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 $\theta$ 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

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

   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 $\mathbf{M}$,

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

2. Consider a binary classification problem 2018 Theory 1

$\theta \in \{0,1\}$ denotes class label

$X|\theta=0 \sim N_p(\mu_0, \Sigma)$ $X|\theta=1 \sim N_p(\mu_1, \Sigma)$ $\Sigma$ is positive definite

Suppose 0-1 loss is used & the prior distribution of $\theta$ is $P(\theta=0)=P(\theta=1)=1/2$

(a) Derive the Bayes rule for classifying a new observation $x \in \mathbb{R}^p$

Under 0-1 loss, the Bayes rule is the posterior MODE

$\therefore$ $\theta=0$ if $f(\theta=0|x) \geq f(\theta=1|x)$

$\theta=1$ if $f(\theta=1|x) > f(\theta=0|x)$

$p(\theta|x) \propto p(x|\theta)\pi(\theta)$

$f(\theta=1|x) = P(x|\theta_1=1)\pi(\theta=1) = \{(2\pi)^p|\Sigma|\}^{-1/2} \exp\{-\frac{1}{2}(x-\mu_1)'\Sigma^{-1}(x-\mu_1)\}\cdot\frac{1}{2}$

$f(\theta=0|x) = P(x|\theta_2=0)\pi(\theta=0) = \{(2\pi)^p|\Sigma|\}^{-1/2} \exp\{-\frac{1}{2}(x-\mu_0)'\Sigma^{-1}(x-\mu_0)\}\cdot\frac{1}{2}$

$f(\theta=0|x) \geq f(\theta=1|x) \Rightarrow \dfrac{f(\theta=0|x)}{f(\theta=1|x)} \geq 1 \Rightarrow \log\left(\dfrac{f(\theta=0|x)}{f(\theta=1|x)}\right) \geq 0$

$$\frac{f(\theta=0|x)}{f(\theta=1|x)} = \frac{\exp\{-\frac{1}{2}(x-\mu_0)'\Sigma^{-1}(x-\mu_0)\}}{\exp\{-\frac{1}{2}(x-\mu_1)'\Sigma^{-1}(x-\mu_1)\}}$$

$$\log\left(\frac{f(\theta=0|x)}{f(\theta=1|x)}\right) = -\frac{1}{2}(x-\mu_0)'\Sigma^{-1}(x-\mu_0)+\frac{1}{2}(x-\mu_1)'\Sigma^{-1}(x-\mu_1) > 0$$

$$(x-\mu_1)'\Sigma^{-1}(x-\mu_1)-(x-\mu_0)'\Sigma^{-1}(x-\mu_0) > 0$$

$$\cancel{x'\Sigma^{-1}x} - x'\Sigma^{-1}\mu_1 - \mu_1'\Sigma^{-1}x + \mu_1'\Sigma^{-1}\mu_1 - \cancel{x'\Sigma^{-1}x} + x'\Sigma^{-1}\mu_0 + \mu_0'\Sigma^{-1}x - \mu_0'\Sigma^{-1}\mu_0 > 0$$

$$\mu_1'\Sigma^{-1}\mu_1 - \mu_0'\Sigma^{-1}\mu_0 + 2x'\Sigma^{-1}(\mu_1-\mu_0) > 0$$

$$(\mu_1+\mu_0)'\Sigma^{-1}(\mu_1-\mu_0) + 2x'\Sigma^{-1}(\mu_1-\mu_0) > 0$$

$$[(\mu_1+\mu_0)' - 2x']\Sigma^{-1}(\mu_1-\mu_0) > 0$$

$$\left[\left(\frac{\mu_1+\mu_0}{2}\right)' - x'\right]\Sigma^{-1}(\mu_0-\mu_1) > 0$$

$$\left[x' - \left(\frac{\mu_1+\mu_0}{2}\right)'\right]\Sigma^{-1}(\mu_1-\mu_0) > 0$$

$$(\mu_1-\mu_0)'\Sigma^{-1}\left(x - \frac{\mu_1+\mu_0}{2}\right) > 0$$

let $\delta = (\mu_1-\mu_0)$

$\frac{\delta}{2} = \bar{\mu}$ $\delta'\Sigma^{-1}(x-\bar{\mu}) > 0$

$\therefore$ the Bayes Rule is

$$\begin{cases} \theta=0 & \text{if } \delta'\Sigma^{-1}(x-\bar{\mu}) \geq 0 \\ \theta=1 & \text{if } \delta'\Sigma^{-1}(x-\bar{\mu}) < 0 \end{cases}$$

