BASIC PHD WRITTEN EXAMINATION

THEORY, SECTION 1

(9:00 AM–1:00 PM, July 29, 2020)

INSTRUCTIONS:

(a) This is a **CLOSED-BOOK** examination.

(b) The time limit for this examination is four hours.

(c) Answer both questions that follow.

(d) Put the answers to different questions on separate sets of paper.

(e) Put your exam code, **NOT YOUR NAME**, on each page. The same code is used for Section 1 and Section 2 of the PhD Theory Exam. Please keep the code confidential and do not share this information with any students or faculty. Sharing your code with either students or faculty is viewed as a violation of the UNC honor code.

(f) Return the examination with a signed statement of the UNC honor pledge, separately from your answers. The pledge statement is given on the last page of the exam handout.

(g) In the questions to follow, you are required to answer only what is asked, and not to tell all you know about the topics involved.

2020 S1 Q1

$N \sim Poi(\lambda)$. $Z_1, Z_2, \cdots \sim$ iid $Exp(1/\mu) \overset{d}{=} Gamma(1, 1/\mu)$

(a) $Pr(X \le t) = Pr(\mathbb{1}\{N > 0\} \max_{1 \le j \le N} Z_j \le t)$. $X \ge 0$ always.

① $t = 0$

$Pr(X \le 0) = Pr(X = 0) = Pr(\mathbb{1}\{N>0\} = 0) = Pr(N = 0) = e^{-\lambda}$

② $t > 0$

$Pr(X \le t) = Pr(X \le t, N = 0) + Pr(X \le t, N > 0)$

$= Pr(N=0) + Pr(\max_{1\le j \le N} Z_j \le t, N > 0) = Pr(N=0) + \sum_{m=1}^{\infty} Pr(\max_{1 \le j \le N} Z_j \le t \mid N = m) \cdot P(N=m)$

Let $m \in \mathbb{N}$.

Then, $Pr(\max_{1 \le j \le m} Z_j \le t) = Pr(\forall Z_j \le t, 1 \le j \le m) = Pr(Z_1 \le t)^m = \left(\int_0^t \mu e^{-\mu x} dx\right)^m$

$= \left(\left[-e^{-\mu x}\right]_0^t\right)^m$

$= (1 - e^{-\mu t})^m$

$\therefore Pr(X \le t) = e^{-\lambda} + \sum_{m=1}^{\infty} (1 - e^{-\mu t})^m \cdot e^{-\lambda} \frac{\lambda^m}{m!}$

$= e^{-\lambda} \cdot \sum_{m=0}^{\infty} \frac{1}{m!} \left(\lambda(1 - e^{-\mu t})\right)^m = e^{-\lambda} \cdot \exp\left(\lambda(1 - e^{-\mu t})\right)$

$= \exp(-\lambda e^{-\mu t})$

$\therefore Pr\{X \le t\} = \exp(-\lambda e^{-\mu t}), \ \forall t \in [0, \infty)$

(b) $\hat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i = 0\}$, $\hat{\beta}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \le 1\}$

Let $Y_i = \begin{bmatrix} \mathbb{1}\{X_i = 0\} \\ \mathbb{1}\{X_i \le 1\} \end{bmatrix}$. $\mathbb{E} Y_i = \begin{bmatrix} Pr\{X_i = 0\} \\ Pr\{X_i \le 1\} \end{bmatrix} = \begin{bmatrix} \exp(-\lambda) \\ \exp(-\lambda e^{-\mu}) \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

$Var(Y_i) = \begin{bmatrix} Pr(X_i=0)(1 - Pr(X_i=0)) & Pr(X_i=0) - Pr(X_i=0) \cdot Pr(X_i \le 1) \\ Pr(X_i=0) - Pr(X_i=0) Pr(X_i \le 1) & Pr(X_i \le 1)(1 - Pr(X_i \le 1)) \end{bmatrix}$

$= \begin{bmatrix} \alpha(1-\alpha) & \alpha(1-\beta) \\ \alpha(1-\beta) & \beta(1-\beta) \end{bmatrix}$: finitely exist

(b) cont'ed

CLT: $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}Y_1\right) \to_d N(0, \mathrm{Var}\, Y_1)$.

(c) SLLN: $\hat{\alpha}_n = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i = 0\} \to_{a.s} \mathbb{E}\, 1\{X_i=0\} = \alpha$

$\hat{\beta}_n = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \le 1\} \to_{a.s} \mathbb{E}\, 1\{X_i \le 1\} = \beta$

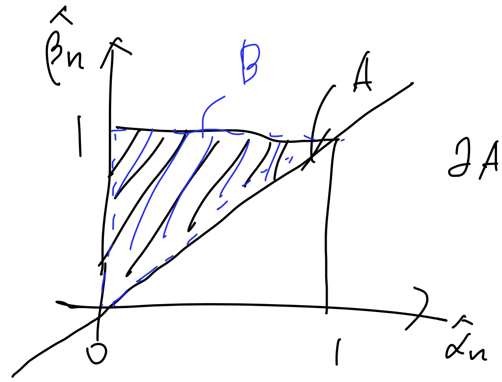
Continuous Mapping Thm: $g(x,y) = \left(-\log x,\ -\log\left[\log(y)/\log(x)\right]\right)$

$(x,y) \in (0,1)\times(0,1) \setminus \{(x,x) : x \in (0,1)\}$

$\Pr(B) \to_n 1$

$\Pr(A \setminus B) \le \Pr(\hat{\alpha}_n = 0) + \Pr(\hat{\beta}_n = 1)$

$+ \Pr(\hat{\alpha}_n = \hat{\beta}_n)$



$\partial A = B$

$\hat{\alpha}_n \le \hat{\beta}_n$

$\Pr(\hat{\alpha}_n = 0) \le \Pr\left(|\hat{\alpha}_n - \alpha| \ge \alpha\right) \to 0$

$\Pr(\hat{\beta}_n = 1) \le \Pr\left(|\hat{\beta}_n - \beta| \ge 1 - \beta\right) \to 0$

$\Pr(\hat{\alpha}_n = \hat{\beta}_n) = \Pr\left(\forall X_i \in \{0\} \cup (1, \infty)\right)$

$= \left\{\Pr(X=0) + \Pr(X>1)\right\}^n$

$= \{\underbrace{e^{-\lambda} + 1 - \exp(-\lambda e^{-\mu})}_{<1}\}^n$

$\to 0$.

