BASIC PHD WRITTEN EXAMINATION

THEORY, SECTION 1

(9:00 AM–1:00 PM, July 24, 2018)

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.

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 $N(0, \sigma^2)$ random variables, independent of N, with $0 < \sigma^2 < \infty$. Let

$$X = 1\{N > 0\} \sum_{j=1}^{N} Z_j,$$

where $1\{A\}$ is the indicator of A. Let $X_1, \ldots, X_n$ be i.i.d. realizations of X, and let $U_i = 1\{X_i = 0\}$, $1 \le i \le n$. Do the following:

(a) (4 points) Show that $EU_i = e^{-\lambda}$, $EX_i = 0$, $EX_i^2 = \lambda\sigma^2$, and $EX_i^4 = 3(\lambda + \lambda^2)\sigma^4$.

(b) (5 points) Show that $\hat{T}_n = -\log\left(n^{-1}\sum_{i=1}^{n} U_i\right)$ is almost surely consistent for λ, and that $\hat{W}_n = n^{-1}\sum_{i=1}^{n} X_i^2/\hat{T}_n$ is almost surely consistent for σ^2, as $n \to \infty$.

(c) (5 points) Show that

$$\sqrt{n}\begin{pmatrix} n^{-1}\sum_{i=1}^{n} U_i - e^{-\lambda} \\ n^{-1}\sum_{i=1}^{n} X_i^2 - \lambda\sigma^2 \end{pmatrix} \to_d N(0, \tau_1^2),$$

as $n \to \infty$, and give the form of τ_1^2.

(d) (6 points) Show that

$$\sqrt{n}\begin{pmatrix} \hat{T}_n - \lambda \\ \hat{W}_n - \sigma^2 \end{pmatrix} \to_d N(0, \tau_2^2),$$

as $n \to \infty$, where

$$\tau_2^2 = \begin{pmatrix} e^{\lambda} - 1 & -(e^{\lambda} - \lambda - 1)\sigma^2/\lambda \\ -(e^{\lambda} - \lambda - 1)\sigma^2/\lambda & \left(\frac{e^{\lambda}-1}{\lambda^2} + 2 + \frac{1}{\lambda}\right)\sigma^4 \end{pmatrix}.$$

(e) (5 points) Show that $\hat{W}_n \pm z_{1-\alpha/2}\hat{\rho}_n/\sqrt{n}$, where

$$\hat{\rho}^2 = \left(\frac{e^{\hat{T}_n} - 1}{\hat{T}_n^2} + 2 + \frac{1}{\hat{T}_n}\right)\hat{W}_n^2$$

and z_q is the qth-quantile of a standard normal, is an asymptotically valid $1 - \alpha$ level confidence interval for σ^2.

2. (25 points) Consider a binary classification problem that $\theta \in \{0, 1\}$ denotes the class label, $\mathbf{X}|(\theta = 0) \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ and $\mathbf{X}|(\theta = 1) \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, where N_p denotes the p-dimensional multivariate normal distribution and $\boldsymbol{\Sigma}$ is a positive definite matrix. Suppose 0-1 loss is used, and the prior distribution of θ is $P(\theta = 0) = 1/2$ and $P(\theta = 1) = 1/2$.

   (a) (4 points) Derive the Bayes rule for classifying a new observation $\boldsymbol{x} \in \mathcal{R}^p$.

   (b) (4 points) Derive the misclassification rate R^* of the Bayes rule.

   (c) (4 points) Let $\mathbf{X}_{0i}$ $(i = 1, \ldots, n_0)$ be independent and identically distributed (i.i.d) samples from the class of $\theta = 0$ and $\mathbf{X}_{1i}$ $(i = 1, \ldots, n_1)$ be i.i.d samples from the class of $\theta = 1$, and $\mathbf{X}_{0i}$ is independent of $\mathbf{X}_{1i}$. Derive the maximum likelihood estimators $(\widehat{\boldsymbol{\mu}}_0, \widehat{\boldsymbol{\mu}}_1, \widehat{\boldsymbol{\Sigma}})$ of $(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\Sigma})$.

   (d) (4 points) If we replace $(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ in the Bayes rule with $(\widehat{\boldsymbol{\mu}}_0, \widehat{\boldsymbol{\mu}}_1, \widehat{\boldsymbol{\Sigma}})$, prove that the misclassification rate of the resulting rule, i.e., the probability of classifying $\boldsymbol{x}$ to a wrong class given the training data $\{\mathbf{X}_{0i}\}_{i=1}^{n_0}$ and $\{\mathbf{X}_{1i}\}_{i=1}^{n_1}$, is given by

$$\frac{1}{2}\Phi\left(\frac{\widehat{\boldsymbol{\delta}}^T \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}_1 - \widehat{\boldsymbol{\mu}})}{\sqrt{\widehat{\boldsymbol{\delta}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\delta}}}}\right) + \frac{1}{2}\Phi\left(-\frac{\widehat{\boldsymbol{\delta}}^T \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}_0 - \widehat{\boldsymbol{\mu}})}{\sqrt{\widehat{\boldsymbol{\delta}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\delta}}}}\right),$$

   where $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\mu}}_0 - \widehat{\boldsymbol{\mu}}_1$ and $\widehat{\boldsymbol{\mu}} = (\widehat{\boldsymbol{\mu}}_0 + \widehat{\boldsymbol{\mu}}_1)/2$.

   (e) (5 points) We propose another classification rule that assigns $\boldsymbol{x}$ to the class of $\theta = 0$ if and only if $\widehat{\boldsymbol{\beta}}^T(\boldsymbol{x} - \widehat{\boldsymbol{\mu}}) \geq 0$, where $\widehat{\boldsymbol{\mu}} = (\widehat{\boldsymbol{\mu}}_0 + \widehat{\boldsymbol{\mu}}_1)/2$ and $\widehat{\boldsymbol{\beta}}$ solves the following problem

$$\widehat{\boldsymbol{\beta}} = \operatorname{argmin}_{\boldsymbol{\beta} \in \mathcal{R}^p} \frac{1}{2}\boldsymbol{\beta}^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{\beta} - (\widehat{\boldsymbol{\mu}}_0 - \widehat{\boldsymbol{\mu}}_1)^T \boldsymbol{\beta} + \lambda \sum_{j=1}^{p} |\beta_j|.$$

   Derive the Majorization-Minimization algorithm for solving $\widehat{\boldsymbol{\beta}}$. Give an explicit choice of step size and closed-form expressions on how iterations need to be done.

