Xiao Zhehao A0235422N

1. a) $\alpha(c)=P(T\in C;\mu=0)=P(X>c;\mu=0)=1-P(X\le c;\mu=0)=1-\Phi(c)$

b) $\beta(c)=P(T\notin C;\mu=1)=P(X\le c;\mu=1)=\Phi(c;\mu=1,\sigma=1)=\Phi(c-1;\mu=0,\sigma=1)=\Phi(c-1)$

c) To find c^*, we find where $\frac{d}{dc}R(c)=0$, assuming $R(c)$ is continuous and convex. $\frac{d}{dc}(1-\Phi(c)+\Phi(1-c))=\frac{1}{\sqrt{2\pi\sigma^2}}\left(-e^{-\frac{c^2}{2\sigma^2}}+e^{-\frac{(1-c)^2}{2\sigma^2}}\right)=0$ $c^2=(1-c)^2=1-2c+c^2$ $c=\frac{1}{2}$.
Hence, $c^*=\frac{1}{2}$ and $R(c^*)=1-\Phi(\frac{1}{2})+\Phi(-\frac{1}{2})=1-(1-\Phi(-\frac{1}{2}))+\Phi(-\frac{1}{2})=2\Phi(-\frac{1}{2})$

2. a) By Neyman Pearson's Lemma, $L(\theta)=\prod_{i=1}^{10}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{X_i^2}{2\sigma^2}}=(2\pi\sigma^2)^{-5}e^{-\frac{1}{2\sigma^2}\sum X_i^2}$ and $\frac{L(\sigma^2=1)}{L(\sigma^2=2)}=\left(\frac{2\pi}{4\pi}\right)^{-5}e^{-\frac{1}{2}\sum X_i^2+\frac{1}{4}\sum X_i^2}=32e^{-\frac{1}{4}\sum X_i^2}\le k$ $-\frac{1}{4}\sum X_i^2\le\ln\frac{1}{32}k$ $\sum X_i^2>-4\ln\frac{k}{32}=c$
As $\sum X_i^2$ increases, the likelihood decreases so the critical region is of form $C=\{\sum X_i^2>c\}$.
Hence, we find $P(\sum X_i^2>c;\sigma^2=1)=0.05$. Since $\sum X_i^2\sim\chi^2(10)$, $\chi^2_{0.05}(10)=18.31$ so the best critical region is $\{\sum X_i^2>18.31\}$.

b) $K(2)=1-\beta(18.31)=1-P(\sum X_i^2\le 18.31;\sigma^2=2)$. Under $\sigma^2=2$, $\sum X_i\sim 2\cdot\chi^2(10)$. Hence $1-P(2\cdot\chi^2(10)\le 18.31)=1-P(\chi^2(10)\le 9.155)\approx 1-0.4824=0.518$

c) Since the critical region controls Type I error under H_0, and H_0 is the same as a), the critical region is also same at $\{\sum X_i^2>18.31\}$

d) $K(4)=1-\beta(18.31)=1-P(\sum X_i^2\le 18.31;\sigma^2=4)=1-P(4\cdot\chi^2(10)\le 18.31)=1-P(\chi^2(10)\le 4.5775)=0.918$

e) Since the critical region controls Type I error under H_0, and H_0 is the same as a), the critical region is also same at $\{\sum X_i^2>18.31\}$

3. a) As $\bar{x}-\bar{y}\sim N(\mu_1,\frac{400}{n})-N(\mu_2,\frac{225}{n})=N(\mu_1-\mu_2,\frac{400+225}{n})$, $K(\theta)=1-\beta=1-P(\bar{x}-\bar{y}<c;\mu_1-\mu_2>0)=1-P(N(\mu_1-\mu_2,\frac{400+225}{n})<c)=1-\Phi\left(\frac{c-\theta}{\sqrt{625/n}}\right)$

b) Under H_0, $\mu_1-\mu_2=0$ so $1-\Phi\left(\frac{c}{\sqrt{625/n}}\right)=0.05$. $K(10)=1-\Phi\left(\frac{c-10}{\sqrt{625/n}}\right)=0.9$ so $\frac{c}{\sqrt{625/n}}=1.645$ & $\frac{c-10}{\sqrt{625/n}}=-1.28$. $1.645\sqrt{\frac{625}{n}}=10-1.28\sqrt{\frac{625}{n}}$ $10=2.925\sqrt{\frac{625}{n}}$ $n=53.472\approx 54$ $c=5.596$

