מודלים סטטיסטיים למתקדמים א' – סיכום הרצאות

מרצה: פרופ' דוד צוקר
המחלקה לסטטיסטיקה, הפקולטה למדעי החברה, האוניברסיטה העברית

סיכום: אריאל וישנה, ariel.vishne@gmail.com

כל טעות אשר נפלה בסיכום זה הינה על אחריות המסכם בלבד

תוכן עניינים

52801 מודלים סטטיסטיים למתקדמים 2023א
מרצה: פרופ' דוד צוקר
סיכום: אריאל וישנה

שיעור 1

נושא 0 - מבוא

מבוא 1.1 מבנה הקורס

1. ניתוח רב משתני
2. השוואות מרובות ונושאים נלווים

מבוא 1.2 מטלות

1. מבחן סופי – 65%
2. תרגילים – 15%
3. בחני בית – 20%

הקדמה 1.3 ניתוח נתונים רב-משתני – ניתוח רב-משתני מתרחש כאשר ישנן n תצפיות $x_1, \ldots, x_n$ וכל תצפית היא וקטור $x_i \in \mathbb{R}^m$ כלומר $x_i = \begin{pmatrix} x_{i1} \\ \cdots \\ x_{im} \end{pmatrix}$.

דוגמה 1.4 – ציונים של תלמידים במספר קורסים או מספר מבחנים באותו קורס

דוגמה 1.5 – מדידות חוזרות על פני זמן

את הקורס נתחיל ממקרה פשוט של דגימה מהתפלגות רב-נורמלית. לאחר מכן נעבור למצבים מורכבים יותר כמו מודלים מעורבים (mixture model) וכך הלאה.

לפני כן נעשה רענון בסיסי על שלושה נושאים הנדרשים להמשך הקורס:

1. ערכים עצמיים ווקטורים עצמיים
2. וקטוריים מקריים
3. התפלגות רב-נורמלית

העמקה אלגברה 1 – ערכים עצמיים ווקטורים עצמיים

הגדרה 1.6 – ערכים עצמיים, וקטורים עצמיים ופירוק ספקטראלי – תהי מטריצה $A \in \mathbb{R}^{m \times m}$. אם מתקיים השוויון $Ax = \lambda x$ עבור סקלר $\lambda \in \mathbb{R}$ כלשהו ווקטור $x \in \mathbb{R}^m$ כלשהו השונה מאפס אזי x מוגדר להיות וקטור עצמי של A עם ערך עצמי λ.

במצב זה ניתן לרשום:

$$Ax = \lambda x \Rightarrow (A - \lambda I)x = 0$$

כיוון שקיים פתרון למשוואה $Ax = \lambda x$ אזי המטריצה $(A - \lambda I)$ היא מטריצה סינגולרית (אם יש למטריצה kernel/nullspacre לא טריוויאלי אזי היא בהכרח סינגולרית) ומכאן שהדטרמיננטה שווה ל-0:

$$\det(A - \lambda I) = |A - \lambda I| = 0$$

הגדרה 1.7 – פולינום אופייני – עבור המטריצה A מהסעיף הקודם נגדיר

$$\chi_A(\lambda) = \det(A - \lambda I)$$

בתור הפולינום האופייני של A.

הגדרה 1.8 – פירוק ספקטראלי נניח $A \in \mathbb{R}^{m \times m}$ מטריצה סימטרית. אזי קיים פירוק

$$A = U\Lambda U^T$$

כאשר:

1. $\Lambda \in \mathbb{R}^{m\times m}$ מטריצה אלכסונית שעל האלכסון שלה נמצאים הערכים העצמיים של A
2. $U \in \mathbb{R}^{m\times m}$ מטריצת עמודות כאשר עמודה j הינה וקטור עצמי של A התואם לערך העצמי λ_j.

מתקיים גם $U^TU = UU^T = I$ כלומר U מטריצה אורתוגונלית.

הערה 1.9 – אם

- A מטריצה סימטרית
- $Ax_1 = \lambda_1 x_1$
- $Ax_2 = \lambda_2 x_2$
- $\lambda_1 \neq \lambda_2$

אזי $x_1 \perp x_2$ כלומר $x_1^T x_2 = 0$. כלומר כל שני וקטורים עצמיים עם ערכים עצמיים שונים של A הם מאונכים זה לזה.

