Statistical Inference Exercises
Chapter 8

XJZ (SID: 224071400028)

October 7, 2024

a

1 Simulation

```r
# For multivariate normal generation
library(MASS)
# Define the T(X) function with omega = 1/d
compute_T <- function(X) {
  T_val <- mean(tan((2 * pnorm(abs(X)) - 3/2) * pi))
  return(T_val)
}

# Function to generate the AR(1) correlation matrix
generate_ar_matrix <- function(rho, d) {
  #using outer
  Sigma <- outer(1:d, 1:d, function(i, j) rho^abs(i - j))
  return(Sigma)
}

# Set parameters
d_values <- c(5, 20, 50, 100, 300, 500)  # Different dimensionalities
rho_values <- c(0.2, 0.4, 0.6, 0.8, 0.99)  # Different values of rho
alpha_values <- c(0.1, 0.01, 0.001)  # Different alpha levels
n_samples <- 10^5  # Number of Monte Carlo samples

# Loop through each combination of d, rho, and alpha
results <- data.frame()
set.seed(2024)
use_time <- system.time({
  for (d in d_values) {
    for (rho in rho_values) {
      # Generate the AR(1) correlation matrix for this rho and d
      Sigma <- generate_ar_matrix(rho, d)

      # Generate Monte Carlo samples
      X_matrix <- mvrnorm(n = n_samples, mu = rep(0, d), Sigma = Sigma)

      # Compute T(X) for each sample (each row of X_matrix)
      T_vals <- apply(X_matrix, 1, compute_T)

      for (alpha in alpha_values) {
        # Calculate the upper alpha-quantile of the standard Cauchy distribution
        t_alpha <- qcauchy(alpha,lower.tail = FALSE)  # Upper alpha-quantile of the standard Cauchy

        # Calculate the empirical probability
        P_empirical <- mean(T_vals > t_alpha)

        # Store the results
        results <- rbind(results, data.frame(d = d, rho = rho, alpha = alpha, P = P_empirical))
      }
    }
  }
})
# Print the time taken
print(use_time)
   user  system elapsed
 57.234   4.470  46.920
# Display results
```

```r
library(dplyr)
library(ggplot2)
library(magrittr)
results %>%
  mutate(
    ratio=P/alpha,
    alpha=factor(alpha,levels=c(0.1,0.01,0.001))
    ) %>%
  ggplot(aes(x=alpha,y=ratio))+
  geom_boxplot()+
  theme_minimal()+
  xlab("Significance level")+
  ylab("(Emprical size)/(Significance level)")
```

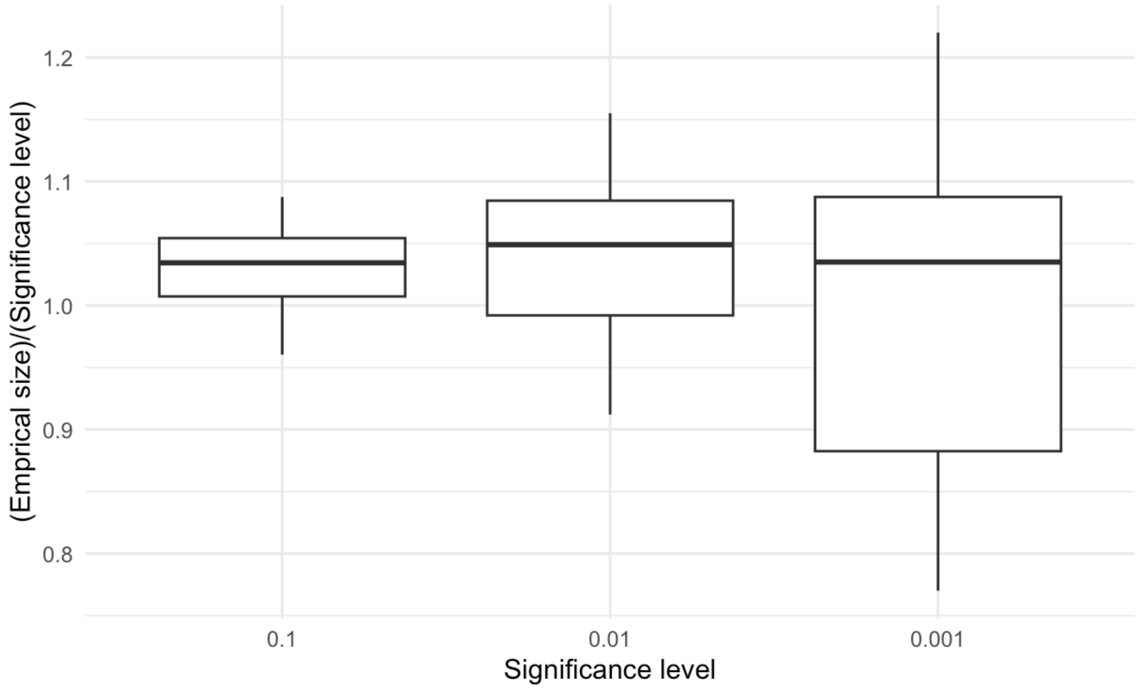

Figure 1: The ratio of empirical size to significance level summarized by boxplots

2 Exercises

Question 1. A special case of a normal family is one in which the mean and the variance are related, the $N(\theta, a\theta)$ family. If we are interested in testing this relationship, regardless of the value of θ, we are again faced with a nuisance parameter problem.

a) Find the LRT of $H_0 : a = 1$ versus $H_1 : a \neq 1$ based on a sample $X_1, \ldots, X_n$ from a $N(\theta, a\theta)$ family, where θ is unknown.

b) A similar question can be asked about a related family, the $N\left(\theta, a\theta^2\right)$ family. Thus, if $X_1, \ldots, X_n$ are iid $N\left(\theta, a\theta^2\right)$, where θ is unknown, find the LRT of $H_0 : a = 1$ versus $H_1 : a \neq 1$

Solution:

We first determine the maximum likelihood estimators (MLE) under both unrestricted and restricted conditions. Beginning with the unrestricted case, the likelihood function of (θ, a) is given by:

$$L(\theta, a) = (2\pi a\theta)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2a\theta}\sum_{i=1}^{n}(x_i - \theta)^2\right\},$$

and the corresponding log-likelihood function is:

$$\begin{aligned}\log L(\theta,a) &= -\frac{n}{2}\log(2\pi a\theta) - \frac{1}{2a\theta}\sum_{i=1}^{n}(x_i-\theta)^2 \\ &= -\frac{n}{2}\log(2\pi a\theta) - \frac{1}{2a\theta}\left(\sum_{i=1}^{n}(x_i-\theta)^2 + \sum_{i=1}^{n}(\bar{x}-\theta)^2\right) \quad \text{let } T=\sum_{i=1}^{n}(x_i-\bar{x})^2 \\ &= -\frac{n}{2}\log(2\pi a\theta) - \frac{1}{2a\theta}\left(T+n(\bar{x}-\theta)^2\right)\end{aligned}$$

To obtain the MLE in the unrestricted case, we take partial derivatives of the log-likelihood function with respect to a and θ and set them to zero.

$$\frac{\partial \log L(\theta,a)}{\partial a} = -\frac{n}{2a} + \frac{T+n(\bar{x}-\theta)^2}{2\theta a^2} \overset{\text{set}}{=} 0 \tag{1}$$

$$\frac{\partial \log L(\theta,a)}{\partial \theta} = -\frac{n}{2\theta} - \frac{1}{2a\theta^2}(T+n(\bar{x}-\theta)^2 + 2n\theta(\bar{x}-\theta)) \overset{\text{set}}{=} 0 \tag{2}$$

From the equation (1), we solve for a:

$$a = \frac{T+n(\bar{x}-\theta)^2}{n\theta} = \frac{\hat{\sigma}^2+(\bar{x}-\theta)^2}{\theta}.$$

Substituting this into the equation (2) simplifies to:

$$\begin{aligned}\frac{\partial \log L(\theta,a)}{\partial \theta} &= -\frac{n}{2\theta} - \frac{1}{2a\theta^2}\left(na\theta + 2n\theta(\bar{x}-\theta)\right) \\ &= \frac{na\theta - na\theta - 2n\theta(\bar{x}-\theta)}{2a\theta^2} = 0,\end{aligned}$$

This implies $\hat{\theta}_{MLE} = \bar{x}$ and $\hat{a}_{MLE} = \frac{\hat{\sigma}^2}{\bar{x}}$ in the unrestricted case.

For the restricted case where $a=1$, we differentiate the log-likelihood function with respect to θ:

$$\begin{aligned}\frac{\partial \log L(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta}\left(-\frac{n}{2}\log\theta - \frac{1}{2\theta}(T+n(\bar{x}-\theta)^2)\right) \\ &= -\frac{n}{2\theta} - \frac{1}{2\theta^2}\left(T+n(\bar{x}-\theta)^2 + 2n\theta(\bar{x}-\theta)\right) \\ &= \frac{-\theta^2-\theta+\left(\frac{T}{n}+\bar{x}\right)}{2n\theta^2} \overset{\text{set}}{=} 0.\end{aligned}$$

Since $\theta>0$, we have $\hat{\theta}_0 = \frac{-1+\sqrt{1+4(\hat{\sigma}^2+\bar{x}^2)}}{2}$. (To verify that these are indeed maxima, we need to check the Hessian matrix. However, due to the large calculations, we will omit the details here). The likelihood ratio test (LRT) statistic is given by:

$$\begin{aligned}\lambda(x) &= \frac{L(\hat{\theta}_{MLE},\hat{a}_{MLE}|x)}{L(\hat{\theta}_0, a=1|x)} \\ &= \frac{(2\pi\hat{\theta}_{MLE}\hat{a}_{MLE})^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\hat{\theta}_{MLE}\hat{a}_{MLE}}\sum_{i=1}(x_i-\hat{\theta}_{MLE})^2\right\}}{(2\pi\hat{\theta}_0)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\hat{\theta}_0}\sum_{i=1}(x_i-\hat{\theta}_0)^2\right\}} \\ &= \left(\frac{\hat{\theta}_0}{\hat{\sigma}^2}\right)^{\frac{n}{2}}\exp\left\{-\frac{n}{2}+\frac{1}{2\hat{\theta}_0}\sum_{i=1}^{n}(x_i-\hat{\theta}_0)^2\right\}.\end{aligned}$$

