11/11/2019

BIOS 667 Midterm Review 2

1) · To get valid inferences
 · To get more efficient estimates

2)

Y	P(Y=y)
0	0.3
1	0.4
2	0.3

$m=2$

$E(Y) = \mu = 1$

Extra-binomial variation => what is theoretical variance of random variable X w/ same mean as Y, if X was binomial

$X \sim \text{Binomial}(m, \pi)$

$E(X) = m\pi$

$Var(X) = m\pi(1-\pi)$

$m=2$

$\pi = \frac{\mu}{m} = \frac{1}{2}$

$Var(X) = 2 \cdot \frac{1}{2}\left(1-\frac{1}{2}\right) = \frac{1}{2}$

$Var(Y) = \underbrace{E(Y^2)}_{\to y^2 \cdot P(Y=y)} - [E(Y)]^2$

$= [0^2 \cdot 0.3 + 1^2 \cdot 0.4 + 2^2 \cdot 0.3] - 1^2$

$= 1.6 - 1 = 0.6$

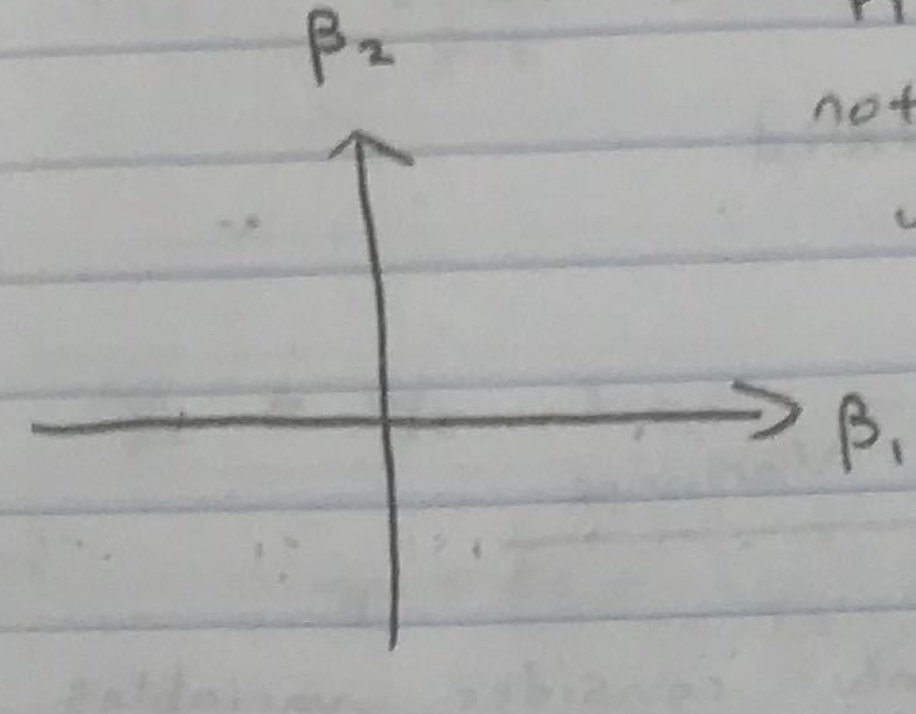
$Var(Y) > Var(X)$ => Yes, extra-binomial variance

Example of functional unrelatedness,

but cov of the two $\neq 0$

$$E[Y_i] = \beta_1 + \beta_2 x_i$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^2$$



β_1 & β_2 are not generally functionally unrelated (so, given a β_1, β_2 could be anything)

But, $cov(\hat{\beta}_1, \hat{\beta}_2) \neq 0$ in general

$cov(\hat{\beta})$

$\sigma^2 (X^T X)^{-1}$

If $\sum_{i=1}^{n} x_i \neq 0$, then $cov(\hat{\beta})$ not diagonal

☆ Consequence of Problem 5: ☆

☆ Adjusting for baseline in binary outcomes not necessarily good idea (might increase variance of estimates)