דוגמה 1.10

$$A = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$$

הפולינום האופייני מוגדר על ידי:

$$\chi_A(\lambda) = \det(A - \lambda I) = \left|\begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right| = \begin{vmatrix} 2-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} = (2-\lambda)(5-\lambda) - 4$$
$$= \lambda^2 - 7\lambda + 6 \Rightarrow \lambda_1 = 1, \lambda_2 = 6$$

נרצה לפתור את השוויון כדי למצוא את הווקטור העצמי עבור $\lambda_1 = 1$:

$$A - \lambda_1 I = 0 \rightarrow \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0 \rightarrow \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \rightarrow \begin{matrix} x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{matrix} = \begin{matrix} 0 \\ 0 \end{matrix} \rightarrow \begin{matrix} x_1 = -2a \\ x_2 = a \end{matrix}$$

עבור סקלר כלשהו $a \in \mathbb{R}$.

כלומר מצאנו עמודות למטריצה U כדי לנרמל את הערכים נקבל:

$$a = \frac{1}{\sqrt{5}} \rightarrow \|x_1\|^2 = 4a^2 + a^2 = 5a^2$$

ולכן:

$$u_1 = \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}$$

עבור הערך העצמי $\lambda_2 = 6$:

$$\chi_A(\lambda) = \det(A - \lambda I) = \left|\begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} - 6\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right| = \begin{vmatrix} -4 & 2 \\ 2 & -1 \end{vmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \Leftrightarrow \begin{matrix} -4x_1 + 2x_2 \\ 2x_1 - x_2 \end{matrix} = \begin{matrix} 0 \\ 0 \end{matrix}$$

שני השווינות הם למעשה אותו שוויון לכן:

$$x_2 = \begin{pmatrix} a/2 \\ a \end{pmatrix}$$

גם כאן נרצה לנרמל כך שהנורמה תהיה שווה ל-1

$$\|x_2\|^2 = 1 \Leftrightarrow \frac{1}{4}a^2 + a^2 = \frac{5}{4}a^2 \Leftrightarrow a = \frac{2}{\sqrt{5}}$$

כלומר:

$$u_2 = \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix}$$

ובסך הכל:

$$U = \begin{pmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}$$

והפירוק הספקטראלי הוא:

$$A = \begin{pmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} -2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}$$

במקרה שלנו $U = U^T$ מטריצה סימטרית (לא בהכרח קורה במקרה הכללי של פירוק ספקטראלי).

הגדרה 1.11 – חזקה של מטריצה סימטרית - באמצעות הפירוק הספקטראלי ניתן לראות כי עבור מטריצה סימטרית A מתקיים:

$$A^2 = U\Lambda U^T U\Lambda U^T = U\Lambda I\Lambda U^T = U\Lambda^2 U^T$$

ובאופן כללי לכל $k \in \mathbb{R}^n$ מתקיים:

$$A^k = U\Lambda^k U^T$$

הגדרה 1.12 שורש סימטרי של מטריצה סימטרית – אם כל הערכים העצמיים $\lambda_1, \dots, \lambda_m \geq 0$ אי-שליליים אז ניתן להגדיר עבור מטריצה סימטרית A:

$$\sqrt{A} = A^{1/2} = U\Lambda^{1/2} U^T$$

נשים לב כי:

$$A = A^{1/2}A^{1/2} = U\Lambda^{1/2}U^T U\Lambda^{1/2}U^T = U\Lambda^{1/2}I\Lambda^{1/2}U^T = U\Lambda^{1/2}\Lambda^{1/2}U^T = U\Lambda U^T$$

טענה 1.13 – מטריצה חיובית לחלוטין וחיובית למחצה – תהי $A \in \mathbb{R}^{m\times m}$ מטריצה סימטרית. אזי:

- A חיובית לחלוטין (Positive Definite) אם לכל וקטור $0 \neq c \in \mathbb{R}^m$ מתקיים $c^T Ac > 0$.
- A חיובית למחצה (Positive Semi-Definite) אם לכל וקטור $c \in \mathbb{R}^m$ מתקיים $c^T Ac \geq 0$.

טענה 1.14 –

- A חיובית לחלוטין אם ורק אם כל הערכים העצמיים של A הם חיוביים $\lambda_1, \dots, \lambda_m > 0$.
- A חיובית למחצה אם ורק אם כל הערכים העצמיים של A הם אי-שליליים $\lambda_1, \dots, \lambda_m \geq 0$.