With the same steps above, the log likelihood function of model $N(\theta, a\theta^2)$ is

$$\log L(\theta,a|x) = -\frac{n}{2}\log 2\pi a\theta^2 - \frac{1}{2a\theta^2}\sum_{i=1}^{n}(x_i-\theta)^2$$

Thus

$$\frac{\partial \log L(\theta,a)}{\partial a} = -\frac{n}{2a} + \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\theta^2 a^2} \overset{\text{set}}{=} 0 \tag{3}$$

$$\frac{\partial \log L(\theta,a)}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{a\theta^3} + \frac{n\bar{x}-n\theta}{a\theta^2} \overset{\text{set}}{=} 0 \tag{4}$$

Solving equation (3) we get

$$a = \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{n\theta^2}.$$

Again, substitute it into equation (4) to get

$$\frac{\partial \log L(\theta,a)}{\partial \theta} = \frac{n\bar{x}-n\theta}{\frac{\sum_{i=1}^{n}(x_1-\theta)^2}{n}} \overset{\text{set}}{=} 0$$

which implies that $\hat{\theta} = \bar{x}$ then $\hat{a} = \frac{\hat{\sigma}^2}{\bar{x}^2}$ (Again, it is tedious to operate large calculations to verify it is indeed maxima, so we omit the details here). Under the null hypothesis,

$$\begin{aligned}\frac{\partial \log L(\theta|x)}{\partial \theta} &= -\frac{n}{\theta} + \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{\theta^3} + \frac{n\bar{x}-n\theta}{\theta^2} \\ &= \frac{-\theta^2+\hat{\sigma}^2+\bar{x}^2-\bar{x}\theta}{\theta^3/n} \overset{\text{set}}{=} 0\end{aligned}$$

which have two solutions $\theta_1 = \frac{-\bar{x}+\sqrt{5\bar{x}^2+4\hat{\sigma}^2}}{2}$ and $\theta_2 = \frac{-\bar{x}-\sqrt{5\bar{x}^2+4\hat{\sigma}^2}}{2}$. In the interval $(-\infty,\theta_2]$, $\log L(\theta|x)$ is monotonic decreasing, while in $[\theta_2,\theta_1]$ it is monotonic increasing, and in $[\theta_1,+\infty)$ it is monotonic decreasing again. Note that $\lim_{\theta\to-\infty}\log L(\theta|x) = -\infty$, so the maximum occurs at $\theta=\theta_1$. Therefore, the estimate of $\hat{\theta}_0$ is, $\hat{\theta}_0 = \frac{-\bar{x}+\sqrt{5\bar{x}^2+4\hat{\sigma}^2}}{2}$. So, the LRT is

$$\lambda(x) = \frac{L(\hat{\theta},\hat{a}|x)}{L(\hat{\theta}_0,a=1|x)} = \left(\frac{\hat{\theta}_0}{\hat{\sigma}}\right)^2 \exp\left\{-\frac{n}{2}+\frac{1}{2\hat{\theta}_0}\sum_{i=1}^{n}(x_i-\hat{\theta}_0)^2\right\}.$$

Question 2. Let X_1, X_2 be iid uniform $(\theta, \theta+1)$. For testing $H_0: \theta = 0$ versus $H_1: \theta > 0$, we have two competing tests:

$$\begin{aligned}\phi_1(X_1):&\ \text{Reject } H_0 \text{ if } X_1 > .95\\ \phi_2(X_1,X_2):&\ \text{Reject } H_0 \text{ if } X_1+X_2 > C\end{aligned}$$

a) Find the value of C so that ϕ_2 has the same size as ϕ_1.
b) Calculate the power function of each test. Draw a well-labeled graph of each power function.
c) Prove or disprove: ϕ_2 is a more powerful test than ϕ_1.
d) Show how to get a test that has the same size but is more powerful than ϕ_2.

Solution:

We first derive the power function for the tests ϕ_1 and ϕ_2, respectively. For ϕ_1, the power function is given by

$$\begin{aligned}\beta_1(\theta) &= P_\theta(X_1 > 0.95)\\ &= 1 - P_\theta(X_1 \le 0.95)\\ &= 1 - P(X_1 - \theta \le 0.95 - \theta)\\ &= 1 - P(U \le 0.95 - \theta),\end{aligned}$$

where U follows a standard uniform distribution.Thus,

$$\begin{aligned}\beta_1(\theta) =& 1 - \begin{cases} 0, & 0.95-\theta < 0; \\ 0.95-\theta, & 0 \le 0.95-\theta \le 1; \\ 1, & 0.95-\theta > 1. \end{cases} \\ =& \begin{cases} 0, & \theta < -0.05; \\ \theta+0.05, & -0.05 \le \theta \le 0.95; \\ 1, & \theta > 0.95. \end{cases}\end{aligned}$$

For ϕ_2, the power function is:

$$\begin{aligned}\beta_2(\theta) &= P_\theta(X_1+X_2>C) \quad (5)\\ &= \int_{x_1+x_2>C} \mathbf{1}_{\theta\le x_1\le\theta+1}\mathbf{1}_{\theta\le x_2\le\theta+1} dx_1 dx_2 \quad (6)\\ &= \begin{cases} 1, & c \le 2\theta; \\ 1-\frac{(c-2\theta)^2}{2}, & 2\theta < c < 2\theta+1; \\ \frac{(2\theta+2-c)^2}{2}, & 2\theta+1 \le c \le 2\theta+2; \\ 0, & c \geqslant 2\theta+2. \end{cases} \quad (7)\end{aligned}$$

To find the critical value C, we set $\theta = 0$ in equation (7):

$$\beta_2(0) = \begin{cases} 1, & c \le 0; \\ 1-\frac{c^2}{2}, & 0 < c < 1; \\ \frac{(2-c)^2}{2}, & 1 \le c \le 2; \\ 0, & c \geqslant 2. \end{cases}$$

Since $\beta_2(0)$ is monotonic decreasing in C, we solve for C such that $\beta_2(0) = \beta_1(0) = 0.05$, giving $C = 2-\frac{1}{\sqrt{10}} \approx 1.68$.

The power curves for both tests are shown in Figure 2. From this figure, we observe that both ϕ_1 and ϕ_2 control the type-I error well. Test ϕ_1 is more powerful than ϕ_2 for values of θ near 0, while ϕ_2 becomes more powerful for larger values of θ.

Additionally, we note that $X_1 > 1$ and $X_2 > 1$ occur with probability 0 under the null hypothesis. Therefore, if $X_1 > 1$ or $X_2 > 1$, we can conclude that $\theta > 0$. Based on this, we construct a new test, ϕ_3, as follows:

$$\phi_3(X_1, X_2) : \text{Reject } H_0 \text{ if } X_1 > 1 \text{ or } X_2 > 1 \text{ or } X_1 + X_2 > C.$$

ϕ_3 is a test of size 0.05 since

$$\begin{aligned}0.05 = P_{\theta=0}(X_1+X_2>C) <& P_{\theta=0}(X_1>1 \text{ or } X_2>1 \text{ or } X_1+X_2>C) \\ <& P_{\theta=0}(X_1>1) + P_{\theta=0}(X_2>1) + P_{\theta=0}(X_1+X_2>C) \\ =& P_{\theta=0}(X_1+X_2>C) \quad X_1>1 \text{ and } X_2>1 \text{ are events with probability 0} \\ =& 0.05\end{aligned}$$

And clearly more powerful than ϕ_2.

Question 3. One very striking abuse of α levels is to choose them after seeing the data and to choose them in such a way as to force rejection (or acceptance) of a null hypothesis. To see what the true Type I and Type II Error probabilities of such a procedure are, calculate size and power of the following two trivial tests:

a) Always reject H_0, no matter what data are obtained (equivalent to the practice of choosing the α level to force rejection of H_0).
b) Always accept H_0, no matter what data are obtained (equivalent to the practice of choosing the α level to force acceptance of H_0).

Solution:

If we always reject H_0, then

$$\text{Size} = P(\text{reject } H_0 | H_0 \text{ is true}) = 1;$$

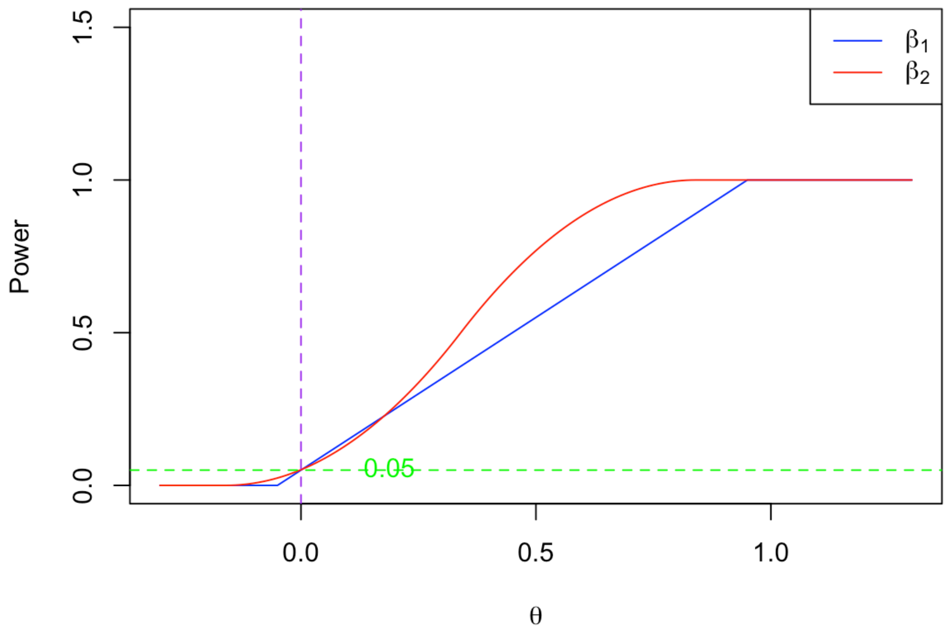

Figure 2: Power of ϕ_1 and ϕ_2

$$\text{Power} = P(\text{reject } H_0 | H_0 \text{ is false}) = 1$$

So we have type-I error 1 but type-II error 0.