2.1b) Derive the misclassification rate $R^*$ of the Bayes Rule

2018 Theory 1

$$R^* = \left[P(\delta'\Sigma^{-1}(x-\bar{\mu}) \geq 0 \mid \theta=1) + P(\delta'\Sigma^{-1}(x-\bar{\mu}) < 0 \mid \theta=0)\right]\frac{1}{2}$$

$$P(\delta'\Sigma^{-1}(x-\mu) \geq 0 \mid \theta=1) = P(\delta'\Sigma^{-1}(x-\tfrac{\mu_0+\mu_1}{2}) \geq 0 \mid \theta=1)$$

$$= P(\delta'\Sigma^{-1}(x-\mu_1-\tfrac{\mu_0-\mu_1}{2}) > 0 \mid \theta=1)$$

$$= P(\delta'\Sigma^{-1}(x-\mu_1-\tfrac{\delta}{2}) > 0 \mid \theta=1)$$

$$= P(\delta'\Sigma^{-1}(x-\mu_1) - \tfrac{1}{2}\delta'\Sigma^{-1}\delta > 0 \mid \theta=1)$$

$$= P(\delta'\Sigma^{-1}(x-\mu_1) > \tfrac{1}{2}\delta'\Sigma^{-1}\delta \mid \theta=1)$$

$$= P\left(\frac{\delta'\Sigma^{-1}(x-\mu_1)}{\sqrt{\delta'\Sigma^{-1}\delta}} > \frac{\delta'\Sigma^{-1}\delta}{2\sqrt{\delta'\Sigma^{-1}\delta}} \mid \theta=1\right)$$

$$= P\left(Z > \tfrac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right) = \Phi\left(-\tfrac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right)$$

when $\theta=1$, $x-\mu_1 \sim N(0,\Sigma)$

$\delta'\Sigma^{-1}(x-\mu_1) \sim N(0, \delta'\Sigma^{-1}\delta)$

$\frac{\delta'\Sigma^{-1}(x-\mu_1)}{\sqrt{\delta'\Sigma^{-1}\delta}} \sim N(0,1)$

let $Z \equiv N(0,1)$

let $\Phi(\cdot)$ be the CDF of a Standard Normal.

$$P(\delta'\Sigma^{-1}(x-\mu) < 0 \mid \theta=0) = P(\delta'\Sigma^{-1}(x-\tfrac{\mu_0+\mu_1}{2}) < 0 \mid \theta=0)$$

$$= P(\delta'\Sigma^{-1}(x-\mu_0+\tfrac{\delta}{2}) < 0 \mid \theta=0)$$

$$= P(\delta'\Sigma^{-1}(x-\mu_0) + \frac{\delta'\Sigma^{-1}\delta}{2} < 0 \mid \theta=0)$$

$$= P(\delta'\Sigma^{-1}(x-\mu_0) < -\tfrac{1}{2}\delta'\Sigma^{-1}\delta \mid \theta=0)$$

$$= P\left(Z < -\tfrac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right)$$

$$= \Phi\left(-\tfrac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right)$$

similarly when $\theta=0$

$x-\mu_0 \sim N(0,\Sigma)$

$\frac{\delta'\Sigma^{-1}(x-\mu_0)}{\sqrt{\delta'\Sigma^{-1}\delta}} \sim N(0,1)$

$$\therefore R^* = \tfrac{1}{2}\Phi\left(-\tfrac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right) + \tfrac{1}{2}\Phi\left(-\tfrac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right) = \Phi\left(-\tfrac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right)$$

2018 Theory 1

2. (c) Let $X_{oi}$ $(i=1,\dots,n_0)$ be iid samples from $\theta=0$
and $X_{ii}$ $(i=1,\dots,n_i)$ be iid samples from $\theta=1$ $\quad X_{oi} \perp\!\!\!\perp X_{ii}$

Derive the MLEs $(\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma})$ of $(\mu_0, \mu_1, \Sigma)$

Both involve $\Sigma$, so we need the joint likelihood of $(\bar{X}_0, \bar{X}_1)$

$$L(\mu_0,\mu_1,\Sigma) = \prod_{i=1}^{n_0}\left\{(2\pi)^p|\Sigma|\right\}^{-1/2}\exp\left\{-\tfrac{1}{2}(x_{oi}-\mu_0)\Sigma^{-1}(x_{oi}-\mu_0)\right\}\prod_{j=1}^{n_1}\left\{(2\pi)^p|\Sigma|\right\}^{-1/2}\exp\left\{-\tfrac{1}{2}(x_{ij}-\mu_1)\Sigma^{-1}(x_{ij}-\mu_1)\right\}$$

$$\ell(\mu_0,\mu_1,\Sigma) = -\frac{n_0}{2}\left[p\log(2\pi)+\log(|\Sigma|)\right]-\frac{1}{2}\sum_{i=1}^{n_0}(x_{oi}-\mu_0)\Sigma^{-1}(x_{oi}-\mu_0)$$
$$-\frac{n_1}{2}\left[p\log(2\pi)+\log(|\Sigma|)\right]-\frac{1}{2}\sum_{j=1}^{n_1}(x_{ij}-\mu_1)\Sigma^{-1}(x_{ij}-\mu_1)$$