4. a) $L(\mu_1)=\prod_{i=1}^{m}f_N(X_i;\mu_1,1)=\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}}e^{-\frac{(X_i-\mu_1)^2}{2}}=\frac{1}{(2\pi)^{m/2}}e^{-\frac{1}{2}\sum_{i=1}^{m}(X_i-\mu_1)^2}$ and similarly $L(\mu_2)=\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\sum_{j=1}^{n}(Y_j-\mu_2)^2}$
Since the samples are independent, $L(\mu_1,\mu_2)=L(\mu_1)\cdot L(\mu_2)=\frac{1}{(2\pi)^{m/2}}e^{-\frac{1}{2}\sum_{i=1}^{m}(X_i-\mu_1)^2}\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\sum_{j=1}^{n}(Y_j-\mu_2)^2}=\frac{1}{(2\pi)^{(m+n)/2}}e^{-\frac{1}{2}\sum_{i=1}^{m}(X_i-\mu_1)^2}e^{-\frac{1}{2}\sum_{j=1}^{n}(Y_j-\mu_2)^2}$

b) $\arg\max_{\mu}L(\mu)$ is at $\frac{d}{d\mu}L(\mu)$ and also $\frac{d}{d\mu}\ln(L(\mu))$ as $\ln$ is a monotonically increasing function over $L(\mu)$'s range.
$\frac{d}{d\mu}\ln(L(\mu))=-\frac{1}{2}\sum_{i=1}^{m}\frac{d}{d\mu}(X_i-\mu)^2=-\frac{1}{2}\sum(-2(X_i-\mu))=\sum_{i=1}^{m}(X_i-\mu)=-m\mu+\sum_{i=1}^{m}X_i=0$ and $\mu=\frac{1}{m}\sum_{i=1}^{m}X_i=\bar{X}$. Hence $\hat{\mu}_1=\frac{1}{m}\sum_{i=1}^{m}X_i=\bar{X}$ and $\hat{\mu}_2=\frac{1}{n}\sum_{j=1}^{n}Y_j=\bar{Y}$

c) Assuming $\mu_1=\mu_2=\mu$, $L(\mu_1,\mu_2)=L(\mu)=\frac{1}{(2\pi)^{(m+n)/2}}e^{-\frac{1}{2}\sum_{i=1}^{m}(X_i-\mu)^2}e^{-\frac{1}{2}\sum_{j=1}^{n}(Y_j-\mu)^2}$.
By b), $\arg\max_{\mu}L(\mu)$ is at $\frac{d}{d\mu}\ln(L(\mu))=\sum_{i=1}^{m}X_i-m\mu+\sum_{j=1}^{n}Y_j-n\mu=0$ $(m+n)\mu=\sum_{i=1}^{m}X_i+\sum_{j=1}^{n}Y_j$ and $\hat{\mu}=\frac{\sum_{i=1}^{m}X_i+\sum_{j=1}^{n}Y_j}{m+n}=\frac{m\bar{X}+n\bar{Y}}{m+n}$

