תרגיל בית 6 - התפלגות דגימה :

שאלה 1 :

(1) מכיוון שהתוחלת אינה ידועה אז האומד לשונות (מהרצאה).
יתבסס על :

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$\frac{n\cdot\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 \sim \chi^2_{(n-1)}$$

$$P\left(\left|\frac{\hat{\sigma}^2-\sigma^2}{\sigma^2}\right|\le 0.5\right) = P\left(\left|\frac{\chi^2_{(n-1)}}{n}-1\right|\le 0.5\right) =$$

$$P\left(1-0.5 \le \frac{\chi^2_{(n-1)}}{n} \le 1+0.5\right) = P\left(n(1-0.5) \le \chi^2_{(n-1)} \le n(1+0.5)\right)$$

$$= P\left(\chi^2_{(n-1)} \le n(1+0.5)\right) - P\left(\chi^2_{(n-1)} \le n(1-0.5)\right) = 0.743$$

(2) $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$\frac{(n-1)}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}) \sim \chi^2_{(n-1)}$$

$$P\left(\left|\frac{S^2-\sigma^2}{\sigma^2}\right|\le 0.5\right) = P\left(\left|\frac{\chi^2_{(n-1)}}{n-1}-1\right|\le 0.5\right) =$$

$$P\left(1-0.5 \le \frac{\chi^2_{(n-1)}}{n-1} \le 1+0.5\right) = P\left((n-1)(1-0.5) \le \chi^2_{(n-1)} \le (n-1)(1+0.5)\right)$$

$$= P\left(\chi^2_{(n-1)} \le (n-1)(1+0.5)\right) - P\left(\chi^2_{(n-1)} \le (n-1)(1-0.5)\right) = 0.734$$

אומד נראות מירבית:

(3) $E(\hat{\sigma}^2) = E\left(\frac{\sigma^2}{n}\chi^2_{(n-1)}\right) = \frac{\sigma^2}{n}E\left(\chi^2_{(n-1)}\right) = \frac{n-1}{n}\sigma^2$

$bias_{\hat{\sigma}^2}(\sigma^2) = E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$

$var(\hat{\sigma}^2) = var\left(\frac{\sigma^2}{n}\chi^2_{(n-1)}\right) = \frac{\sigma^4}{n^2}var\left(\chi^2_{(n-1)}\right) = \frac{\sigma^4}{n^2}\cdot 2(n-1)$

$MSE_{\hat{\sigma}^2}(\sigma^2) = var(\hat{\sigma}^2) + bias_{\hat{\sigma}^2}(\sigma^2)^2 = \frac{\sigma^4}{n^2}\cdot 2(n-1) + \frac{\sigma^4}{n^2} = \frac{\sigma^4}{n^2}(2n-1)$

:S^2

$E(s^2) = E\left(\frac{\sigma^2}{n-1}\chi^2_{(n-1)}\right) = \frac{\sigma^2}{n-1}E\left(\chi^2_{(n-1)}\right) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$

$bias_{s^2}(\sigma^2) = E(s^2) - \sigma^2 = \sigma^2 - \sigma^2 = 0$. אומד חסר הטיה.

$MSE_{s^2}(\sigma^2) = var(s^2) = var\left(\frac{\sigma^2}{n-1}\chi^2_{(n-1)}\right) = \frac{\sigma^4}{(n-1)^2}var\left(\chi^2_{(n-1)}\right) = \frac{\sigma^4}{(n-1)^2}(n-1)\cdot 2$

$= \frac{2\sigma^4}{(n-1)}$

לכל n כי:

$\frac{(2n-1)\sigma^4}{n^2} < \frac{2\sigma^4}{(n-1)}$

$MSE_{\hat{\sigma}^2}(\sigma^2) < MSE_{s^2}(\sigma^2)$

לכן נעדיף אומד נראות מירבית $\hat{\sigma}^2$ על s^2 מבחינת MSE.

(4) קיבלנו כי $p(s^2) < p(\hat{\sigma}^2)$ כלומר ההסתברות לטס"פ

קטנה מ 0.5 של האומד נראות מירבית גדול מההסתברות

שמתקבלת על ידי האומד s^2 והגודל האמפירי כי ב $\hat{\sigma}^2$ ~~[illegible]~~

קיבלנו ההסתברות גדולה לטס"פ לנכון שהוא עקיב יותר מ s^2

כלומר כמה הדיוק שלו גבוה יותר.