הגדרה 1.15 – פירוק חולסקי (Cholesky) – תהי מטריצה $A \in \mathbb{R}^{m\times m}$ סימטרית וחיובית לחלוטין. אזי נוכל להציג את A על ידי:

$$A = LL^T$$

כאשר L מטריצה משולשית תחתונה. פירוק זה הוא פירוק יחיד עם הערכים באלכסון של L אם כל הערכים הם ערכים חיוביים.

המטריצה L מכונה שורש $Cholesky$ של A.

הערה 1.16 – ב-R הפונקציה $L = chol(A)$ מייצגת את L^T.

דוגמה 1.17 – עבור $m = 3$:

$$L = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \to LL^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$
$$= \begin{pmatrix} L_{11}^2 & L_{11}L_{21} & L_{11}L_{31} \\ L_{11}L_{21} & L_{21}^2 + L_{22}^2 & L_{21}L_{31} + L_{22}L_{32} \\ L_{11}L_{31} & L_{21}L_{31} + L_{22}L_{32} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

העמקה אלגברה 2 – וקטורים מקריים

הגדרה 1.18 – וקטור מקרי – וקטור $Z = \begin{pmatrix} z_1 \\ \cdots \\ z_m \end{pmatrix}$ הוא וקטור שבו כל קואורדינטה z_j היא משתנה מקרי.

הגדרה 1.19 – מטריצה מקרית – באופן דומה $U = \begin{pmatrix} u_{11} & \cdots & u_{1k} \\ \cdots & \cdots & \cdots \\ u_{m1} & \cdots & u_{mk} \end{pmatrix}$ מטריצה מקרית כאשר כל משתנה u_{rs} הוא משתנה מקרי.

הגדרה 1.20 – תוחלת של וקטור מקרי ושל מטריצה מקרית היא וקטור ומטריצת התוחלות בהתאמה

$$\mathbb{E}[Z] = \mathbb{E}\begin{bmatrix} z_1 \\ \cdots \\ z_m \end{bmatrix} = \begin{bmatrix} \mathbb{E}[z_1] \\ \cdots \\ \mathbb{E}[z_m] \end{bmatrix}$$
$$\mathbb{E}[U] = \mathbb{E}\begin{bmatrix} u_{11} & \cdots & u_{1k} \\ \cdots & \cdots & \cdots \\ u_{m1} & \cdots & u_{mk} \end{bmatrix} = \begin{bmatrix} \mathbb{E}[u_{11}] & \cdots & \mathbb{E}[u_{1k}] \\ \cdots & \cdots & \cdots \\ \mathbb{E}[u_{m1}] & \cdots & \mathbb{E}[u_{mk}] \end{bmatrix}$$

טענה 1.21 – אם $Z, W \in \mathbb{R}^m$ וקטורים מקריים אזי:

$$\mathbb{E}[Z + W] = \mathbb{E}[Z] + \mathbb{E}[W]$$

טענה 1.22 – אם $Z \in \mathbb{R}^m$ וקטור מקרי ו-$A \in \mathbb{R}^{m\times k}$ מטריצה קבועה אזי:

$$\mathbb{E}[AZ] = A \cdot \mathbb{E}[Z]$$

טענה 1.23 – עבור U מטריצה מקרית ו-A, B מטריצות קבועות, אזי:

$$\mathbb{E}[AUB] = A \cdot \mathbb{E}[U] \cdot B$$

הגדרה 1.24 – מטריצת שונויות – יהי וקטור מקרי $Z \in \mathbb{R}^m$. אזי מטריצת השונויות המשותפות של Z היא מטריצה ריבועית סימטרית בגודל $m \times m$ ומוגדרת על ידי:

$$[Cov(Z)]_{rs} = Cov(Z_r, Z_s)$$

הגדרה 1.25 – עבור כל וקטורים מקריים $Z \in \mathbb{R}^m$ ו-$W \in \mathbb{R}^k$ מתקיים כי מטריצת השונויות המשותפות היא מטריצה בגודל $m \times k$ המוגדרת על ידי:

$$[Cov(Z, W)]_{rs} = Cov(Z_r, W_s)$$

נשים לב כי מתוך ההגדרה נובע:

$$Cov(Z, Z) = Cov(Z)$$

טענה 1.26 – תכונות מטריצת שונויות משותפות – יהיו וקטורים מקריים $Z, W \in \mathbb{R}^m$, מטריצות קבועות $A, B \in \mathbb{R}^{m\times k}$ ווקטור קבוע $a \in \mathbb{R}^m$. אזי:

$$Cov(AZ) = A \cdot Cov(Z) \cdot A^T$$

$Cov(AZ,BW)=A\cdot Cov(Z,W)\cdot B^T$

$Cov(a^TZ)=Var(a^TZ)=a^TCov(Z)a$

כלומר המטריצה $Cov(Z)$ היא מטריצה חיובית למחצה.

