Ann Marie
Weideman

2. (25 points) Suppose that $y_1, \ldots, y_n$ are independent binary random variables, where

$$P(y_i = 1|\beta_0, \beta_1, x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)},$$

where $x_1, \ldots, x_n$ are fixed covariates and they are not all equal.

(a) (6 points) Suppose that (β_0, β_1) are both unknown and suppose we wish to test

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 \neq 0.$$

Derive the Uniformly Most Powerful Unbiased (UMPU) α level test for this hypothesis and express the rejection region and critical value in the simplest possible form. Please note that there need not be a closed form for the distribution of the test statistic.

(b) (5 points) Using the UMPU conditional test from part (a), compute an explicit closed form for its conditional mean and conditional variance under the null hypothesis to find an explicit form for an asymptotically correct approximation to the UMPU test. You are allowed to assume that the conditional test statistic is asymptotically normal.

(c) (4 points) Derive the score test for the hypothesis in part (a), and compare its form to the approximate UMPU test derived in part (b).

(d) (6 points) Now consider the more general problem in which we have p covariates, and

$$P(y_i = 1|\boldsymbol{\beta}, \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})},$$

where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ is a $p \times 1$ vector of covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of regression coefficients. Suppose we wish to test

$$H_0 : \ell'\boldsymbol{\beta} = \theta_0 \quad \text{versus} \quad H_1 : \ell'\boldsymbol{\beta} \neq \theta_0,$$

where θ_0 is a specified scalar and ℓ is a specified and non-zero $p \times 1$ vector. Derive the UMPU size α test for this hypothesis and express the rejection region and critical value in the simplest possible form.

(e) (4 points) Describe in detail a non-parametric bootstrap algorithm for computing the exact p-value based on the UMPU test of part (d).

0

2.a) Both (β_0, β_1) unknown.

Want to test $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

Derive UMPU level α test and express rejection region and critical value in simplest form.

Does not have to have a closed form.

Given y_i iid binary RVs where $P(y_i = 1 \mid \beta_0, \beta_1, x_i) = \underbrace{\dfrac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}_{p}$

$$\Rightarrow P(y_i \mid \beta_0, \beta_1, x_i) = \prod_{i=1}^{n} p^{y_i} (1-p)^{1-y_i}$$

$$= \prod_{i=1}^{n} \left[\frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})}\right]^{y_i} \left[1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right]^{1 - y_i}$$

$$= \prod_{i=1}^{n} \left[\frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})}\right]^{y_i} \left[\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}\right]^{1 - y_i}$$

$$= \prod_{i=1}^{n} \frac{e^{\beta_0 y_i + \beta_1 x_i y_i}}{(1 + e^{\beta_0 + \beta_1 x_i})} = \underbrace{\underbrace{\left(\prod_{i=1}^{n} (1 + e^{\beta_0 + \beta_1 x_i})\right)^{-1}}_{c(\beta)} \underbrace{e^{\beta_1 \sum_i x_i y_i + \beta_0 \sum_i y_i}}_{\exp(\theta U + \sum_{i=1}^{K} \xi_i T_i)}}_{\text{multiparameter exponential family according to form given in 761 slides}}$$

$\Rightarrow$ θ = parameter of interest = β_1

ξ = nuisance parameter = β_0

U = CSS of parameter of interest = $\sum_i x_i y_i$

T = CSS of nuisance parameter = $\sum_i y_i$

Thus, the UMPU level α test takes the form,

$$\phi(U) = \begin{cases} 1 & \text{if } U < c_1(t) \text{ or } U > c_2(t) \\ \gamma_1 & \text{if } U = c_1(t) \\ \gamma_2 & \text{if } U = c_2(t) \\ 0 & \text{else} \end{cases} = \begin{cases} 1 & \text{if } \sum_i x_i y_i < \overbrace{c_1(\sum_i y_i)}^{c_1} \text{ or } \sum_i x_i y_i > \overbrace{c_2(\sum_i y_i)}^{c_2} \\ \gamma_1 & \text{if } \sum_i x_i y_i = c_1(\sum_i y_i) \\ \gamma_2 & \text{if } \sum_i x_i y_i = c_2(\sum_i y_i) \\ 0 & \text{else} \end{cases}$$