↳ not guaranteed more efficient like w/ multivariate normal ☆

☆ Report correlations as correlations $\cdot$ 100 (generally, report 2 decimals first, then multiply by 100) $\Rightarrow$ more readable

↳ or generally, just present 3 significant digits (but still multiply by 100) ☆

Given $Y_{i1} = 0$ $(P = 0.9)$,
what is prob of $Y_{i2} = 1$

$$P(Y_{i2} = 1 \mid Y_{i1} = 0) = \frac{0.132}{0.9}$$

restrict to row where $Y_{i1} = 0$, then numerator is cell where $Y_{i2} = 1$

$$P(Y_{i2} = 1 \mid Y_{i1} = 1) = \frac{0.068}{0.1}$$

Solve for γ_1 & γ_2 like you did before

$\gamma_1 = P(Y_{i2} = 1 \mid Y_{i1} = 0) = 0.147$

$\gamma_2 = P(Y_{i2} = 1 \mid Y_{i1} = 1) - P(Y_{i2} = 1 \mid Y_{i1} = 0)$

$= 0.68 - 0.147$

b) One variance => $0.147 \cdot (1 - 0.147)$

Point (?)

(?)

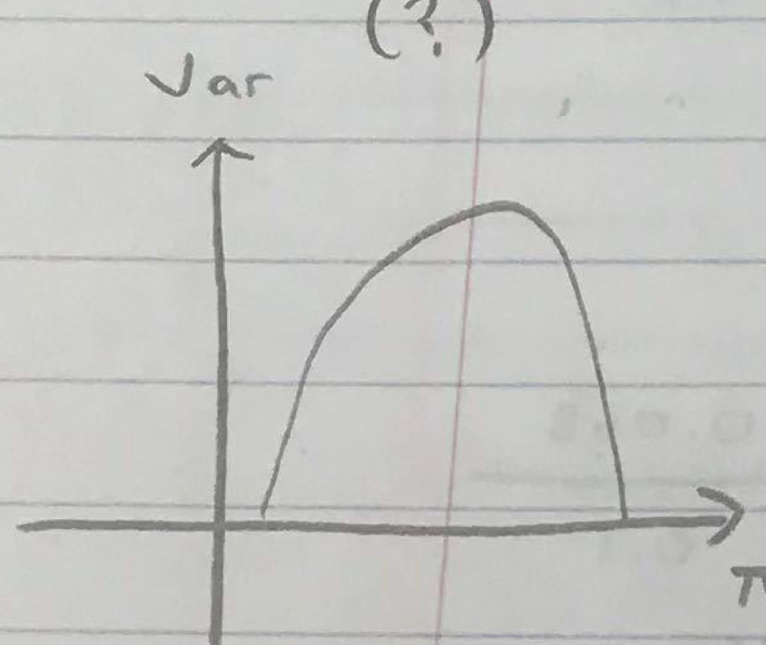
Other => $0.68 \cdot (1 - 0.68)$

calculated above

$Y_{i2} = 1 \mid Y_{i1} = 0 \sim$ Binomial( )

$Y_{i2} = 1 \mid Y_{i1} = 1 \sim$ Binomial( )

calculated above



Something about how at some point Var of conditional will be > than marginal & < at others

b/c quadratic form

c) Take 2x2 table in a) & calculate OR

$$\frac{0.768 \cdot 0.068}{0.032 \cdot 0.132}$$

Notation:

$P(Y_{i1} = 1)$ → Valid

on event

$E[Y_{i1} = 1]$ → Not valid

$E[Y_{i1}]$ → Valid

Extra- Binomial Variance

↳ Usually only consider variables w/ finite range

But: Ex: $Y \sim N(0, 1)$

$T \sim Bin(m, \pi) \Rightarrow E[T] = \underbrace{m\pi} = 0 = E(Y)$

Either $m=0$, $\pi = 0$ or both

- If $m = 0$, degenerate case (not meaningful)
- If $\pi = 0$, then $var(T) = 0$

↳ Technically, Y would have extra-binomial variance ($Var(Y) = 1 > Var(T) = 0$), but you'd never really be asked this