$$\frac{\partial\ell(\mu_0,\mu_1,\Sigma)}{\partial\mu_0} = \sum_{i=1}^{n_0}\Sigma^{-1}(x_{oi}-\mu_0)\overset{\text{set}}{=}0 \quad\Rightarrow\quad \sum_{i=1}^{n_0}(x_{oi}-\mu_0)=0$$

$$\sum_{i=1}^{n_0}x_{oi}-n_0\mu_0=0 \;\Rightarrow\; \hat{\mu}_0=\frac{1}{n_0}\sum_{i=1}^{n_0}x_{oi}\overset{\text{let}}{=}\bar{X}_0$$

similarly, $\hat{\mu}_1=\frac{1}{n_1}\sum_{j=1}^{n_1}x_{ij}\overset{\text{let}}{=}\bar{X}_1$

let $\Omega=\Sigma^{-1}$

$$\frac{\partial\ell(\mu_0,\mu_1,\Sigma)}{\partial\Omega} = \frac{n_0}{2}\Sigma+\frac{n_1}{2}\Sigma-\frac{1}{2}\sum_{i=1}^{n_0}(x_{oi}-\mu_0)(x_{oi}-\mu_0)'-\frac{1}{2}\sum_{j=1}^{n_1}(x_{ij}-\mu_1)(x_{ij}-\mu_1)'\overset{\text{set}}{=}0$$

$$=\frac{n_0+n_1}{2}\Sigma-\frac{1}{2}\left(\sum_{i=1}^{n_0}(x_{oi}-\hat{\mu}_0)(x_{oi}-\hat{\mu}_0)'+\sum_{j=1}^{n_1}(x_{ij}-\hat{\mu}_1)(x_{ij}-\hat{\mu}_1)'\right)\overset{\text{set}}{=}0$$

$$\Rightarrow\hat{\Sigma}=\frac{1}{n_0+n_1}\left(\sum_{i=1}^{n_0}(x_{oi}-\hat{\mu}_0)(x_{oi}-\hat{\mu}_0)'+\sum_{j=1}^{n_1}(x_{ij}-\hat{\mu}_1)(x_{ij}-\hat{\mu}_1)'\right)$$

Derivation of MVN MLEs $\hat{\Sigma}$ using tr(.). redo this!

2(d) If we replace $(\mu_0, \mu_1, \Sigma)$ in the Bayes Rule w/ $(\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma})$ prove the misclassification rate of the resulting rule is:

$$\frac{1}{2}\Phi\left(\frac{\hat{\delta}^T\hat{\Sigma}^{-1}(\mu_1-\hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)+\frac{1}{2}\Phi\left(-\frac{\hat{\delta}^T\Sigma^{-1}(\mu_0-\hat{\mu})}{\sqrt{\hat{\delta}^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right) \quad \text{where } \hat{\delta}=\hat{\mu}_0-\hat{\mu}_1 \text{ and } \hat{\mu}=\frac{\hat{\mu}_0+\hat{\mu}_1}{2}$$

let $\tilde{R}=\frac{1}{2}P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\hat{\mu})\geq 0\mid\theta=1)+\frac{1}{2}P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\hat{\mu})<0\mid\theta=0)$

under $\theta=1$

$x-\mu_1\sim N(0,\Sigma)$

$\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_1)\sim N(0,\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta})$

$$P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\hat{\mu})\geq 0\mid\theta=1)=P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_1+(\mu_1-\hat{\mu}_1))\geq 0\mid\theta=1)$$

$$=P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_1)+\hat{\delta}'\hat{\Sigma}^{-1}(\mu_1-\hat{\mu}_1)\geq 0\mid\theta=1)$$

$$=P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_1)\geq-\hat{\delta}'\hat{\Sigma}^{-1}(\mu_1-\hat{\mu}_1)\mid\theta=1)$$

$$=P\left(z\geq-\frac{\hat{\delta}'\hat{\Sigma}^{-1}(\mu_1-\hat{\mu}_1)}{\sqrt{\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)=\Phi\left(\frac{\hat{\delta}'\hat{\Sigma}^{-1}(\mu_1-\hat{\mu}_1)}{\sqrt{\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)$$

under $\theta=0$

$x-\mu_0\sim N(0,\Sigma)$

$\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_0)\sim N(0,\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta})$

$$P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\hat{\mu})<0\mid\theta=0)=P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_0+\mu_0-\hat{\mu})<0\mid\theta=0)$$