where $\underbrace{\alpha = E_0[\phi(U) \mid T = t]}_{\text{i)}}$ & $\underbrace{E_0[U \phi(U) \mid T = t]}_{\text{ii)}} = \alpha \underbrace{E_0[U \mid T = t]}_{\text{iii)}}$

Let $A = \{y : \sum y_i = t\}$

$B(c_1, c_2) = \{y : \sum y_i x_i \notin [c_1, c_2]\}$

$B_1 = \{y : \sum y_i x_i = c_1\}$

$B_2 = \{y : \sum y_i x_i = c_2\}$

cont'd next pg. $\longrightarrow$

AMW

2 a) cont'd

Then, (i) $E[\phi(\sum_i x_i Y_i) \mid \sum_i Y_i = t] = 1 \cdot P_0(\sum_i x_i Y_i \notin [c_1, c_2] \mid \sum_i Y_i = t)$
$+ \gamma_1 \cdot P_0(\sum_i x_i Y_i = c_1 \mid \sum_i Y_i = t)$
$+ \gamma_2 \cdot P_0(\sum_i x_i Y_i = c_2 \mid \sum_i Y_i = t)$

$$= \frac{P_0(Y \in (A \cap B)) + \gamma_1 P_0(Y \in (A \cap B_1)) + \gamma_2 P_0(Y \in (A \cap B_2))}{P_0(Y \in A)}$$

$$= \frac{\sum_{Y \in A} c(\beta_0) e^{\overset{0 \text{ under } H_0}{\cancel{\beta_1 \sum_i x_i Y_i}} + \beta_0 \sum_i Y_i} [I(Y \in B) + \gamma_1 I(Y \in B_1) + \gamma_2 I(Y \in B_2)]}{\sum_{Y \in A} c(\beta_0) e^{\underset{0 \text{ under } H_0}{\cancel{\beta_1 \sum_i x_i Y_i}} + \beta_0 \sum_i Y_i}}$$

where $c(\beta_0) = \left[\prod_{i=1}^n (1 + e^{\beta_0})\right]^{-1}$

$$= \frac{\cancel{c(\beta_0) e^{\beta_0 \sum_i Y_i}} \sum_{Y \in A} [I(Y \in B) + \gamma_1 I(Y \in B_1) + \gamma_2 I(Y \in B_2)]}{\cancel{c(\beta_0) e^{\beta_0 \sum_i Y_i}} \sum_{Y \in A} 1_n}$$

$$= \frac{\sum_{Y \in A} [I(Y \in B) + \gamma_1 I(Y \in B_1) + \gamma_2 I(Y \in B_2)]}{\sum_{Y \in A} 1_n}$$

NOTE:

Allowed to cancel these terms b/c we conditioned $\sum_i Y_i = t$

(ii) $E_0[(\sum_i x_i Y_i) \phi(\sum_i x_i Y_i) \mid \sum_i Y_i = t]$

$$= \sum_{Y \in A} (\sum_i x_i Y_i) \cdot \underbrace{\phi(\sum_i x_i Y_i) P_0(Y = y \mid \sum_i Y_i = t)}$$

found in part (i)

$$= \frac{\sum_{Y \in A} (\sum_i x_i Y_i) [I(Y \in B) + \gamma_1 I(Y \in B_1) + \gamma_2 I(Y \in B_2)]}{\sum_{Y \in A} 1_n}$$

