רגרסיה ומודלים סטטיסטיים- תרגיל 3

שאלה 1:

יהי $Z = (Z_1, \dots, Z_n)^T$ וקטור מקרי מהתפלגות משותפת f_Z. עבור המודלים הבאים, מצאו את וקטור התוחלות ומטריצת השונויות של Z:

א. $Z_i \sim N(i, i^2), i = 1, .., n$ ב״ת.

ב. $Z_1 = \eta_1$ ונגדיר $Z_i = 0.5 \cdot \eta_{i-1} + \eta_i, i = 2, \dots, n$ כאשר $\eta_i \sim N(0,1), iid$.

ג. בעבור $j = 1, .. k$ נגדיר

$P(X_j = 1) = p, P(X_j = 2) = q, P(X_j = 3) = 1 - p - q$. וכן:

$Z_i = \sum_{j=1}^{k} 1_{\{X_j = i\}}, i = 1,2,3$.

פתרון: על מנת למצוא את וקטור התוחלות ואת מטריצת השונויות נחשב את התוחלות והשונויות השוליות של כל אחד מהמשתנים המקריים ואת השונות המשותפת בין כל זוג.

א.

$$E(Z) = \left(E(Z_1), \dots, E(Z_n)\right)^t = (1, \dots, n)^T$$
$$Var(Z_i) = i^2$$
$$Cov(Z_i, Z_j) = 0$$

כי ב״ת. לכן:

$$cov(Z)_{ij} = \begin{cases} 0, & i \neq j \\ i^2, & i = j \end{cases}$$

ב.

$$E(Z_i) = 0.5E(\eta_{i-1}) + E(\eta_i) = 0$$

$$Var(Z_i) = 0.5^2 Var(\eta_{i-1}) + Var(\eta_i) = 1.25, i > 1$$
$$Var(Z_i) = 1, i = 1$$

$$Cov(Z_i, Z_j) = Cov(0.5\eta_{i-1} + \eta_i, 0.5\eta_{j-1} + \eta_j)$$

נפריד למקרים:

כאשר $i, j > 1$ עבור $|j - i| > 1$ נקבל 0, ואילו עבור $|j - i| = 1$ נקבל 0.5.

כאשר $i \ or \ j = 1$ נקבל 0.

דרך נוספת:

נגדיר $\eta = (\eta_1, .., \eta_n)^T$ אז כבר ראינו כי וקטור התוחלות הוא וקטור ה-0 ומטריצת השונויות היא מטריצת היחידה.

נגדיר $B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0 & 0 \\ 0 & 0.5 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 1 & 0 \\ 0 & 0 & 0 & 0.5 & 1 \end{pmatrix}$ אז שימו לב ש-$Z = B\eta$. ולכן:

$$cov(Z) = Bcov(\eta)B^T = BB^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 1.25 & 0 & 0 & 0 \\ 0 & 0.5 & 1.25 & 0 & 0 \\ 0 & 0 & 0.5 & 1.25 & 0 \\ 0 & 0 & 0 & 0.5 & 1.25 \end{pmatrix}$$

ג. ההתפלגות השולית של כל אחד מהרכיבים היא התפלגות בינומית שכן מדובר בסכום של אינדיקטורים מקריים ב"ת. לכן:

$$E(Z) = \left(kp, kq, k(1-p-q)\right)^T$$

$$Cov(Z_1, Z_2) = Cov\left(\sum_{i=1}^{k} 1_{\{X_i=1\}}, \sum_{j=1}^{k} 1_{\{X_j=2\}}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} Cov(1_{\{X_i=1\}}, 1_{\{X_j=2\}})$$

$$= \sum_{i=1}^{k} cov\left(1_{\{X_i=1\}}, 1_{\{X_i=2\}}\right) = k\left[E\left(1_{\{X_1=1,X_1=2\}}\right) - E\left(1_{\{X_1=1\}}\right)E\left(1_{\{X_1=2\}}\right)\right]$$

$$= -kpq$$

ובאותו האופן עבור שאר הכניסות. כאשר המעבר האחרון נובע מכך שהמאורעות $\{X_1 = 1\}, \{X_1 = 2\}$ זרים.

שאלה 2:

יהיו $Z, W \in R^p$ וקטורים מקריים. הראו שהבאים שקולים:

(1) $\forall v \in R^p, Var(v^T Z) \geq Var(v^T W)$

(2) $B \coloneqq Var(Z) - Var(W)$ היא מטריצה חיובית למחצה.