שאלה 2:

$X_1,\dots,X_n \sim N(\mu_x,\sigma^2)$]
$Y_1,\dots,Y_m \sim N(\mu_y,\sigma^2)$] המדגמים בלתי תלויים.

א) נתון S^2 חסר הטיה ל-σ^2 כלומר $E(S^2)=\sigma^2$ ולכן:

$$E(S^2)=E\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{K}\right)=\frac{\sigma^2}{K}E\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}+\frac{\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{\sigma^2}\right)$$

$$=\frac{\sigma^2}{K}\left[E\underbrace{\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\right)}_{\sim\chi^2(n-1)}+E\underbrace{\left(\frac{\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{\sigma^2}\right)}_{\sim\chi^2(m-1)}\right]=\frac{\sigma^2}{K}\left[(n-1)+(m-1)\right]$$

$$=\frac{\sigma^2}{K}(n+m-2)\underset{\text{חסר הטיה}}{\overset{\text{צריך}}{=}}\sigma^2 \Longleftrightarrow \boxed{K=n+m-2}$$

ב)

$$\omega=\frac{(n+m-2)}{\sigma^2}S^2=\frac{(n+m-2)}{\sigma^2}\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{n+m-2}\right]$$

$$=\underbrace{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}}_{\sim\chi^2(n-1)}+\underbrace{\frac{\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{\sigma^2}}_{\sim\chi^2(m-1)}\overset{(*)}{\sim}\chi^2_{(n+m-2)}$$

(*) מתכונת הרפרודוקטיביות של התפלגות חי-בריבוע.

ג)

$$\left.\begin{array}{l}X\sim N(\mu_x,\sigma^2)\Rightarrow\bar{X}\sim N\left(\mu_x,\frac{\sigma^2}{n}\right)\\ Y\sim N(\mu_y,\sigma^2)\Rightarrow\bar{Y}\sim N\left(\mu_y,\frac{\sigma^2}{m}\right)\end{array}\right]\Rightarrow \bar{X}-\bar{Y}\sim N\left(\mu_x-\mu_y,\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)\right)$$

$$E(\bar{X}-\bar{Y})=E(\bar{X})-E(\bar{Y})=\mu_x-\mu_y$$

$$var(\bar{X}-\bar{Y})=var(\bar{X})+var(\bar{Y})-\underset{0}{cov(\bar{X},\bar{Y})}=var(\bar{X})+var(\bar{Y})=\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)$$

$\bar{X},\bar{Y}$ מתפלגים נורמלים בלתי תלויים ולכן הם בלתי מתואמים.

ניגון לפי נוסח הבלה הנורמלי:

$$\frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim N(0,1)$$

$$S=\sqrt{S^2}=\sigma\sqrt{\frac{W}{(n+m-2)}} \sim \sigma\sqrt{\frac{\chi^2_{(n+m-2)}}{n+m-2}}$$

מסעיף (ב)

$$M=\frac{Z}{\sqrt{\frac{W_{n+m-2}}{n+m-2}}}=\frac{\dfrac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sigma\sqrt{\dfrac{W(n+m-2)}{n+m-2}}} \quad \begin{matrix}\sim N(0,1)\\ \\ \sim \sqrt{\dfrac{\chi^2_{(n+m-2)}}{(n+m-2)}}\end{matrix}$$

(*) הם בלתי תלויים.

ולכן לפי משפט $\boxed{M \sim t_{(m+n-2)}}$

(*) נראה שהמונה והמכנה בלתי תלויים:

- לפי משפט מההרצאה $\bar{X}$ ו- $X_i-\bar{X}$ הם בלתי תלויים לכל $i=1,\dots,n$
וגם $\bar{Y}$ ו- $Y_i-\bar{Y}$ הם בלתי תלויים לכל $i=1,\dots,m$
- $\bar{X}$ ו- $Y_i-\bar{Y}$ הם בלתי תלויים לכל $i=1,\dots,m$ וגם
$\bar{Y}$ ו- $X_i-\bar{X}$ הם בלתי תלויים לכל $i=1,\dots,n$.
- הווקטור: $(X_1-\bar{X},\dots,X_n-\bar{X},Y_1-\bar{Y},\dots,Y_m-\bar{Y},\bar{X},\bar{Y})$
מתפלג רב נורמלית כי כל קומבינציה לינארית של רכיביו היא קומבינציה לינארית של משתנים נורמליים בלתי תלויים. ולפי משפט כל קומבינציה מתפלגת נורמלית.
- לפי סעיף 2 של משפט שראינו בהרצאה $(\bar{X},\bar{Y})$ ו- $(X_1-\bar{X},\dots,X_n-\bar{X},Y_1-\bar{Y},\dots,Y_m-\bar{Y})$
הם בלתי תלויים.