d) We see $\Sigma(x_i-\hat{\mu})^2-\Sigma(x_i-\bar{x})^2=\Sigma(x_i^2-2\hat{\mu}x_i+\hat{\mu}^2-x_i^2+2x_i\bar{x}-\bar{x}^2)=-2\hat{\mu}\Sigma x_i+m\hat{\mu}^2+2\bar{x}\Sigma x_i-m\bar{x}^2=-2m\hat{\mu}\bar{x}+m\hat{\mu}^2+2m\bar{x}^2-m\bar{x}^2=m(\hat{\mu}^2-2\hat{\mu}\bar{x}+\bar{x}^2)=m(\bar{x}-\hat{\mu})^2=m\left(\bar{x}-\frac{m\bar{x}+n\bar{y}}{m+n}\right)^2=m\left(\frac{n}{m+n}(\bar{x}-\bar{y})\right)^2$
$\Lambda=\frac{\max_{\mu}L(\mu)}{\max_{\mu_1,\mu_2}L(\mu_1,\mu_2)}=e^{-\frac{1}{2}\sum_{i=1}^{m}(X_i-\hat{\mu})^2}e^{-\frac{1}{2}\sum_{j=1}^{n}(Y_j-\hat{\mu})^2}\Big/e^{-\frac{1}{2}\sum_{i=1}^{m}(X_i-\bar{x})^2}e^{-\frac{1}{2}\sum_{j=1}^{n}(Y_j-\bar{y})^2}=\exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{m}(X_i-\hat{\mu})^2-\sum_{i=1}^{m}(X_i-\bar{x})^2+\sum_{j=1}^{n}(Y_j-\hat{\mu})^2-\sum_{j=1}^{n}(Y_j-\bar{y})^2\right)\right\}=\exp\left\{-\frac{1}{2}(m(\bar{x}-\hat{\mu})^2+n(\bar{y}-\hat{\mu})^2)\right\}$
$=\exp\left\{-\frac{1}{2}\left[m\left(\frac{n}{m+n}(\bar{x}-\bar{y})\right)^2+n\left(\frac{m}{m+n}(\bar{x}-\bar{y})\right)^2\right]\right\}=\exp\left\{-\frac{1}{2}(\bar{x}-\bar{y})^2\left[\frac{n^2m+m^2n}{(m+n)^2}\right]\right\}=\exp\left\{-\frac{1}{2}(\bar{x}-\bar{y})^2\left[\frac{nm(m+n)}{(m+n)^2}\right]\right\}=\exp\left\{-\frac{1}{2}\left(\frac{nm}{n+m}\right)(\bar{x}-\bar{y})^2\right\}$

e) So we reject H_0 for $\Lambda\le k$ $\Lambda=\exp\left\{-\frac{1}{2}\left(\frac{nm}{n+m}\right)(\bar{x}-\bar{y})^2\right\}=k$, $-\frac{1}{2}\frac{nm}{n+m}(\bar{x}-\bar{y})^2\le\ln k$, $\frac{nm}{n+m}(\bar{x}-\bar{y})^2\ge 2\ln k$ Since both distributions have $\sigma^2=1$ $Var(\bar{x})=\frac{1}{m}$ $Var(\bar{y})=\frac{1}{n}$ $Var(\bar{x}-\bar{y})=\frac{1}{m}+\frac{1}{n}=\frac{n+m}{nm}$
Hence, $(\bar{x}-\bar{y})/\sqrt{\frac{n+m}{nm}}\sim N(0,1)$ and if we let $z_{\alpha/2}=\sqrt{2\ln k}$, the critical region is $C=\left\{|\bar{x}-\bar{y}|/\sqrt{\frac{n+m}{nm}}\ge z_{\alpha/2}\right\}$. Abs as we use a right tailed test for $\mu_1\neq\mu_2$.

ST2132 Mathematical Statistics
Assignment 6
Due date: 21 April 2025 23:59 on Canvas
Please submit in pdf format.

Please work on the assignment by yourself only. We follow a strict rule on academic honesty.

Generative AI policy: The use of generative AI tools (e.g., ChatGPT, Gemini, etc.) is strictly prohibited for structural or conceptual questions in this course. However, these tools may be used to assist with coding tasks, provided their use is clearly documented.

1. (Risk minimization and a simplified binary classifier) In lecture, we learnt about the Neyman-Pearson framework: we seek to find a "best" critical region that maximizes the power of the test while maintaining a given significance level. Another major paradigm in the literature to define a "best" critical region is known as risk minimization that we now introduce. Let $X \sim N(\mu, 1)$. We are interested in testing

$$H_0 : \mu = 0, \quad H_1 : \mu = 1.$$

As we only have a single sample X, we intend to use the test statistic T and critical region C to be respectively

$$T = X, \quad C = \{X > c\},$$

where $c > 0$ is a constant. Let Φ be the standard normal cdf.

(a) Prove that the probability of type I error is

$$\alpha(c) = 1 - \Phi(c).$$

The bracket of c on the left hand side is to indicate the dependence on c of α, that is, α is a function of c.

(b) Prove that the probability of type II error is

$$\beta(c) = \Phi(c - 1).$$

The bracket of c on the left hand side is to indicate the dependence on c of β, that is, β is a function of c.

(c) Define the risk R to be the sum of the probability of type I and type II error, that is,

$$R(c) = \alpha(c) + \beta(c) = 1 - \Phi(c) + \Phi(c - 1).$$

In risk minimization, we seek to find an optimal critical region by minimizing $R(c)$ with respect to c. Let c^* be the resulting minimizer, that is,

$$c^* = \arg\min_c R(c).$$

Prove that

$$c^* = \frac{1}{2}.$$

and hence

$$R(c^*) = 2\Phi(-1/2).$$

As a result, the optimal critical region from risk minimization is $\{X > \frac{1}{2}\}$ with a probability of type I error $\alpha(1/2)$ and probability of type II error $\beta(1/2)$.