$$=P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_0)+\hat{\delta}'\hat{\Sigma}^{-1}(\mu_0-\hat{\mu})<0\mid\theta=0)$$

$$=P(\hat{\delta}'\hat{\Sigma}^{-1}(x-\mu_0)<-\hat{\delta}'\hat{\Sigma}^{-1}(\mu_0-\hat{\mu})\mid\theta=0)$$

$$=P\left(z<-\frac{\hat{\delta}'\Sigma^{-1}(\mu_0-\hat{\mu})}{\sqrt{\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)=\Phi\left(-\frac{\hat{\delta}'\Sigma^{-1}(\mu_0-\hat{\mu})}{\sqrt{\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)$$

$$\therefore \tilde{R}=\frac{1}{2}\Phi\left(\frac{\hat{\delta}'\hat{\Sigma}^{-1}(\mu_1-\hat{\mu})}{\sqrt{\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)+\frac{1}{2}\Phi\left(-\frac{\hat{\delta}'\Sigma^{-1}(\mu_0-\hat{\mu})}{\sqrt{\hat{\delta}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\delta}}}\right)$$

2(e). We propose another classification rule that assigns $x$ to $\theta=0$ if & only if $\hat{\beta}^T(x-\hat{\mu})\geq 0$ where $\hat{\mu}=\left(\frac{\hat{\mu}_0+\hat{\mu}_1}{2}\right)$ and $\hat{\beta}$ solves

2018 Theory

$$\hat{\beta}=\underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\ \frac{1}{2}\beta'\hat{\Sigma}\beta-(\hat{\mu}_0-\hat{\mu}_1)'\beta+\lambda\sum_{j=1}^{p}|\beta_j|$$

Derive the majorization-minimization algorithm for solving $\hat{\beta}$.
as the penalty is $\lambda\sum_{j=1}^{p}|\beta_j|$ we know we will have a soft-thresholding in our soln.

$$\ell(\beta)=\ell(\tilde{\beta})+\nabla\ell(\tilde{\beta})'(\beta-\tilde{\beta})+\frac{1}{2}(\beta-\tilde{\beta})\nabla^2\ell(\tilde{\beta})(\beta-\tilde{\beta})$$
$$\leq\ell(\tilde{\beta})+\nabla\ell(\tilde{\beta})'(\beta-\tilde{\beta})+\frac{c}{2}(\beta-\tilde{\beta})(\beta-\tilde{\beta})$$

where $c\geq\sup_{\beta}\lambda_{max}(\nabla^2\ell(\beta))$

let $\ell(\beta)=\frac{1}{2}\beta'\hat{\Sigma}\beta-(\hat{\mu}_0-\hat{\mu}_1)'\beta$

$$\nabla\ell(\beta)=\hat{\Sigma}\beta-(\hat{\mu}_0-\hat{\mu}_1)\qquad\nabla^2\ell(\beta)=\hat{\Sigma}$$

$$\therefore\ \ell(\beta)=\ell(\tilde{\beta})+\left(\hat{\Sigma}\beta-(\hat{\mu}_0-\hat{\mu}_1)\right)'(\beta-\tilde{\beta})+\frac{c}{2}(\beta-\tilde{\beta})(\beta-\tilde{\beta})$$

$$\therefore\ \tilde{\beta}^{(new)}=\underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\ \frac{c}{2}\|\beta-\tilde{\beta}\|_2^2+\hat{\Sigma}\tilde{\beta}-(\hat{\mu}_0-\hat{\mu}_1)+\lambda\sum_{j=1}^{p}|\beta_j|$$

The KKT condition is given by:

$$c(\beta-\tilde{\beta})_j+(\hat{\Sigma}\tilde{\beta}-(\hat{\mu}_0-\hat{\mu}_1))_j+\text{sign}(\beta_j)=0\quad\text{for }\beta_j\neq 0$$
$$\left|c(\beta-\tilde{\beta})_j+(\hat{\Sigma}\tilde{\beta}-|\hat{\mu}_0-\hat{\mu}_1|)_j\right|\leq\lambda\quad\text{for }\beta_j=0$$

} maybe not write out on exam

$$\Rightarrow\ \tilde{\beta}_j^{(new)}=s\left(\tilde{\beta}_j-\frac{1}{c}(\hat{\Sigma}\beta-(\hat{\mu}_0-\hat{\mu}_1)),\ \lambda/c\right)\quad\text{with stepsize } 1/c\quad c\geq\sup_{\beta}\lambda_{max}(\nabla^2\ell(\beta))$$

The algorithm is as follows

1. Initialize $\beta$ at $\beta^{(0)}$
2. At the $k^{th}$ iteration, let $\beta^{(k)}=s\left(\beta_j^{(k-1)}-\frac{1}{c}(\hat{\Sigma}\beta^{(k-1)}-\hat{\delta}),\frac{\lambda}{c}\right)$
3. Update $\nabla\ell(\beta)$ with $\nabla\ell(\beta^{(k)})$

Iterate steps 2 & 3 until convergence.

