EI035 Econometrics I
The Graduate Institute, Fall 2024, Marko Milkota

PS2 Solutions

Jingle Fu

Solution (a).

Step 1: Write the pdf of observations $x_i|\theta$

Since $x_i = \theta + u_i$, and we assume $u_i \sim \mathcal{N}(0, \sigma^2)$, then $x_i \sim \mathcal{N}(\theta, \sigma^2)$, and we have:

$$p(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}\right\}.$$

Step 2: Define Likelihood Function

We have assumed that observations in the sample are independent. Thus,

$$L_n(\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

Log-linearize the function, and we define the log-likelihood function:

$$\ell_n(\theta) = n\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2$$

Step 3: Define the Likelihood Estimation problem and find the $\hat{\theta}$

For maximum likelihood estimation, we need to solve the following problem:

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\max_{\theta\in\Theta} \ell_n(\theta)$$

So, we need to maximize:

$$\ell_n(\theta) = n\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2$$

Take the derivative of $\ell_n(\theta)$ with respect to θ, and set it to zero for maximization,

$$\begin{aligned}\frac{\partial \ell_n(\theta)}{\partial \theta} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta) \\ &= \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\theta) \\ &= 0\end{aligned}$$

Thus, we have

$$\hat{\theta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Solution (b).

Step 1: Find the likelihood function

For $\mathcal{H}_0 : \theta = \theta_0$, the likelihood function is:

$$L_n(\theta_0) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_0)^2\right\}$$

For $\mathcal{H}_1 : \theta \neq \theta_0$, the maximum likelihood estimator is $\hat{\theta}_{ML} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \hat{\theta}$. The likelihood function is:

$$L_n(\theta_1) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\theta})^2\right\}$$

Step 2: Define the likelihood ratio test and c_α

The likelihood ratio and the test is firstly defined as follows(we'll simplify to another version later):

$$\varphi_{LR}(x) = \mathbf{1}\left\{LR_n < c\right\}$$

$$\begin{aligned}LR_n &= \frac{L_n(\theta_1)}{L_n(\theta_0)} \\ &= \exp\left\{\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\theta_0)^2 - \sum_{i=1}^{n}(x_i-\hat{\theta})^2\right]\right\}\end{aligned}$$

Denote

$$D = \sum_{i=1}^{n}(x_i-\theta_0)^2 - \sum_{i=1}^{n}(x_i-\hat{\theta})^2$$

Using the identity:

$$\sum_{i=1}^{n}(x_i-\theta_0)^2=\sum_{i=1}^{n}(x_i-\hat{\theta}+\hat{\theta}-\theta_0)^2=\sum_{i=1}^{n}(x_i-\hat{\theta})^2+n(\hat{\theta}-\theta_0)^2$$

Thus,

$$\begin{aligned}
D&=\sum_{i=1}^{n}(x_i-\theta_0)^2-\sum_{i=1}^{n}(x_i-\hat{\theta})^2\\
&=\sum_{i=1}^{n}(x_i-\hat{\theta})^2+n(\hat{\theta}-\theta_0)^2-\sum_{i=1}^{n}(x_i-\hat{\theta})^2\\
&=n(\hat{\theta}-\theta_0)^2
\end{aligned}$$

So, the likelihood ratio LR_n is:

$$LR_n=\exp\left\{\frac{n}{2\sigma^2}(\hat{\theta}-\theta_0)^2\right\}$$

Then, we simplify the expression and define the test statistic $T(x)$ as below:

$$T(x)=2\log\left(LR_n\right)=\frac{n}{\sigma^2}(\hat{\theta}-\theta_0)^2$$

And our LR test would be:

$$\varphi_{LR}(x)=\mathbf{1}\left\{T(x)=\frac{n}{\sigma^2}(\hat{\theta}-\theta_0)^2<c'\right\}$$

where $c'=2\log(c)$. To get a size α test, we find c' so as to set the Type I error to α, which is:

$$\mathbb{P}\left[T(x)\geq c'|\mathcal{H}_0\right]=\alpha$$

we can denote that $c'=c_\alpha$.