If we always accept H_0, then

$$\begin{aligned} \text{Size} &= P(\text{reject } H_0 | H_0 \text{ is true}) = 0; \\ \text{Power} &= P(\text{reject } H_0 | H_0 \text{ is false}) = 0 \end{aligned}$$

So we have type-I error 0 but type-II error 1.

Question 4. Let X be a random variable whose pmf under H_0 and H_1 is given by

x	1	2	3	4	5	6	7
$f(x \mid H_0)$	.01	.01	.01	.01	.01	.01	.94
$f(x \mid H_1)$	.06	.05	.04	.03	.02	.01	.79

Use the Neyman-Pearson Lemma to find the most powerful test for H_0 versus H_1 with size $\alpha = .04$. Compute the probability of Type II Error for this test.

SOLUTION:

To use Neyman-Pearson Lemma in a discrete variable, we first calculate $\frac{f(x|H_1)}{f(x|H_0)}$ for each x:

x	1	2	3	4	5	6	7
$\frac{f(x\|H_1)}{f(x\|H_0)}$	6	5	4	3	2	1	0.84

Since the ratio $\frac{f(x|H_1)}{f(x|H_0)}$ decreases as x increases, the most powerful test will reject H_0 for smaller values of x, where the likelihood ratio is larger. To control the size at $\alpha = 0.04$, we observe that $P(x \leq 4|H_0) = 0.04$, meaning we reject H_0 if $x \leq 4$. The probability of a type-II error in this test is $P(x > 4|H_1) = 0.82$.

Question 5. Suppose X is one observation from a population with beta$(\theta, 1)$ pdf.

a) For testing $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$, find the size and sketch the power function of the test that rejects H_0 if $X > \frac{1}{2}$.

b) Find the most powerful level α test of $H_0 : \theta = 1$ versus $H_1 : \theta = 2$.

c) Is there a UMP test of $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$? If so, find it. If not, prove so.

SOLUTION:

The power function of the given test is derived as follows:

$$\begin{aligned}\beta(\theta) &= \int_{1/2}^{1} \frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)} x^{\theta-1}(1-x)^{1-1} dx \\ &= 1 - \frac{1}{2^{\theta}}.\end{aligned}$$

The size of a test is the supremum of the power function over the null hypothesis H_0, so the size α is:

$$\alpha = \sup_{\theta \leq 1} \beta(\theta) = \frac{1}{2}.$$

To construct the most powerful test of $H_0 : \theta = 1$ against $H_1 : \theta = 2$, we use the Neyman-Pearson Lemma. We first calculate the likelihood ratio:

$$\lambda = \frac{f(x|\theta_1)}{f(x|\theta_0)} = \frac{\frac{1}{B(\theta_1,1)} x^{\theta_1-1}}{\frac{1}{B(\theta_0,1)} x^{\theta_0-1}} \propto x^{\theta_1-\theta_0} \tag{8}$$

which is monotone increasing in x if $\theta_1 - \theta_0$, i.e. it is MLR. When $\theta_1 = 2$ and $\theta_0 = 1$, $\lambda \propto x$. According to the Neyman-Pearson Lemma, we reject H_0 if $\lambda > k$ for some threshold k, which is equivalent to rejecting H_0 if $x > k'$ for some corresponding k'. The Neyman-Pearson Lemma says this test is the most powerful test. Now we determine k' with α.

$$\alpha = P_{\theta=1}(X > k') = \int_{k'}^{1} \frac{1}{B(1,1)} x^{1-1}(1-x)^{1-1} dx = 1 - k'$$

so the most powerful test is the one we reject H_0 if $X > 1 - \alpha$. Since it is MLR, so there is a UMP test of $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$ from the Karlin-Rubin theorem. In fact, the rejection region is also $X > 1 - \alpha$.

Remark 1. We now consider testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ where $\theta_1 > \theta_0$. From equation (8), we know that the likelihood ratio is given by:

$$\lambda(x) \propto x^{\theta_1-\theta_0}.$$

This ratio is monotonic in x whenever $\theta_1 > \theta_0$, meaning that rejecting H_0 for large values of x (i.e. when $x > k\prime$ for some threshold $k\prime$) is optimal. According to the Neyman-Pearson Lemma, this rejection rule provides the most powerful test (MPT). Importantly, the rejection rule $x > k'$ is independent of the specific values of θ_1 and θ_0, as long as $\theta_1 > \theta_0$.

Thus, this test can be used for any θ_1 as long as $\theta_1 > \theta_0$. In fact, if ϕ is an MPT of size α for testing $H_0 : \theta = \theta_0$ against any $\theta_1 \in \Theta_1$, then ϕ remains an MPT of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta_1$.

From this discussion, we find that the test ϕ, which rejects H_0 if $X > 1-\alpha$, is the uniformly most powerful test (UMPT) for the following three scenarios:

$$\begin{aligned} &H_0 : \theta = 1 \text{ versus } H_1 : \theta = 2, \\ &H_0 : \theta = 1 \text{ versus } H_1 : \theta > 1, \\ &H_0 : \theta \leq 1 \text{ versus } H_1 : \theta > 1. \end{aligned}$$

We can explain this observation through two key points:

1. Under any $H_0 : \theta \in \Theta_0$ above, testing $\theta = 1$ against any H_1 above is the most difficult (because $\theta = 1$ is the closest to Θ_1). Hence, controlling the type-I error at $\theta = 1$ is sufficient to control the size.
2. As long as Θ_1 is on the same side relative to $\theta = 1$, the test is the same and does not depend on Θ_1. Hence, testing H_0 against H_1 is equivalent to testing H_0 against any simple alternative hypothesis $\theta = \theta_1$.

Hence, all three hypothesis testing problems above are reduced to the problem of testing simple hypotheses.

Question 6. Let $f(x \mid \theta)$ be the Cauchy scale pdf

$$f(x \mid \theta) = \frac{\theta}{\pi} \frac{1}{\theta^2 + x^2}, \quad -\infty < x < \infty, \quad \theta > 0$$

a) Show that this family does not have an MLR.

b) If X is one observation from $f(x \mid \theta)$, show that $|X|$ is sufficient for θ and that the distribution of $|X|$ does have an MLR.

SOLUTION:

For $\theta_1 > \theta_0 > 0$ the likelihood ratio is given by

$$\lambda = \frac{f(x|\theta_1)}{f(x|\theta_0)} = \frac{\theta_1}{\theta_0} \frac{\theta_0^2 + x^2}{\theta_1^2 + x^2}.$$

Taking its derivative with respect to x, we obtain

$$\frac{\partial \lambda}{\partial x} = \frac{\theta_1}{\theta_0} \frac{\theta_1^2 - \theta_0^2}{\left(\theta_1^2 + x^2\right)^2} x,$$

which have positive value for $x > 0$ and negative value for $x < 0$, hence, λ is not monotone, i.e., this family does not have an MLR.

Given that $f(x \mid \theta) = \frac{\theta}{\pi} \frac{1}{\theta^2 + |x|^2}$, by the factorization theorem, $|x|$ is a sufficient statistic. The pdf of $|x|$ is

$$f(y \mid \theta) = f(y \mid \theta) = \frac{2\theta}{\pi} \frac{1}{\theta^2 + y^2}, \quad y > 0, \quad \theta > 0.$$

Now, take the derivative of the likelihood ratio of $|x|$ to get

$$\frac{\partial \lambda}{\partial y} = \frac{\theta_1}{\theta_0} \frac{\theta_1^2 - \theta_0^2}{\left(\theta_1^2 + y^2\right)^2} y > 0, \qquad y > 0, \quad \theta_1 > \theta_0 > 0.$$

so $|X|$ has a MLR.

Question 7. Let $X_1, \ldots, X_n$ be a random sample from the uniform $(\theta, \theta + 1)$ distribution. To test $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, use the test

$$\text{reject } H_0 \text{ if } Y_n \geq 1 \text{ or } Y_1 \geq k$$

where k is a constant, $Y_1 = \min\{X_1, \ldots, X_n\}, Y_n = \max\{X_1, \ldots, X_n\}$.

a) Determine k so that the test will have size α.

b) Find an expression for the power function of the test in part (a).

c) Prove that the test is UMP level α.

d) Find values of n and k so that the UMP .10 level test will have power at least .8 if $\theta > 1$.

SOLUTION:

We first derive the PDF of Y_1 and Y_n. The joint PDF is given by:

$$f(y_1, y_n) = n(n-1)(y_n - y_1)^{n-2}, \quad \theta < y_1 \leq y_n < \theta + 1.$$

and the marginal PDF for Y_1 and Y_n are:

$$\begin{aligned} f(y_1) =& n(1 - (y_1 - \theta))^{n-1}, \quad \theta < y_1 < \theta + 1; \\ f(y_n) =& n(y_n - \theta)^{n-1}, \quad \theta < y_1 < \theta + 1. \end{aligned}$$

Under H_0, the event $Y_n \geq 1$ occurs with probability 0, so we compute the significance level α as follows:

$$\begin{aligned}\alpha &= P_{\theta=0}(Y_n \geq 1 \text{ or } Y_1 \geq k)\\ &= P_{\theta=0}(Y_1 \geq k)\\ &= P_{\theta=0}(X_i \geq k \text{ for any } i = 1 \cdots n)\\ &= (1-k)^n \qquad (k < 1, \text{ otherwise } \alpha = 0)\end{aligned}$$

so $k = 1 - \alpha^{\frac{1}{n}}$.

To find the power of the test, we divide the range of θ into four disjoint intervals: $\theta \leq -\alpha^{\frac{1}{n}}$, $-\alpha^{\frac{1}{n}} < \theta \leq 0$, $0 < \theta \leq 1 - \alpha^{\frac{1}{n}}$, and $\theta > 1 - \alpha^{\frac{1}{n}}$. For each interval, we compute $P_\theta(Y_n \geq 1 \text{ or } Y_1 \geq k)$.