(3) קיימת המטריצה $B^{\frac{1}{2}}$.

פתרון:

נראה את השקילות בכל מעבר בין (1) ל-(2):

לכל וקטור שונה מ-0 $v \in R^p$:

$$Var(v^T Z) \geq Var(v^T W) \Leftrightarrow v^T Var(Z) v \geq v^T Var(W) v \Leftrightarrow v^T Var(Z) v - v^T Var(W) v \geq 0$$
$$\Leftrightarrow v^T\left(Var(Z) - Var(W)\right)v \geq 0 \ \Leftrightarrow Var(Z) - Var(W) \ is\ s.p.d\ matrix$$

עבור $(2) \Leftrightarrow (3)$:

המטריצה B סימטרית כיוון שהפרש של מטריצות סימטריות. מכאן שקיים לה פירוק ספקטרלי:

$$B = U\Lambda U^T$$

כעת אם המטריצה אי שלילית אז כל ערכיה העצמיים אי שליליים ולכן הפעולה $\lambda^{\frac{1}{2}}$ מוגדרת. ונקבל כי קיימת המטריצה $B^{\frac{1}{2}} = U\Lambda^{\frac{1}{2}}U^T$. באופן דומה בכיוון ההפוך.

שאלה 3:

יהיו X, Y וקטורים מקריים עם התפלגות משותפת עם תוחלות μ_x, μ_y בהתאמה. נגדיר:

$$\Sigma_X = E[(X-\mu_x)(X-\mu_x)^T]$$

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)^T]$$

הוכיחו את התכונות הבאות:

(a) $\Sigma_x = \mathbb{E}XX^T - \mu_x\mu_x^T$

(b) $\Sigma_x \geq 0$ (The covariance matrix of X is positive semidefinite)

(c) $cov(AX+b) = A\Sigma_x A^T$

(d) $cov(X,Y) = cov(Y,X)^T$

(e) $cov(X_1+X_2,Y) = cov(X_1,Y) + cov(X_2,Y)$

(f) $cov(AX,BY) = A[cov(X,Y)]B^T$

פתרון:

.1

$$\Sigma_X = E(XX^T - \mu_X X^T - \mu_X^T X + \mu_X\mu_X^T) = E(XX^T) - 2\mu_X E(X)^T + \mu_X\mu_X^T = E(XX^T) - \mu_X\mu_X^T$$

.2

לכל $v \neq 0$ מתקיים:

$$v^T E[(X-\mu_x)(X-\mu_x)^T]v = E[v^T(X-\mu_x)(X-\mu_x)^T v] = E\left[\left|\left|(X-\mu_x)^T v\right|\right|^2\right] \geq 0$$

כי תוחלת של מ"מ אי שלילי.

.3

$$\begin{aligned} Cov(AX+b) &= E[(AX+b-A\mu_x-b)(AX+b-A\mu_x-b)^T] \\ &= E[(AX-A\mu_x)(AX-A\mu_x)^T] = E[(A(X-\mu_x)(A(X-\mu_x))^T] \\ &= E[A(X-\mu_x)(X-\mu_x)^T A^T] = ACov(X)A^T \end{aligned}$$

.4 כפי שראינו בתרגול, $Cov(X,Y)_{ij} = cov(X_i,Y_j) = cov(Y_j,X_i) = Cov(Y,X)_{ji}$

.5 $E((X_1+X_2)Y^T) - (\mu_{X_1}+\mu_{X_2})\mu_y^T = E(X_1Y^T) - (\mu_{X_1}\mu_y^T) + E(X_2Y^T) - (\mu_{X_2}\mu_y^T) =$ $Cov(X1,Y) + Cov(X2,Y)$

.6 בדיוק באותו האופן כמו בשאלה 3.

שאלה 4:

הניחו מודל רגרסיה לינארית פשוט (תרגיל 1 שאלה 4) בו המשתנה המוסבר הוא לחץ הדם (Y) והמשתנה המסביר הוא הגיל בשנים (X). הניחו שבידיכם דגימות של 100 מטופלים כאשר לכל מטופל מופיע הגיל ולחץ הדם שנמדד עבורו. הראו (מתמטית!) אילו מהנחות המודל הלינארי מופרות בכל אחד מהמקרים הבאים:

א. משיקולי תקציב הנדגמים הגיעו מ-20 משפחות בנות 5 נפשות שנדגמו באופן מקרי מכלל האוכלוסיה. ידוע כי ישנו קשר בין הגנטיקה ובין נטייה למחלות הקשורות ללחץ הדם.
ב. ידוע שהמינון של תרופות המפחיתות לחץ דם עולה עם הגיל.