Step 3: Determine the distribution of $T(x)$ under $\mathcal{H}_0$ and find the value of c_α

Under $\mathcal{H}_0$, $\hat{\theta} \sim \mathcal{N}(\theta_0, \frac{\sigma^2}{n})$, because:

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[x_i] = \frac{1}{n}\cdot n\theta_0 = \theta_0$$
$$\mathbb{V}[\hat{\theta}] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[x_i] = \frac{1}{n}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Then, standardizing $\hat{\theta}$, we'll have:

$$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1)$$

Using the hint, we know that

$$Z^2 = \left(\frac{\hat{\theta} - \theta_0}{\sqrt{\sigma^2/n}}\right)^2 = \frac{n}{\sigma^2}(\hat{\theta} - \theta_0)^2 \sim \chi_1^2$$

Therefore, under $\mathcal{H}_0$,

$$T(x) = \frac{n}{\sigma^2}(\hat{\theta} - \theta_0)^2 = Z^2 \sim \chi_1^2$$

Given $\alpha = 0.05$,

$$c_\alpha = \chi^2_{1,0.95} \approx 3.8415$$

Step 4: Set the decision rule

- Reject $\mathcal{H}_0$: $T(x) > c_\alpha = 3.8415$
- Do not reject $\mathcal{H}_0$: $T(x) \leq c_\alpha = 3.8415$

Solution (c).

We have $\sigma^2 = 6$, $n = 4$, $x_1 = 178$, $x_2 = 161$,$x_3 = 168$, $x_4 = 172$, $\theta_0 = 175$, so $\hat{\theta} = 169.75$.

Put this data back into our $T(x)$ and LR test, we have:

$$T(x) = \frac{n}{\sigma^2}(\hat{\theta} - \theta_0)^2 = \frac{4}{6}(169.75 - 175)^2 = 18.735 > 3.8415$$

We reject $\mathcal{H}_0$.

Solution (d).

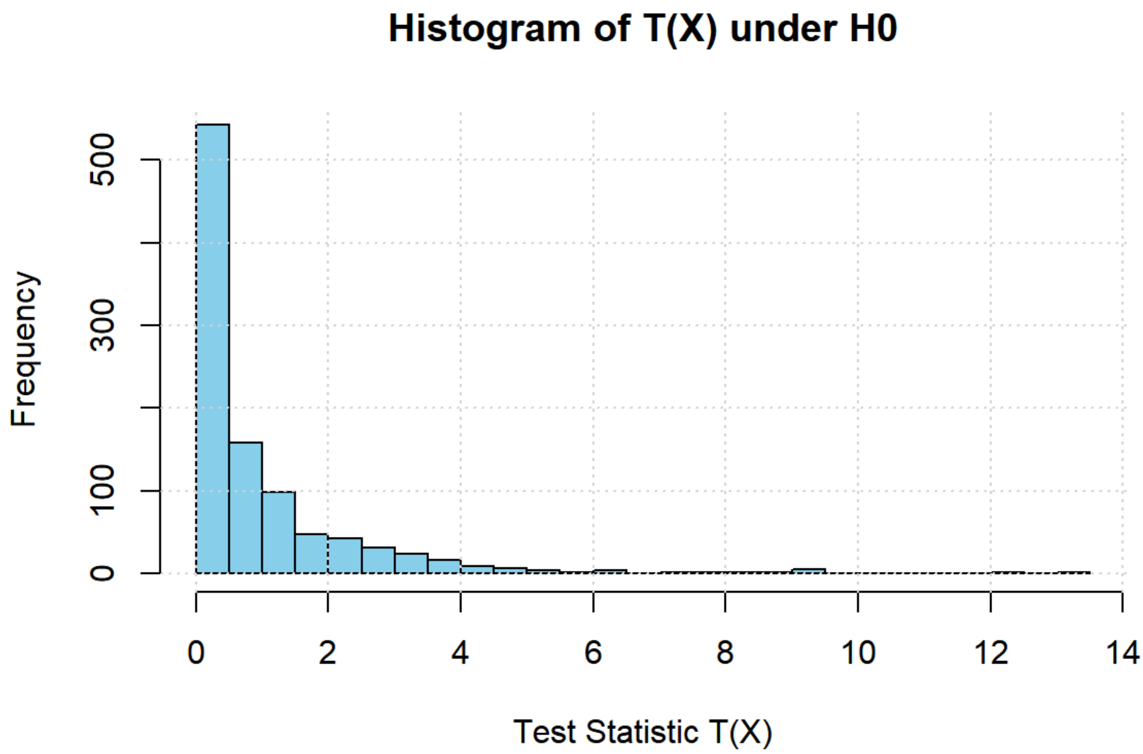
```r
rm(list = ls())
set.seed(2024)
```

```r

theta_0 <- 175
sigma_squared <- 6
sigma <- sqrt(sigma_squared)
n <- 4
alpha <- 0.05
M <- 1000
T_values <- numeric(M)

for (m in 1:M) {
  x_m <- rnorm(n, mean = theta_0, sd = sigma)
  x_bar_m <- mean(x_m)
  T_m <- (n / sigma_squared) * (x_bar_m - theta_0)^2
  T_values[m] <- T_m
}

hist(T_values, breaks = 30, col = 'skyblue', border = 'black', freq = FALSE,
     main = 'Histogram of T(X) under H0', xlab = 'Test Statistic T(X)')
grid()

T_values_sorted <- sort(T_values)
c_alpha_index <- ceiling(M * (1 - alpha))
c_alpha <- T_values_sorted[c_alpha_index]

cat(sprintf('Numerical approximation of c_alpha: %.4f\n', c_alpha))

c_alpha_analytical <- qchisq(1 - alpha, df = 1)
cat(sprintf('Analytical c_alpha from chi-squared distribution: %.4f\n', c_alpha_analytical))

difference <- abs(c_alpha - c_alpha_analytical)
cat(sprintf('Difference between numerical and analytical c_alpha: %.4f\n', difference))
```