ולכן משווני משפט סעיף 3 נסיק:

① $\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{m}(Y_i-\bar{Y})^2$

בלתי תלויים.

② $\dfrac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$

קומבינציה של $\bar{X}-\bar{Y}$

הסבר :

ⓑ עבור $m=n$

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

$$=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}+\frac{\sum_{i=1}^{n}(y_i-\bar{y})}{\sigma^2}\sim\chi^2_{(2n-2)}$$

$$=\frac{(n-1)}{\sigma^2}\left[\underbrace{\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{(n-1)}}_{\text{אומד לשונות X}}+\underbrace{\sum_{i=1}^{n}\frac{(y_i-\bar{y})^2}{(n-1)}}_{\text{אומד לשונות Y}}\right]$$

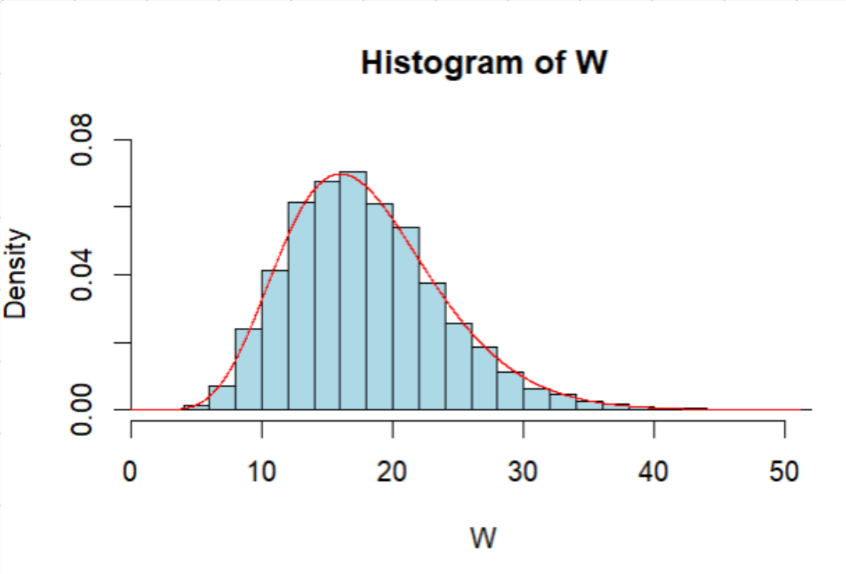
אומד לשונות Y = אומד לשונות X

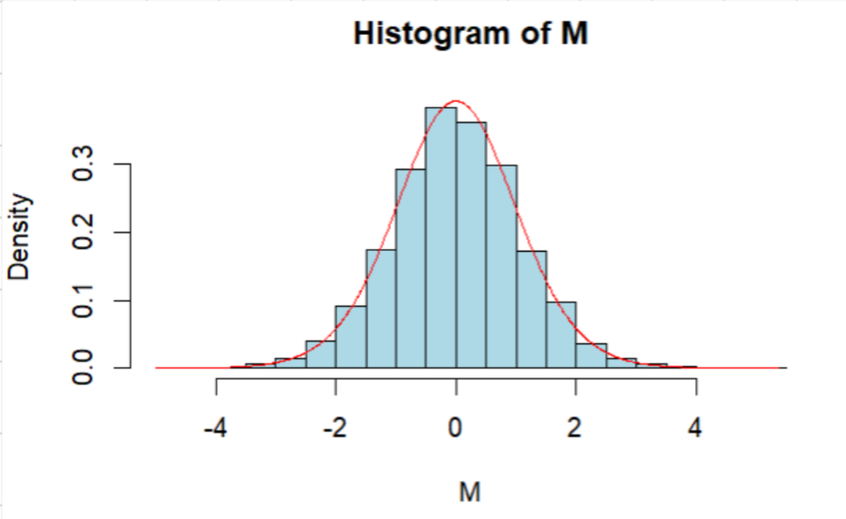
ⓗ הסבר :

$$M=\frac{\dfrac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{1}{n}+\frac{1}{n}}}}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}{2n-2}}}=\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}}\cdot\sqrt{\frac{2n-2}{\frac{2}{n}}}$$