   (f) (4 points) Let R_n denote the misclassification rate of the rule described in (e). Suppose we can show that $\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)$ as $n \to \infty$. Use this result to prove $R_n \xrightarrow{P} R^*$.

You may use the following facts: (i) The density of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is $\{(2\pi)^p|\boldsymbol{\Sigma}|\}^{-1/2} \exp\{-(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2\}$; (ii) For symmetric matrices $\mathbf{A}$ and $\mathbf{M}$,

$$\frac{\partial \mathrm{tr}(\mathbf{AM})}{\partial \mathbf{M}} = \frac{\partial \mathrm{tr}(\mathbf{MA})}{\partial \mathbf{M}} = \mathbf{A}, \qquad \frac{\partial \log |\mathbf{M}|}{\partial \mathbf{M}} = \mathbf{M}^{-1}.$$

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

1. 3:20 start

$N \sim Poi(\lambda)$. $Z_1, Z_2, \cdots \sim iid\ N(0, \sigma^2)$. $N \perp\!\!\!\perp Z_i$

$X = 1\{N>0\} \cdot \sum_{j=1}^{N} Z_j$. $X_1, \cdots, X_n \overset{d}{\equiv} X$. $U_i = 1\{X_i = 0\}$

(a) $\mathbb{E} U_i = P(X_i = 0) = P(N=0) + P(N>0, \sum_{j=1}^{N} Z_j = 0)$

$= P(N=0) + \sum_{N_0=1}^{\infty} P(\sum_{j=1}^{N} Z_j = 0 \mid N = N_0) \cdot P(N = N_0)$

$= e^{-\lambda}$ $\quad (\because)\ \sum_{j=1}^{N} Z_j \mid N = N_0 \sim N(0, N_0\sigma^2)$: conti

$$\mathbb{E}[e^{tX} \mid N = N_0] = \begin{cases} \mathbb{E}[e^{t\sum_{j=1}^{N_0} Z_j} \mid N = N_0], & N_0 > 0 \\ \mathbb{E}[e^{t\cdot 0} \mid N = 0], & N_0 = 0 \end{cases}$$

$$= \begin{cases} \mathbb{E}[e^{t\sum_{j=1}^{N_0} Z_j}], & N_0 > 0 \\ \mathbb{E}[1 \mid N = 0], & N_0 = 0 \end{cases}$$

$$= \begin{cases} \exp(\frac{1}{2} N_0 \sigma^2 t^2), & N_0 > 0 \\ 1, & N_0 = 0 \end{cases} \quad (\because)\ \sum_{j=1}^{N_0} Z_j \sim N(0, N_0\sigma^2)$$

$\therefore\ \mathbb{E}[e^{tX}] = \sum_{N_0=0}^{\infty} \mathbb{E}[e^{tX} \mid N = N_0] \cdot P(N = N_0)$

$= 1 \cdot e^{-\lambda} + \sum_{N_0=1}^{\infty} \exp(\frac{1}{2} N_0 \sigma^2 t^2) \cdot \frac{e^{-\lambda}\lambda^{N_0}}{N_0!}$

$= \sum_{N_0=0}^{\infty} \frac{e^{-\lambda} \cdot (\lambda \cdot e^{\frac{1}{2}\sigma^2 t^2})^{N_0}}{N_0!} = \exp(-\lambda + \lambda \cdot e^{\frac{1}{2}\sigma^2 t^2})$

$\therefore\ \mathbb{E}X = \frac{d}{dt} \mathbb{E}[e^{tX}] \Big|_{t=0} = \exp(-\lambda + \lambda \cdot e^{\frac{1}{2}\sigma^2 t^2}) \cdot \lambda \cdot e^{\frac{1}{2}\sigma^2 t^2} \cdot \sigma^2 t \Big|_{t=0} = 0$

$\mathbb{E}X^2 = \frac{d^2}{dt^2} \mathbb{E}[e^{tX}] \Big|_{t=0} = \exp(-\lambda + \lambda \cdot e^{\frac{1}{2}\sigma^2 t^2}) \cdot (\quad)^2 + \exp(-\lambda + \lambda \cdot e^{\frac{1}{2}\sigma^2 t^2}) \cdot \lambda e^{\frac{1}{2}\sigma^2 t^2} \cdot (\sigma^2 t)^2$

$+ \exp(-\lambda + \lambda \cdot e^{\frac{1}{2}\sigma^2 t^2}) \cdot \lambda e^{\frac{1}{2}\sigma^2 t^2} \cdot \sigma^2 \Big|_{t=0}$

$= \exp(-\lambda + \lambda) \cdot \lambda \cdot \sigma^2 = \lambda \sigma^2$.

$$\mathbb{E}[X^4 \mid N = N_0] = \begin{cases} \mathbb{E}[(\sum_{j=1}^{N_0} Z_j)^4 \mid N = N_0], & N_0 > 0 \\ \mathbb{E}[0 \mid N = 0], & N_0 = 0 \end{cases} = \begin{cases} \mathbb{E}[(\sum_{j=1}^{N_0} Z_j)^4], & N_0 > 0 \\ 0, & N_0 = 0 \end{cases} \quad (\because)\ N \perp\!\!\!\perp \sum_{j=1}^{N_0} Z_j$$

From $\sum_{j=1}^{N_0} Z_j \sim N(0, N_0\sigma^2) \Rightarrow \mathbb{E} e^{t \cdot \sum_{j=1}^{N_0} Z_j} = \exp(\frac{1}{2} N_0 \sigma^2 t^2)$

$\therefore\ \mathbb{E}[(\sum_{j=1}^{N_0} Z_j)^4] = \frac{d^4}{dt^4} \exp(\frac{1}{2} N_0 \sigma^2 t^2) \Big|_{t=0}$.