Numerical approximation of c_α: 3.6266

Analyticalc_α from chi-squared distribution: 3.8415

Difference between numerical and analytical c_α: 0.2148, which is about 5.6% of the analytical c_α, so our approximation is not very close to the true value c_α.



I expect the estimated approximation to get closer to the real analytical value of c_α as M is larger.

Since the $T(x)$ we get is 18.735 which is greatly larger than 3.84 and 3.62, which is our numerical result, the conclusion from previous exercise doesn't change, we still reject $\mathcal{H}_0$.

Solution (e).

Based on our previous LR test, we have:

$$\begin{aligned}
\varphi_{LR}(x) &= \mathbf{1}\left\{T(x) = \frac{n}{\sigma^2}(\hat{\theta}-\theta_0)^2 < c_\alpha\right\} \\
&= \mathbf{1}\left\{(\hat{\theta}-\theta_0)^2 < \frac{c_\alpha \sigma^2}{n}\right\} \\
&= \mathbf{1}\left\{-\sqrt{\frac{c_\alpha \sigma^2}{n}} < (\hat{\theta}-\theta_0) < \sqrt{\frac{c_\alpha \sigma^2}{n}}\right\} \\
&= \mathbf{1}\left\{\hat{\theta}-\sqrt{\frac{c_\alpha \sigma^2}{n}} < \theta_0 < \hat{\theta}+\sqrt{\frac{c_\alpha \sigma^2}{n}}\right\}
\end{aligned}$$

Thus, we can define $C(X)$ as:

$$C(X) = \left[\hat{\theta} - \sqrt{\frac{c_\alpha \sigma^2}{n}}, \hat{\theta} + \sqrt{\frac{c_\alpha \sigma^2}{n}}\right]$$

Apply our previous data: $\sigma^2 = 6$, $n = 4$, $x_1 = 178$, $x_2 = 161$,$x_3 = 168$, $x_4 = 172$, $\theta_0 = 175$, $\hat{\theta} = 169.75$, and $c_\alpha = 3.8415$, we have:

