Solve $X^T(\underset{\sim}{Y}-\underset{\sim}{\mu})$

get $\hat{\beta}$ that is consistent

- Don't assume that model is Poisson. use Poisson likelihood to get $\hat{\beta}$, then find R.V.E. for $\widehat{SE}$

HW 4 $\boxed{\text{Review}}$

$\Sigma = \underset{b}{\rho} J + \underbrace{(1-\rho)}_{a} I$

positive definite when $\frac{-1}{n-1} < \rho < 1$

has to be to guarantee that the variances of linear combinations of Y are positive

$W = \Sigma^{-1} = dJ + cI$

$\{\rho \underset{\sim}{1}\underset{\sim}{1}^T + (1-\rho)I\}\{d\underset{\sim}{1}\underset{\sim}{1}^T + cI\} = I \quad \leftarrow W \cdot \Sigma$

$\rho dn \underset{\sim}{1}\underset{\sim}{1}^T + c\rho \underset{\sim}{1}\underset{\sim}{1}^T + (1-\rho)d\underset{\sim}{1}\underset{\sim}{1}^T + c(1-\rho)I = I$

Diagonal: $\rho dn + c\rho + (1-\rho)d + c(1-\rho) = 1$

Off-diagonal: $\rho dn + c\rho + (1-\rho)d + 0 = 0$

get c, d, then calculate

$\{X^T(d\underset{\sim}{1}\underset{\sim}{1}^T + cI)X\}^{-1}$

$= \underbrace{\{cX^TX + d\underset{\sim}{s}\underset{\sim}{s}^T\}^{-1}}$ $\qquad \underset{p\times 1}{\underset{\sim}{s}} = \sum_{i=1}^n \underset{\sim}{x}_i$

$X^T\underset{\sim}{1} = s, \quad 1^TX = (X^T1)^T = \underset{\sim}{s}^T$

pattern: $X^TX \to$ full rank, invertible. $\underset{\sim}{s}\underset{\sim}{s}^T \to$ rank 1 ?

$(X^TX)^{-1}$ known, then if you add a matrix of rank 1, then you can find the inverse of their sum. Name: rank 1 update of an inverse

In the long way you need to use the fact that X has the intercept too, can't escape it.

Long Way

Rank 1 update of an inverse: your punishment for plugging in without thinking.

The thinking way:

We are thinking about two things: $\hat{\beta}_{OLS}, \hat{\beta}_{WLS}$

Don't think $\hat{\beta} =$, but $\hat{\beta}$ solves

$\hat{\beta}_{OLS}$ equation: $X^T(\underset{\sim}{Y}-\underset{\sim}{\mu}) = \underset{\sim}{0}_{p\times 1}$, $\mu = \hat{\beta}X$

$\hat{\beta}_{WLS}$ $\hat{\beta}$ solves: $X^T W(\underset{\sim}{Y}-\underset{\sim}{\mu}) = \underset{\sim}{0}_{p\times 1}$

Does the $\hat{\beta}$ that solves the $\hat{\beta}_{OLS}$ equation also solve the $\hat{\beta}_{WLS}$ equation?

$\hat{\beta}_{OLS}$	$\hat{\beta}_{WLS}$
$X^T(\underset{\sim}{Y}-\underset{\sim}{\mu}) = \underset{\sim}{0}$	$X^T W(\underset{\sim}{Y}-\underset{\sim}{\mu}) = \underset{\sim}{0}$
There is a linear combination of 1's that makes this 0.	$X^T(cI + d\underset{\sim}{1}\underset{\sim}{1}^T)(\underset{\sim}{Y}-\underset{\sim}{\mu}) = 0$
	$\underbrace{cX^T(\underset{\sim}{Y}-\underset{\sim}{\mu})}_{=0 \text{ from OLS}} + dX^T\underbrace{\underset{\sim}{1}^T(\underset{\sim}{Y}-\underset{\sim}{\mu})\underset{\sim}{1}}_{\text{Sum of residuals}} = \underset{\sim}{0}$
	$\to = 0$ since there is an intercept from OLS

$\underset{\sim}{1}^T(\underset{\sim}{Y}-\underset{\sim}{\mu}) = 0$

X design matrix contains the intercept ← have to use this fact!

~~Not~~ This result is not true if X doesn't contain intercept

Example of X that doesn't contain intercepts: $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 8 \\ 3 & 9 & 27 \\ 4 & 16 & 64 \end{bmatrix}$

Implicit intercept: some

So, there is a row of 1's that makes this 0 in $\hat{\beta}_{OLS}$, so $\hat{\beta}_{OLS}$ also makes this 0.

① $Y_1, \dots Y_n$ independent Bern

$\underset{\sim}{\mu_1}, \dots \underset{\sim}{\mu_n}$

$\text{logit}\ \mu_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta}$

- Deviance depends on y only thru $\hat{\beta}$
- Deviance does not $\xrightarrow{d} \chi^2_{n-p}$ as $n \to \infty$

(we're not talking about likelihood ratio test here. Just talking about this model, not any other sub models and hypotheses.)

Other Bernoullis:

$Y_i \sim Bin(m_i, \pi_i)$, $i = 1, \dots, K \to K$ is fixed

② K groups (educational level, cross-classification → edu level × gender...)

$m_i \to \infty$ for each m_i

Deviance $\xrightarrow{d} \chi^2_{K-p}$ K fixed, $m_i \to \infty$

Doesn't appear here

If you have continuous variables, you can't group to go from Bern ① to Bern ②

If all X's are categorical and you group them into a small and finite number of groups, then you can move from Bernoulli ① to Bernoulli ②.

When you group the data, the deviance will be different.

midterm question → - Set of observations with the same X (so same values for the variables) → covariate class

---

Last quiz question: Extra Binomial Variation

$Y_i \in [0, m]$

$EY = \mu$, $var(Y) = \mu(1 - \frac{\mu}{m})$

Options? $var(Y) = \phi\mu(1 - \frac{\mu}{m})$

$0 < \phi \leq m$ from what we did last time

	Y_i	m_i
1^{st} observation →	1	3
2^{nd} " →	2	5
3^{rd} " →	5	15
	11	27

$var(Y_i) = \varphi \mu_i (1 - \frac{\mu_i}{m_i})$

$0 < \varphi \leq m_i$

What is one φ that I can use for all of these observations? The smallest one in the dataset is what restricts it.

So, the largest φ can be for 1^{st} observation is 3, so that's the biggest it can be for all of the data.

So, take-away is when the m_i's vary wildly, then this model with one φ becomes very restrictive.

Instead can have model: $Var(Y_i) = \varphi_i \mu_i (1 - \frac{\mu_i}{m_i})$

$\varphi_i \mu_i (1 - \frac{\mu_i}{m_i})$

$\varphi_i = 1 + \tau(m_i - 1)$

positive (+ve) [negative: −ve]

Beta binomial has this form

But, you don't have to use the beta binomial → you can use a different likelihood function to get the estimates + use this ~~form~~ form of the variation.

No such thing as extra Bernoulli variation. B/c the variance is completely determined by the mean, ~~mathemat~~ mathematically. You can't have something different

$\pi(1-\pi)$

0	$1-\pi$
1	π