(iii) $E_0[\sum_i x_i Y_i \mid \sum_i Y_i = t] = \sum_{Y \in A} (\sum_i x_i Y_i) P_0(Y = y \mid \sum_i Y_i = t) = \dfrac{\sum_{Y \in A} (\sum_i x_i Y_i) \overbrace{P_0(Y = y, \sum_i Y_i = t)}}{P_0(Y \in A)}$

doesn't provide any more info

$= P_0(Y_1, \ldots, Y_n, \sum_i Y_i)$
$= P_0(Y_1, \ldots, Y_n)$

$$= \frac{\sum_{Y \in A} (\sum_i x_i Y_i) \underbrace{c(\beta_0) \cdot e^{\beta_0 \sum_i Y_i}}}{\sum_{Y \in A} \underbrace{c(\beta_0) \cdot e^{\beta_0 \sum_i Y_i}}} = \frac{\sum_{Y \in A} (\sum_i x_i Y_i)}{\sum_{Y \in A} 1_n}$$

recall, fixed $\sum_i Y_i$ since conditioned on it

cont'd next pg

2a) cont'd

Thus, choose c_1, c_2, γ_1, and γ_2 to satisfy above requirements $\ni$ $(\gamma_1, \gamma_2) \in [0,1] \times [0,1]$ for a chosen level of α.

Enumerate all possible vectors Y then restrict only to those samples in A.

Vary c_1 and c_2 within these samples to find suitable cutpoints.

2 b) Compute an explicit closed form for the conditional mean & conditional variance under H_0 to find an explicit form for the asymptotically correct approximation to the UMPU test.

Allowed to assume conditional test statistic is asymptotically normal.

Aim : Normalize the conditional test statistic $U|T = \sum_i x_i Y_i \mid \sum_i Y_i = t$.

This will allow us to:

1. Remove the randomization terms γ_1 and γ_2 in the UMPU level α test b/c the asymptotic distn. will be asymptotically normal.
2. Simplify the calculation of c_1 and c_2.

i) First, derive $E_0[\sum_i x_i Y_i \mid \sum_i Y_i = t] \overset{\text{part a)}}{=} \dfrac{\sum_{Y \in A} (\sum_i x_i Y_i)}{\sum_{Y \in A} 1_n}$ $\quad \mu$

ii) 2nd, derive $Var_0[\sum_i x_i Y_i \mid \sum_i Y_i = t]$

$$= E_0[(\sum_i x_i Y_i)^2 \mid \sum_i Y_i = t] - \{\underbrace{E_0[\sum_i x_i Y_i \mid \sum_i Y_i = t]}\}^2$$

$$= \underbrace{\frac{\sum_{Y \in A} (\sum_i x_i Y_i)^2}{\sum_{Y \in A} 1_n}}_{\text{Same process as done in part a), left to reader ☺}} - \left[\frac{\sum_{Y \in A} (\sum_i x_i Y_i)}{\sum_{Y \in A} 1_n}\right]^2 \quad \sigma^2$$

cont'd next pg. →

2 b) cont'd

Then, the UMPU level α test is

$$\phi(u) = \begin{cases} 1 & \text{if } \sum_i x_i y_i < c_1(\sum_i y_i) \text{ or } \sum_i x_i y_i > c_2(\sum_i y_i) \\ 0 & \text{else} \end{cases}$$

Then, $\alpha = E_0[\phi(\sum_i x_i y_i) \mid \sum_i y_i = t] = 1 \cdot P(\sum_i x_i y_i < \overbrace{c_1(\sum_i y_i)}^{c_1} \text{ or } \sum_i x_i y_i > \overbrace{c_2(\sum_i y_i)}^{c_2} \mid \sum_i y_i = t)$

$$= 1 - P(c_1 < \underbrace{\sum_i x_i y_i}_{z} < c_2 \mid \sum_i y_i = t)$$

Since, $\underbrace{\sum_i x_i y_i \mid \sum_i y_i = t}_{z} \sim N(\mu, \sigma^2)$

$$= 1 - \int_{c_1}^{c_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

$\Rightarrow \alpha = 1 - \Phi(c_2) + \Phi(c_1)$ where Φ is the cdf of a $N(\mu, \sigma^2)$ distn.