When $\theta \leq -\alpha^{\frac{1}{n}}$, we have $\theta + 1 \leq k$ so $P(Y_1 \geq k) = 0$, hence $P(Y_n \geq 1 \text{ or } Y_1 \geq k) = 0$. When $-\alpha^{\frac{1}{n}} < \theta \leq 0$, implying that $P_\theta(Y_n \geq 1) = 0$, so

$$P_\theta(Y_n \geq 1 \text{ or } Y_1 \geq k) = P_\theta(Y_1 \geq k) = (\theta + 1 - k)^n.$$

For $0 < \theta \leq 1 - \alpha^{\frac{1}{n}}$, we have

$$P_\theta(Y_n \geq 1 \text{ or } Y_1 \geq k) = P_\theta(Y_n \geq 1) + P_\theta(Y_1 \geq k) - P_\theta(Y_n \geq 1, Y_1 \geq k).$$

We calculate them one by one:

$$P_\theta(Y_n \geq 1) = \int_1^{\theta+1} n(y_n - \theta)^{n-1} dy_n = 1 - (1-\theta)^n.$$

$$P_\theta(Y_1 \geq k) = \int_k^{\theta+1} n(1 - (y_1 - \theta))^{n-1} dy_1 = \alpha.$$

$$P_\theta(Y_n \geq 1, Y_1 \geq k) = \int_1^{\theta+1} \int_k^{y_n} n(n-1)(y_n - y_1)^{n-2} dy_1 dy_n = (\theta + \alpha^{\frac{1}{n}})^n - \alpha.$$

Thus

$$P_\theta(Y_n \geq 1 \text{ or } Y_1 \geq k) = 1 - (1-\theta)^n + 2\alpha - (\theta + \alpha^{\frac{1}{n}})^n.$$

For $\theta > 1 - \alpha^{\frac{1}{n}}$, it is clear that $P_\theta(Y_n \geq 1 \text{ or } Y_1 \geq k) = 1$.

Note that

$$f(x_{1:n}|\theta) = \mathbf{1}_{\{\theta < y_1\}} \mathbf{1}_{\{y_n < \theta+1\}} = \mathbf{1}_{\{\theta < y_1 \leq y_n < \theta+1\}}$$

by the factorization theorem, (Y_1, Y_n) are sufficient statistics for θ. We can, therefore apply Corollary 8.3.13 to find the UMP (Uniformly Most Powerful) test. The likelihood ratio is

$$\lambda = \frac{L(y_1, y_n|\theta_1)}{L(y_1, y_n|0)} = \frac{\mathbf{1}_{\{\theta_1 < y_1 \leq y_n < \theta+1\}}}{\mathbf{1}_{\{0 < y_1 \leq y_n < 1\}}}.$$

To illustrate how the value of λ changes with different (y_1, y_n), we provide two figures, based on θ, in Figure 3 and Figure 4. When $\theta > 1$, λ takes two values (if $\frac{0}{0}$ is defined, the other regions take a value of 1). When $0 < \theta < 1$, it takes three values (with other regions similarly defined as 1).

The rejection decision depends on k and θ. If $\theta > k$, as shown in Figure 5, when $\lambda > 0$, we always reject H_0, and when $\lambda < 0$, we never reject H_0 (the rejection region is $y_1 > k$ and $y_n > 1$). In Figure 6, if $\lambda > 1$, we always reject H_0, and if $\lambda < 1$, we never reject H_0.

Thus, by Corollary 8.3.13, the test is uniformly most powerful (UMP) at level α for a given θ. Since θ is arbitrary, the test is UMP at level α for $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.

We have known that when $\theta > 1 - \alpha^{\frac{1}{n}}$, $P_\theta(Y_n \geq 1 \text{ or } Y_1 \geq k) = 1$, so $\beta(\theta) = 1$ for any $\theta > 1$ so any n is OK and $k = 1 - 0.1^{\frac{1}{n}}$.

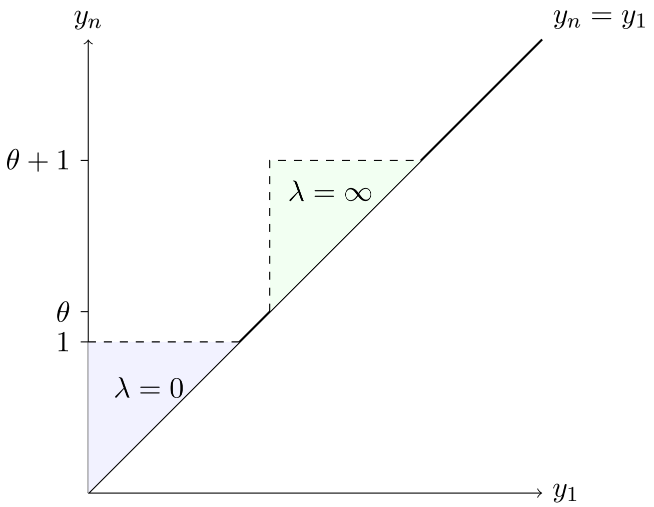

Figure 3: $\theta > 1$

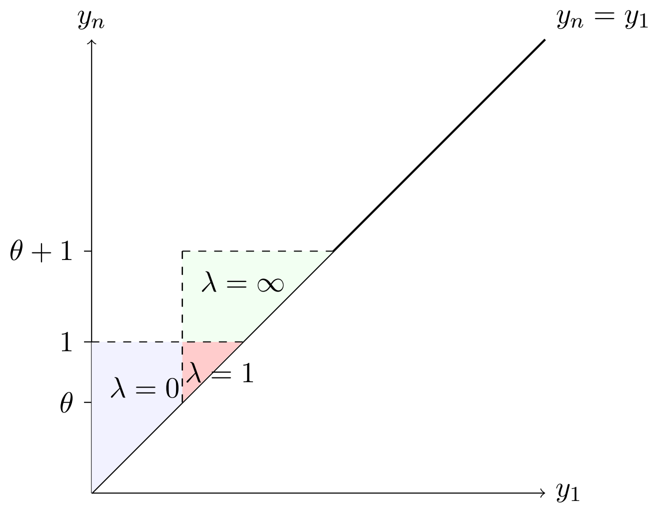

Figure 4: $0 < \theta < 1$

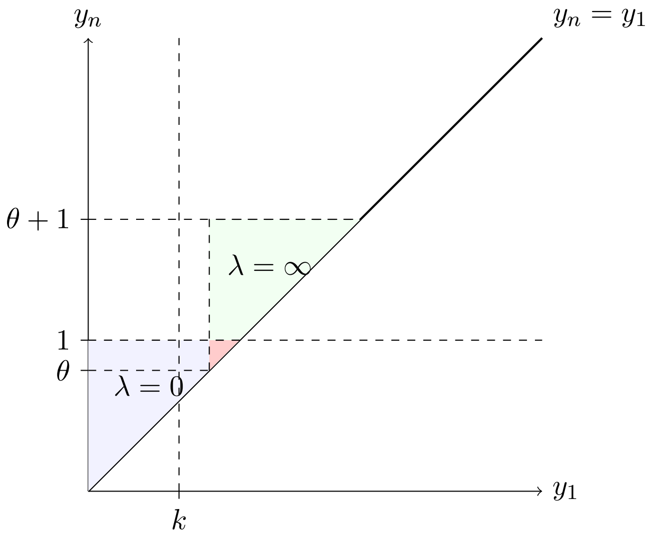

Figure 5: $\theta > k$

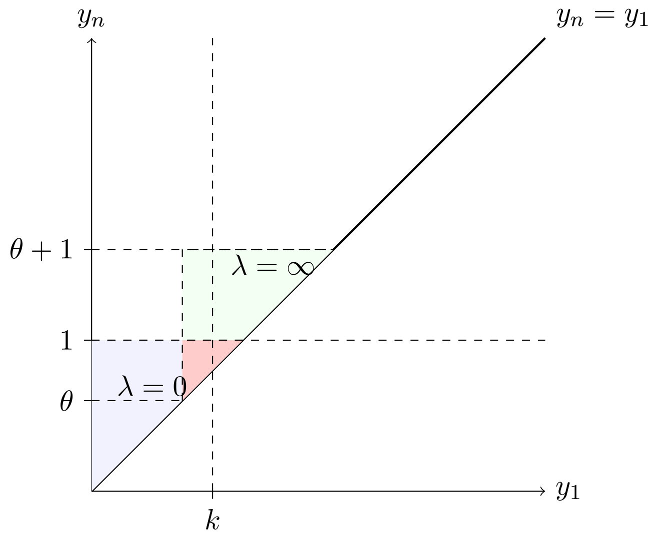

Figure 6: $0 < \theta < k$

Question 8. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate normal distribution with parameters $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$. We are interested in testing

$$H_0 : \mu_X = \mu_Y \quad \text{versus} \quad H_1 : \mu_X \neq \mu_Y$$

a) Show that the random variables $W_i = X_i - Y_i$ are iid $N\left(\mu_W, \sigma_W^2\right)$.

b) Show that the above hypothesis can be tested with the statistic

$$T_W = \frac{\bar{W}}{\sqrt{\frac{1}{n} S_W^2}}$$

where $\bar{W} = \frac{1}{n}\sum_{i=1}^n W_i$ and $S_W^2 = \frac{1}{(n-1)}\sum_{i=1}^n \left(W_i - \bar{W}\right)^2$. Furthermore, show that, under $H_0, T_W \sim$ Student's t with $n-1$ degrees of freedom. (This test is known as the paired-sample t test.)

Solution:

We know that $(X_i, Y_i)^T \sim N\left(\binom{\mu_X}{\mu_Y}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix}\right)$, then $X_i - Y_i = (1, -1)(X_i, Y_i)^T$, which implies that

$$X_i - Y_i \sim N\left((1, -1)\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, (1, -1)\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix}(1, -1)^T\right).$$

i.e. $X_i - Y_i \sim N(\mu_W, \sigma_W^2)$ where $\mu_W = \mu_X - \mu_Y, \sigma_W^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$. Clearly, $X_i - Y_i$ are independent.

Testing $H_0 : \mu_X = \mu_Y$ is equivalent to testing H_0': $\mu_W = \mu_X - \mu_Y = 0$, so we can base the test on W_i. Using the likelihood ratio test (LRT) for H_0'': $\mu_W = \mu_0$ against $H_1'' : \mu_W \neq \mu_0$, we generalize the mean test for a normal distribution, where $\mu_0 = 0$ is a special case in this context.