(a) cont'ed

Let $M(t) = \exp(\frac{1}{2}At^2)$

$$\frac{d}{dt}M(t) = \exp(\tfrac{1}{2}At^2)\cdot At$$

$$\frac{d^2}{dt^2}M(t) = \exp(\tfrac{1}{2}At^2)\cdot A^2t^2 + \exp(\tfrac{1}{2}At^2)\cdot A$$

$$\frac{d^3}{dt^3}M(t) = \exp(\tfrac{1}{2}At^2)\cdot A^3t^3 + \underline{\exp(\tfrac{1}{2}At^2)\cdot 2A^2t + \exp(\tfrac{1}{2}At^2)\cdot A^2t} \quad \exp(\tfrac{1}{2}At^2)\cdot 3A^2t$$

$$\frac{d^4}{dt^4}M(t) = \exp(\tfrac{1}{2}At^2)\cdot A^4t^4 + \exp(\tfrac{1}{2}At^2)\cdot 3A^3t^2 + \exp(\tfrac{1}{2}At^2)\cdot 3A^3t^2 + \exp(\tfrac{1}{2}At^2)\cdot 3A^2$$

$$\frac{d^4}{dt^4}M(t)\Big|_{t=0} = 3A^2$$

$$\therefore \mathbb{E}\Big[\Big(\sum_{j=1}^{N_0} Z_j\Big)^4\Big] = 3N_0^2\sigma^4$$

$$\therefore \mathbb{E}X_i^4 = \mathbb{E}\big[\mathbb{E}[X^4|N]\big] = \sum_{N_0=0}^{\infty}\mathbb{E}[X^4|N=N_0]\cdot p(N=N_0)$$

$$= \sum_{N_0=0}^{\infty} 3N_0^2\sigma^4\cdot\frac{e^{-\lambda}\lambda^{N_0}}{N_0!}$$

$$= 3\sigma^4\cdot\sum_{N_0=0}^{\infty} N_0^2\cdot\frac{e^{-\lambda}\lambda^{N_0}}{N_0!}$$

$$= 3\sigma^4\cdot\mathbb{E}[N^2] = 3\sigma^4(\lambda^2+\lambda)$$

(b) $U_i \overset{iid}{\sim} Ber(e^{-\lambda})$

SLLN: $\mathbb{E}U_i = e^{-\lambda}$, $\mathbb{E}U_i^k < \infty$, $\forall k\in\mathbb{N}$.

$$\frac{1}{n}\sum_{i=1}^n U_i \longrightarrow_{as} \mathbb{E}U_i = e^{-\lambda}$$

~~CLT: $\sqrt{n}(\frac{1}{n}\sum_{i=1}^n U_i - e^{-\lambda}) \to_d N(0, e^{-\lambda}(1-e^{-\lambda}))$~~

~~Delta-Method: $g(x) = -\log(x)$, $g'(x) = -x^{-1}$~~

~~$\sqrt{n}(-\log(n^{-1}\sum_{i=1}^n U_i) - \lambda) \to_d N(0, e^{\lambda}-1)$~~

Conti. Map. Thm: $g(x) = -\log(x)$, $x\in(0,\infty)$

$$\hat{T}_n = -\log(n^{-1}\sum_{i=1}^n U_i) \longrightarrow_{as} \lambda$$

SLLN: $\mathbb{E}X_i^2 = \lambda\sigma^2$, $\mathbb{E}(X_i^2)^2 = 3(\lambda+\lambda^2)\sigma^4 < \infty$

$$n^{-1}\sum_{i=1}^n X_i^2 \longrightarrow_{as} \mathbb{E}X_i^2 = \lambda\sigma^2$$

Slutsky:
$$\frac{n^{-1}\sum_{i=1}^n X_i^2}{\hat{T}_n} \longrightarrow_{as} \frac{\lambda\sigma^2}{\lambda} = \sigma^2.$$

(c) $\mathbb{E}\begin{bmatrix} U_i \\ X_i^2 \end{bmatrix} = \begin{bmatrix} \mathbb{E}U_i \\ \mathbb{E}X_i^2 \end{bmatrix} = \begin{bmatrix} e^{-\lambda} \\ \lambda\sigma^2 \end{bmatrix}$

$\mathrm{Var}(U_i) = e^{-\lambda}(1-e^{-\lambda})$

$\mathrm{Var}(X_i^2) = \mathbb{E}X_i^4 - (\mathbb{E}X_i^2)^2 = 3(\lambda+\lambda^2)\sigma^4 - (\lambda\sigma^2)^2 = (3\lambda+2\lambda^2)\sigma^4$

$\mathrm{Cov}(U_i, X_i^2) = \mathbb{E}U_iX_i^2 - \mathbb{E}U_i \cdot \mathbb{E}X_i^2 = -e^{-\lambda}\cdot\lambda\sigma^2$ $\quad (\because)\ \mathbb{E}U_iX_i^2 = \mathbb{E}\left[\mathbb{E}[U_iX_i^2 \mid U_i]\right]$

$= \mathbb{E}[X_i^2 \mid U_i=1]\cdot P(U_i=1) + \mathbb{E}[0 \mid U_i=0]\cdot P(U_i=0)$

$= 0 \quad (\because)\ U_i=1 \Leftrightarrow X_i=0$

$\therefore \mathrm{Var}\begin{bmatrix} U_i \\ X_i^2 \end{bmatrix} = \begin{bmatrix} e^{-\lambda}(1-e^{-\lambda}) & -e^{-\lambda}\cdot\lambda\sigma^2 \\ -e^{-\lambda}\cdot\lambda\sigma^2 & (3\lambda+2\lambda^2)\sigma^4 \end{bmatrix}$: exists.