$$=\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}}\sqrt{n(n-1)}$$

3) אנו רואים כי ההיסטוגרמה
תואמת את צורת
ההתפלגות



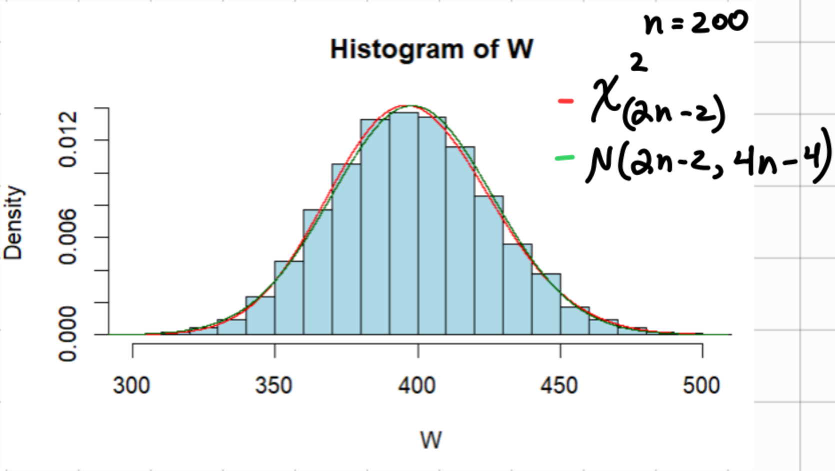



```r
N = 10000
n = 10


mu_x = 2
mu_y = 2
sigma = 3

W = rep(0, N)
M = rep(0, N)

S = function(sampleX, sampleY) {
  return((sum((sampleX - mean(sampleX))^2) + sum((sampleY - mean(sampleY))^2)) / (length(sampleX) + length(sampleY) - 2))
}

for (i in 1:N) {
  sampleX = rnorm(n, mean=mu_x, sd=sigma)
  sampleY = rnorm(n, mean=mu_y, sd=sigma)

  W[i] = (2 * n - 2) * S(sampleX, sampleY) / sigma^2
  M[i] = ((mean(sampleY) - mean(sampleX)) - (mu_y - mu_x)) / sqrt(S(sampleX, sampleY) * 2 / n)
}

interW = seq(0, max(W), 0.01)
hist(W, freq=FALSE, col='light blue', breaks=20, ylim=c(0, 0.08))
lines(interW, dchisq(interW, 2 * n - 2), col='red')

interM = seq(min(M), max(M), 0.01)
hist(M, freq=FALSE, col='light blue', breaks=20)
lines(interM, dt(interM, 2 * n - 2), col='red')
```

ה) ראינו בהרצאה כי $\chi^2_{(k)} \xrightarrow[k\to\infty]{} N(\mu,\sigma^2)$

כך ש $\mu = E(\chi^2_{(k)}) = k$

$\sigma^2 = var(\chi^2_{(k)}) = 2k$

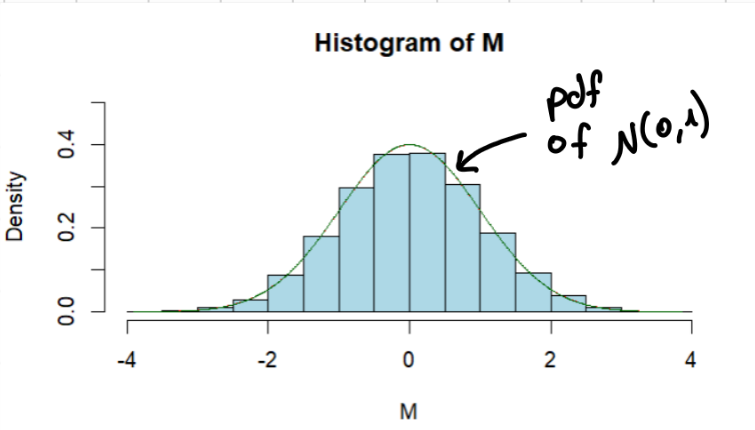


רואים כי ההתפלגות של $\chi^2_{(k)}$ קרובה מאוד לנורמלית

עם $\mu = E(\chi^2_{(k)})$

$\sigma^2 = var(\chi^2_{(k)})$

כך ש $k = 2n-2$



מההרצאה ראינו כי: $t_{(n)} \xrightarrow[n\to\infty]{D} N(0,1)$

והסימולציה תומכת בהנחה הזו

```
interW = seq(0, max(W), 0.01)
hist(W, freq=FALSE, col='light blue', breaks=20)
lines(interW, dchisq(interW, 2 * n - 2), col='red')
lines(interW, dnorm(interW, mean=2*n-2, sd=sqrt(2 * (2*n-2))), col='dark green')

interM = seq(min(M), max(M), 0.01)
hist(M, freq=FALSE, col='light blue', breaks=20, ylim=c(0, 0.5))
lines(interM, dnorm(interM), col='dark green')
```