Similarly, $\alpha E_0[\sum_i x_i y_i \mid \sum_i y_i = t] = E_0[\sum_i x_i y_i \phi(\sum_i x_i y_i) \mid \sum_i y_i = t]$

$\Rightarrow \alpha \mu = E_0[(\sum_i x_i y_i) \phi(\sum_i x_i y_i) \mid \sum_i y_i = t]$ where $A = \{y : \sum_i y_i = t\}$

$= E_0[z \overbrace{\phi(z)}^{[I(z < c_1) + I(z > c_2)]}]$ where $z = \sum_i x_i y_i \mid \sum y_i$

$= E_0[z I(z < c_1)] + E_0[z I(z > c_2)]$

$+ \int_{-\infty}^{c_1} z f_z \, dz + \int_{c_2}^{\infty} z f_z \, dz = \int_{-\infty}^{\infty} z f_z \, dz - \int_{c_1}^{c_2} z f_z \, dz$

$$= \mu - \int_{c_1}^{c_2} \frac{z}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

$\Rightarrow \alpha \mu = \mu - \Phi(c_2) + \Phi(c_1)$

Thus, we use the two equations from above, $\alpha = 1 - \Phi(c_2) + \Phi(c_1)$, $\alpha\mu = \mu - \Phi(c_2) + \Phi(c_1)$ to solve for c_1 and c_2 given a specific level of α.

We would pick an explicit value for c_1 and solve for c_2 or vice versa. The combination of c_1 and c_2 should result in the desired level of α. ⌋

AMW

2.c) Derive the score test for the hypothesis in a), and compare its form to the approximate UMPU test derived in b).

$$SC_n = [\nabla l(\theta)]' (I_n(\theta))^{-1} [\nabla l(\theta)] \Big|_{\theta = \tilde{\theta}}$$

As in a), $\mathcal{L}(\beta_0, \beta_1 | \underline{Y}) = \left(\prod_{i=1}^{n} (1 + e^{\beta_0 + \beta_1 x_i}) \right)^{-1} e^{\beta_1 \sum_i x_i Y_i + \beta_0 \sum_i Y_i}$

$$\Rightarrow l(\beta_0, \beta_1 | \underline{Y}) = -\sum_{i=1}^{n} \log(1 + e^{\beta_0 + \beta_1 x_i}) + \beta_1 \sum_i x_i Y_i + \beta_0 \sum_i Y_i$$

$$\Rightarrow \frac{\partial l}{\partial \beta_0} = -\sum_{i=1}^{n} \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})} + \sum_i Y_i$$

$$\Rightarrow \frac{\partial^2 l}{\partial \beta_0^2} = -\sum_{i=1}^{n} \underbrace{\left[\frac{-e^{2(\beta_0 + \beta_1 x_i)}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} + \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})} \right]}_{\text{from chain rule}} = -\sum_{i=1}^{n} \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2}$$

$$\Rightarrow \frac{\partial l}{\partial \beta_1} = -\sum_{i=1}^{n} \frac{x_i e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})} + \sum_i x_i Y_i$$

$$\Rightarrow \frac{\partial^2 l}{\partial \beta_1^2} = -\sum_{i=1}^{n} \left[\frac{-x_i^2 e^{2(\beta_0 + \beta_1 x_i)}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} + \frac{x_i^2 e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})} \right] = -\sum_{i=1}^{n} \frac{x_i^2 e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2}$$

$$\Rightarrow \frac{\partial l}{\partial \beta_1 \partial \beta_0} = \frac{\partial}{\partial \beta_1} \left[-\sum_{i=1}^{n} \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})} \right] = -\sum_{i=1}^{n} \frac{-x_i e^{2(\beta_0 + \beta_1 x_i)}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} + \frac{x_i e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})}$$

$$= -\sum_{i=1}^{n} \frac{x_i e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2}$$

$$\Rightarrow I_n(\theta) = \begin{bmatrix} \sum_{i=1}^{n} \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} & \sum_{i=1}^{n} \frac{x_i e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \\ \sum_{i=1}^{n} \frac{x_i e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} & \sum_{i=1}^{n} \frac{x_i^2 e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \end{bmatrix}$$

cont'd. →

2c) cont'd.