$\therefore$ CLT:

$$\sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\begin{bmatrix} U_i \\ X_i^2 \end{bmatrix} - \begin{bmatrix} e^{-\lambda} \\ \lambda\sigma^2 \end{bmatrix}\right) \to_d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tau_1^2\right),$$

where

$$\tau_1^2 = \mathrm{Var}\begin{bmatrix} U_1 \\ X_1^2 \end{bmatrix} = \begin{bmatrix} e^{-\lambda}(1-e^{-\lambda}) & -e^{-\lambda}\cdot\lambda\sigma^2 \\ -e^{-\lambda}\cdot\lambda\sigma^2 & (3\lambda+2\lambda^2)\sigma^4 \end{bmatrix}$$

(d)

$$\sqrt{n}\left(\begin{bmatrix} n^{-1}\Sigma_i U_i \\ n^{-1}\Sigma_i X_i^2 \end{bmatrix} - \begin{bmatrix} e^{-\lambda} \\ \lambda\sigma^2 \end{bmatrix}\right) \to_d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tau_1^2\right)$$

Delta-Method: $g(x,y) = \left(-\log(x),\ y/(-\log(x))\right)^T$, $\nabla g = \begin{bmatrix} -x^{-1} & +y(\log x)^{-2}\cdot x^{-1} \\ 0 & -(\log x)^{-1} \end{bmatrix}$

$$\sqrt{n}\left(g\begin{pmatrix} n^{-1}\Sigma_i U_i \\ n^{-1}\Sigma_i X_i^2 \end{pmatrix} - g\begin{pmatrix} e^{-\lambda} \\ \lambda\sigma^2 \end{pmatrix}\right) \to_d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \nabla g(e^{-\lambda}, \lambda\sigma^2)^T \tau_1^2 \nabla g(e^{-\lambda}, \lambda\sigma^2)\right)$$

$$\sqrt{n}\left(\begin{bmatrix} \hat{\tau}_n \\ \hat{w}_n \end{bmatrix} - \begin{bmatrix} \lambda \\ \sigma^2 \end{bmatrix}\right) \to_d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tau_2^2\right)$$

$$\tau_2^2 = \begin{bmatrix} -e^{\lambda} & 0 \\ \frac{\lambda\sigma^2}{\lambda^2\cdot e^{-\lambda}} & -\frac{1}{-\lambda} \end{bmatrix}^T \tau_1^2 \begin{bmatrix} -e^{\lambda} & 0 \\ \frac{\lambda\sigma^2}{\lambda^2 e^{-\lambda}} & -\frac{1}{-\lambda} \end{bmatrix}$$

$$= \begin{bmatrix} -e^{\lambda} & 0 \\ \frac{\sigma^2 e^{\lambda}}{\lambda} & \lambda \end{bmatrix} \begin{bmatrix} e^{-\lambda}(1-e^{-\lambda}) & -\lambda\sigma^2 e^{-\lambda} \\ -\lambda\sigma^2 e^{-\lambda} & (3\lambda+2\lambda^2)\sigma^4 \end{bmatrix} \begin{bmatrix} -e^{\lambda} & \frac{\sigma^2 e^{\lambda}}{\lambda} \\ 0 & \lambda \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} -e^{\lambda} & 0 \\ \frac{\sigma^2 e^{\lambda}}{\lambda} & \frac{1}{\lambda} \end{bmatrix} \begin{bmatrix} e^{-\lambda}(1-e^{-\lambda}) & -e^{-\lambda}\cdot\lambda\sigma^2 \\ -e^{-\lambda}\cdot\lambda\sigma^2 & (3\lambda+2\lambda^2)\sigma^4 \end{bmatrix} \begin{bmatrix} -e^{\lambda} & \frac{\sigma^2 e^{\lambda}}{\lambda} \\ 0 & \frac{1}{\lambda} \end{bmatrix}$$

$[1,1]:\quad -e^{\lambda}\cdot e^{-\lambda}(1-e^{-\lambda})\cdot(-e^{\lambda}) = e^{\lambda}-1$

$[1,2]:\quad -e^{\lambda}\cdot e^{-\lambda}(1-e^{-\lambda})\cdot\frac{\sigma^2 e^{\lambda}}{\lambda} + (-e^{\lambda})\cdot(-e^{-\lambda}\cdot\lambda\sigma^2)\cdot\frac{1}{\lambda} = -(e^{\lambda}-1)\frac{\sigma^2}{\lambda}+\sigma^2 = -(e^{\lambda}-\lambda-1)\sigma^2/\lambda$

$[2,1]:\quad \frac{\sigma^2 e^{\lambda}}{\lambda}\cdot e^{-\lambda}(1-e^{-\lambda})\cdot(-e^{\lambda}) + \frac{1}{\lambda}\cdot(-e^{-\lambda}\cdot\lambda\sigma^2)\cdot(-e^{\lambda}) = -(e^{\lambda}-1)\frac{\sigma^2}{\lambda}+\sigma^2 = \quad ''$

$[2,2]:\quad \frac{\sigma^2 e^{\lambda}}{\lambda}\cdot e^{-\lambda}(1-e^{-\lambda})\cdot\frac{\sigma^2 e^{\lambda}}{\lambda} + \frac{\sigma^2 e^{\lambda}}{\lambda}\cdot(-e^{-\lambda}\cdot\lambda\sigma^2)\cdot\frac{1}{\lambda} + \frac{1}{\lambda}\cdot(-e^{-\lambda}\cdot\lambda\sigma^2)\cdot\frac{\sigma^2 e^{\lambda}}{\lambda} + \frac{1}{\lambda}\cdot(3\lambda+2\lambda^2)\sigma^4\cdot\frac{1}{\lambda}$

$$= \frac{\sigma^4}{\lambda^2}(e^{\lambda}-1) + \frac{\sigma^4}{\lambda^2}(-\lambda) + \frac{\sigma^4}{\lambda^2}(-\lambda) + \frac{\sigma^4}{\lambda^2}(3\lambda+2\lambda^2)$$

$$= \left(\frac{e^{\lambda}-1}{\lambda^2} - \frac{2}{\lambda} + \frac{3+2\lambda}{\lambda}\right)\sigma^4$$

$$= \left(\frac{e^{\lambda}-1}{\lambda^2} + 2 + \frac{1}{\lambda}\right)\sigma^4 .$$

(e) From (d), we have

$$\sqrt{n}(\hat{W}_n - \sigma^2) \to_d N\left(0, \left(\frac{e^{\lambda}-1}{\lambda^2}+2+\frac{1}{\lambda}\right)\sigma^4\right)$$

From (b), we have

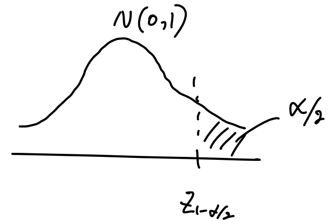
$\hat{T}_n \to_{a.s} \lambda, \quad \hat{W}_n \to_{a.s} \sigma^2$