חשבו את התוחלת והשונות של $\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY$ בכל אחד מהמקרים.

א. כיוון שיש קשר בין הגנטיקה ובין הנטייה למחלות שקשורות ללחץ הדם, אך הדגימה היא דגימה מקרית מכלל האוכלוסיה (דגימת אשכולות) עדיין ניתן להניח כי $E(\epsilon|X) = 0$, אך ישנם מתאם בין הפרטים מאותה המשפחה. לכן ההנחה כי $cov(\epsilon_i,\epsilon_j|X) = 0$ מופרת. (אפילו אם נניח שהשונות שווה בין משפחות). נסמן $cov(\epsilon_i,\epsilon_j|i \in family_I, j \in family_I) = \rho, I = 1,..,20$

כלומר בתוך כל משפחה I מטריצת השונויות היא (זהה עבור כל המשפחות):

$$\Sigma_I = \begin{pmatrix} \sigma^2 & \rho & \rho & \rho & \rho \\ \rho & \sigma^2 & \rho & \rho & \rho \\ \rho & \rho & \sigma^2 & \rho & \rho \\ \rho & \rho & \rho & \sigma^2 & \rho \\ \rho & \rho & \rho & \rho & \sigma^2 \end{pmatrix}$$

ומטריצת השונויות היא מטריצת בלוקים אלכסונית:

$$\Sigma = \begin{pmatrix} \Sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_{20} \end{pmatrix}$$

$$E(\hat{\beta}^{OLS}) = E((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^TE(Y) = (X^TX)^{-1}X^T(X\beta + \epsilon) = \beta$$
$$cov(\hat{\beta}^{OLS}) = (X^TX)^{-1}X^Tcov(Y)X(X^TX)^{-1} = (X^TX)^{-1}X^T\Sigma X(X^TX)^{-1}$$

שימו לב שחוסר ההטייה לא נפגע.

ב. במקרה הזה ישנו משתנה מושמט (לשם נוחות נכתוב זאת באופן מטריציוני עבור המקרה הכללי):

$$Y = X\beta + Z\alpha + \epsilon$$

כאשר Z היא מטריצה קבועה של משתנים מסבירים שאינם כלולים במטריצה X ו-α הוא וקטור המקדמים שלהם, ולכן הנחת הליניאריות מופרת. נקבל:

$$E(\hat{\beta}^{OLS}) = (X^TX)^{-1}X^T(X\beta + Z\alpha + \epsilon) = \beta + (X^TX)^{-1}X^TZ\alpha$$

במקרה הזה $\hat{\beta}^{OLS}$ יהיה חסר הטיה אם ורק אם אחד מהשניים מתקיים:

- $\alpha = 0$- כלומר התרופות לא משפיעות על לחץ הדם, אך נתון שזה לא המקרה.
- $Z \in IM(X)^{\perp}$- אבל נתון שהמינון של התרופות (Z) עולה בגיל (X), ולכן גם זה אינו המקרה.

$$cov(\hat{\beta}^{OLS}) = (X^TX)^{-1}X^Tcov(X\beta + Z\alpha + \epsilon)(X^TX)^{-1}X^T = (X^TX)^{-1}X^Tcov(\epsilon)X(X^TX)^{-1}$$
$$= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$$

כלומר האומד מוטה אך שונותו עדיין זהה למקרה בו ההנחות מתקיימות.

שאלה 5:

בשאלה זו נבחן את הנחות המודל הליניארי דרך סימולציה:

א. ראשית הניחו כי כל הנחות המודל הליניארי מתקיימות. מצאו ביטוי להטייה ולמטריצת השונויות של $\hat{\beta}_{OLS}$.

ב. באמצעות הספרייה MASS והפקודה mvrnorm(num_vectors, mu, Sigma) הגרילו מהתפלגות נורמלית 5-מימדית 500 וקטורים ב"ת, כאשר:
Sigma- מטריצת היחידה ממימד 5.
mu- הוקטור $(0,1,1,2,2)^T$.
לאורך השאלה שמרו את המטריצה המתקבלת קבועה. זו תשמש אותנו להיות המטריצה X.