טענה 1.27 –

$Cov(Z)=\mathbb{E}[(Z-\mathbb{E}[Z])(Z-\mathbb{E}[Z])^T]$
$Cov(Z,W)=\mathbb{E}[(Z-\mathbb{E}[Z])(W-\mathbb{E}[W])^T]$

נושא 1 – התפלגות רב-נורמלית

הגדרה 1.28 התפלגות רב-נורמלית –וקטור $W\in\mathbb{R}^k$ מתפלג רב-נורמלי סטנדרטי אם ניתן לכתוב את W על-ידי:

$W=\vec{\mu}+AZ$

כאשר $\mu\in\mathbb{R}^k$ וקטור קבוע של תוחלות, $A\in\mathbb{R}^{k\times m}$ מטריצה קבועה ו-$Z\in\mathbb{R}^m$ וקטור מקרי שבו $Z_1,\dots,Z_m\overset{iid}{\sim}N(0,1)$.

טענה 1.29 – אם W מתפלג רב-נורמלית כלומר $W=\mu+AZ$ אזי:

$\mathbb{E}[W]=\mathbb{E}[\mu+AZ]=\mu+A\mathbb{E}[Z]=\mu+A\cdot 0=\mu$
$Cov[W]=Cov[\mu+AZ]=Cov[AZ]=A\cdot Cov(Z)\cdot A^T=A\cdot I\cdot A^T=AA^T$

שיעור 2

תזכורת 2.0 - התפלגות רב-נורמלית –וקטור $W\in\mathbb{R}^k$ מתפלג רב-נורמלי סטנדרטי אם ניתן לכתוב את W על-ידי:

$W=\vec{\mu}+AZ$

כאשר $\mu\in\mathbb{R}^k$ וקטור קבוע של תוחלות, $A\in\mathbb{R}^{k\times m}$ מטריצה קבועה ו-$Z\in\mathbb{R}^m$ וקטור מקרי שבו $Z_1,\dots,Z_m\overset{iid}{\sim}N(0,1)$.

תזכורת 2.1 - אם W מתפלג רב-נורמלית כלומר $W=\mu+AZ$ אזי:

$\mathbb{E}[W]=\mu$
$Cov[W]=AA^T$

טענה 2.2 – פונקציה יוצרת מומנטים של W הינה:

$$M_W(t)=\exp\left\{\mu^Tt+\frac{1}{2}t^TAA^Tt\right\}\overset{V=AA^T}{=}\exp\left\{\mu^Tt+\frac{1}{2}t^TVt\right\}$$

נזכור כי לשני וקטורים מקריים בעלי אותה פונקציית יוצרת מומנטים יש אותה התפלגות. כתוצאה מכך שהפונקציה יוצרת המומנטים של התפלגות רב-נורמלית תלויה רק ב-μ ו-V אנחנו מגיעים למסקנה הבאה:

טענה 2.3 – אם

$W_1=\mu+A_1Z$
$W_2=\mu+A_2Z$
$A_1A_1^T=A_2A_2^T=V$

אזי ההתפלגות של W_1 ו-W_2 זהה:

$$W_1 \sim W_2 \sim N(\mu, V)$$

טענה 2.4 – פונקציית צפיפות של התפלגות רב-נורמלית – אם $W \sim N_d(\mu, V)$ ו-V הפיכה אזי פונקציית הצפיפות של W הינה:

$$f_W(x) = (2\pi)^{-d/2} \cdot |V|^{-1/2} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^T V^{-1}(x-\mu)\right\}$$

טענה 2.5 – אם $W \sim N(\mu, V)$ כאשר $W \in \mathbb{R}^d$ וקיים וקטור קבוע $c \in \mathbb{R}^d$ אזי:

$$c^T W \sim N_d(c^T\mu, c^T V c)$$

טענה 2.6 - $W \sim N_d(\mu, V)$ אם ורק אם לכל $c \in \mathbb{R}^d$ מתקיים

$$c^T W \sim N(c^T\mu, c^T V c)$$

טענה 2.7 (מקרה פרטי של 2.6) – אם $W \sim N_d(\mu, V)$ אזי לכל $1 \le r \le d$ מתקיים $W_r \sim N(\mu_r, V_{rr})$