$\lambda > \lambda e^{-\mu}$

$-\lambda < -\lambda e^{-\mu}$

$\exp(-\lambda) < \exp(-\lambda e^{-\mu})$

	$1\{X_i=0\}$	$1\{X_i \le 1\}$
$X_i = 0$	1	1
$1 \ge X_i > 0$	0	1
$X_i > 1$	0	0

$\therefore\ g(\hat{\alpha}_n, \hat{\beta}_n) = \left(-\log \hat{\alpha}_n,\ -\log\left[\log(\hat{\beta}_n)/\log(\hat{\alpha}_n)\right]\right) = (\hat{\lambda}_n, \hat{\mu}_n)$

$\to_{a.s} g(\alpha, \beta) = (\lambda, \mu)$.

(d) $g: \mathbb{R}^2 \to \mathbb{R}^2$

$$\underset{2\times 1}{g(x+h)} = \underset{2\times 1}{g(x)} + \underset{2\times 2}{\frac{dg}{dx^T}} h + R(x,h)$$

Δ - method

$$\sqrt{n}\left(\begin{bmatrix}\hat{\alpha}_n \\ \hat{\beta}_n\end{bmatrix} - \begin{bmatrix}\alpha \\ \beta\end{bmatrix}\right) \to_d N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\alpha(1-\alpha) & \alpha(1-\beta) \\ \alpha(1-\beta) & \beta(1-\beta)\end{bmatrix}\right)$$

$$\frac{dg}{d(\alpha,\beta)^T} = \begin{bmatrix} \frac{\partial(-\log\alpha)}{\partial\alpha} & \frac{\partial(-\log\alpha)}{\partial\beta} \\ \frac{\partial -\log[\log(\beta)/\log(\alpha)]}{\partial\alpha} & \frac{\partial -\log[\log(\beta)/\log(\alpha)]}{\partial\beta} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{\alpha} & 0 \\ \frac{1}{\log\alpha}\cdot\frac{1}{\alpha} & -\frac{1}{\log\beta\cdot\beta} \end{bmatrix}$$

$$\sqrt{n}\left(g\begin{pmatrix}\hat{\alpha}_n \\ \hat{\beta}_n\end{pmatrix} - g\begin{pmatrix}\alpha \\ \beta\end{pmatrix}\right) \longrightarrow_d N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \frac{dg}{d(\alpha,\beta)^T}\begin{bmatrix}\alpha(1-\alpha) & \alpha(1-\beta) \\ \alpha(1-\beta) & \beta(1-\beta)\end{bmatrix}{}^T\right)$$

$\Rightarrow \sqrt{n}\,($

1. (25 points) Let N be Poisson distributed with parameter $0 < \lambda < \infty$, and let $Z_1, Z_2, \ldots$ be an i.i.d. sequence of exponential random variables with mean $1/\mu$, where $0 < \mu < \infty$, and which are independent of N. Let

$$X = 1\{N > 0\} \max_{1 \le j \le N} Z_j,$$

where $1\{A\}$ is the indicator of A. Let $X_1, \ldots, X_n$ be i.i.d. realizations of X, and define

$$\hat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i = 0\} \text{ and } \hat{\beta}_n = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \le 1\}.$$

Do the following:

(a) Show that $\text{pr}\{X \le t\} = \exp\left(-\lambda e^{-\mu t}\right)$, for all $0 \le t < \infty$.

(b) Show that

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_n - \alpha \\ \hat{\beta}_n - \beta \end{pmatrix} \to_d N\left(0, \begin{bmatrix} \alpha(1-\alpha) & \alpha(1-\beta) \\ \alpha(1-\beta) & \beta(1-\beta) \end{bmatrix}\right),$$

where $\alpha = e^{-\lambda}$ and $\beta = \exp\left(-\lambda e^{-\mu}\right)$, as $n \to \infty$.

(c) Let $\hat{\lambda}_n = -\log(\hat{\alpha}_n)$ and $\hat{\mu}_n = -\log\left[-\log(\hat{\beta}_n)/\hat{\lambda}_n\right]$. Show that $\hat{\lambda}_n$ and $\hat{\mu}_n$ converge almost surely to λ and μ, respectively, as $n \to \infty$.

(d) Let $\theta = \lambda - \mu$ and $\hat{\theta}_n = \hat{\lambda}_n - \hat{\mu}_n$. Show that $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2)$, as $n \to \infty$, and give the form of σ^2 in terms of λ and μ.

(e) Construct an asymptotically valid hypothesis test of $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

Points: (a) 5; (b) 5; (c) 4; (d) 6; (e) 5.

2. (25 points) Consider a linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{X} \in \mathcal{R}^{n\times p}$ is a fixed design matrix and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ are i.i.d samples with $\mathrm{E}(\epsilon_1) = 0$, $\mathrm{Var}(\epsilon_1) = \sigma^2$ and $\mathrm{E}|\epsilon_1|^3 < \infty$. $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_p^*)'$ is the vector of true covariate coefficients. Suppose we consider estimating $\boldsymbol{\beta}^*$ by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ n^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda_1\boldsymbol{\beta}'\boldsymbol{\beta} + \lambda_2\sum_{j=1}^{p} I(\beta_j \neq 0), \tag{1}$$

where λ_1 and λ_2 are positive numbers, and $I(\cdot)$ is the indicator function.