ג. הגרילו 500 משתנים מקריים נורמליים סטנדרטיים ב"ת (חד מימדיים), והגדירו:
$$Y_i = 2 - 3X_{i1} + 2X_{i2} + X_{i3} + 6X_{i4} - 2X_{i5} + \epsilon_i$$

נסחו את המודל בתצורה מטריציונית. הציגו את 10 השורות הראשונות של המטריצה $X \in R^{n\times(p+1)}$, את וקטור המקדמים (מה הוא $p + 1$?), ואת 10 הכניסות הראשונות של הוקטורים Y ו-ϵ. ללא שימוש

בפונקציות מובנות, אלא רק בפעולות פשוטות על מטריצות, אמדו את $\hat{\beta}_{OLS}$, ושמרו את ערכי המקדמים.

ד. חזרו על הסעיף הקודם 10,000 פעמים, כך שברשותכם יהיו 10,000 אומדים ל-$\hat{\beta}_{OLS}$. חשבו את וקטור הממוצעים שלהם, ואת מטריצת השונויות האמפירית. השוו אותם לפרמטרים התיאורטיים שמתקבלים מהחישוב שביצעתם בסעיף א׳. הסבירו את התוצאות.

ה. חזרו על סעיפים ג׳ ו-ד׳ פעמיים נוספות, אך כעת תחת תוך דגימה של הרעש המקרי באופנים הבאים:

1. $\epsilon_i \sim N\left(0, \left|\left|X_i\right|\right|^2\right)$. כאשר הכוונה היא ש-$X_i \in R^{p+1}$ הוא השורה ה-i של המטריצה X.
2. $\epsilon_i \sim N(1,1)$.

אילו הנחות מופרות בכל אחד מהמקרים? כיצד זה השפיע על התוחלת והשונות של האומדים ביחס לסעיף הקודם? (כדי להשוות את השונות, השתמשו בקריטריון משאלה 2).

הערה: הפקודות מסעיף ב׳ הן ב-*R* אבל אתם יכולים להשתמש בשפת תכנות לבחירתכם, וכן בכל חומר עזר/מודל שפה ג׳נרטיבי. הקפידו לצרף את הקוד <u>והסברים מילוליים משלכם</u>. הגישו קובץ *PDF* אחד הכולל את הפתרון שלכם לכל השאלות וכן את הקוד והתוצאות של השאלה האחרונה.

Exercise 4

Niv Brosh

2 6 2024

Q.1.

a. As we saw in class:

$$E(\hat{\beta}_{OLS}) = E((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^TE(Y) = (X^TX)^{-1}X^TX\beta = \beta$$

$$Var(\hat{\beta}_{OLS}) = (X^TX)^{-1}X^TVar(Y)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$$

b.

```
library(MASS)
options(scipen = 999)

num_vectors <- 500

dimension <- 5

mu <- c(0,1,1,2,2)
Sigma <- diag(dimension)
Sigma=Sigma*10

X <- mvrnorm(num_vectors, mu, Sigma)
ones = rep(1, nrow(X))
X =cbind(ones,X)
beta = c(2,-3,2,1,6,-2)
epsilon = rnorm(num_vectors, 0, 1)
Y = X%*%beta +epsilon
show = list(X_10 = head(X,10),beta=beta,Y_10= head(Y,10),epsilon_10 =head(epsilon,10)  )
show
```

```
## $X_10
##       ones
##  [1,]    1 -6.4268228 -2.43178426  0.002825552  1.848830  2.7150275
##  [2,]    1 -3.6687149 -0.48761816  7.289962847  4.621083  1.7591071
##  [3,]    1  0.6044354 -4.58588687  4.941767011 -2.248502  5.1721951
##  [4,]    1  0.2163416 -0.80354817 -3.987751865  2.167787  5.0930243
##  [5,]    1  1.0510499 -3.91037592 -0.021927031 -3.140650  1.0482499
##  [6,]    1  1.3293282 -0.42797190  2.492569352  3.759522  3.1552190
##  [7,]    1  4.5752927  3.44754824  1.982310671  4.340535  4.0670846
##  [8,]    1  1.2472463  3.38182367 -1.284490332  2.310364 -0.8713118
##  [9,]    1 -4.3158238 -0.90491106 -1.555306666  8.551542 -0.6328714
```

```
## [10,]    1  2.2847866  0.01299272 -3.046687532  1.751308  5.3826120
##
## $beta
## [1]  2 -3  2  1  6 -2
##
## $Y_10
##             [,1]
##  [1,]  20.672106
##  [2,]  43.923786
##  [3,] -28.949190
##  [4,]  -3.537736
##  [5,] -28.608809
##  [6,]  16.234421
##  [7,]  14.324497
##  [8,]  18.038805
##  [9,]  63.383103
## [10,]  -9.391438
##
## $epsilon_10
##  [1] -1.4105472  0.3946281 -1.0704766 -2.1145364  1.3274178  0.3390865
##  [7] -0.7360747 -1.3034189 -0.7742328 -1.2590003
```