$$C(X) = [169.75 - 2.4, 169.75 + 2.4] = [167.35, 172.15]$$

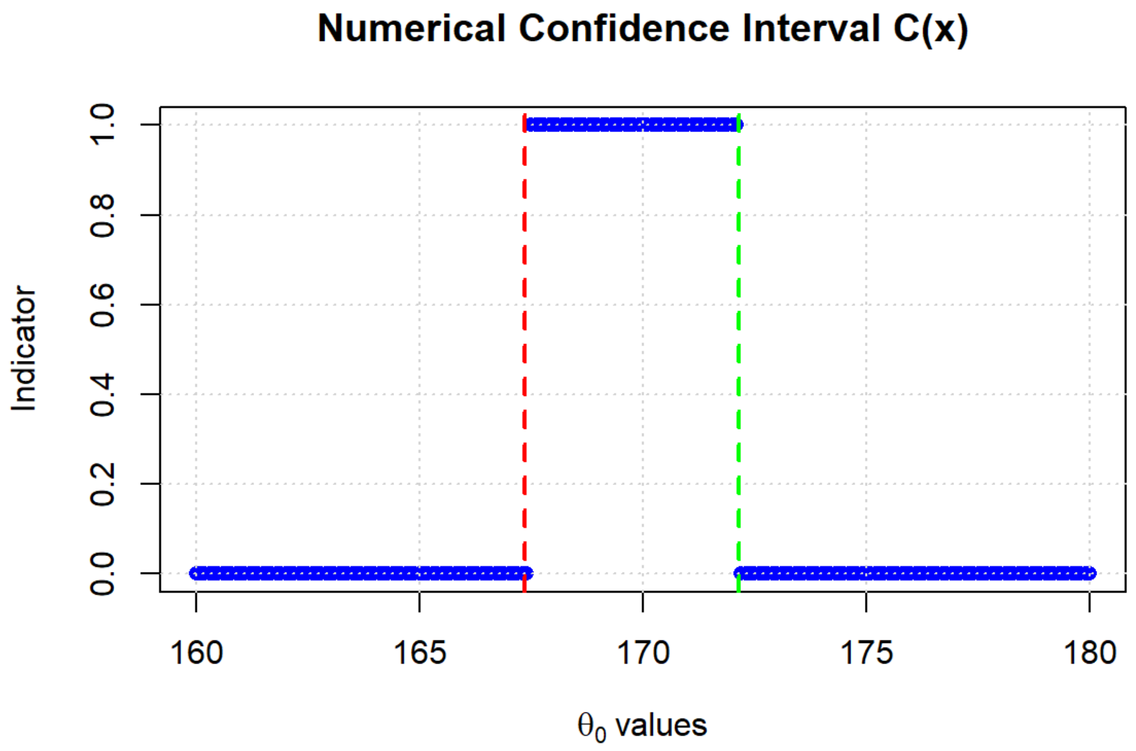
$\theta_0 = 175$ is not in this interval.

Because we rejected $\mathcal{H}_0 : \theta = 175$, it's consistent that 175 is not within the 95% confidence interval.

Solution (f).

```
rm(list = ls())
set.seed(2024)
x <- c(178, 161, 168, 172)
n <- length(x)
sigma_squared <- 6
sigma <- sqrt(sigma_squared)
alpha <- 0.05
M <- 1000
T_values <- seq(160, 180, by = 0.1)
vc <- numeric(length(T_values))

for (i in seq_along(T_values)) {
  theta_0 <- T_values[i]
  T_m <- numeric(M)
  for (m in 1:M) {
    x_m <- rnorm(n, mean = theta_0, sd = sigma)
    T_m_value <- (n / sigma_squared) * (mean(x_m) - theta_0)^2
    T_m[m] <- T_m_value
  }
  T_m_sorted <- sort(T_m)
  c_alpha_theta_index <- ceiling(M * (1 - alpha))
  c_alpha_theta <- T_m_sorted[c_alpha_theta_index]

```

```
  x_bar <- mean(x)
  T_x_theta <- (n / sigma_squared) * (x_bar - theta_0)^2

  if (T_x_theta < c_alpha_theta) {
    vc[i] <- 1
  } else {
    vc[i] <- 0
  }
}

theta_in_Cx <- T_values[vc == 1]
if (length(theta_in_Cx) > 0) {
  numerical_Cx <- c(min(theta_in_Cx), max(theta_in_Cx))
  cat(sprintf("Numerical Confidence Interval C(x): [%.2f, %.2f]\n",
              numerical_Cx[1], numerical_Cx[2]))
} else {
  cat("No values of Theta_0 are included in the numerical confidence
    interval C(x).\n")
}

plot(T_values, vc, type = "p", col = "blue", pch = 16,
     main = "Numerical Confidence Interval C(x)",
     xlab = expression(theta[0] ~ "values"),
     ylab = "Indicator")
grid()

c_alpha <- qchisq(1 - alpha, df = 1)
margin <- sqrt((c_alpha * sigma_squared) / n)
analytical_Cx <- c(x_bar - margin, x_bar + margin)

abline(v = analytical_Cx[1], col = 'red', lty = 2, lwd = 2)
abline(v = analytical_Cx[2], col = 'green', lty = 2, lwd = 2)

cat(sprintf("Numerical C(x): [%.2f, %.2f]\n", numerical_Cx[1], numerical
    _Cx[2]))
cat(sprintf("Analytical C(x): [%.2f, %.2f]\n",
            analytical_Cx[1], analytical_Cx[2]))
```



Numerical $C(x)$: [167.50, 172.10]

Analytical $C(x)$: [167.35, 172.15]

The numerical confidence interval $C(x)$ aligns closely with the analytical interval, with the lower bound a bit larger than the analytical result and upper bound a bit smaller than the analytical result, confirming our earlier findings that 175 is not included.