Then, under $H_0: \beta_1 = 0$, have

$$\mathcal{L}(\beta_0, \beta_1 = 0 \mid \underset{\sim}{y}) = \left[\prod_{i=1}^{n}(1+e^{\beta_0})\right]^{-1} e^{\beta_0 \sum_i y_i}$$

$$\Rightarrow \ell(\beta_0, \beta_1 = 0 \mid \underset{\sim}{y}) = -\sum_{i=1}^{n} \log(1+e^{\beta_0}) + \beta_0 \sum_i y_i$$

$$\Rightarrow \frac{\partial \ell}{\partial \beta_0} = -\sum_{i=1}^{n} \frac{e^{\beta_0}}{(1+e^{\beta_0})} + \sum_i y_i \overset{set}{=} 0$$

$$\Rightarrow \underbrace{\sum_{i=1}^{n} \frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})} = \sum_i y_i}_{} \quad \tilde{\beta}_0 = \text{logit}(\bar{y})$$

or

$$\text{Thus, } \tilde{\beta} = \begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{pmatrix} = \begin{pmatrix} \text{logit}(\bar{y}) \\ 0 \end{pmatrix} \quad \text{or simply } \sum_{i=1}^{n} \frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})} = \sum_{i=1}^{n} y_i$$

$$\Rightarrow \nabla \ell(\tilde{\beta})' = \Bigg(\underbrace{-\sum_{i=1}^{n} \frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})}}_{\sum_{i=1}^{n} y_i} + \sum_i y_i, \; -\sum_{i=1}^{n} \frac{x_i e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})} + \sum_{i=1}^{n} x_i y_i\Bigg)$$

$$= \left(0, \; -\frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i y_i\right)$$

and
$$I_n(\tilde{\beta})^{-1} = \begin{bmatrix} \sum_{i=1}^{n} \frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})^2} & \sum_{i=1}^{n} \frac{x_i e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})^2} \\ \sum_{i=1}^{n} \frac{x_i e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})^2} & \sum_{i=1}^{n} \frac{x_i^2 e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})^2} \end{bmatrix}^{-1}. \quad \text{Let } \tilde{c} = \frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})^2}, \text{ then}$$

$$I_n(\tilde{\beta})^{-1} = \frac{1}{\tilde{c}(n\sum_i x_i^2 - (\sum_i x_i)^2)} \begin{bmatrix} \tilde{c}\sum_{i=1}^{n} x_i^2 & -\tilde{c}\sum_{i=1}^{n} x_i \\ -\tilde{c}\sum_{i=1}^{n} x_i & n\tilde{c} \end{bmatrix}$$

$$\text{Then, } SC_N = \frac{1}{\tilde{c}(n\sum_i x_i^2 - (\sum_i x_i)^2)} \left(0, -\frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})}\sum_i x_i + \sum_i x_i y_i\right) \begin{bmatrix} \tilde{c}\sum_{i=1}^{n} x_i^2 & -\tilde{c}\sum_{i=1}^{n} x_i \\ -\tilde{c}\sum_{i=1}^{n} x_i & n\tilde{c} \end{bmatrix} \begin{pmatrix} 0 \\ -\frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})}\sum_i x_i \\ + \sum_i x_i y_i \end{pmatrix}$$

$$= \frac{1}{\tilde{c}(n\sum_i x_i^2 - (\sum_i x_i)^2)} \left(-\tilde{c}\sum_i x_i \left[-\frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})}\sum_i x_i + \sum_i x_i y_i\right] + n\tilde{c}\left[-\frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})}\sum_i x_i + \sum_i x_i y_i\right]\right) \begin{pmatrix} 0 \\ -\frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})}\sum_i x_i \\ + \sum_i x_i y_i \end{pmatrix}$$

$$= \frac{-n\cancel{\tilde{c}}}{\cancel{\tilde{c}}(n\sum_i x_i^2 - (\sum_i x_i)^2)} \left[-\frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})}\sum_i x_i + \sum_i x_i y_i\right]^2$$

cont'd next pg.