(a) To solve (1), we start with a univariate problem. Let z be a real number. Prove that the function

$$f(\theta) = (z-\theta)^2 + \lambda_1\theta^2 + \lambda_2 I(\theta \neq 0),$$

is minimized at $\theta = z(1+\lambda_1)^{-1}I(|z| > \sqrt{(1+\lambda_1)\lambda_2})$.

(b) Derive the Majorization-Minimization algorithm to solve (1). Give closed-form expressions on how iterations need to be done.

In the following questions, we assume p is fixed; $\max_{i,j}|X_{ij}| \le 1$, where X_{ij} is the (i,j)th element of $\boldsymbol{X}$; $(1/n)\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$, where $\boldsymbol{I}$ is an identity matrix; and we choose $\lambda_1 = 0$ and $\lambda_2 = n^{-1}$.

(c) Prove that the solution to (1) is given by $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$, where

$$\hat{\beta}_j = (n^{-1}\sum_{i=1}^{n} Y_i X_{ij}) I(|n^{-1}\sum_{i=1}^{n} Y_i X_{ij}| > n^{-1/2}). \tag{2}$$

(d) Let $\mathcal{M} = \{j : \beta_j^* > 0\}$. Prove that for $j \in \mathcal{M}$, $P(|n^{-1}\sum_{i=1}^{n} Y_i X_{ij}| > n^{-1/2}) \to 1$ and

$$\sqrt{n}\{I(|n^{-1}\sum_{i=1}^{n} Y_i X_{ij}| > n^{-1/2}) - 1\} \xrightarrow{P} 0.$$

(e) Derive the limiting distribution of $\sqrt{n}(\hat{\beta}_j - \beta_j^*)$ for $j \in \mathcal{M}$, where $\hat{\beta}_j$ is given in (2). Hint: Write $\sqrt{n}(\hat{\beta}_j - \beta_j^*)$ as

$$\sqrt{n}(\hat{\beta}_j - \beta_j^*) = \sqrt{n}\left(n^{-1}\sum_{i=1}^{n} Y_i X_{ij} - \beta_j^*\right) I(|n^{-1}\sum_{i=1}^{n} Y_i X_{ij}| > n^{-1/2})$$
$$+ \sqrt{n}\beta_j^*\left\{I(|n^{-1}\sum_{i=1}^{n} Y_i X_{ij}| > n^{-1/2}) - 1\right\}.$$

Points: (a) 5; (b) 5; (c) 4; (d) 6; (e) 5.

Statement of the UNC honor pledge:

"In recognition of and in the spirit of the honor code, I certify that I have neither given nor received aid on this examination and that I will report all Honor Code violations observed by me."

(Signed) ______________________________
NAME

(Printed) ______________________________
NAME

2020 S1

Q1.

$N \sim Poi(\lambda)$ $\quad Z_1, \dots \sim$ iid $Exp(\frac{1}{\mu}) = Ga(1, \frac{1}{\mu})$

$X = 1\{N>0\} \max_{1\le j\le N} Z_j$. $\rightarrow X_1, \dots, X_n$: iid X. $\quad \hat{\alpha}_n = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i = 0\}$ $\quad \hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \le 1\}$

(a) $P\{X \le t\} = \left[\begin{array}{ll} P\{X\le 0\}, & t=0 \\ P\{X \le t\}, & t>0 \end{array}\right. = \left[\begin{array}{ll} P\{N=0\}, & t=0 \\ \mathbb{E}\left[P\left(1\{N>0\} \max_{1\le j\le N} Z_j \le t \mid N\right)\right], & t>0 \end{array}\right.$

$P\left(1\{N>0\} \max_{1\le j\le N} Z_j \le t \mid N = N_0\right)$ ← realisation of N

$= P\left(1\{N_0>0\} Z_{(N_0)} \le t\right) \quad (t>0)$

$= \left[\begin{array}{ll} P(0\le t), & N_0 = 0 \\ P(Z_{(N_0)} \le t), & N_0 > 0 \end{array}\right. = \left[\begin{array}{ll} 1, & N_0 = 0 \\ P(Z_1 \le t)^{N_0}, & N_0>0 \end{array}\right. = P(Z_1 \le t)^{N_0} = \left[\int_0^t \mu e^{-\mu z} dz\right]^{N_0}$

$= \left(\left[-e^{-\mu z}\right]_0^t\right)^{N_0}$

$= (1 - e^{-\mu t})^{N_0}$

$\Rightarrow P\{X \le t\} = \left[\begin{array}{ll} P\{N=0\}, & t=0 \\ \mathbb{E}\left[(1-e^{-\mu t})^N\right], & t>0 \end{array}\right.$

$= \left[\begin{array}{ll} e^{-\lambda}, & t=0 \\ \sum_{N_0=0}^{\infty} (1-e^{-\mu t})^{N_0} \cdot \frac{e^{-\lambda}\lambda^{N_0}}{N_0!}, & t>0 \end{array}\right.$

$= \left[\begin{array}{ll} e^{-\lambda}, & t=0 \\ \exp(-\lambda + \lambda(1-e^{-\mu t})), & t>0 \end{array}\right.$

$= \exp(-\lambda e^{-\mu t}), \quad \forall t \in [0, \infty)$

(b) $Y_i := \begin{bmatrix} 1\{X_i = 0\} \\ 1\{X_i \le 1\} \end{bmatrix}$. $\quad \mathbb{E} Y_i = \begin{bmatrix} P(X_i=0) \\ P(X_i\le 1) \end{bmatrix} = \begin{bmatrix} P(X_i \le 0) \\ P(X_i \le 1) \end{bmatrix} = \begin{bmatrix} e^{-\lambda} \\ \exp(-\lambda e^{-\mu}) \end{bmatrix} = \begin{bmatrix} \alpha \\ p \end{bmatrix}$