Slutsky:

$$\frac{\sqrt{n}(\hat{W}_n - \sigma^2)}{\left\{\left(\frac{e^{\hat{T}_n}-1}{\hat{T}_n^2}+2+\frac{1}{\hat{T}_n}\right)\hat{W}_n^2\right\}^{1/2}} \to_d N(0,1)$$

$\therefore$ Asymptotic $(1-\alpha)$ level C.I for σ^2 is

$$\sigma^2 : \left| \frac{\sqrt{n}(\hat{W}_n - \sigma^2)}{\hat{\rho}_n} \right| \le z_{1-\alpha/2}$$



$$\Leftrightarrow \sigma^2 : \sigma^2 \in [\hat{W}_n - z_{1-\alpha/2}\cdot\hat{\rho}_n/\sqrt{n},\ \hat{W}_n + z_{1-\alpha/2}\cdot\hat{\rho}_n/\sqrt{n}]$$

2.

$\theta \sim Ber(\frac{1}{2})$

$X|\theta=0 \sim N_p(\mu_0, \Sigma)$

$X|\theta=1 \sim N_p(\mu_1, \Sigma)$

(a) Given a new observation x,

$$P(\theta|X=x) \propto P(X|\theta) \cdot \pi(\theta)$$

$$\Rightarrow P(\theta=0|X=x) = \frac{P(X=x|\theta=0)\cdot\pi(\theta=0)}{P(X=x|\theta=0)\cdot\pi(\theta=0) + P(X=x|\theta=1)\cdot\pi(\theta=1)}$$

$$= \frac{\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)}{\exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right) + \exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)}$$

$$P(\theta=1|X=x) = 1 - P(\theta=0|X=x).$$

Posterior Expected Loss given $X=x$, where $a \in \{0,1\}$ is classifying action (i.e) $a=0 \Rightarrow$ decide $\theta=0$.

$$E_\Theta[\ell(\theta,a)|X=x] = \begin{cases} P(\theta=1|X=x), & a=0 \\ P(\theta=0|X=x), & a=1 \end{cases}$$

The Bayes rule is the minimizer a of the posterior expected loss.

Thus, Bayes rule

$$d_B(x) = \begin{cases} 0 \text{ (i.e) obs belongs to class 0} & \text{, if } P(\theta=1|X=x) < P(\theta=0|X=x) \\ 1 \text{ (i.e) '' '' '' '' 1} & \text{, if '' > ''} \\ \text{any action} & \text{, if } = \end{cases}$$

For simplicity,

$$P(\theta=1|X=x) < P(\theta=0|X=x) \Leftrightarrow -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) < -\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)$$

$$\Leftrightarrow x^T\Sigma^{-1}x - 2\mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 > x^T\Sigma^{-1}x - 2\mu_0^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0$$

$$\Leftrightarrow (\mu_0-\mu_1)^T\Sigma^{-1}x > \frac{1}{2}\left(\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1\right)$$

$\therefore$ Bayes rule

$$d_B(x) = \begin{cases} 0 & \text{, if } (\mu_0-\mu_1)^T\Sigma^{-1}x > \frac{1}{2}(\quad) \\ 1 & < \\ \text{any action} & = \end{cases}$$

(b) $R^* = P(d_B(x) \neq \theta)$

$= 1 - P(d_B(x) = \theta)$

$= 1 - \{ P(d_B(x)=1 \mid \theta=1) \cdot P(\theta=1) + P(d_B(x)=0 \mid \theta=0) \cdot P(\theta=0) \}$

$= 1 - \{ P((\mu_0-\mu_1)^T \Sigma^{-1} x < \frac{1}{2}(\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1) \mid x \sim N(\mu_1, \Sigma)) \cdot \frac{1}{2} +$

$P((\mu_0-\mu_1)^T \Sigma^{-1} x > \frac{1}{2}(\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1) \mid x \sim N(\mu_0, \Sigma)) \cdot \frac{1}{2} \}$

Let $a^T := (\mu_0-\mu_1)^T\Sigma^{-1}$, $b := \frac{1}{2}(\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1)$

$P(a^Tx < b \mid x \sim N(\mu_1, \Sigma))$

$= P(a^Tx < b \mid a^Tx \sim N(a^T\mu_1, a^T\Sigma a))$

$= P\left(\frac{a^Tx - a^T\mu_1}{(a^T\Sigma a)^{1/2}} < \frac{b - a^T\mu_1}{(a^T\Sigma a)^{1/2}} \;\middle|\; \frac{a^Tx - a^T\mu_1}{(a^T\Sigma a)^{1/2}} \sim N(0,1)\right)$

$= \Phi\left(\frac{b - a^T\mu_1}{(a^T\Sigma a)^{1/2}}\right)$, where Φ: cdf of $N(0,1)$

Similarly,

$P(a^Tx > b \mid x \sim N(\mu_0, \Sigma)) = P\left(\frac{a^Tx - a^T\mu_0}{(a^T\Sigma a)^{1/2}} > \frac{b - a^T\mu_0}{(a^T\Sigma a)^{1/2}} \;\middle|\; \frac{a^Tx - a^T\mu_0}{(a^T\Sigma a)^{1/2}} \sim N(0,1)\right)$

$= 1 - \Phi\left(\frac{b - a^T\mu_0}{(a^T\Sigma a)^{1/2}}\right)$

$\therefore R^* = 1 - \frac{1}{2}\left\{ \Phi\left(\frac{b - a^T\mu_1}{(a^T\Sigma a)^{1/2}}\right) + 1 - \Phi\left(\frac{b - a^T\mu_0}{(a^T\Sigma a)^{1/2}}\right) \right\}$

$= 1 - \frac{1}{2}\left\{ 1 - \Phi\left(\frac{a^T\mu_1 - b}{(a^T\Sigma a)^{1/2}}\right) + 1 - \Phi\left(\frac{b - a^T\mu_0}{(a^T\Sigma a)^{1/2}}\right) \right\}$

