_{j+1}')^{x_{j+1}}\cdots(p_K')^{x_K}\cancel{(1-p_j)^{\overbrace{x_1+\dots+x_{j-1}+x_{j+1}+\dots+x_K}^{n'}}}}{\cancel{(1-p_j)^{n'}}}$$

$$= \text{Multinom}(n', p')$$

~~Wht is expectation? (mean?)~~

~~$\mu := E(X) = \sum_{x \in Supp(X)} x\, p(x)$~~

~~$X \sim \text{Binom}(7, \frac{1}{2})$~~



~~Balance pt.~~

~~Long run average!~~

Recall
from last lecture.. ~~$E[g(X)] = \sum_{x \in S_p(x)} g(x)\, p(x)$~~ , $E[g(X_1, \dots X_n)] = \sum_{x_1 \in S_p(x)} \dots \sum_{x_n \in S_p(x_n)} g(x_1, \dots x_n)\, p(x_1, \dots x_n)$

$E(\sum X_i) = \sum_{i=1}^{n} E(X_i) = n\mu$ always (proof in 241)

if $X_i \overset{d}{=} X_j \; \forall i,j$

$E(aX) = aE(X)$

$E(X+c) = E(X) + c$

$E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E(X_i)$ if $X_1, \dots, X_n \overset{ind}{\sim}$

$= \mu^n$ if $X_1, \dots X_n \overset{iid}{\sim}$

Proof:

$$E[\underset{\underset{g(X_1, \dots X_n)}{\uparrow}}{\prod X_i}] = \sum_{x_1} \dots \sum_{x_n} x_1 \cdot x_2 \cdots x_n \underbrace{p(x_1, \dots, x_n)}_{p(x_1) \cdots p(x_n)} = \sum_{x_1} x_1 p(x_1) \sum_{x_2} x_2 p(x_2) \cdots \sum_{x_n} x_n p(x_n)$$

$$= E(X_1) E(X_2) \cdots E(X_n)$$

$$\sigma^2 := Var(X) := E[\underset{\underset{g(X)}{\uparrow}}{(X-\mu)^2}] = \sum_{x \in S_p(x)} g(x)\, p(x) = \sum (x-\mu)^2 p(x)$$

$$= \sum x^2 p(x) + \sum (-2x\mu) p(x) + \sum \mu^2 p(x)$$

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

expected squared error
difference from the expectation

Rules

$Var[X+a] = Var(X)$

$Var(cX) = c^2 Var(X)$

$Var[X_1 + X_2] = E[((X_1 + X_2) - (\mu_1 + \mu_2))^2]$

$$= E\left[X_1^2 + X_2^2 + \mu_1^2 + \mu_2^2 - 2\mu_1 X_1 - 2\mu_1 X_2 - 2\mu_2 X_1 - 2\mu_2 X_2 + 2X_1X_2 + 2\mu_1\mu_2\right]$$

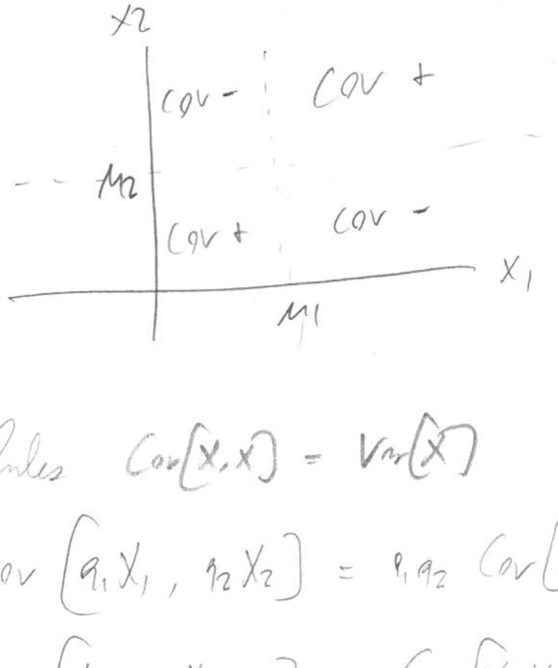
$$= E(X_1^2) + E(X_2^2) + \cancel{\mu_1^2} + \cancel{\mu_2^2} \underset{-\mu_1^2}{\cancel{-2\mu_1^2}} - 2\mu_1\mu_2 - \cancel{2\mu_1\mu_2} \underset{-\mu_2^2}{\cancel{-2\mu_2^2}} + 2E[X_1X_2] + \cancel{2\mu_1\mu_2}$$