$\mathrm{Var}\, Y_i = \begin{bmatrix} P(X_i=0) - P(X_i=0)^2 & \mathbb{E}\, 1\{X_i=0\}\cdot 1\{X_i\le 1\} - P(X_i=0)\cdot P(X_i \le 1) \\ ''' & P(X_i \le 1) - P(X_i\le 1)^2 \end{bmatrix}$

$= \begin{bmatrix} \alpha(1-\alpha) & \alpha - \alpha p \\ \alpha - \alpha p & p(1-p) \end{bmatrix}$ $\quad (1\{X_i=0\}\cdot 1\{X_i \le 1\} = 1\{X_i = 0\})$

∵ exists

CLT: $\sqrt{n}\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} Y_i}_{= \begin{bmatrix} \hat{\alpha}_n \\ \hat{p}_n \end{bmatrix}} - \mathbb{E} Y_1\right) \rightarrow_d N(0, \mathrm{Var}\, Y_1)$

(c) SLLN: $\hat{\alpha}_n = \frac{1}{n}\sum_i 1\{X_i=0\} \xrightarrow{a.s} E1\{X_1=0\} = P(X_1=0) = \alpha > 0$

$$\hat{\beta}_n = \frac{1}{n}\sum_i 1\{X_i \le 1\} \xrightarrow{a.s} E1\{X_1\le 1\} = P(X_1 \le 1) = \beta > 0$$

CMT: $\hat{\lambda}_n = -\log(\hat{\alpha}_n) \xrightarrow{a.s} -\log(\alpha) = \lambda$

$$-\log(\hat{\beta}_n) \xrightarrow{a.s} -\log(\beta) = \lambda e^{-\mu}$$

Then, $\begin{bmatrix} \hat{\lambda}_n \\ -\log(\hat{\beta}_n) \end{bmatrix} \xrightarrow{a.s} \begin{bmatrix} \lambda \\ \lambda e^{-\mu} \end{bmatrix}$.

CMT: (with $g(x,y) = -\log[y/x]$)

$$g(\hat{\lambda}_n, -\log(\hat{\beta}_n)) \xrightarrow{a.s} g(\lambda, \lambda e^{-\mu}) = \mu$$

$$\Rightarrow \quad -\log[-\log(\hat{\beta}_n)/\hat{\lambda}_n] \xrightarrow{a.s} \mu.$$

Note that $P(\hat{\alpha}_n = 0) = P(\forall X_i > 0) = P(X_1 > 0)^n = (1 - P(X_1 \le 0))^n = (1-e^{-\lambda})^n \xrightarrow[n\to\infty]{} 0$

$$P(\hat{\beta}_n = 0) = P(\forall X_i > 1) = P(X_1>1)^n = (1-P(X_1\le 1))^n = (1-\exp(-\lambda e^{-\mu}))^n \xrightarrow[n\to\infty]{} 0$$

$$P\left(\frac{-\log(\hat{\beta}_n)}{\hat{\lambda}_n} = 0\right) = P(\hat{\beta}_n = 1) = P(\forall X_i \le 1) = P(X_1\le 1)^n = (\exp(-\lambda e^{-\mu}))^n \xrightarrow[n\to\infty]{} 0$$

$$P(\hat{\lambda}_n = 0) = P(\hat{\alpha}_n = 1) = P(\forall X_i = 0) = P(X_1 = 0)^n = e^{-\lambda n} \xrightarrow[n\to\infty]{} 0$$

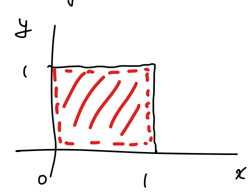
$\Rightarrow$ Domain of the Continuous ft^n well established

(d)

$$\sqrt{n}\left(\begin{bmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{bmatrix} - \begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right) \xrightarrow{d} N(0, \Sigma)$$

Let $g(x,y) = -\log x + \log[\log(y)/\log x]$, where $(x,y) \in [0,1]^2$, $x \ne 0$, $y \ne 0$, $\frac{\log(y)}{\log x} \ne 0$, $\log x \ne 0$

Then, g: differentiable and ∇g is continuous on $(0,1)^2$, (i.e) $(x,y) \in (0,1)^2$



and $\frac{\partial g}{\partial (x,y)} = \begin{bmatrix} -x^{-1} - (\log x)^{-1} x^{-1} \\ (\log y)^{-1} y^{-1} \end{bmatrix}$

Since $P\left(\begin{bmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{bmatrix} \in (0,1)^2\right) \xrightarrow[n\to\infty]{} 1$, we can apply Δ-method.

Note that $g\left(\begin{bmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{bmatrix}\right) = -\log\hat{\alpha}_n + \log[\log(\hat{\beta}_n)/\log(\hat{\alpha}_n)] = \hat{\lambda}_n - \hat{\mu}_n$

$$g\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right) = -\log\alpha + \log[\log(\beta)/\log(\alpha)] = \lambda - \mu.$$

$\therefore \sqrt{n}\left((\hat{\lambda}_n - \hat{\mu}_n) - (\lambda - \mu)\right) \xrightarrow{d} N(0, \sigma^2)$, where