$= \frac{1}{2}\Phi\left(\frac{a^T\mu_1 - b}{(a^T\Sigma a)^{1/2}}\right) + \frac{1}{2}\Phi\left(\frac{b - a^T\mu_0}{(a^T\Sigma a)^{1/2}}\right)$

$a^T\mu_1 - b = \mu_0\mu_1 - \mu_1^2 - \frac{1}{2}\mu_0^2 + \frac{1}{2}\mu_1^2$

$= -\frac{1}{2}(\mu_0^2 + \mu_1^2 - 2\mu_0\mu_1)$

$b - a^T\mu_0 = \frac{1}{2}\mu_0^2 - \frac{1}{2}\mu_1^2 - \mu_0^2 + \mu_1\mu_0$

$= -\frac{1}{2}\mu_0^2 - \frac{1}{2}\mu_1^2 + \mu_1\mu_0$

Here, $a^T\mu_1 - b = -\frac{1}{2}(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1)$

$b - a^T\mu_0 = -\frac{1}{2}(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1)$

$a^T\Sigma a = (\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1)$.

$\therefore R^* = \Phi\left(-\frac{1}{2}\{(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1)\}^{1/2}\right)$

(c) Likelihood

$$L(\mu_0,\mu_1,\Sigma) = \prod_{i=1}^{n_0} |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x_{0i}-\mu_0)^T\Sigma^{-1}(x_{0i}-\mu_0)\right)$$

$$\cdot \prod_{j=1}^{n_1} |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x_{1j}-\mu_1)^T\Sigma^{-1}(x_{1j}-\mu_1)\right)$$

Profiling method (fix Σ) : $\hat{\mu}_0 = \overline{x}_{0\cdot} = \frac{1}{n_0}\sum_{i=1}^{n_0} x_{0i}$, $\hat{\mu}_1 = \overline{x}_{1\cdot} = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j}$

Then,

$$L(\hat{\mu}_0,\hat{\mu}_1,\Sigma) \propto |\Sigma|^{-(n_0+n_1)/2} \cdot \exp\left(-\frac{1}{2}\sum_i (x_{0i}-\hat{\mu}_0)^T\Sigma^{-1}(x_{0i}-\hat{\mu}_0) - \frac{1}{2}\sum_j \underline{(x_{1j}-\hat{\mu}_1)^T\Sigma^{-1}(x_{1j}-\hat{\mu}_1)}\right)$$

$$\propto |\Sigma|^{-(n_0+n_1)/2} \cdot \exp\left(-\frac{1}{2}\sum_i tr\left((x_{0i}-\hat{\mu}_0)^T\Sigma^{-1}(x_{0i}-\hat{\mu}_0)\right) - \frac{1}{2}\sum_j tr\left(\quad\downarrow\quad\right)\right) \quad (\because) \text{ scalar}$$

$$\propto \quad '' \quad \cdot \exp\left(-\frac{1}{2}\sum_i tr\left(\Sigma^{-1}(x_{0i}-\hat{\mu}_0)^{\otimes 2}\right) - \frac{1}{2}\sum_j tr\left(\Sigma^{-1}(x_{1j}-\hat{\mu}_1)^{\otimes 2}\right)\right) \quad (\because) \ tr(ABC) = tr(BCA)$$

$$\propto \quad '' \quad \cdot \exp\left(-\frac{1}{2} tr\left(\Sigma^{-1}\sum_i (x_{0i}-\hat{\mu}_0)^{\otimes 2}\right) - \frac{1}{2} tr\left(\Sigma^{-1}\sum_j (x_{1j}-\hat{\mu}_1)^{\otimes 2}\right)\right) \quad (\because) \text{ linearity of trace}$$

$$\propto \quad '' \quad \cdot \exp\left(-\frac{1}{2} tr\left(\Sigma^{-1}\cdot\left(\sum_i (x_{0i}-\hat{\mu}_0)^{\otimes 2} + \sum_j (x_{1j}-\hat{\mu}_1)^{\otimes 2}\right)\right)\right)$$

Profiled log-likelihood wrt Σ

$$l(\Sigma) = \frac{n_0+n_1}{2}\cdot \log|\Sigma^{-1}| - \frac{1}{2} tr\left(\Sigma^{-1}\cdot\left(\sum_i \quad + \sum_j \quad\right)\right)$$

$$\frac{\partial l}{\partial \Sigma^{-1}} = \frac{n_0+n_1}{2}\cdot(\Sigma^{-1})^{-1} - \frac{1}{2}\left(\sum_i (x_{0i}-\hat{\mu}_0)^{\otimes 2} + \sum_j (x_{1j}-\hat{\mu}_1)^{\otimes 2}\right) \overset{set}{=} 0$$

$$\Rightarrow \widehat{\Sigma^{-1}} = \left(\frac{1}{n_0+n_1}\left(\sum_i (x_{0i}-\hat{\mu}_0)^{\otimes 2} + \sum_j (x_{1j}-\hat{\mu}_1)^{\otimes 2}\right)\right)^{-1}$$

By invariance mapping thm of MLE,

$$\hat{\Sigma} = (\widehat{\Sigma^{-1}})^{-1} = \frac{1}{n_0+n_1}\left(\sum_i (x_{0i}-\hat{\mu}_0)^{\otimes 2} + \sum_j (x_{1j}-\hat{\mu}_1)^{\otimes 2}\right) //$$

(d) From (a),

Bayes rule is given by

$$d_B(x) = \begin{cases} 0 & , (\mu_0-\mu_1)^T\Sigma^{-1}x > \frac{1}{2}(\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1) \\ 1 & , \quad < \\ \text{any action} & , \quad = \end{cases}$$

Replacing (μ_0, μ_1, Σ) with $(\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma})$,

$$\hat{d}_B(x) = \begin{cases} 0 & , (\hat{\mu}_0-\hat{\mu}_1)^T\hat{\Sigma}^{-1}x > \frac{1}{2}(\hat{\mu}_0^T\hat{\Sigma}^{-1}\hat{\mu}_0 - \hat{\mu}_1\hat{\Sigma}^{-1}\hat{\mu}_1) \\ 1 & \quad < \\ \text{any action} & \quad = \end{cases}$$