(1)

$$P\left(\left|\frac{S^2-\sigma^2}{\sigma^2}\right| \le 0.5\right) = P\left(\left|\frac{S^2}{\sigma^2}-1\right| \le 0.5\right) =$$

$$P\left(\left|\frac{\sigma^2 \cdot \omega}{\sigma^2(2n-2)}-1\right| \le 0.5\right) = P\left((1-0.5) \le \frac{\omega}{2n-2} \le (1+0.5)\right)$$

$$P\left(\left|\frac{\omega}{2n-2}-1\right| \le 0.5\right) \xrightarrow{n\to\infty} P(0 \le 0.5) = 1$$

לפי החוק החלש של המספרים הגדולים

מכיוון ש- $\omega \sim \chi^2(2n-2)$ אז

$$\frac{\omega}{(2n-2)} \xrightarrow[P]{n\to\infty} 1$$

התשובה שקיבלנו אינטואיטיבית כי אנחנו מצפים שככל ש n

גדל אזי? האומד S^2 יתקרב לאומד σ^2. כלומר

$$\frac{S^2}{\sigma^2} \xrightarrow[n\to\infty]{} 1$$

עבור $n=10$:

$$P = P\left((2n-2)(1-0.5) \le \omega \le (2n-2)(1+0.5)\right)$$

$$P = P\left((2\cdot10-2)(1-0.5) \le \omega \le (2\cdot10-2)(1+0.5)\right)$$

$$= P(\omega \le 27) - P(\omega \le 9) \underset{\omega\sim\chi^2_{(18)}}{=} \text{pchisq}(27,18) - \text{pchisq}(9,18)$$

$$= 0.8807$$

עבור $n=100$:

$$P = P\left((2\cdot100-2)(1-0.5) \le \omega \le (2\cdot100-2)(1+0.5)\right)$$

$$= P(\omega \le 297) - P(\omega \le 99) \underset{\omega\sim\chi^2_{(198)}}{=} \text{pchisq}(297,198) - \text{pchisq}(99,198)$$

$$= 0.999$$

ב. נתון $bias_{s^2}(\sigma^2)=0$ ולכן

$$MSE_{s^2}(\sigma^2) = var(s^2) = var\left(\frac{\sigma^2}{m+n-2}\omega\right) = var\left(\frac{\sigma^2}{2n-2}\omega\right) =$$

$$\frac{\sigma^4}{(2n-2)^2} var(\omega) = \frac{\sigma^4 2(2n-2)}{(2n-2)^2} = \frac{\sigma^4}{n-1}$$

$$\omega \sim \chi^2_{(2n-2)}$$

השוואה: עבור $n \geq 3$:

$$\frac{\sigma^4}{n-1} < \frac{(2n-1)}{n^2}\sigma^4 < \frac{2\sigma^4}{(n-1)}$$

האומדים שהצענו היו לפחות בגודל של זה ולכן נעדיף
MSE המינימלי הכי של השלושה.

3 ridke

$$vec = \left(\sum_{i=1}^{10} (X_i - 1)^2 , \dots , \sum_{j=1}^{10} (X_j - 1)^2 \right)$$

$$vec = \left(4 \cdot W_{10_1} , \dots \quad 4 W_{10_{1000}} \right)$$

$$X_i \sim N(1, 4)$$

$$X_i - 1 \sim N(0, 4)$$

$$\left(\frac{X_i - 1}{2} \right) \sim N(0, 1)$$

$$W_{1_i} = \left(\frac{X_i - 1}{2} \right)^2 \sim \chi^2_{(1)}$$

$$\sum_{i=1}^{10} W_{1_i} \sim \chi^2_{(10)}$$

$$Var(vec) = \frac{\sum_{j=1}^{1000} 4 \cdot W_{10_j}}{1000 - 1} = Var\left(4 W_{10_j}\right) = 16 \, Var\left(W_{10_j}\right)$$

$$= 16 \cdot 2 \cdot 10 = 320$$