$\sigma^2 = \nabla g(\alpha,\beta)^T \Sigma \nabla g(\alpha,\beta)$, and $\nabla g(\alpha,\beta) = \begin{bmatrix} -\alpha^{-1} - (\log\alpha)^{-1}\alpha^{-1} \\ (\log\beta)^{-1}\beta^{-1} \end{bmatrix} = \begin{bmatrix} -e^{\lambda} + \frac{1}{\lambda}e^{\lambda} \\ (-\lambda e^{-\mu})^{-1}\exp(-\lambda e^{-\mu}) \end{bmatrix} = \begin{bmatrix} (\frac{1}{\lambda}-1)e^{\lambda} \\ -\frac{1}{\lambda}\exp(\mu + \lambda e^{-\mu}) \end{bmatrix}$

$$\therefore \Sigma_1 = \begin{bmatrix} -\frac{1}{\alpha} - \frac{1}{\alpha \log \alpha} \\ \frac{1}{\beta \log \beta} \end{bmatrix}^T \begin{bmatrix} \alpha(1-\alpha) & \alpha(1-\beta) \\ \alpha(1-\beta) & \beta(1-\beta) \end{bmatrix} \begin{bmatrix} -\frac{1}{\alpha} - \frac{1}{\alpha \log \alpha} \\ \frac{1}{\beta \log \beta} \end{bmatrix}$$

$$= \begin{bmatrix} \quad '' \quad \end{bmatrix}^T \begin{bmatrix} -(1-\alpha) - \frac{1-\alpha}{\log \alpha} + \frac{\alpha(1-\beta)}{\beta \log \beta} \\ -(1-\beta) - \frac{1-\beta}{\log \alpha} + \frac{1-\beta}{\log \beta} \end{bmatrix}$$

$$= \cancel{\frac{1-\alpha}{\alpha}} + \cancel{\frac{1-\alpha}{\alpha \log \alpha}} - \cancel{\frac{1-\beta}{\beta \log \beta}} + \frac{1-\alpha}{\alpha \log \alpha} + \cancel{\frac{1-\alpha}{\alpha (\log \alpha)^2}} - \frac{1-\beta}{\beta \log \beta \cdot \log \alpha}$$

$$- \cancel{\frac{1-\beta}{\beta \log \beta}} - \frac{1-\beta}{\beta \log \beta \log \alpha} + \cancel{\frac{1-\beta}{\beta (\log \beta)^2}}$$

$$= \frac{1-\alpha}{\alpha} + \frac{2(1-\alpha)}{\alpha \log \alpha} + \frac{1-\alpha}{\alpha (\log \alpha)^2} - \frac{2(1-\beta)}{\beta \log \beta} + \frac{1-\beta}{\beta (\log \beta)^2} - \frac{2(1-\beta)}{\beta \log \beta \log \alpha}$$

$\alpha = e^{-\lambda}$, $\beta = \exp(-\lambda e^{-\mu})$

$$\sigma^2 = (1-e^{-\lambda})e^{\lambda} + \frac{2(1-e^{-\lambda})}{e^{-\lambda} \cdot (-\lambda)} + \frac{1-e^{-\lambda}}{e^{-\lambda}(-\lambda)^2} - \frac{2(1-\exp(-\lambda e^{-\mu}))}{\exp(-\lambda e^{-\mu}) \cdot (-\lambda e^{-\mu})} + \frac{1-\exp(-\lambda e^{-\mu})}{\exp(-\lambda e^{-\mu})(-\lambda e^{-\mu})^2}$$

$$- \frac{2(1-\exp(-\lambda e^{-\mu}))}{\exp(-\lambda e^{-\mu}) \cdot (-\lambda e^{-\mu}) \cdot (-\lambda)}$$

$$= e^{\lambda} - 1 - \frac{2}{\lambda}(e^{\lambda}-1) + \frac{1}{\lambda^2}(e^{\lambda}-1) + \frac{2}{\lambda}(1-\exp(-\lambda e^{-\mu})) \cdot \exp(\mu + \lambda e^{-\mu})$$

$$+ \frac{1}{\lambda^2}(1-\exp(-\lambda e^{-\mu})) \cdot \exp(\lambda e^{-\mu} + 2\mu) - \frac{2}{\lambda^2}(1-\exp(-\lambda e^{-\mu})) \cdot \exp(\lambda e^{-\mu} + \mu)$$

$$= (e^{\lambda}-1)\left(1-\frac{1}{\lambda}\right)^2 + \frac{2}{\lambda} e^{\mu}(\exp(\lambda e^{-\mu}) - 1) + \frac{1}{\lambda^2} e^{2\mu}(\exp(\lambda e^{-\mu}) - 1)$$

$$- \frac{2}{\lambda^2} e^{\mu}(\exp(\lambda e^{-\mu}) - 1)$$

(e) We have $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2)$.

$\Rightarrow \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma} \to_d N(0,1) \Rightarrow P\left(\left|\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma}\right| > Z_{\frac{\alpha}{2}}\right) \approx \alpha$.

Under $H_0: \theta = 0$, $P\left(\left|\frac{\sqrt{n}\hat{\theta}_n}{\sigma}\right| > Z_{\frac{\alpha}{2}}\right) \overset{H_0}{\doteq} \alpha$ $H_1: \theta = 0$ indicates that $|\hat{\theta}_n|$ tends to be large.

$\therefore$ Reject H_0 if $|\hat{\theta}_n| > \frac{1}{\sqrt{n}} Z_{\frac{\alpha}{2}} \cdot \sigma$ if asymp valid test.