2(f). Let $R_n$ denote the misclassification rate of (e) | 2018 Theory 1

Suppose we can show $\hat{\beta} \xrightarrow{p} \Sigma^{-1}(\mu_0 - \mu_1)$ as $n \to \infty$. Use this result to prove $R_n \xrightarrow{p} R^*$

$$R_n = \frac{1}{2} P\left(\hat{\beta}'(x-\hat{\mu}) \geq 0 \mid \theta=1\right) + \frac{1}{2} P\left(\hat{\beta}'(x-\hat{\mu}) < 0 \mid \theta=0\right)$$

$$P\left(\hat{\beta}'(x-\hat{\mu}) \geq 0 \mid \theta=1\right) = P\left(\hat{\beta}'(x-\mu_1+\mu_1-\hat{\mu}) \geq 0 \mid \theta=1\right)$$

$$= P\left(\hat{\beta}'(x-\mu_1) + \hat{\beta}'(\mu_1-\hat{\mu}) \geq 0 \mid \theta=1\right)$$

$$= P\left(\hat{\beta}'(x-\mu_1) \geq -\hat{\beta}'(\mu_1-\hat{\mu}) \mid \theta=1\right)$$

$$= P\left(Z \geq -\frac{\hat{\beta}'(\mu_1-\hat{\mu})}{\sqrt{\hat{\beta}\Sigma\hat{\beta}}}\right) = \Phi\left(\frac{\hat{\beta}'(\mu_1-\hat{\mu})}{\sqrt{\hat{\beta}\Sigma\hat{\beta}}}\right)$$

under $\theta=0$ $x-\mu_1 \sim N(0,\Sigma)$

$\hat{\beta}'(x-\mu_1) \sim N(0, \hat{\beta}'\Sigma\hat{\beta})$

$\frac{\hat{\beta}(x-\mu_1)}{\sqrt{\hat{\beta}'\Sigma\hat{\beta}}} \sim N(0,1)$

Similarly $P\left(\hat{\beta}'(x-\hat{\mu}) < 0 \mid \theta=0\right) = \Phi\left(-\frac{\hat{\beta}'(\mu_0-\hat{\mu})}{\sqrt{\hat{\beta}\Sigma\hat{\beta}}}\right)$

$$\therefore R_n = \frac{1}{2}\Phi\left(\frac{\hat{\beta}'(\mu_1-\hat{\mu})}{\sqrt{\hat{\beta}'\Sigma\hat{\beta}}}\right) + \frac{1}{2}\Phi\left(-\frac{\hat{\beta}'(\mu_0-\hat{\mu})}{\sqrt{\hat{\beta}'\Sigma\hat{\beta}}}\right)$$

recall $R^* = \Phi\left(-\frac{1}{2}\sqrt{\delta'\Sigma^{-1}\delta}\right)$

$= \Phi\left(-\frac{1}{2}\sqrt{(\mu_0-\mu_1)\Sigma^{-1}(\mu_0-\mu_1)}\right)$

By LLN $\hat{\mu} \xrightarrow{p} \mu$ $\hat{\beta}' \to (\mu_0-\mu_1)'\Sigma^{-1}$

$$R_n \to \frac{1}{2}\Phi\left(\frac{(\mu_0-\mu_1)'\Sigma^{-1}(\mu_1-\mu)}{\sqrt{\delta'\Sigma\delta}}\right) + \frac{1}{2}\Phi\left(-\frac{(\mu_0-\mu_1)\Sigma^{-1}(\mu_0-\mu)}{\sqrt{\delta'\Sigma\delta}}\right)$$

$$\frac{1}{2}\Phi\left(\frac{\frac{1}{2}\delta'\Sigma\delta}{\sqrt{\delta'\Sigma\delta}}\right) + \frac{1}{2}\Phi\left(\frac{-\frac{1}{2}\delta'\Sigma\delta}{\sqrt{\delta'\Sigma\delta}}\right) = \Phi\left(-\frac{1}{2}\sqrt{\delta'\Sigma\delta}\right)$$

$$\therefore R_n \xrightarrow{p} R^*$$

2018 #1

Kosorok Session 7/9

(a) Use lots of conditioning to get through it

Factorial moments ① $n$ ② $n(n-1)$ ③ $n(n-1)(n-2)$

possibly use the mgf?