The likelihood function for (μ_W, σ_W^2) is

$$L(\mu_W, \sigma_W^2) = (2\pi\sigma_W^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_W^2}\sum_{i=1}^{n}(W_i - \mu_W)^2\right\}.$$

The unrestricted and H_0-restricted MLEs of (μ_W, σ_W^2) are $(\hat{\mu}_W, \hat{\sigma}_W^2)$ and $(\mu_0, \hat{\sigma}_0^2)$, respectively, where

$$\hat{\mu}_W = \bar{W} = \frac{1}{n}\sum_{i=1}^{n} W_i, \quad \hat{\sigma}_W^2 = \frac{1}{n}\sum_{i=1}^{n}(W_i - \bar{W})^2, \quad \hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(W_i - \mu_0)^2.$$

So, the LR is

$$\begin{aligned}
\lambda &= \frac{L(\hat{\mu}_W, \hat{\sigma}_W^2)}{L(\mu_0, \hat{\sigma}_0^2)} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_W^2}\right)^{\frac{n}{2}} \\
&= \left(\frac{\sum_{i=1}^{n}(W - \bar{W})^2 + n(\bar{W} - \mu_0)^2}{\sum_{i=1}^{n}(W - \bar{W})^2}\right)^{\frac{n}{2}} \\
&= \left(1 + \frac{1}{n-1}\frac{\frac{(\bar{W}-\mu_0)^2}{\sigma_W^2/n}}{\frac{\sum_{i=1}^{n}(W-\bar{W})^2}{(n-1)\sigma_W^2}}\right)^{\frac{n}{2}} \\
&= \left(1 + \frac{1}{n-1}T^2\right)^{\frac{n}{2}} = \left(1 + \frac{1}{n-1}\frac{U^2}{V^2}\right)^{\frac{n}{2}}.
\end{aligned}$$

Where

$$T = \frac{U}{V}, \qquad U = \frac{\sqrt{n}(\bar{W} - \mu_0)}{\sigma_W}, \qquad V = \frac{\sqrt{S_W^2}}{\sigma_W}.$$

Clearly, λ is increasing with T^2. The rejection region is $T^2 > c \iff |T| > c'$. In fact $|T|$ is $|T_W|$. Under H_0, $U \sim N(0,1)$ and $(n-1)V^2 \sim \chi^2_{n-1}$ and they are independent, hence $T_W = T = \frac{U}{V} \sim t_{n-1}$.

Remark 2. In addition to testing using T_W, we can also use T_W^2 which follows a $F_{(1,n-1)}$ distribution. The power function for this test can be derived as follows:

$$\begin{aligned}
\beta(\mu_W, \sigma_W^2) &= P_{\mu_W,\sigma_W^2}\{|T_W| > c\} \\
&= P_{\mu_W,\sigma_W^2}\left\{\frac{|U|}{V} > c\right\} \\
&= P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W} - \mu_0)}{\sigma_W} < -cV\right) + P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W} - \mu_0)}{\sigma_W} > cV\right) \\
&= P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W} - \mu_W)}{\sigma_W} < -cV + \frac{\theta_0 - \mu_W}{\sigma/\sqrt{n}}\right) + P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W} - \mu_0)}{\sigma_W} > cV + \frac{\theta_0 - \mu_W}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

Note that, generally, $Z := \sqrt{n}(\bar{W} - \mu_W)/\sigma_W \sim N(0,1)$ and $V^2 \sim \chi^2_{n-1}/(n-1)$ are independent. The distribution of Z, V are free of μ. By the law of iterative expectations, we have

$$\begin{aligned}
P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W} - \mu_W)}{\sigma_W} < -cV + \frac{\theta_0 - \mu_W}{\sigma/\sqrt{n}}\right) =& E\left\{P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W} - \mu_W)}{\sigma_W} < -cV + \frac{\theta_0 - \mu_W}{\sigma/\sqrt{n}} \mid V\right)\right\} \\
=& E\Phi\left(-cV + \frac{\theta_0 - \mu_W}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

$$P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W}-\mu_W)}{\sigma_W} > cV + \frac{\theta_0-\mu_W}{\sigma/\sqrt{n}}\right) = E\left\{P_{\mu_W,\sigma_W^2}\left(\frac{\sqrt{n}(\bar{W}-\mu_W)}{\sigma_W} > cV + \frac{\theta_0-\mu_W}{\sigma/\sqrt{n}} \mid V\right)\right\}$$
$$= 1 - E\Phi\left(cV + \frac{\theta_0-\mu_W}{\sigma/\sqrt{n}}\right)$$

So

$$\beta(\mu_W,\sigma_W^2) = 1 + E\Phi\left(-cV + \frac{\theta_0-\mu_W}{\sigma/\sqrt{n}}\right) - E\Phi\left(cV + \frac{\theta_0-\mu_W}{\sigma/\sqrt{n}}\right)$$

Figure 7 illustrates this relationship. As seen in the figure, the power increases as ρ increases, which is due to the fact that the variance of W decreases as ρ grows.

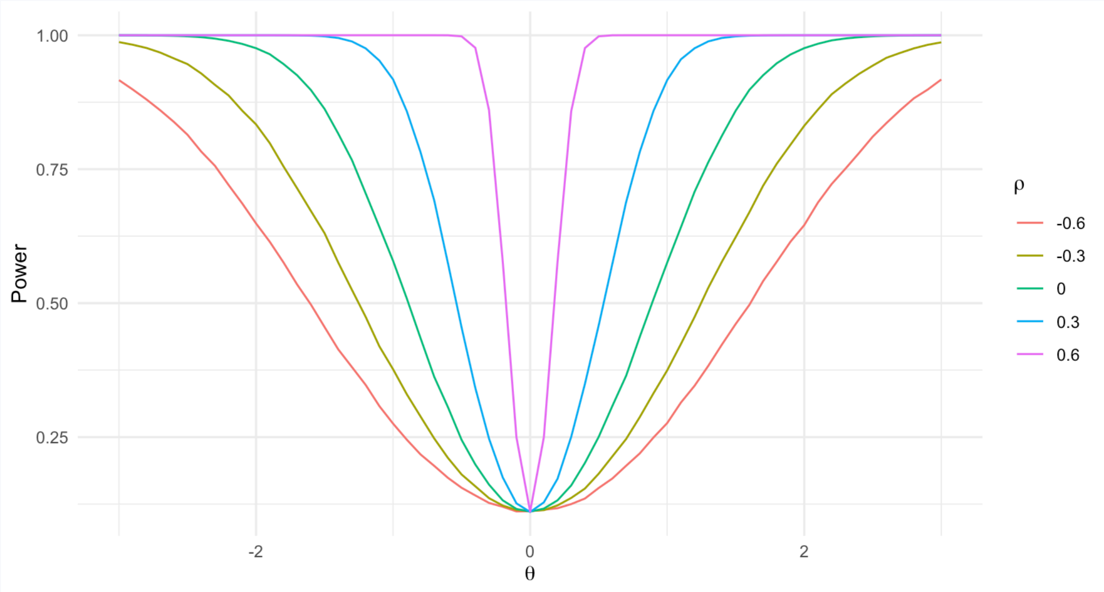

Figure 7: Power of t-test with different ρ

Question 9. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate normal distribution with parameters $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$

a) Derive the LRT of

$$H_0 : \mu_X = \mu_Y \quad \text{versus} \quad H_1 : \mu_X \neq \mu_Y$$

b) Show that the test derived in part (a) is equivalent to the paired t test of Exercise 8.39.

(Hint: Straightforward maximization of the bivariate likelihood is possible but somewhat nasty. Filling in the gaps of the following argument gives a more elegant proof.) Make the transformation $u = x - y, v = x + y$. Let $f(x, y)$ denote the bivariate normal pdf, and write

$$f(x, y) = g(v \mid u)h(u)$$

where $g(v \mid u)$ is the conditional pdf of V given U, and $h(u)$ is the marginal pdf of U. Argue that (1) the likelihood can be equivalently factored and (2) the piece involving $g(v \mid u)$ has the same maximum whether or not the means are restricted. Thus, it can be ignored (since it will cancel) and the LRT is based only on $h(u)$. However, $h(u)$ is a normal pdf with mean $\mu_X - \mu_Y$, and the LRT is the usual one-sample t test, as derived in Exercise 8.38.

Before proceeding with the proof, we introduce a helpful lemma:

Lemma 1 (Conditional distribution of normal)**.** *Let* $X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}\begin{matrix} r \\ p-r \end{matrix} \sim N_p(\mu, \Sigma)(\Sigma > 0)$, *then the distribu-*

tion of $X^{(1)}$ conditioning on $X^{(2)}$ is

$$\left(X^{(1)} \mid X^{(2)}\right) \sim N_r\left(\mu_{1\cdot 2}, \Sigma_{11\cdot 2}\right)$$

where

$$\begin{aligned}\mu_{1\cdot 2} =& \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\left(x^{(2)} - \mu^{(2)}\right);\\ \Sigma_{11\cdot 2} =& \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.\end{aligned}$$

$X^{(1)}$ and $X^{(2)}$ are independent. The proof is provided in Appendix A.