Thus, misclassification rate of this rule is

$$\hat{R}^* = P(\hat{d}_B(X) \neq \theta)$$

$$= P(\hat{d}_B(X)=0 \mid \theta=1)\cdot P(\theta=1) + P(\hat{d}_B(X)=1 \mid \theta=0)\cdot P(\theta=0)$$

$$= P\left((\hat{\mu}_0-\hat{\mu}_1)^T\hat{\Sigma}^{-1}X > \frac{1}{2}(\hat{\mu}_0^T\hat{\Sigma}^{-1}\hat{\mu}_0 - \hat{\mu}_1^T\hat{\Sigma}^{-1}\hat{\mu}_1) \mid X\sim N(\mu_1,\Sigma)\right)\cdot\frac{1}{2} +$$

$$P\left(\quad '' \quad < \quad '' \quad \mid X\sim N(\mu_0,\Sigma)\right)\cdot\frac{1}{2}$$

$$= \frac{1}{2}\cdot P\left(\hat{\delta}^T\hat{\Sigma}^{-1}X > \hat{\delta}^T\hat{\Sigma}^{-1}\hat{\mu} \mid X\sim N(\mu_1,\Sigma)\right) +$$

$$\frac{1}{2}\cdot P\left(\quad '' \quad < \quad '' \quad \mid X\sim N(\mu_0,\Sigma)\right)$$

$(\because)$ $\hat{\delta}^T\hat{\Sigma}^{-1}\hat{\mu} = \frac{1}{2}(\hat{\mu}_0-\hat{\mu}_1)^T\hat{\Sigma}^{-1}(\hat{\mu}_0+\hat{\mu}_1) = \frac{1}{2}\Big(\hat{\mu}_0^T\hat{\Sigma}^{-1}\hat{\mu}_0 - \hat{\mu}_1^T\hat{\Sigma}^{-1}\hat{\mu}_1 \underbrace{- \hat{\mu}_1^T\hat{\Sigma}^{-1}\hat{\mu}_0 + \hat{\mu}_0^T\hat{\Sigma}^{-1}\hat{\mu}_1}_{\text{same scalar}}\Big)$

$$= \frac{1}{2}\cdot P\left(\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(X-\mu_1)}{\{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}\}^{1/2}} > \frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\hat{\mu}-\mu_1)}{\{\cdots\}^{1/2}} \;\Big|\; \frac{\hat{\delta}^T\hat{\Sigma}^{-1}(X-\mu_1)}{\{\cdots\}^{1/2}} \sim N(0,1)\right) +$$

$$\frac{1}{2}\cdot P\left(\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(X-\mu_0)}{\{\cdots\}^{1/2}} < \frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\hat{\mu}-\mu_0)}{\{\cdots\}^{1/2}} \;\Big|\; \frac{\hat{\delta}^T\hat{\Sigma}^{-1}(X-\mu_0)}{\{\cdots\}^{1/2}} \sim N(0,1)\right)$$

$$= \frac{1}{2}\,\Phi\left(-\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\hat{\mu}-\mu_1)}{\{\cdots\}^{1/2}}\right) + \frac{1}{2}\cdot\Phi\left(\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\hat{\mu}-\mu_0)}{\{\cdots\}^{1/2}}\right) \,_{\prime\prime}$$

(e) let $L(\beta) = \frac{1}{2}\beta^T \hat{\Sigma} \beta - (\hat{\mu}_0 - \hat{\mu}_1)^T \beta$

Given $\tilde{\beta} \in \mathbb{R}^p$ fixed, for $\beta \in \mathbb{R}^p$,

$$L(\beta) = L(\tilde{\beta}) + \dot{L}(\tilde{\beta})^T (\beta - \tilde{\beta}) + \frac{1}{2}(\beta - \tilde{\beta})^T \ddot{L}(\tilde{\beta}^*)(\beta - \tilde{\beta}), \quad \exists\ \tilde{\beta}^* = t\beta + (1-t)\tilde{\beta}$$

$$\leq L(\tilde{\beta}) + \dot{L}(\tilde{\beta})^T (\beta - \tilde{\beta}) + \frac{1}{2} C (\beta - \tilde{\beta})^T (\beta - \tilde{\beta}),$$

where $C = \lambda_{max}(\ddot{L}(\tilde{\beta}^*)) = \lambda_{max}(\hat{\Sigma})$: the greatest eigen value of $\hat{\Sigma}$.

Here, $\dot{L}(\beta) = \hat{\Sigma}\beta - (\hat{\mu}_0 - \hat{\mu}_1)$, $\ddot{L}(\beta) = \hat{\Sigma}$.

Letting $L(\beta|\tilde{\beta}) := L(\tilde{\beta}) + \dot{L}(\tilde{\beta})^T(\beta - \tilde{\beta}) + \frac{1}{2}C(\beta - \tilde{\beta})^T(\beta - \tilde{\beta})$, this is the majorization of the original objective ftn $L(\beta)$.

Now, we want to solve

$$\tilde{\beta}^{new} = \underset{\beta}{\operatorname{argmin}}\ L(\beta|\tilde{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|.$$

Since it is a convex problem, we apply Coordinate Gradient Descent algorithm.

By fixing all coordinates but β_j, letting them $\hat{\beta}_k$, $k \neq j$

$$\hat{\beta}_j^{new} = \underset{\beta_j}{\operatorname{argmin}}\ L(\hat{\beta}_{-j}, \beta_j | \tilde{\beta}) + \lambda \sum_{k \neq j} |\hat{\beta}_k| + \lambda |\beta_j|$$

$$= \underset{\beta_j}{\operatorname{argmin}}\ (\hat{\Sigma}\tilde{\beta} - (\hat{\mu}_0 - \hat{\mu}_1))^T (\hat{\beta}_1, \cdots, \beta_j, \cdots, \hat{\beta}_p)^T + \frac{1}{2} C \cdot (\beta_j - \tilde{\beta}_j)^2 + \lambda |\beta_j|$$