$X_i^2 | N$ is $\chi^2$

(b) This works for all modes of convergence

$X_n \xrightarrow{a.s.} X$ $Y_n \xrightarrow{a.s.} Y$ $g(X,Y)$ is cont on a set $A: P_{X,Y}(A)=1$

∴ the function is continuous almost everywhere WRT $X$ and $Y$

∴ By CMT, $g(X_n, Y_n) \xrightarrow{a.s.} g(X,Y)$

$\frac{X_n}{Y_n} \xrightarrow{a.s.} \frac{X}{Y}$ when $P(Y=0)=0$

$X_n \xrightarrow{d} X$ $Y_n \xrightarrow{d} Y$ if $X$ or $Y$ are constant then $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ Y \end{pmatrix}$

If $X_n \to X$ $Y_n \to Y$ marginally, then they converge jointly automatically when converging a.s. & in p. But not necessarily in dist. in distribution, marginal conv. $\not\Rightarrow$ joint convergence unless independent or ⊛

(c) make sure when using CLT state the 2nd moment is bounded

(e) to show $\hat{\rho}^2$ is consistent, use CMT. for $\rho$

note, we want an asymptotic pivot "classic Slutsky's"

let $\rho^2 = \tau^2_{2,22}$

$$\frac{\sqrt{n}(\hat{W}_n - \sigma^2)}{\hat{\rho}_n} \xrightarrow{d} N(0,1) \quad \text{as} \quad \frac{\sqrt{n}(\hat{W}_n - \sigma^2)}{\sigma^2 \rho}$$

2018 Theory 1

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:

$U_i \sim \text{Bern}(e^{-\lambda})$

(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 $\hat{T}_n = -\log(n^{-1}\sum_{i=1}^n U_i)$ is almost surely consistent for $\lambda$, and that $\hat{W}_n = n^{-1}\sum_{i=1}^n X_i^2/\hat{T}_n$ is almost surely consistent for $\sigma^2$, 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 $\tau_1^2$.

(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 $q$th-quantile of a standard normal, is an asymptotically valid $1 - \alpha$ level confidence interval for $\sigma^2$.

1. Let $N$ be Poisson($\lambda$) and $z_1, z_2, \ldots \overset{iid}{\sim} N(0,\sigma^2)$ $\quad 0<\lambda<\infty \quad 0<\sigma^2<\infty$ | 2018 Theory 1

Let $X = I\{N>0\}\sum_{j=1}^{N} z_j$ and $x_1, \ldots, x_n$ be realizations of $X$

$= \sum_{j=1}^{N} z_j$ when $N>0$ 0ow.

let $u_i = I\{x_i = 0\}$ $\quad 1 \le i \le n$

(a) Show that $E[u_i] = e^{-\lambda}$, $E[x_i] = 0$, $E[x_i^2] = \lambda\sigma^2$, and $E[x_i^4] = 3(\lambda+\lambda^2)\sigma^2$

$$E[x_i] = E\left[I\{N>0\}\sum_{j=1}^{N} z_j\right] = E_N\left[E_z\left[I\{N>0\}\sum_{j=1}^{N} z_j \mid N\right]\right] = E_N\left[I\{N>0\}\sum_{j=1}^{N} E[z_j|N]\right]$$

$$= E_N\left[I\{N>0\}\sum_{j=1}^{N} 0\right] = E_N[0] = 0$$

$$E[u_i] = E[I\{x_i=0\}] = P(x_i=0) = P\left(I\{N>0\}\sum_{i=1}^{N} z_j = 0\right)$$

$$= P\left(I\{N>0\}=0 \cup \sum_{i=1}^{N} z_j = 0\right) = P(N=0) \cup P\left(\sum_{i=1}^{N} z_j = 0\right)$$

$$= f_N(0) + P\left(\sum_{i=1}^{N} z_j = 0\right) - P\left(N=0 \text{ AND } \sum_{i=1}^{N} z_j = 0\right)$$

$$= \frac{e^{-\lambda}\lambda^0}{0!} + 0 - 0 = e^{-\lambda}$$

$$x_i^2 = \left[I\{N>0\}\sum_{j=1}^{N} z_j\right]^2 = I\{N>0\}\left(\sum_{j=1}^{N} z_j\right)^2$$

$$E[x_i^2] = E\left[I[N>0]\left(\sum_{j=1}^{N} z_j\right)^2\right] = E_N\left[E_{x|N}\left[I[N>0]\left(\sum_{j=1}^{N} z_j\right)^2 \mid N\right]\right]$$

$$= E_N\left[I[N>0]\, E_{x|N}\left[\left(\sum_{j=1}^{N} z_j\right)^2 \mid N\right]\right] = E_N[N\sigma^2] = \sigma^2 E[N] = \lambda\sigma^2$$

$$x_i^4 = \left[I\{N>0\}\sum_{j=1}^{N} z_j\right]^4 = I\{N>0\}\left(\sum_{j=1}^{N} z_j\right)^4$$