Solution:

Denote the PDF of (X, Y) by $f(x, y|\theta)$, where $\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)^T$. Then the likelihood of θ is:

$$L_1(\theta|x_{1:n}, y_{1:n}) = \prod_{i=1}^{n} f(x_i, y_i|\theta)$$

We now apply the transformation h as follows:

$$\begin{aligned}h :& (x, y) \mapsto (u, v)\\ u =& x + y\\ v =& x - y\end{aligned}$$

Denote the distribution of (U, V) by $g(u, v)$. By the transformation theorem, we have $f(x, y|\theta) = g(h(x, y)) \left|\frac{\partial(u,v)}{\partial(x,y)}\right|$, and (U, V) is also bivariate normal:

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_X + \mu_Y \\ \mu_X - \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\right].$$

Since h is a one-to-one mapping, the likelihood can be written as:

$$L_2(\theta|v_{1:n}, u_{1:n}) = \prod_{i=1}^{n} g(v_i, u_i|\theta) = \prod_{i=1}^{n} g(h(x_i, y_i)|\theta) = |\frac{\partial(u, v)}{\partial(x, y)}|^{-n} \prod_{i=1}^{n} f(x_i, y_i|\theta) \propto L_1(\theta|x_{1:n}, y_{1:n}),$$

which implies that

$$\arg\max_{\theta} L_2(\theta|v_{1:n}, u_{1:n}) = \arg\max_{\theta} L_1(\theta|x_{1:n}, y_{1:n}).$$

Now consider the likelihood ratio

$$\begin{aligned}\lambda =& \frac{\sup_{\theta\in\Theta} L_1(\theta|x_{1:n}, y_{1:n})}{\sup_{\theta\in\Theta_0} L_1(\theta|x_{1:n}, y_{1:n})}\\ =& \frac{\sup_{\theta\in\Theta} L_2(\theta|u_{1:n}, v_{1:n})}{\sup_{\theta\in\Theta_0} L_2(\theta|u_{1:n}, v_{1:n})}\\ =& \frac{\sup_{\theta\in\Theta} \prod_{i=1}^{n} g(v_i, u_i|\theta)}{\sup_{\theta\in\Theta_0} \prod_{i=1}^{n} g(v_i, u_i|\theta)}\\ =& \frac{\sup_{\theta\in\Theta} \prod_{i=1}^{n} g(u_i|v_i) \prod_{i=1}^{n} g_1(v_i)}{\sup_{\theta\in\Theta_0} \prod_{i=1}^{n} g(u_i|v_i) \prod_{i=1}^{n} g_1(v_i)}\end{aligned}$$

We will now show that the maximum of $\prod_{i=1}^{n} g(u_i|v_i, \theta)$ is the same whether or not the means are restricted. Here, $g(u, v|\theta)$ is the conditional distribution of U given V, and $g_1(v)$ is the marginal distribution of V.

To maximize $\prod_{i=1}^{n} g(u_i|v_i, \theta)g_1(v_i)$, we take the logarithm:

$$\log \prod_{i=1}^{n} g(u_i|v_i, \theta)g_1(v_i) = \sum_{i=1}^{n} \log g(u_i|v_i, \theta) + \sum_{i=1}^{n} \log g_1(v_i) \tag{9}$$

From Lemma 1, we know:

$$(U \mid V) \sim N\left(\mu_{1\cdot 2}, \Sigma_{11\cdot 2}\right),$$

where

$$\mu_{1\cdot 2} = \mu_X + \mu_Y + \Sigma_{12}\Sigma_{22}^{-1}(v - \mu_X + \mu_Y) := kv + b;$$
$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} := \sigma^2.$$

Since $U|V$ and V are independent, we can consider each part of the log-likelihood separately. The unrestricted maximum likelihood estimate (MLE) is found by:

$$\arg\max_{k,b,\sigma^2} \sum_{i=1}^{n} \log g(u_i|v_i,\theta) = \arg\max_{k,b,\sigma^2} -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(u_i - kv_i - b)^2 \tag{10}$$

Under H_0,

$$\mu_{1\cdot 2} = 2\mu_X + \Sigma_{12}\Sigma_{22}^{-1}v := kv + b';$$
$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} := \sigma^2.$$

the restricted MLE becomes:

$$\arg\max_{k,b',\sigma^2} \sum_{i=1}^{n} \log g(u_i|v_i,\theta) = \arg\max_{k,b',\sigma^2} -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(u_i - kv_i - b'\right)^2. \tag{11}$$

Since equation (11) and equation (10) have the same form, they yield the same maximum, implying:

$$\lambda = \frac{\sup_{\theta\in\Theta}\prod_{i=1}^{n} g_1(v_i)}{\sup_{\theta\in\Theta_0}\prod_{i=1}^{n} g_1(v_i)}$$

which is based only on $x - y$. This completes the desired result.

Remark 3. The equation (11) and equation (10) are the solutions for linear regression using likelihood methods.

Question 10. The assumption of equal variances, which was made in Exercise 8.41, is not always tenable. In such a case, the distribution of the statistic is no longer a t. Indeed, there is doubt as to the wisdom of calculating a pooled variance estimate. (This problem of making inference on means when variances are unequal, is, in general, quite a difficult one. It is known as the Behrens-Fisher Problem.) A natural test to try is the following modification of the two-sample t test: Test

$$H_0 : \mu_X = \mu_Y \quad \text{versus} \quad H_1 : \mu_X \neq \mu_Y$$

where we do not assume that $\sigma_X^2 = \sigma_Y^2$, using the statistic

$$T' = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)}}$$

where

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \quad \text{and} \quad S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2$$

The exact distribution of T' is not pleasant, but we can approximate the distribution using Satterthwaite's approximation (Example 7.2.3).

a) Show that

$$\frac{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \sim \frac{\chi_\nu^2}{\nu} \quad \text{(approximately)}$$

where ν can be estimated with.

$$\hat{\nu} = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}$$

b) Argue that the distribution of T' can be approximated by a t distribution with $\hat{v}$ degrees of freedom.

c) Re-examine the data from Exercise 8.41 using the approximate t test of this exercise; that is, test if the mean age of the core is the same as the mean age of the periphery using the T' statistic.

d) Is there any statistical evidence that the variance of the data from the core may be different from the variance of the data from the periphery? (Recall Example 5.4.1.)

SOLUTION:

$$\begin{aligned}
\frac{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} =& \frac{\sigma_X^2}{n(n-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)}\frac{(n-1)S_X^2}{\sigma_X^2} \\
&+\frac{\sigma_Y^2}{m(m-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)}\frac{(m-1)S_Y^2}{\sigma_Y^2} \\
:=& a_1Y_1+a_2Y_2
\end{aligned}$$

where $Y_1=\frac{(n-1)S_X^2}{\sigma_X^2}\sim\chi^2_{n-1}$ and $Y_2=\frac{(m-1)S_Y^2}{\sigma_Y^2}\sim\chi^2_{m-1}$ and they are independent. According to Satterthwaites approximation, if $Y_i\sim\chi^2_{ri}$ and they are independent, then $\sum_i a_iY_i\sim\frac{\chi^2_{\hat{\nu}}}{\hat{\nu}}$, where $\hat{\nu}=\frac{\left(\sum_i a_iY_i\right)^2}{\sum_i a_i^2Y_i^2/r_i}$. Thus,

$$\frac{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\sim\frac{\chi^2_{\hat{\nu}}}{\hat{\nu}},\ \text{where }\hat{\nu}=\frac{\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2(n-1)}+\frac{S_Y^4}{m^2(m-1)}}.$$

The statistic T' can be written as

$$T'=T'=\frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)}}=\frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)}}=\frac{(\bar{X}-\bar{Y})/\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}{\sqrt{\frac{(S_X^2/n+S_Y^2/m)}{(\sigma_X^2/n+\sigma_Y^2/m)}}},$$

where $(\bar{X}-\bar{Y})/\sqrt{\sigma_X^2/n+\sigma_Y^2/m}\sim N(0,1)$ and $\sqrt{\frac{(S_X^2/n+S_Y^2/m)}{(\sigma_X^2/n+\sigma_Y^2/m)}}\sim\frac{\chi^2_{\hat{\nu}}}{\hat{\nu}}$ approximately. Thus, T' can be approximated by a t distribution with $\hat{v}$ degrees of freedom.

Upon examining the data, the p-value is greater than 0.05, so under the significance level of 0.05, we cannot reject H_0.

```
> Core <- c(1294, 1279, 1274, 1264, 1263, 1254, 1251, 1251, 1248, 1240, 1232, 1220, 1218, 1210)
> Periphery <- c(1284, 1272, 1256, 1254, 1242, 1274, 1264, 1256, 1250)
> t.test(Core,Periphery,var.equal = FALSE)

    Welch Two Sample t-test

 data:  Core and Periphery
 t = -1.4599, df = 20.636, p-value = 0.1594
 alternative hypothesis: true difference in means is not equal to 0
 95 percent confidence interval:
  -27.841668   4.889287
 sample estimates:
 mean of x mean of y
  1249.857  1261.333
```

To compare the variance of two populations, we derive the LRT for $H_0:\sigma_X=\sigma_Y$ against $H_1:\sigma_X\neq\sigma_Y$. First, the likelihood of $\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2$ is

$$L(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2\mid x_{1:n},y_{1:m})=\left(2\pi\sigma_X^2\right)^{-\frac{n}{2}}\left(2\pi\sigma_Y^2\right)^{-\frac{m}{2}}\exp\left\{-\frac{1}{2\sigma_X^2}\sum_{i=1}^n(x_i-\mu_X)^2-\frac{1}{2\sigma_Y^2}\sum_{i=1}^m(y_i-\mu_Y)^2\right\}$$

In the unrestricted scenario, maximizing the likelihood is equivalent to finding the maxima of $L(\mu_X,\sigma_X^2)$ and $L(\mu_Y,\sigma_Y^2)$ individually. So the unrestricted MLEs of $\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2$ are

$$\hat{\mu}_X=\bar{x},\quad\hat{\mu}_Y=\bar{y},\quad\hat{\sigma}_X^2=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2,\quad\hat{\sigma}_Y^2=\frac{1}{m}\sum_{i=1}^m(y_i-\bar{y})^2,$$

respectively. Under H_0, the restricted MLEs of μ_X and μ_Y are

$$\hat{\mu}_X = \bar{x}, \quad \hat{\mu}_Y = \bar{y},$$

since maximizing μ_X does not involve μ_Y and vice visa. Finally the restricted MLE of σ^2 can be derived by derivating $\log L(\bar{x}, \bar{y}, \sigma^2)$:

$$\frac{\partial \log L(\bar{x}, \bar{y}, \sigma^2)}{\partial \sigma^2} = -\frac{m+n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2\right) \overset{\text{set}}{=} 0,$$

which implies that

$$\hat{\sigma}^2 = \frac{1}{m+n}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2\right).$$