Ex 2: (Extra- Poisson variance)

↓

only relevant for non-negative support

$Y \in \{0, 1, 2, 3, \ldots\}$

or

$X \in [0, \infty)$

$E(X) = 5$ $Var(X) = 25$

Yes, extra-Poisson variance for X b/c $Var(X) > E(X)$

So, $P(Y_{i1}=1, Y_{i2}=1) = P(Y_{i1}Y_{i2}) = E[Y_{i1}Y_{i2}]$

$$= E[Y_{i1}]\cdot E[Y_{i2}] + \underline{cov(Y_{i1}, Y_{i2})}$$

We are given $corr(Y_{i1}, Y_{i2})$, so easy to get $cov(Y_{i1}, Y_{i2})$

$$\overbrace{corr(Y_{i1}, Y_{i2})}^{\text{Given}} = \frac{cov(Y_{i1}, Y_{i2})}{\sqrt{\underbrace{Var(Y_{i1})\,Var(Y_{i2})}_{\text{Given}}}}$$

$$E[Y_{i1}Y_{i2}] = 0.02 + 0.4\sqrt{0.09(0.16)} = \overset{P(Y_{i1}=1, Y_{i2}=1)}{\overset{//}{0.068}}$$

So, $P(Y_{i1}=0, Y_{i2}=1) = 0.2 - 0.068 = 0.132$

So, $P(Y_{i2}=1 \mid Y_{i1}=0) = \dfrac{0.132}{0.9}$

Repeat for $P(Y_{i2}=1 \mid Y_{i1}=1) = \dfrac{0.068}{0.1}$

<u>Short-Way</u> Don't have to remember Bayes'

Given marginals → Find one cell like above, then fill in remainder of cells given marginals

		Y_{i2} = 0	Y_{i2} = 1	
Y_{i1}	0	0.768	0.132	0.9
	1	0.032	$E(Y_{i1})E(Y_{i2}) + cov(Y_{i1}, Y_{i2})$	0.1
		0.8	0.2	1

5a) $\text{Corr}(Y_{i1}, Y_{i2}) = 0.4$

$P(Y_{i1}=1) = 0.1$ $\quad$ $P(Y_{i2}=1) = 0.2$

$P(Y_{i2}=1 \mid Y_{i1}=0) =$

&

$P(Y_{i2}=1 \mid Y_{i1}=1)$

Long-Way

$P(A|B) = \dfrac{\overbrace{P(A \cap B)}^{\text{Need to find}}}{\underbrace{P(B)}_{\text{Given}}}$

$A = \{Y_{i2}=1\}$ $\quad$ $B = \{Y_{i1}=0\}$

Need $P(Y_{i1}=0 \cap Y_{i2}=1) = P(Y_{i1}=0, Y_{i2}=1)$

$P(Y_{i1}=0 \cap Y_{i2}=1) + \underline{P(Y_{i1}=1 \cap Y_{i2}=1)}$
$= P(Y_{i2}=1) = 0.2$

$\underline{P(Y_{i1}=1, Y_{i2}=1)} = \underline{P(Y_{i1} \cdot Y_{i2}) = 1}$

Equivalent events, so probs are equal

$Y_{i1} \cdot Y_{i2} \rightarrow$ binary b/c Y_{i1} & Y_{i2} are binary

3) all these estimators are generally biased (whether MLE or REML)

Asymptotically, full ML & REML are equivalent
=> neither is asymptotically more efficient than the other

4) $\hat{\text{cov}}(\hat{\beta}_M, \hat{\theta}_M) = 0$ Matrix $\qquad$ θ & β are functionally unrelated

☆ Look at information matrix (Fisher) → expected information ☆

$$\begin{array}{c c} & \begin{array}{cc} \beta & \theta \end{array} \\ \begin{array}{c} \beta \\ \theta \end{array} & \left[\begin{array}{cc} & 0 \\ 0 & \end{array} \right] \end{array}$$

Asymptotic covariance matrix is inverse of Fisher information matrix

Distribution of response must be multivariate normal

Orthogonal parameters