$$E[x_i^4] = E\left[I(N>0)\left(\sum_{j=1}^{N} z_j\right)^4\right] = E_N\left[E_{x|N}\left[I(N>0)\left(\sum_{j=1}^{N} z_j\right)^4 \mid N\right]\right]$$

$$= E_N\left[I(N>0)\, E_{x|N}\left[\left(\sum_{j=1}^{N} z_j\right)^4 \mid N\right]\right] = E_N\left[I(N>0)\left(\text{var}_{z|N}\left(\sum z_j^2\right) + E\left[\left(\sum z_j^2\right)^2\right] \mid N\right]\right]$$

$$= E_N\left[2N^2(\sigma^2)^2 + (N\sigma^2)^2\right] = E_N\left[3N^2\sigma^4\right] = 3E_N[N^2]\sigma^4$$

$$= 3\sigma^4\left[\text{var}(N) + E(N)^2\right] = 3\sigma^4(\lambda+\lambda^2) \checkmark$$

$\sum_{i=1}^{N} z_j \mid N \sim N(0, N\sigma^2)$

$\frac{1}{\sqrt{N\sigma^2}}\sum_{i=1}^{N} z_j \mid N \sim N(0,1)$

$\frac{1}{N\sigma^2}\left(\sum_{i=1}^{N} z_j\right)^2 \mid N \sim \chi_1^2$

$\text{var}\left(\frac{1}{N\sigma^2}\sum z_j^2\right) \mid N = 2$

$\frac{1}{N^2\sigma^4}\text{var}\left(\sum z_j^2\right) \mid N = 2$

$\text{Var}\left(\sum z_j^2\right)^2 \mid N = 2N^2\sigma^4$

1(b). Show that $\hat{T}_n = -\log\left(\frac{1}{n}\sum_{i=1}^{n} U_i\right) \xrightarrow{a.s.} \lambda$

and $\hat{W}_n = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^2}{\hat{T}_n} \xrightarrow{a.s.} \sigma^2$ as $n \to \infty$

$U_i$ are iid w/ mean $e^{-\lambda}$

$\therefore$ by SLLN, $\frac{1}{n}\sum_{i=1}^{n} U_i \xrightarrow{a.s.} e^{-\lambda}$

$\therefore$ by CMT $\log\left(\frac{1}{n}\sum_{i=1}^{n} U_i\right) \xrightarrow{a.s.} -\lambda$

$-\log\left(\frac{1}{n}\sum_{i=1}^{n} U_i\right) \xrightarrow{a.s.} \lambda \Rightarrow \hat{T}_n \xrightarrow{a.s.} \lambda$

$X_i^2$ are iid w/ mean $\lambda\sigma^2$

$\therefore$ by SLLN $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{a.s.} \lambda\sigma^2$

because marginal $\Rightarrow$ joint convergence a.s.

$\therefore \begin{pmatrix} \frac{1}{n}\sum X_i^2 \\ \hat{T}_n \end{pmatrix} \xrightarrow{a.s.} \begin{pmatrix} \lambda\sigma^2 \\ \lambda \end{pmatrix} \Rightarrow$ by CMT

by CMT, $\frac{1}{\hat{T}_n} \xrightarrow{a.s.} \frac{1}{\lambda}$

$\therefore \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\hat{T}_n} \xrightarrow{a.s.} \frac{\lambda\sigma^2}{\lambda} \equiv \sigma^2 \Rightarrow \hat{W}_n \xrightarrow{a.s.} \sigma^2$

when $P(\lambda = 0) = 0$ which is true as $0 < \lambda < \infty$

2018 Theory 1

1(c). Show that $\sqrt{n}\begin{pmatrix} 1/n \sum U_i - e^{-\lambda} \\ 1/n \sum X_i^2 - \lambda\sigma^2 \end{pmatrix} \xrightarrow{d} N(0, \tau_i^2)$ as $n \to \infty$, give the form of $\tau_i^2$

by CLT we know $\sqrt{n}\left(\frac{1}{n}\sum U_i - e^{-\lambda}\right) \xrightarrow{d} N(0, \text{var}(U_i))$

$\sqrt{n}\left(\frac{1}{n}\sum X_i^2 - \lambda\sigma^2\right) \xrightarrow{d} N(0, \text{var}(X_i^2))$

$U_i \sim \text{Bernoulli}(e^{-\lambda})$

$\text{var}(U_i) = E[U_i^2] - E[U_i]^2$

$= E[I(X_i=0)^2] - (e^{-\lambda})^2 = E[I(X_i=0)] - (e^{-\lambda})^2 = (e^{-\lambda}) - (e^{-\lambda})^2$