Further $\frac{\partial \log L(\bar{x},\bar{y},\sigma^2)}{\partial(\sigma^2)^2} < 0$, so $\hat{\sigma}^2$ is indeed maxima. So, the LRT can be written as:

$$\begin{aligned}
\lambda =& \frac{L(\hat{\mu}_X, \hat{\mu}_Y, \hat{\sigma}_X^2, \hat{\sigma}_Y^2)}{L(\hat{\mu}_X, \hat{\mu}_Y, \hat{\sigma}^2)} = \frac{(\hat{\sigma}^2)^{\frac{n}{2}}(\hat{\sigma}^2)^{\frac{m}{2}}}{(\hat{\sigma}_X^2)^{\frac{n}{2}}(\hat{\sigma}_Y^2)^{\frac{m}{2}}} \\
=& \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}}{(n+m)^{\frac{n+m}{2}}}\left(1 + \frac{\sum_{i=1}^{m}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{\frac{n}{2}}\left(1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2}\right)^{\frac{m}{2}} \\
=& \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}}{(n+m)^{\frac{n+m}{2}}}\left(1 + \frac{m-1}{n-1}\frac{\frac{\sum_{i=1}^{m}(y_i-\bar{y})^2}{m-1}}{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\right)^{\frac{n}{2}}\left(1 + \frac{n-1}{m-1}\frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}{\frac{\sum_{i=1}^{m}(y_i-\bar{y})^2}{m-1}}\right)^{\frac{m}{2}} \\
=& \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}}{(n+m)^{\frac{n+m}{2}}}\left(1 + \frac{m-1}{n-1}\frac{\hat{\sigma}_Y^2}{\hat{\sigma}_X^2}\right)^{\frac{n}{2}}\left(1 + \frac{n-1}{m-1}\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2}\right)^{\frac{m}{2}} \\
=& \frac{n^{\frac{n}{2}}m^{\frac{m}{2}}}{(n+m)^{\frac{n+m}{2}}}\left(1 + \frac{m-1}{n-1}\frac{1}{F}\right)^{\frac{n}{2}}\left(1 + \frac{n-1}{m-1}F\right)^{\frac{m}{2}} \qquad \text{let } \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} = F > 0.
\end{aligned}$$

Now we show the monotonicity of $\log \lambda$ with F:

$$\frac{\partial \log \lambda}{\partial F} = \frac{m(n-1)(F-1)}{2(n-1)F^2 + 2(m-1)F}$$

so λ is increasing in $(1, +\infty)$ and decreasing in $(0, 1)$. To get a size α test, we try to find k_1 and k_2 satisfying

$$P_{H_0}(F < k_1 \text{ or } F > k_2) = \alpha \quad \text{and} \quad \lambda(k_1) = \lambda(k_2) \tag{12}$$

Under H_0 $F \sim F_{(n-1,m-1)}$, but equation (12) is difficult to find a close form, we can calculate it numerically or use $k_1 = F_{(n-1,m-1)}(\alpha/2)$ and $k_1' = F_{(m-1,n-1)}(\alpha/2)$, where $F_{(n-1,m-1)}(\cdot)$ is the lower quantile function. According to the results below, we can't reject H_0.

```
> n=length(Core)
> m=length(Periphery)
> alpha=.05
> (k1 <- qf(alpha/2,n-1,m-1))
[1] 0.2951605
> (k2 <- qf(1-alpha/2,n-1,m-1))
[1] 4.16217
> (F=var(Core)/var(Periphery))
[1] 3.360015 ##k1<F<k2 so do not reject H0
> 2*min(1-pf(F,n-1,m-1),pf(F,n-1,m-1))#p-value
[1] 0.09200881 ##p-value tells us the same that do not reject H0
```

Question 11. Let $X_1, \ldots, X_n$ be iid $N(\theta, \sigma^2)$, σ^2 known, and let θ have a double exponential distribution, that is, $\pi(\theta) = e^{-|\theta|/a}/(2a), a$ known. A Bayesian test of the hypotheses $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$ will decide in favor of H_1 if its posterior probability is large.

a) For a given constant K, calculate the posterior probability that $\theta > K$, that is, $P(\theta > K \mid x_1, \ldots, x_n, a)$.

b) Find an expression for $\lim_{a\to\infty} P(\theta > K \mid x_1, \ldots, x_n, a)$.

c) Compare your answer in part (b) to the p-value associated with the classical hypothesis test.

To begin with, we introduce a lemma and the definition of the truncated normal distribution.

Lemma 2 (Normal kernel). *For any $A > 0$ and $B \in \mathbf{R}$, we have*

$$f(\theta) \propto \exp\left\{-\frac{1}{2}(A\theta^2 - 2B\theta)\right\} \implies \theta \sim N\left(\frac{B}{A}, \frac{1}{A}\right).$$

This is straightforward to prove, so we omit the details here.

Definition 1. *Suppose X has a normal distribution with mean μ and variance σ^2 and lies within the interval (a, b), with $-\infty \le a < b \le \infty$. Then X conditional on $a < X < b$ has a truncated normal distribution, denoted by $\mathrm{TN}(\mu, \sigma^2, a, b)$*

Lemma 3.

$$\int_a^b \exp\left\{-\frac{1}{2}(A\theta^2 - 2B\theta)\right\} d\theta = \exp\left\{\frac{B^2}{2A}\right\} \sqrt{2\pi(1/A)}\Phi\left(Ab - B\right) - \Phi\left(Aa - B\right)$$

See Appendix B for the proof.

SOLUTION:

Now we calculate the posterior of θ:

$$\begin{aligned}
\pi(\theta|x_{1:n}) &\propto \pi(\theta) \times f(x_{1:n}|\theta) \\
&= \frac{1}{2a}\exp\left\{-\frac{|\theta|}{a}\right\} \times \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2\right\} \\
&= \frac{1}{2a}\exp\left\{-\frac{|\theta|}{a}\right\} \times \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2 - \frac{1}{2\sigma^2}n(\bar{x} - \theta)^2\right\} \\
&\propto \exp\left\{-\frac{|\theta|}{a} - \frac{1}{2\sigma^2}n(\bar{x} - \theta)^2\right\} \\
&\propto \exp\left\{-\frac{|\theta|}{a} + \frac{2n\bar{x}\theta}{2\sigma^2} - \frac{\theta^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} - \frac{1}{a}\right)\theta\right)\right\}\mathbf{1}_{\{\theta>0\}} + \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} + \frac{1}{a}\right)\theta\right)\right\}\mathbf{1}_{\{\theta\le 0\}}
\end{aligned}$$

Let

$$\begin{aligned}
B &= \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} - \frac{1}{a}\right)\theta\right)\right\}\mathbf{1}_{\{\theta>0\}} + \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} + \frac{1}{a}\right)\theta\right)\right\}\mathbf{1}_{\{\theta\le 0\}} d\theta \\
&= \int_0^{+\infty} \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} - \frac{1}{a}\right)\theta\right)\right\} d\theta + \int_{-\infty}^0 \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} + \frac{1}{a}\right)\theta\right)\right\} d\theta \\
&= \exp\left\{\frac{n(n\bar{x}/\sigma^2 - 1/a)^2}{2\sigma^2}\right\}\sqrt{2\pi\sigma^2/n}\left(1 - \Phi(1/a - n\bar{x}/\sigma^2)\right) \\
&\quad + \exp\left\{\frac{n(n\bar{x}/\sigma^2 + 1/a)^2}{2\sigma^2}\right\}\sqrt{2\pi\sigma^2/n}\Phi(-1/a - n\bar{x}/\sigma^2) \\
&:= B_1 + B_2
\end{aligned}$$

Thus, the posterior distribution is

$$\pi(\theta|x_{1:n}) = \frac{1}{B}\exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} - \frac{1}{a}\right)\theta\right)\right\}\mathbf{1}_{\{\theta>0\}} + \frac{1}{B}\exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} + \frac{1}{a}\right)\theta\right)\right\}\mathbf{1}_{\{\theta\le 0\}}$$

In fact $\pi(\theta|x)$ is a mixed distribution, since

$$\pi(\theta|x) = \pi(\theta|x, \theta < 0)P(\theta < 0) + \pi(\theta|x, \theta > 0)P(\theta > 0)$$

Now $\pi(\theta|x,\theta<0)$ and $\pi(\theta|x,\theta>0)$ are truncated normal distribuion $\mathrm{TN}(\bar{x}+\frac{\sigma^2}{na},\frac{\sigma^2}{n},-\infty,0)$ and $\mathrm{TN}(\bar{x}-\frac{\sigma^2}{na},\frac{\sigma^2}{n},0,+\infty)$, respectively, (Lemma 2 confirms that they are normal and truncated). And $P(\theta>0)=B_1/(B_1+B_2)$, $P(\theta<0)=B_2/(B_1+B_2)$. To calculate $P(\theta>K\mid x_1,\ldots,x_n,a)$, we integrate $\pi(\theta|x_{1:n})$ from K to $+\infty$, and for $K>0$

$$\begin{aligned}P(\theta>K\mid x_1,\ldots,x_n,a)=&\frac{1}{B_1+B_2}\int_K^{+\infty}\exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}\theta^2-2(\frac{n\bar{x}}{\sigma^2}-\frac{1}{a})\theta\right)\right\}d\theta\\=&\frac{1}{B_1+B_2}\exp\left\{\frac{n(n\bar{x}/\sigma^2-1/a)^2}{2\sigma^2}\right\}\sqrt{2\pi\sigma^2/n}\left(1-\Phi(1/a+n(K-\bar{x})/\sigma^2)\right)\end{aligned}\tag{13}$$

As $a\to\infty$,

$$\begin{aligned}\lim_{a\to\infty}\pi(\theta|x)=&\lim_{a\to\infty}\mathrm{TN}(\bar{x}+\frac{\sigma^2}{na},\frac{\sigma^2}{n},-\infty,0)P(\theta<0)+\lim_{a\to\infty}\mathrm{TN}(\bar{x}-\frac{\sigma^2}{na},\frac{\sigma^2}{n},0,+\infty)P(\theta>0)\\=&\mathrm{TN}(\bar{x},\frac{\sigma^2}{n},-\infty,0)P(\theta<0)+\mathrm{TN}(\bar{x},\frac{\sigma^2}{n},0,+\infty)P(\theta>0)\\=&N(\bar{x},\frac{\sigma^2}{n})\end{aligned}$$

so

$$\begin{aligned}\lim_{a\to\infty}P(\theta>K\mid x_1,\ldots,x_n,a)&=\lim_{a\to\infty}\int_K^{\infty}\pi(\theta|x_{1:n})dx\\&=\int_K^{\infty}\lim_{a\to\infty}\pi(\theta|x_{1:n})dx\\&=\int_K^{\infty}N(\bar{x},\frac{\sigma^2}{n})d\theta=1-\Phi(\frac{\sqrt{n}(K-\bar{x})}{\sigma}),\end{aligned}$$

which is $1-p$-value when $K=0$ (Let $a\to\infty$ in equation (13), the same results can be obtained). With a increase, the distribution of the prior $\pi(\theta)$ is more and more flat so that the prior information is intending to 0. When $a\to\infty$, this is the so-called non-informative prior.