The above can be considered as a simplified machine learning algorithm known as classification. Let X be a normalized weight of an animal, which is either a cat or a dog. If we fail to reject H_0, we say that X belongs to "group 0" (think of dog). If we reject H_0, we say that X belongs to "group 1" (think of cat). In other words, if the normalized weight of an animal is $X \leq 1/2$, we classify the animal as a dog. If the normalized weight is $X > 1/2$, we classify the animal as a cat.

2. Let $X_1, \ldots, X_{10}$ be a random sample of $n = 10$ from a normal distribution $N(0, \sigma^2)$.

   (a) Find a best critical region of size 0.05 for testing

   $$H_0 : \sigma^2 = 1, \quad H_1 : \sigma^2 = 2.$$

   (b) Deduce the power of the test in part (a), that is, compute the power function $K(2)$. Feel free to use any computing language to help you compute the power.

   (c) Find a best critical region of size 0.05 for testing

   $$H_0 : \sigma^2 = 1, \quad H_1 : \sigma^2 = 4.$$

   (d) Deduce the power of the test in part (c), that is, compute the power function $K(4)$.

   (e) Find a best critical region of size 0.05 for testing

   $$H_0 : \sigma^2 = 1, \quad H_1 : \sigma^2 = \sigma_1^2,$$

   where $\sigma_1^2 > 1$.

3. Consider two independent normal distributions $N(\mu_1, 400)$ and $N(\mu_2, 225)$. Let $\theta = \mu_1 - \mu_2$. Let $\overline{x}$ and $\overline{y}$ denote the observed means of two independent random samples, each of size n, from these two distributions. To test

   $$H_0 : \theta = 0, \quad H_1 : \theta > 0,$$

   we use the critical region

   $$C = \{\overline{x} - \overline{y} \geq c\}.$$

   (a) Express the power function $K(\theta)$ in terms of the standard normal distribution cdf Φ. It depends on n and c.

   (b) Find n and c so that the probability of type I error is 0.05, and the power at $\theta = 10$ is 0.9, approximately. Assume that $z_{0.10} = 1.28$.

4. (Pooled z-test as likelihood ratio test) Let $X_1, \ldots, X_m \overset{i.i.d.}{\sim} N(\mu_1, 1)$ and $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} N(\mu_2, 1)$. Suppose that these two samples are independent. We would like to test

   $$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$

   using the likelihood ratio test.

   (a) Prove that the likelihood function can be written as

   $$L(\mu_1, \mu_2) = \frac{1}{(2\pi)^{(m+n)/2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} (x_i - \mu_1)^2 \right\} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} (y_j - \mu_2)^2 \right\}.$$

   (b) Prove that the maximum likelihood estimators of μ_1 and μ_2 are respectively the sample mean of X and Y, that is,

   $$\widehat{\mu}_1 = \overline{X} = \frac{1}{m} \sum_{i=1}^{m} X_i, \quad \widehat{\mu}_2 = \overline{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j.$$

(c) Under $H_0 : \mu_1 = \mu_2$, let us write $\mu_1 = \mu_2 = \mu$. Prove that the maximum likelihood estimator of μ, assuming H_0 is true, is the pooled estimator

$$\widehat{\mu} = \frac{\sum_{i=1}^{m} X_i + \sum_{j=1}^{n} Y_j}{m+n} = \frac{m\overline{X} + n\overline{Y}}{m+n}.$$

(d) Using part (a), (b) and (c), prove that the likelihood ratio can be written as

$$\frac{\max_{\mu_1=\mu_2=\mu} L(\mu,\mu)}{\max_{\mu_1,\mu_2} L(\mu_1,\mu_2)} = \exp\left[-\frac{1}{2}\frac{mn}{(m+n)}(\bar{x}-\bar{y})^2\right].$$

(e) Using part (d), prove that the critical region of the likelihood ratio test with significance level α is of the form

$$C = \left\{\sqrt{\frac{mn}{m+n}}|\overline{x}-\overline{y}| \geq z_{\alpha/2}\right\}.$$