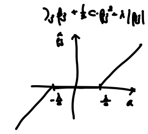
$$= \underset{\beta_j}{\operatorname{argmin}}\ \frac{1}{2} C(\beta_j - \tilde{\beta}_j)^2 + [\hat{\Sigma}\tilde{\beta} - (\hat{\mu}_0 - \hat{\mu}_1)]_j\ \beta_j + \lambda |\beta_j|$$

$$= \underset{\beta_j}{\operatorname{argmin}}\ \frac{1}{2} C(\beta_j - a_j)^2 + \lambda |\beta_j|,$$

$M_j(\beta) = ($

$\underset{\beta_j}{\operatorname{min}} M_j(\beta_j) = S(a, \frac{\lambda}{c})$

$\frac{1}{2}C(\beta_j - a)^2 + \lambda|\beta_j|$ + const



$\beta_j = a - \frac{\lambda}{c}$

$|ac| \leq \lambda$

$|a| \leq \frac{\lambda}{c}$

$\hat{\beta}_j = a - \frac{\lambda}{c}$, $a > \frac{\lambda}{c}$

where $[\hat{\Sigma}\tilde{\beta} - (\hat{\mu}_0 - \hat{\mu}_1)]_j$: j^{th} component of $\hat{\Sigma}\tilde{\beta} - (\hat{\mu}_0 - \hat{\mu}_1)$,

$$a_j := \tilde{\beta}_j - \frac{1}{C}[\hat{\Sigma}\tilde{\beta} - (\hat{\mu}_0 - \hat{\mu}_1)]_j.$$

This is a univariate convex optimization problem.

KKT condition: $\frac{1}{2}C(\hat{\beta}_j - a_j) + \lambda\, sgn(\hat{\beta}_j) \ni 0$, where $sgn(\hat{\beta}_j) = \begin{cases} 1, & \hat{\beta}_j > 0 \\ -1 & < 0 \\ [-1, 1] & = 0 \end{cases}$

Solving by case : $\hat{\beta}_j = S(a_j, \frac{\lambda}{c}) = sgn(a_j) \cdot (|a_j| - \frac{\lambda}{c})_+$

By applying CGD on $j = 1, \cdots, p$, we get

$$\hat{\beta}^{new} = (\hat{\beta}_1^{new}, \cdots, \hat{\beta}_p^{new})$$

from previous $\hat{\beta}$

(e) Cont'ed.

Now, algorithm is as follow:

1. Initialize $\hat{\beta}$. (set by $\underset{\beta}{\operatorname{argmin}} \frac{1}{2}\beta^T\hat{\Sigma}\beta - (\hat{\mu}_0-\hat{\mu}_1)^T\beta$)

2. Given previous $\hat{\beta}$, get new update of j^{th} coordinate $(j=1,\cdots,p)$

$$\hat{\beta}_j^{new} = \operatorname{sgn}(a_j)\cdot\left(|a_j| - \frac{\lambda}{c}\right)_+,$$

where $a_j = \hat{\beta}_j - \frac{1}{c}\left[\hat{\Sigma}\hat{\beta} - (\hat{\mu}_0-\hat{\mu}_1)\right]_j$,

$[\hat{\Sigma}\hat{\beta}-(\hat{\mu}_0-\hat{\mu}_1)]_j$: j^{th} component of $\hat{\Sigma}\hat{\beta}-(\hat{\mu}_0-\hat{\mu}_1)$

$c = \lambda_{max}(\hat{\Sigma})$

3. Iterative until $\|\hat{\beta}^{new}-\hat{\beta}\| < \varepsilon$: pre-determined convergence threshold.

(f) $R_n = P(\hat{\beta}^T(X-\hat{\mu}) \geq 0, \theta=1) + P(\hat{\beta}^T(X-\hat{\mu}) < 0, \theta=0)$

$= P(\hat{\beta}^T(X-\hat{\mu}) \geq 0 \mid \theta=1)\cdot P(\theta=1) + P(\hat{\beta}^T(X-\hat{\mu}) < 0 \mid \theta=0)\cdot P(\theta=0)$

$= P(\hat{\beta}^T(X-\hat{\mu}) \geq 0 \mid X\sim N(\mu_1,\Sigma))\cdot\frac{1}{2} + P(\hat{\beta}^T(X-\hat{\mu}) < 0 \mid X\sim N(\mu_0,\Sigma))\cdot\frac{1}{2}$

From $\hat{\beta} \to_p \Sigma^{-1}(\mu_0-\mu_1)$ and $\hat{\mu} \to_p \mu := \frac{1}{2}(\mu_0+\mu_1)$ $(\because)$ WLLN,

$X\sim N(\mu_1,\Sigma) \Rightarrow X-\hat{\mu} \to_d N(\frac{1}{2}(\mu_1-\mu_0), \Sigma)$

$\Rightarrow \hat{\beta}^T(X-\hat{\mu}) \to_d (\Sigma^{-1}(\mu_0-\mu_1))^T\cdot N(\frac{1}{2}(\mu_1-\mu_0),\Sigma)$

$\stackrel{d}{=} N(-\frac{1}{2}(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1), (\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1))$

$X\sim N(\mu_0,\Sigma) \Rightarrow X-\hat{\mu} \to_d N(\frac{1}{2}(\mu_0-\mu_1),\Sigma)$

$\Rightarrow \hat{\beta}^T(X-\hat{\mu}) \to_d (\Sigma^{-1}(\mu_0-\mu_1))^T\cdot N(\frac{1}{2}(\mu_0-\mu_1),\Sigma)$

$\stackrel{d}{=} N(\frac{1}{2}(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1), (\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1))$

$\therefore R_n \xrightarrow[n]{} \frac{1}{2}P(N(-\frac{1}{2}\Delta,\Delta) \geq 0) + \frac{1}{2}P(N(\frac{1}{2}\Delta,\Delta) < 0)$, $\Delta := (\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0-\mu_1)$

$= \frac{1}{2}P\left(N(0,1) \geq \frac{\frac{1}{2}\Delta}{\sqrt{\Delta}}\right) + \frac{1}{2}P\left(N(0,1) < \frac{-\frac{1}{2}\Delta}{\sqrt{\Delta}}\right)$

$= \frac{1}{2}\Phi(-\frac{1}{2}\sqrt{\Delta}) + \frac{1}{2}\Phi(-\frac{1}{2}\sqrt{\Delta})$

$= \Phi(-\sqrt{\Delta}) = R^*$