2020 S1 Q2

(a) $f(\theta) = (z-\theta)^2 + \lambda_1\theta^2 + \lambda_2 I(\theta \neq 0)$

① $\theta = 0$: $f(\theta) = z^2$

② $\theta \neq 0$: $f(\theta) = (z-\theta)^2 + \lambda_1\theta^2 + \lambda_2$: quadratic ftn of θ.

$$\Rightarrow \frac{\partial f}{\partial \theta}\Big|_{\theta=\hat\theta} \overset{set}{=} 0$$

$$\Leftrightarrow 2(\hat\theta - z) + 2\lambda_1\hat\theta = 0$$

$$\Leftrightarrow \hat\theta = z(1+\lambda_1)^{-1}$$

Now, compare $f(0)$ vs $f(\hat\theta)$

If $f(0) > f(\hat\theta)$ (i.e) $z^2 > (z - z(1+\lambda_1)^{-1})^2 + \lambda_1 z^2(1+\lambda_1)^{-2} + \lambda_2$

$$\Leftrightarrow 0 > -2z^2(1+\lambda_1)^{-1} + z^2(1+\lambda_1)^{-2} + \lambda_1 z^2(1+\lambda_1)^{-2} + \lambda_2$$

$$\Leftrightarrow 0 > z^2(1+\lambda_1)^{-2}(-2(1+\lambda_1) + 1 + \lambda_1) + \lambda_2$$

$$\Leftrightarrow z^2(1+\lambda_1)^{-1} > \lambda_2$$

$$\Leftrightarrow z^2 > (1+\lambda_1)\lambda_2,$$

then $\hat\theta_{min} = \hat\theta = z(1+\lambda_1)^{-1}$

Else $f(0) \le f(\hat\theta)$ (i.e) $z^2 \le (1+\lambda_1)\lambda_2$,

then $\hat\theta_{min} = 0$

$$\therefore \hat\theta_{min} = z(1+\lambda_1)^{-1} I\left(|z| > \sqrt{(1+\lambda_1)\lambda_2}\right)$$

(b) Let $L(\beta) = n^{-1}(Y - X\beta)^T(Y - X\beta) + \lambda_1\beta^T\beta$, $g_{\lambda_2}(\beta) = \lambda_2 \sum_{j=1}^{p} I(\beta_j \neq 0)$.

Then, $\hat\beta = \underset{\beta}{\operatorname{argmin}}\, L(\beta) + g_{\lambda_2}(\beta)$. Note that $L(\beta)$ is second-differentiable and every 2^{nd} order part. deriv. are conti.

By Taylor Expansion, assume $\tilde\beta$ is an arbitrary value of β,

$$L(\beta) = L(\tilde\beta) + \dot L(\tilde\beta)(\beta - \tilde\beta) + \frac{1}{2}(\beta-\tilde\beta)^T \ddot L(\bar\beta)(\beta - \tilde\beta), \text{ where } \bar\beta = t\beta + (1-t)\tilde\beta, \exists t \in [0,1].$$

$$\le L(\tilde\beta) + \dot L(\tilde\beta)(\beta - \tilde\beta) + \frac{c}{2}\|\beta - \tilde\beta\|^2 =: L(\beta|\tilde\beta) \text{ : majorization of } L(\beta)$$

where $c = \lambda_{max}(\ddot L(\bar\beta))$: the largest eigen value of $\ddot L(\bar\beta)$.

Here, $L(\beta) = n^{-1}(Y^TY - 2Y^TX\beta + \beta^TX^TX\beta) + \lambda_1\beta^T\beta$

$$\Rightarrow \dot L(\beta) = n^{-1}(-2X^TY + 2X^TX\beta) + 2\lambda_1\beta$$

$\ddot L(\beta) = 2n^{-1}X^TX + 2\lambda_1 I$: constant, positive definite. Thus, c is fixed.

Now, Major-Minimi algorithm is to update new solution from the current solution $\tilde\beta$

$$\hat\beta = \underset{\beta}{\operatorname{argmin}}\, L(\beta|\tilde\beta) + g_{\lambda_2}(\beta). = \underset{\beta}{\operatorname{argmin}}\, \dot L(\tilde\beta)\beta + \frac{c}{2}\|\beta - \tilde\beta\|^2 + g_{\lambda_2}(\beta).$$

Since it is convex optimization, we can apply coordinate gradient descent algorithm to get updating formula.

Assume $\tilde{\beta}$ is current solution. $\tilde{\beta}_{(-j)}$:= all $\tilde{\beta}$ components but $\tilde{\beta}_j$. Then, updating scheme for β_j is

$$\hat{\beta}_j = \underset{\beta_j}{\operatorname{argmin}}\ L(\tilde{\beta}_{(-j)}, \beta_j \mid \tilde{\beta}) + g_{\lambda_2}(\beta_{(-j)}, \beta_j)$$

$$= \underset{\beta_j}{\operatorname{argmin}}\ [\dot{L}(\tilde{\beta})]_j \cdot \beta_j + \frac{c}{2}(\beta_j - \tilde{\beta}_j)^2 + \lambda_2 I(\beta_j \neq 0)$$

$$= \underset{\beta_j}{\operatorname{argmin}}\ \frac{c}{2}\left[\beta_j^2 - 2\tilde{\beta}_j\beta_j + \frac{2}{c}[\dot{L}(\tilde{\beta})]_j\beta_j + \frac{2\lambda_2}{c}I(\beta_j \neq 0)\right]$$

$$= \underset{\beta_j}{\operatorname{argmin}}\ (\beta_j - \underbrace{(\tilde{\beta}_j - \frac{1}{c}[\dot{L}(\tilde{\beta})]_j)}_{=:z_j})^2 + 0 \cdot \beta_j^2 + \frac{2\lambda_2}{c} I(\beta_j \neq 0)$$

$$= z_j (1+0)^{-1} I\left(|z_j| > \sqrt{(1+0)\cdot\frac{2\lambda_2}{c}}\right)$$

$$= z_j I\left(|z_j| > \sqrt{\frac{2\lambda_2}{c}}\right).$$

$\therefore$ The whole algorithm is as follows:

1. Initialize $\hat{\beta}$ with minimizer of $n^{-1}(Y - X\beta)^T(Y - X\beta) + \lambda_1 \beta^T\beta$.
2. Given the current solution $\tilde{\beta}$, update its jth component by

$$\hat{\beta}_j = z_j I\left(|z_j| > \sqrt{\frac{2\lambda_2}{c}}\right), \quad \text{where} \quad z_j := \tilde{\beta}_j - \frac{1}{c}[\dot{L}(\tilde{\beta})]_j$$

$$= \tilde{\beta}_j - \frac{1}{c}\left[n^{-1}(-2X^TY + 2X^TX\tilde{\beta}) + 2\lambda_1\tilde{\beta}\right]_j$$

$$C = \lambda_{max}(2n^{-1}X^TX + 2\lambda_1 I)$$

$j = 1, \ldots, p$.