$$\Rightarrow SC_n = \frac{-1}{\left(\sum_i x_i^2 - \underbrace{\tfrac{1}{n}\left(\sum_i x_i\right)^2}_{n\bar{x}^2}\right)} \left[\sum_i x_i Y_i - \frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})} \sum_i x_i \right]^2$$

$$= \frac{1}{\left(n\bar{x}^2 - \sum_i x_i^2\right)} \left[\sum_i x_i Y_i - \frac{e^{\tilde{\beta}_0}}{(1+e^{\tilde{\beta}_0})} \sum_i x_i \right]^2 \xrightarrow[H_0]{d} \chi_1^2$$

Reject H_0 when $SC_n > 3.84$.

As in part b), we reject the null when $\sum_i x_i Y_i$ is too small or too large since the numerator term is squared.

Also, note the connection between the standard normal distn. (asymptotic distribution of normalized estimator in b) and the χ_1^2 distribution.

AMW

2d) Given p covariates.

$$P(y_i = 1 \mid \beta, x_i) = \frac{\exp(x_i'\beta)}{(1+\exp(x_i'\beta))}$$

where $x_i = (x_{i1}, \ldots, x_{ip})'$ is $p \times 1$ and $\beta = (\beta_1, \ldots, \beta_p)'$ is $p \times 1$.

Want to test $H_0: \ell'\beta = \theta_0$ vs. $H_1: \ell'\beta \neq \theta_0$.

Aim: Derive the UMPU level α test and express the rejection region and critical value in simplest form.

Step 1 show a member of multiparameter exp. family and define CSS for both parameter of interest & any nuisance parameters.

Want to test $H_0: \ell'\beta = \theta_0$ vs. $H_1: \ell'\beta \neq \theta_0$

$\iff H_0: \underbrace{\ell'\beta - \theta_0}_{} = 0$ vs. $H_1: \underbrace{\ell'\beta - \theta_0}_{} \neq 0$

$\iff H_0: \theta = 0$ vs. $H_1: \theta \neq 0$

Then, since $\theta = \ell'\beta - \theta_0 \iff \theta + \theta_0 - \ell'\beta = 0 \iff \theta + \theta_0 - \ell_1\beta_1 - \ell_2\beta_2 - \ldots - \ell_n\beta_n = 0$

$$\iff \beta_1 = \frac{\theta + \theta_0 - \ell_{(-1)}'\beta_{(-1)}}{\ell_1}$$

Then, $$\mathcal{L}(\beta \mid \underset{\sim}{X}, \underset{\sim}{Y}) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} = \prod_{i=1}^{n}\left[\frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})}\right]^{y_i}\left[1 - \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})}\right]^{1-y_i}$$

$$= \prod_{i=1}^{n}\left[\frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})}\right]^{y_i}\left[\frac{1}{(1+e^{x_i'\beta})}\right]^{1-y_i} = \underbrace{\left[\prod_{i=1}^{n}(1+e^{x_i'\beta})\right]^{-1}}_{c(\beta)} \exp\left\{\textstyle\sum_i y_i x_i'\beta\right\}$$

$$= c(\beta)\exp\left\{\beta_1 \textstyle\sum_i y_i x_{i1} + \beta_2 \sum_i y_i x_{i2} + \ldots + \beta_n \sum_i y_i x_{in}\right\}$$