c. OLS estimation:

```
beta_hat_func = function(X,Y){
  return(solve(t(X)%*%X)%*%t(X)%*% Y)
}

beta_hat_func(X,Y)
```

```
##            [,1]
## ones  1.9898517
##      -3.0109134
##       1.9956779
##       0.9966917
##       5.9841127
##      -1.9851424
```

d. 10,000 simulations, based on LLN for approximations of the bias and variance:

```
create_random_Y <- function(X, beta, noise_mu, noise_sigma) {
  n <- nrow(X)
  noise <- rnorm(n, noise_mu, noise_sigma)
  Y_samples <- X %*% beta + noise
  return(Y_samples)
}

betasim <- replicate(10000, beta_hat_func(X, create_random_Y(X, beta, 0,1)))
betasim <- matrix(betasim, nrow=length(beta))
cov_matrix <- cov(t(betasim))
var_diff = sum(diag(cov_matrix))-sum(diag(solve(t(X) %*% X)))
bias_sq = sum(rowMeans(betasim)-beta)^2
```

e. Now we'll examine the effect of violation the linear model assumptions. first we'll draw ϵ_i-s such that $\epsilon_i \sim N(0, ||X_i||^2)$, which violate the assumption that the variance of the noise is constant ($\sigma^2 I$) across X (Homoscedasticity assumption).

```
create_random_Y_norm <- function(X, beta, noise_mu) {
  n <- nrow(X)
  noise_variances <- apply(X, 1, function(row) sum(row^2))  # Calculate ||row_X_i||^2 for each row
  noise <- rnorm(n, noise_mu, sqrt(noise_variances))
  Y_samples <- X %*% beta + noise
  return(Y_samples)
}

betasim_norm <- replicate(10000,
                          beta_hat_func(X, create_random_Y_norm(X, beta, 0)))
betasim_norm <- matrix(betasim_norm,
                       nrow=length(beta))
row_means_norm <- rowMeans(betasim_norm)
cov_matrix_norm <- cov(t(betasim_norm))

var_diff_norm= sum(diag(cov_matrix_norm))-sum(diag(solve(t(X) %*% X)))
bias_sq_norm = sum(row_means_norm-beta)^2
```

Finally, we'll compare the effect of violating the $E(\epsilon) = 0$ assumption by sampling from the $N(1, 1)$ distribution:

```
betasim_biased <- replicate(10000, beta_hat_func(X, create_random_Y(X, beta, 1,1)))
betasim_biased <- matrix(betasim_biased, nrow=length(beta))
cov_matrix_biased <- cov(t(betasim_biased))
var_diff_biased = sum(diag(cov_matrix_biased))-sum(diag(solve(t(X) %*% X)))
bias_biased_sq = sum(rowMeans(betasim_biased)-beta)^2
```

summary table of the results:

```
summary_table <- data.frame(
  Distribution = c("Homoscedastic",
                   "Heteroscedastic", "Biased"),
  Bias_Squared = c(round(bias_sq,6),
                   round(bias_sq_norm,6), round(bias_biased_sq,6)),
  Variance_Difference = c(round(var_diff,6),
                          round(var_diff_norm,6), round(var_diff_biased,6))
)
summary_table
```

```
##      Distribution Bias_Squared Variance_Difference
## 1   Homoscedastic     0.000000            0.000033
## 2 Heteroscedastic     0.000015            0.288703
## 3          Biased     1.000098           -0.000062
```

As we can see from the table, when all the assumptions are correct, we get the best results in terms of MSE. Violation of the homoscedasticity assumption leads to higher variance (which also causes slower convergence of the average to the mean according to Chebyshev's inequality), but it does not affect the bias term of the estimator. Violation of the centrality assumption of ϵ yields biased estimators, but does not affect their variance.