3. Iterate until convergence. That is, stop algorithm until

$$\|\hat{\beta} - \tilde{\beta}\|_2 < \varepsilon : \text{prespecified threshold.}$$

(c) $L(\beta) = n^{-1}\|Y - X\beta\|^2 + n^{-1}\sum_{j=1}^{p} I(\beta_j \neq 0)$

$$= n^{-1}(Y^TY - 2Y^TX\beta + \beta^TX^TX\beta) + n^{-1}\sum_{j=1}^{p} I(\beta_j \neq 0)$$

$$= n^{-1}Y^TY - \frac{2}{n}\sum_{j=1}^{p}(X^TY)_j\beta_j + \beta^T\beta + n^{-1}\sum_{j=1}^{p} I(\beta_j \neq 0)$$

$$= n^{-1}Y^TY + \sum_{j=1}^{p}\left(\beta_j^2 - \frac{2}{n}(X^TY)_j\beta_j + \frac{1}{n}I(\beta_j \neq 0)\right)$$

Therefore, minimizing $L(\beta)$ wrt β is equiv. to minimizing $\beta_j^2 - \frac{2}{n}(X^TY)_j\beta_j + \frac{1}{n}I(\beta_j \neq 0)$ for each $j = 1, \ldots, p$

$$\hat{\beta}_j = \underset{\beta_j}{\operatorname{argmin}}\ \beta_j^2 - \frac{2}{n}(X^TY)_j\beta_j + \frac{1}{n}I(\beta_j \neq 0) = \underset{\beta_j}{\operatorname{argmin}}\ (\beta_j - \frac{1}{n}(X^TY)_j)^2 + 0\beta_j^2 + \frac{1}{n}I(\beta_j \neq 0) \overset{\text{by (a)}}{=} \left(\frac{1}{n}(X^TY)_j\right)(1+0)^{-1} \cdot I\left(\left|\frac{1}{n}(X^TY)_j\right| > \frac{1}{\sqrt{n}}\right)$$

(d) $Y = X\beta^* + \varepsilon$

$$\sum_{i=1}^n Y_i x_{ij} = (X^T Y)_j = (X^T X \beta^* + X^T \varepsilon)_j = (n\beta^* + X^T\varepsilon)_j = n\beta_j^* + (X^T\varepsilon)_j$$

$$\varepsilon \sim (0, \sigma^2 I_n) \Rightarrow X^T\varepsilon = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{1p} & \cdots & & x_{np} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_{i1}\varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ip}\varepsilon_i \end{bmatrix}$$

$$\Rightarrow (X^T\varepsilon)_j = \sum_{i=1}^n x_{ij}\varepsilon_i . \quad E(X^T\varepsilon)_j = 0, \quad Var(X^T\varepsilon)_j = [Var(X^T\varepsilon)]_{j,j} = \frac{\sigma^2}{n}$$

<WTS> $P\left(\left|\beta_j^* + \frac{1}{n}\sum_{i=1}^n x_{ij}\varepsilon_i\right| > n^{-1/2}\right) \xrightarrow[n]{} 1$, where $\beta_j^* > 0$.

Enough large n, $n^{-1/2} - \beta_j^* < -\frac{1}{2}\beta_j^*$

Then, $\left|\frac{1}{n}\sum_{i=1}^n x_{ij}\varepsilon_i\right| \le \frac{1}{2}\beta_j^*$

$$\Rightarrow -\frac{1}{2}\beta_j^* \le \frac{1}{n}\sum_i x_{ij}\varepsilon_i \le \frac{1}{2}\beta_j^*$$

$$\Rightarrow \frac{1}{2}\beta_j^* \le \beta_j^* + \frac{1}{n}\sum_i x_{ij}\varepsilon_i \le \frac{3}{2}\beta_j^*$$

$$\Rightarrow \left|\beta_j^* + \frac{1}{n}\sum_i x_{ij}\varepsilon_i\right| \ge \frac{1}{2}\beta_j^* > n^{-1/2}.$$

Thus, $$P\left(\left|\beta_j^* + \frac{1}{n}\sum_i x_{ij}\varepsilon_i\right| > n^{-1/2}\right) \ge P\left(\left|\frac{1}{n}\sum_{i=1}^n x_{ij}\varepsilon_i\right| \le \frac{1}{2}\beta_j^*\right)$$

$$= 1 - P\left(\left|\frac{1}{n}\sum_i x_{ij}\varepsilon_i\right| > \frac{1}{2}\beta_j^*\right)$$

$$\ge 1 - \frac{E\left|\frac{1}{n}\sum x_{ij}\varepsilon_i\right|^2}{\left(\frac{1}{2}\beta_j^*\right)^2} \quad \text{(Chebyshev)}$$

$$= 1 - \frac{Var(\sum x_{ij}\varepsilon_i)}{n^2\left(\frac{1}{2}\beta_j^*\right)^2} = 1 - \frac{\sigma^2}{n^3\left(\frac{1}{2}\beta_j^*\right)^2} \xrightarrow[n\to\infty]{} 1$$

Now, let's show $\sqrt{n}\left\{I\left(\left|n^{-1}\sum_i Y_i x_{ij}\right| > n^{-1/2}\right) - 1\right\} \to_p 0$.