$$= c(\beta)\exp\left\{\left(\frac{\theta + \theta_0 - \ell_{(-1)}'\beta_{(-1)}}{\ell_1}\right)\textstyle\sum_i y_i x_{i1} + \beta_2 \sum_i y_i x_{i2} + \ldots + \beta_n \sum_i y_i x_{in}\right\}$$

$$= c(\beta)\exp\left\{\left[\left(\frac{\theta+\theta_0}{\ell_1}\right)\textstyle\sum_i y_i x_{i1} - \left(\frac{\ell_2\beta_2}{\ell_1}\right)\sum_i y_i x_{i1} - \ldots - \left(\frac{\ell_n\beta_n}{\ell_1}\right)\sum_i y_i x_{i1}\right] + \beta_2\sum_i y_i x_{i2} + \ldots + \beta_n \sum_i y_i x_{in}\right\}$$

$$= c(\beta)\exp\left\{(\theta+\theta_0)\textstyle\sum_i y_i(x_{i1}/\ell_1) + \beta_2 \sum_i y_i\left[x_{i2} - (\ell_2/\ell_1)x_{i1}\right] + \ldots + \beta_p \sum_i y_i\left[x_{ip} - (\ell_p/\ell_1)x_{i1}\right]\right\}$$

$\Rightarrow \theta$ is parameter of interest

$S_1 = \sum_i y_i(x_{i1}/\ell_1)$ is the CSS associated w/ parameter of interest

$S_{-1} = (S_2, \ldots, S_p) = \left(\sum_i y_i(x_{i2} - \ell_2/\ell_1)x_{i1}, \ldots, \sum_i y_i(x_{ip} - \ell_p/\ell_1)x_{i1}\right)$ are the CSS associated w/ nuisance parameters.

AMW

2 d) <u>Step 2</u> Write UMPU level α test

$$\emptyset(U) = \begin{cases} 1 & \text{if } U < C_1(T) \text{ or } U > C_2(T) \\ \gamma_1 & \text{if } U = C_1(T) \\ \gamma_2 & \text{if } U = C_2(T) \\ 0 & \text{else} \end{cases} = \begin{cases} 1 & \text{if } S_1 < C_1(S_{(-1)}) \text{ or } S_1 > C_2(S_{(-1)}) \\ \gamma_1 & \text{if } S_1 = C_1(S_{(-1)}) \\ \gamma_2 & \text{if } S_1 = C_2(S_{(-1)}) \\ 0 & \text{else} \end{cases}$$

where $\alpha = \underbrace{E_0[\emptyset(u) \mid T=t]}_{i)}$ & $\underbrace{E_0[U\emptyset(u) \mid T=t]}_{ii)} = \alpha \underbrace{E_0[U \mid T=t]}_{iii)}$

Let $A = \{y : S_2(y) = s_2, \dots, S_p(y) = s_p\}$

$B = \{y : S_1(y) \notin [c_1, c_2]\}$

$B_1 = \{y : S_1(y) = c_1\}$

$B_2 = \{y : S_2(y) = c_2\}$

Then, (i) $E_0[\emptyset(S_1) \mid S_{(-1)}] = P_0(S_1 \notin (c_1, c_2) \mid S_{(-1)}) + \gamma_1 P_0(S_1 = c_1 \mid S_{(-1)}) + \gamma_2 P_0(S_1 = c_2 \mid S_{(-1)})$

$$= \frac{P_0(Y \in (A \cap B)) + \gamma_1 P_0(Y \in (A \cap B_1)) + \gamma_2 P_0(Y \in (A \cap B_2))}{P_0(Y \in A)}$$

$$= \frac{\sum_{y \in A} P_0(Y=y)\left[I(y \in B) + \gamma_1 I(y \in B_1) + \gamma_2 I(y \in B_2)\right]}{\sum_{y \in A} P_0(Y=y)}$$

$$= \frac{\sum_{y \in A} c(\beta_0) \exp\{(\overset{0 \text{ under } H_0}{\cancel{\beta}} + \theta_0) S_1(y) + \dots + \beta_p S_p(y)\}\left[I(y \in B) + \gamma_1 I(y \in B_1) + \gamma_2 I(y \in B_2)\right]}{\sum_{y \in A} c(\beta_0) \exp\{(\overset{0 \text{ under } H_0}{\cancel{\beta}} + \theta_0) S_1(y) + \dots + \beta_p S_p(y)\}}$$