$= e^{-\lambda} - e^{-2\lambda} = e^{-\lambda}(1 - e^{-\lambda})$

$\text{var}(X_i^2) = E[X_i^4] - E[X_i^2]^2 = 3(\lambda + \lambda^2)\sigma^4 - (\lambda\sigma^2)^2$

$= 3\lambda\sigma^4 + 3\lambda^2\sigma^4 - \lambda^2\sigma^4 = 3\lambda\sigma^4 + 2\lambda^2\sigma^4 = \lambda\sigma^4(3 + 2\lambda)$

$\text{cov}(U_i, X_i^2) = E[U_i X_i^2] - E[U_i]E[X_i^2]$

$E[U_i X_i^2] = E[I(X_i=0) X_i^2] = 0$ because the indicator will make any value of $X_i \neq 0$ $= 0$ in the product, so $\forall$ values of $X_i$, $U_i X_i^2 = 0$

$U_i X_i^2 = I(X_i=0) X_i^2 = \{X_i^2 \text{ when } X_i = 0;\ 0 \text{ where } X_i \neq 0\} = 0 \ \forall$ values of $X$.

$\therefore \text{cov}(U_i, X_i^2) = -E[U_i]E[X_i^2] = -e^{-\lambda}\lambda\sigma^2$

$\therefore \sqrt{n}\begin{pmatrix} \frac{1}{n}\sum U_i - e^{-\lambda} \\ \frac{1}{n}\sum X_i^2 - \lambda\sigma^2 \end{pmatrix} \xrightarrow{d} N(0, \tau_i^2)$

where $\tau_i^2 = \begin{pmatrix} \text{var}(U_i) & \text{cov}(U_i, X_i^2) \\ & \text{var}(X_i^2) \end{pmatrix} = \begin{pmatrix} e^{-\lambda}(1-e^{-\lambda}) & -e^{-\lambda}\lambda\sigma^2 \\ -e^{-\lambda}\lambda\sigma^2 & \lambda\sigma^4(3+2\lambda) \end{pmatrix}$

1(d). Show that $\sqrt{n}\begin{pmatrix}\hat{T}_n-\lambda\\ \hat{W}_n-\sigma^2\end{pmatrix}\xrightarrow{d} N(0,\tau_2^2)$ as $n\to\infty$

let $\frac{1}{n}\sum U_i=\bar{U}$ $\quad \hat{T}_n=-\log\bar{U}$

let $\frac{1}{n}\sum X_i^2=\bar{X}$ $\quad \hat{W}_n=\frac{\bar{X}}{-\log\bar{U}}$ $\quad$ let $g(a,b)=\begin{pmatrix}-\log(a)\\ -\frac{b}{\log(a)}\end{pmatrix}$

$$\nabla g(a,b)=\begin{pmatrix}-\frac{1}{a} & 0\\ \frac{b}{a\log(a)^2} & -\frac{1}{\log(a)}\end{pmatrix}$$

from (c) we know $\sqrt{n}\begin{pmatrix}\bar{U}-e^{-\lambda}\\ \bar{X}-\lambda\sigma^2\end{pmatrix}\xrightarrow{d} N(0,\tau_1^2)$

$\therefore$ by Delta method $\sqrt{n}\begin{pmatrix}\hat{T}_n-\lambda\\ \hat{W}_n-\sigma^2\end{pmatrix}\xrightarrow{d}\nabla g(e^{-\lambda},\lambda\sigma^2)N(0,\tau_1^2)$

$\equiv N(0,\tau_2^2)$

$$\nabla g(e^{-\lambda},\lambda\sigma^2)\begin{pmatrix}-e^{\lambda} & 0\\ \frac{\lambda\sigma^2}{e^{-\lambda}\lambda^2} & \frac{1}{\lambda}\end{pmatrix}=\begin{pmatrix}-e^{\lambda} & 0\\ \frac{e^{-\lambda}\sigma^2}{\lambda} & \frac{1}{\lambda}\end{pmatrix}$$

$$\tau_2^2=\begin{pmatrix}-e^{\lambda} & 0\\ e\frac{\lambda\sigma^2}{\lambda} & \frac{1}{\lambda}\end{pmatrix}\begin{pmatrix}e^{-\lambda}(1-e^{-\lambda}) & -e^{-\lambda}\lambda\sigma^2\\ -e^{-\lambda}\lambda\sigma^2 & \lambda\sigma^4(3+2\lambda)\end{pmatrix}\begin{pmatrix}-e^{\lambda} & e\frac{\lambda\sigma^2}{\lambda}\\ 0 & 1/\lambda\end{pmatrix}$$

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

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

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

1.(e). Show that $\hat{W}_n \pm z_{1-\alpha/2}\hat{\rho}_n/\sqrt{n}$ is an asymptotically valid $1-\alpha$ level CI for $\delta^2$

Go back to Kushal's Solution