Let $\varepsilon > 0$.

$$P\left(\left|\sqrt{n}\left\{I\left(\left|n^{-1}\sum_i Y_i x_{ij}\right| > n^{-1/2}\right) - 1\right\} - 0\right| > \varepsilon\right)$$

$= P\left(\sqrt{n}\left\{1 - I\left(\left|n^{-1}\sum_i Y_i x_{ij}\right| > n^{-1/2}\right)\right\} > \varepsilon\right)$. For enough large n, (i.e. $n > \varepsilon^2$)

$$= P\left(I\left(\left|n^{-1}\sum_i Y_i x_{ij}\right| > n^{-1/2}\right) = 0\right)$$

$$= P\left(\left|n^{-1}\sum_i Y_i x_{ij}\right| \le n^{-1/2}\right) \xrightarrow[n\to\infty]{} 0.$$

(e)

$$\sqrt{n}\left(\hat{\beta}_j - \beta_j^*\right) = \sqrt{n}\left(n^{-1}\sum_i Y_i x_{ij} - \beta_j^*\right) \underbrace{I\left(\left|n^{-1}\sum_i Y_i x_{ij}\right| > n^{-1/2}\right)}_{\to_p 1 \text{ from (d)}}$$

$$+ \underbrace{\sqrt{n}\,\beta_j^* \left\{ I\left(\left|n^{-1}\sum_i Y_i x_{ij}\right| > n^{-1/2}\right) - 1 \right\}}_{\to_p 0 \text{ from (d)}}$$

Find $\sqrt{n}\left(n^{-1}\sum_i Y_i x_{ij} - \beta_j^*\right)$'s limiting dist^n.

$$\overset{\text{"}}{=}\ \sqrt{n}\left(n^{-1}(X^T\epsilon)_j\right) =: Z_n.$$

Characteristic ft^n of Z_n is given by

$$\mathbb{E}\, e^{itZ_n} = \mathbb{E}\exp\left(\frac{it}{\sqrt{n}} \sum_i x_{ij}\epsilon_i\right)$$

$$= \prod_i \mathbb{E}\exp\left(\frac{it}{\sqrt{n}} x_{ij}\epsilon_i\right)$$

$$= \prod_i \left(1 + \mathbb{E}\epsilon_i \cdot \left(\frac{it}{\sqrt{n}} x_{ij}\right) + \frac{1}{2!}\mathbb{E}\epsilon_i^2 \cdot \left(\frac{it}{\sqrt{n}} x_{ij}\right)^2 + o\left(\left(\frac{t}{\sqrt{n}} x_{ij}\right)^2\right)\right)$$

$$\Rightarrow \log \mathbb{E}\, e^{itZ_n} = \sum_i \log\left(1 + 0 + \frac{1}{2}\sigma^2 \frac{-t^2}{n} x_{ij}^2 + o\left(\frac{t^2}{n}\right)\right)$$

$$= \sum_i \log\left(1 - \frac{1}{2}\frac{\sigma^2}{n} x_{ij}^2 t^2 + o\left(\frac{t^2}{n}\right)\right)$$

$$= \sum_i \left(-\frac{1}{2}\frac{\sigma^2}{n} x_{ij}^2 t^2 + o\left(\frac{t^2}{n}\right)\right)$$

$$= -\frac{1}{2}\frac{\sigma^2}{n}\sum_i x_{ij}^2 t^2 + o(1)$$

$$= -\frac{1}{2}\frac{\sigma^2}{n}\cdot\left[X^TX\right]_{jj} t^2 + o(1)$$

$$= -\frac{1}{2}\sigma^2 t^2 + o(1)$$

$\log(1-x) = -\left(x + \frac{1}{2}x^2\right)$

$\log(1+x) \le x - \frac{1}{2},$

$\therefore$ $\mathbb{E}\, e^{itZ_n} \xrightarrow[n]{} \exp\left(-\frac{1}{2}\sigma^2 t^2\right)$: characteristic ft^n of $N(0, \sigma^2)$

$\therefore$ $Z_n \to_d N(0, \sigma^2)$

$\therefore$ $\sqrt{n}\left(\hat{\beta}_j - \beta_j^*\right) \to_d N(0, \sigma^2)$

$x_{1j}\epsilon_1 = w_{11}Z_1$

$x_{1j}\epsilon_1 = w_{21}Z_1$ $\quad x_{2j}\epsilon_2 = w_{22}Z_2$

$\vdots$

$x_{1j}\epsilon_1 = w_{n1}Z_1$ $\quad x_{2j}\epsilon_2 = w_{n2}Z_2$ $\quad \cdots$ $\quad x_{nj}\epsilon_n = w_{nn}Z_n$

$\boxed{\text{Weighted CLT}}$

$(X^T\epsilon)_j = \sum_{i=1}^n x_{ij}\epsilon_i$ $\quad \epsilon_i \sim \text{iid}\ (0, \sigma^2)$

"$w_{ni} = x_{ij}$"

Cond'n: $\dfrac{\max_i x_{ij}^2 \le 1}{\left(\sum_i x_{ij}^2\right) = n} \to 0$

$\therefore$ weighted CLT,

$$\frac{\sum_{i=1}^n x_{ij}(\epsilon_i - 0)}{\left(\sum_{i=1}^n x_{ij}^2\right)^{1/2}\sigma} \xrightarrow[n]{} N(0,1)$$

$$\Rightarrow \frac{1}{\sqrt{n}}(X^T\epsilon)_j \xrightarrow[n]{} N(0, \sigma^2)$$