Remark 4. If we use Bayes inference in the weighted loss function :

$$L(\theta,\hat{\phi})=a_0\mathbf{1}(\phi<\hat{\phi})+a_1\mathbf{1}(\phi>\hat{\phi})=\begin{cases}a_0 & \text{if } \phi=0 \text{ and } \hat{\phi}=1;\\ a_1 & \text{if } \phi=1 \text{ and } \hat{\phi}=0;\\ 0 & \text{otherwise,}\end{cases}$$

where $a_0,a_1\geq 0$. The Bayes estimator is

$$\hat{\phi}_\pi=\mathbf{1}\left\{\hat{p}_0<\frac{a_1}{a_0+a_1}\right\}=\mathbf{1}\left\{\hat{p}_1\geq\frac{a_0}{a_0+a_1}\right\},\qquad \text{where } \hat{p}_i=P(\theta\in\Theta_i|x).$$

We can interpret it in terms of frequentist's languages:

- $\hat{p}_0$ plays the role of the p-value, and $\alpha=\frac{a_1}{a_0+a_1}$ plays the role of the nominal size.
- $\alpha=\frac{a_1}{a_0+a_1}=0.05$ means that the loss of committing type-I error is 19 times that of committing type-II error.

If we use $\alpha=\frac{a_1}{a_0+a_1}=0.05$, the classical test is the same as our Bayes one once we use non-informative prior.

Question 12. Here is another common interpretation of p-values. Consider a problem of testing H_0 versus H_1. Let $W(\mathbf{X})$ be a test statistic. Suppose that for each $\alpha, 0\leq\alpha\leq 1$, a critical value c_α can be chosen so that $\{\mathrm{x}: W(\mathbf{x})\geq c_\alpha\}$ is the rejection region of a size α test of H_0. Using this family of tests, show that the usual p-value $p(\mathbf{x})$, defined by (8.3.9), is the smallest α level at which we could reject H_0, having observed the data $\mathbf{x}$.

Solution:

By definition

$$\alpha = \sup_{\theta\in\Theta_0} P(W(X) \geq c_\alpha)$$
$$p(x) = \sup_{\theta\in\Theta_0} P(W(X) \geq W(x))$$

Here $W(X)$ is a statistic such that a large value of W gives evidence that H_1 is true. If $\alpha < p(x)$, then $W(x) < c_\alpha$, which show that we can not reject H_0. On the other hand, $\alpha \geq p(x)$ implies that $W(x) \geq c_\alpha$, so we can reject H_0.

Question 13. Consider testing $H_0 : \theta \in \bigcup_{j=1}^k \Theta_j$. For each $j = 1, \ldots, k$, let $p_j(\mathbf{x})$ denote a valid p-value for testing $H_{0j} : \theta \in \Theta_j$. Let $p(\mathbf{x}) = \max_{1\leq j\leq k} p_j(\mathbf{x})$.

a) Show that $p(\mathbf{X})$ is a valid p -value for testing H_0.
b) Show that the α level test defined by $p(\mathbf{X})$ is the same as an α level IUT defined in terms of individual tests based on the p_j(x)s.

Solution:

$$P(p(\mathbf{X}) \leq \alpha) = P(\bigcap_{j=1}^k p_j(\mathbf{X}) \leq \alpha) \leq \max_{1\leq j\leq k} P(p_j(\mathbf{X}) \leq \alpha) \leq \alpha$$

The last inequality holds since $p_j(\mathbf{x})$'s are valid p-values.

The rejection region of a level α test defined by $p(\mathbf{X})$ is

$$\{\mathbf{X} : p(\mathbf{X}) \leq \alpha\} = \left\{\mathbf{X} : \bigcap_{j=1}^k (p_j(\mathbf{X}) \leq \alpha)\right\}$$

which shows that it is the same as an α level IUT defined in terms of individual tests based on the p_j(x)s

Question 14. Consider the hypothesis testing problem and loss function given in Example 8.3.31, and let $\sigma = n = 1$. Consider tests that reject H_0 if $X < -z_\alpha + \theta_0$. Find the value of α that minimizes the maximum value of the risk function, that is, that yields a minimax test.

Solution:

The risk function is:

$$\begin{aligned} R(\theta,\delta) =& L(\theta,a_0)(1-\beta(\theta) + L(\theta,a_1)\beta(\theta) \\ =& 8\Phi(-z_\alpha+\theta_0-\theta)\mathbf{1}_{\{\theta\leq\theta_0\}} + 3(1-\Phi(-z_\alpha+\theta_0-\theta))\mathbf{1}_{\{\theta>\theta_0\}}. \end{aligned}$$

Here z_α is the upper α quantile of standard normal distribution. We find that $R(\theta,\delta)$ increase in $(-\infty,\theta]$ and decrease in $(\theta,+\infty)$. Hence, the maximum value of the risk function occurs at θ_0. With α grove, $\lim_{\theta\to\theta_0^+} R(\theta,\delta)$ incline and $\lim_{\theta\to\theta_0^-} R(\theta,\delta)$ increase. So the minimum risk is α satisfies

$$8\Phi(-z_\alpha) = \lim_{\theta\to\theta_0^-} R(\theta,\delta) = \lim_{\theta\to\theta_0^+} R(\theta,\delta) = 3 - 3\Phi(-z_\alpha) \implies \Phi(-z_\alpha) = \frac{3}{11}$$

so $\alpha = \frac{3}{11}$.

Appendix

A Proof of Lemma 1

Let

$$Z=\left[\begin{array}{c} X^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}X^{(2)} \\ X^{(2)} \end{array}\right]=\left[\begin{array}{cc} I_r & -\Sigma_{12}\Sigma_{22}^{-1} \\ O & I_{p-r} \end{array}\right]\left[\begin{array}{c} X^{(1)} \\ X^{(2)} \end{array}\right]=BX.$$

By the property of multi-normal distribution, we have

$$Z=\left[\begin{array}{c} Z^{(1)} \\ Z^{(2)} \end{array}\right]\sim N_p\left(\left[\begin{array}{c} \mu^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)} \\ \mu^{(2)} \end{array}\right],\left[\begin{array}{cc} \Sigma_{11\cdot 2} & O \\ O & \Sigma_{22} \end{array}\right]\right)$$

so $Z^{(1)}$,$Z^{(2)}$ are independent and $Z^{(2)}=X^{(2)}$. The joint distribution of Z is

$$g\left(z^{(1)},z^{(2)}\right)=g_1\left(z^{(1)}\right)g_2\left(z^{(2)}\right)=g_1\left(z^{(1)}\right)f_2\left(z^{(2)}\right),\quad (\text{since } Z^{(2)}=X^{(2)})$$

We can derive the PDF of X by Y:

$$\begin{aligned} f\left(x^{(1)},x^{(2)}\right)&=g(Bx)\left|\frac{\partial z}{\partial x}\right|_+ \\ &=g_1\left(x^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}x^{(2)}\right)g_2\left(x^{(2)}\right)\left|\frac{\partial z}{\partial x}\right|_+ \\ &=g_1\left(x^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}x^{(2)}\right)f_2\left(x^{(2)}\right)\ \text{ since } |B|=1 \end{aligned}$$

so

$$\begin{aligned} f_1\left(x^{(1)}\mid x^{(2)}\right)=&\frac{f\left(x^{(1)},x^{(2)}\right)}{f_2\left(x^{(2)}\right)} \\ =&g_1\left(x^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}x^{(2)}\right) \\ =&\frac{1}{(2\pi)^{r/2}\left|\Sigma_{11\cdot 2}\right|^{1/2}}\exp\left[-\frac{1}{2}\left(x^{(1)}-\mu_{1\cdot 2}\right)'\Sigma_{11\cdot 2}^{-1}\left(x^{(1)}-\mu_{1\cdot 2}\right)\right] \end{aligned}$$

where

$$\begin{aligned} \mu_{1\cdot 2}=&\mu^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}\left(x^{(2)}-\mu^{(2)}\right); \\ \Sigma_{11\cdot 2}=&\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

B Proof of Lemma 3

$$\begin{aligned} \int_a^b \exp\left\{-\frac{1}{2}(A\theta^2-2B\theta)\right\}&=\int_a^b \exp\left\{-\frac{A}{2}(\theta^2-2\frac{B}{A}\theta)\right\}d\theta \\ &=\int_a^b \exp\left\{-\frac{A}{2}(\theta-\frac{B}{A})^2+\frac{B^2}{2A}\right\}d\theta \\ &=\exp\left\{\frac{B^2}{2A}\right\}\sqrt{2\pi(1/A)}\int_a^b\frac{1}{\sqrt{2\pi(1/A)}}\exp\left\{-\frac{1}{2(1/A)}(\theta^2-\frac{B}{A})^2\right\}d\theta \\ &=\exp\left\{\frac{B^2}{2A}\right\}\sqrt{2\pi(1/A)}P(a<\theta<b)\quad (\text{here } \theta\sim N\left(\frac{B}{A},\frac{1}{A}\right)) \\ &=\exp\left\{\frac{B^2}{2A}\right\}\sqrt{2\pi(1/A)}P\left(\frac{a-\frac{B}{A}}{\frac{1}{A}}<Z<\frac{b-\frac{B}{A}}{\frac{1}{A}}\right)\quad (\text{here } Z\sim N(0,1)) \\ &=\exp\left\{\frac{B^2}{2A}\right\}\sqrt{2\pi(1/A)}\Phi\left(Ab-B\right)-\Phi\left(Aa-B\right) \end{aligned}$$