<u>Note</u>: Since we conditioned on $S_{(-1)}$, we can factor the terms involving $S_2(y), \dots, S_p(y)$ from the summation and cancel in numerator & denominator.

$$= \frac{\sum_{y \in A} \exp\{\theta_0 S_1(y)\}\left[I(y \in B) + \gamma_1 I(y \in B_1) + \gamma_2 I(y \in B_2)\right]}{\sum_{y \in A} \exp\{\theta_0 S_1(y)\}}$$

cont'd next pg. $\longrightarrow$

2016 Qual, Section 1

(ii) $E_0[S_1 \phi(S_1) \mid S_{(-1)}] = \dfrac{\sum_{y \in A} S_1(y) \exp\{\theta_0 S_1(y)\} (I(y \in B) + \gamma_1 I(y \in B_1) + \gamma_2 I(y \in B_2))}{\sum_{y \in A} \exp\{\theta_0 S_1(y)\}}$ ← Similar to above work

(iii) $E_0[S_1 \mid S_{(-1)}] = \dfrac{\sum_{y \in A} S_1(y) \exp\{\theta_0 S_1(y)\}}{\sum_{y \in A} \exp\{\theta_0 S_1(y)\}}$

Compute the above by enumerating all possible values of $Y \in A$.

Select the values of c_1, c_2, γ_1, and γ_2 so that the above criteria are met in addition to the requirement that $(\gamma_1, \gamma_2) \in [0,1] \times [0,1]$.

2e) Describe, in detail, a non-parametric bootstrap algorithm for computing the exact p-value based on the the UMPU test of part d).

Define: $Y' = [y_1, \dots, y_n]$; $\tilde{X} = [\tilde{X}_1, \tilde{X}_2]$

where $\tilde{X}_1 = X_1 / l_1$ is an $n \times 1$ matrix

$\tilde{\beta} = (\theta_0, \theta, \beta_2, \dots, \beta_p)$ $\quad \tilde{X}_2 = [X_2 - (l_2/l_1)X_1, \dots, X_p - (l_p/l_1)X_1]$ is a $n \times p-1$ matrix

$\Rightarrow \mathcal{L} = c(\beta) \exp\{Y' \tilde{X} \tilde{\beta}\}$

Then, the non-parametric bootstrap is conducted as follows:

Note: Think of this as like the $[c_1, c_2]$ in the UMPU test.

1) Define a plausible range $[e_1, e_2]$ for the test statistic $S_1 = \sum y_i (x_{i1}/l_1)$

2) Create B independent bootstrap samples by sampling n rows from the matrix $[Y, \tilde{X}_2]$ w/ replacement

3) For each sample, b, perform the following steps:

(3.1) Define $X_b = [X_1, \tilde{X}_{2b}]$ where $\tilde{X}_{2b}$ are the covariate terms that were sampled and X_1 remains in its original ordering. (b-th sample)

In this way, we shuffle the X_1 values to break the association between Y and θ, but still preserve the relationships between Y and the remaining parameters.

(3.2) Compute $S_{1b} = \sum y_i (x_{i1b}/l_1)$ for this sample.

(3.3) Compute $t_b = \min(S_b - e_1, e_2 - S_b)$

Note: Step 1) and this t_b allows us to compute a two-sided test which mirrors the UMPU test.

4) Compute S_1 for the original data, such that $t_0 = \min(S_1 - e_1, e_2 - S_1)$

(5) Compute the exact p-value using $P_{boot} = \dfrac{\sum_{i=1}^{B} I\{t_b \le t_0\}}{B}$