$$= Var(X_1) + Var(X_2) + 2\left(E[X_1X_2] - \mu_1\mu_2\right)$$

$$Cov[X_1, X_2] := E[X_1X_2] - \mu_1\mu_2$$

Sometimes defined as $E[(X_1 - \mu_1)(X_2 - \mu_2)] = E[X_1X_2 - \mu_1 X_2 - \mu_2 X_1 + \mu_1\mu_2]$

$= E[X_1X_2] - \mu_1\mu_2 - \cancel{\mu_1\mu_2} + \cancel{\mu_1\mu_2}$ ✓



$$Corr[X_1, X_2] = \frac{Cov[X_1, X_2]}{SE(X_1)\, SE(X_2)} \in [-1, 1]$$

$\uparrow$

a unitless, interpretable covariance

Rules $Cov[X, X] = Var[X]$

$Cov[a_1 X_1, a_2 X_2] = a_1 a_2 Cov[X_1, X_2]$

$Cov[X_1 + c_1, X_2 + c_2] = Cov[X_1, X_2]$

$Cov[X_2, X_1] = Cov[X_1, X_2]$ Commutativity

$Cov[X + Y, Z] = Cov[X, Z] + Cov[Y, Z]$, $Cov\left[\sum_i X_i, \sum_j Y_j\right] = \sum_i \sum_j Cov[X_i, Y_j]$ Distr?

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

Note

$$
\begin{aligned}
Var[X_1+X_2] &= Var(X_1) + Var(X_2) + 2Cov(Y_1 X_2) \quad \text{(from before)} \\
&= Cov(X_1 X_1) + Cov(X_2 X_2) + Cov(X_1 X_2) + Cov(X_2, X_1) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} Cov[X_i, X_j]
\end{aligned}
$$

$$Var\left(X_1 + X_2 + \dots + X_k\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} Cov[X_i, X_k]$$

If $\vec{X}$ is a vector of r.v of dim K...

$Cov[X_1, X_0]$ // $\nearrow$ symmetric matrix due to comm. prop.

$$\vec{\mu} := E[\vec{X}] := \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E[X_K] \end{bmatrix}, \quad \Sigma := Var[\vec{X}] = \begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & \dots & Cov(X_1,X_K] \\ Cov(X_2,X_1) & Var(X_2) & & \vdots \\ \vdots & & \ddots & \\ Cov(X_K,X_1) & \dots & & Var[X_K] \end{pmatrix} = \left\{Cov[X_i, X_j]\right\}_{\substack{i=1\dots k,\\ j=1\dots k}}$$

$$\Sigma_0 = Corr[\vec{X}] = \begin{pmatrix} 1 & Corr[X_i X_j] & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} = \left\{Corr(X_i, X_j]\right\}_{i=1,\dots,k,\ j=1,\dots,k}$$

Prove $Corr[X_i, X_i] = 1$

[HW]

$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ dim(K) $\in \mathbb{R}^K$

$$E\left[\vec{1}^T \vec{X}\right] = \sum E(X_i)$$

$$Var\left[\vec{1}^T \vec{X}\right] = \sum_{i=1}^{k}\sum_{j=1}^{k} Cov[X_i, X_K]$$

let $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \in \mathbb{R}^{n \times n}$

What is $\vec{1}^T \underbrace{A \vec{1}}$? $\in \mathbb{R}$ $\quad (1 \times n) \times (n \times n) \times (n \times 1)$

$$\vec{1}^T \begin{bmatrix} a_{11} + \dots + a_{1n} \\ \vdots \\ a_{n1} + \dots a_{nn} \end{bmatrix} = a_{11} + \dots a_{1n} + a_{21} + \dots a_{2n} + a_{n1} + \dots a_{nn}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$$