טענה 2.8 – אם $W \sim N(\mu, V)$ ו-C מטריצה קבועה אזי

$$CW \sim N(C\mu, CVC^T)$$

טענה 2.9 – אם $W \sim N(\mu, V)$ ו-$W_2 = \begin{pmatrix} w_{r+1} \\ \cdots \\ w_d \end{pmatrix}$, $W_1 = \begin{pmatrix} w_1 \\ \cdots \\ w_r \end{pmatrix}$ והחלוקה של μ בהתאמה. כמו כן V הפיכה והיא מטריצת בלוקים מהצורה:

$$V = \begin{pmatrix} V_{11} \in \mathbb{R}^{r\times r} & V_{12} \in \mathbb{R}^{r\times(d-r)} \\ V_{21} \in \mathbb{R}^{(d-r)\times r} & V_{22} \in \mathbb{R}^{(d-r)\times(d-r)} \end{pmatrix}$$

אזי ההתפלגות המותנית של W_2 בהינתן W_1 היא:

$$W_2 | W_1 \sim N(\mu_2 + V_{21}V_{11}^{-1}(W_1 - \mu_1), V_{22} - V_{21}V_{11}^{-1}V_{12})$$

במקרה הפרטי שבו $Cov(W_1, W_2) = 0$ אזי: $V_{12} = 0$ ובמקרה זה:

$$W_2 | W_1 \sim N(\mu_2, V_{22})$$

והתפלגות זו לא תלויה ב-W_1 כלומר $W_1 \perp W_2$.

בהכללה – בהתפלגות רב-נורמלית שמורכבת משני וקטורים בעלי שונות משותפת אפס הם בלתי-תלויים. דבר זה נכון רק לגבי התפלגות רב-נורמלית.

הגדרה 2.10 – התפלגות חי-בריבוע χ^2 – נניח כי $Z_1, \dots, Z_d \sim N(0,1)$ מדגם מקרי בהתפלגות נורמלית סטנדרטית. נגדיר

$$Q = \sum_{i=1}^{d} Z_i^2$$

אזי Q מתפלג חי-בריבוע עם d דרגות חופש ומסמנים $Q \sim \chi_d^2$.

הגדרה 2.11 – התפלגות t – נניח כי $Z \sim N(0,1)$ ו-$Q \sim \chi_d^2$ ו-$Z \perp Q$. נגדיר

$$T = \frac{Z}{\sqrt{Q/d}}$$

אזי T מתפלג t עם d דרגות חופש $T \sim t_d$.

משפט 2.12 – נניח $W \sim N_d(0,V)$ כאשר V הפיכה. נגדיר $Q = W^T V^{-1} W$. אזי $Q \sim \chi_d^2$.

הוכחה 2.12 – נציג את V באמצעות הפירוק הספקטראלי

$$V = U\Lambda U^T$$

כאשר נזכור כי $U^T U = UU^T = I$ מטריצה אורתוגונלית ואילו Λ מטריצה אלכסונית כאשר באלכסון מופיעים הערכים העצמיים של V.

עתה, מכיוון ש-V הפיכה הערכים העצמיים $\lambda_1, \dots, \lambda_d > 0$ חיוביים ממש. לכן ניתן להגדיר:

$$\Lambda = diag\{\lambda_1, \dots, \lambda_d\} \Rightarrow \Lambda^{-1} = diag\left\{\lambda_1^{-1}, \dots, \lambda_d^{-1}\right\}$$

לכן

$$(U\Lambda^{-1}U^T)V = U\Lambda^{-1}U^T U\Lambda U^T = U\Lambda^{-1} I \Lambda U^T = U\Lambda^{-1}\Lambda U^T = UIU^T = UU^T = I$$

מכאן יוצא שהמטריצה ההופכית של V היא:

$$V^{-1} = U\Lambda^{-1}U^T$$

כלומר אם נציב בנתון:

$$Q = W^T V^{-1} W = W^T U\Lambda^{-1} U^T W = W^T U \Lambda^{-1/2} \Lambda^{-1/2} U^T W = \widetilde{W}^T \widetilde{W}$$

כאשר

$$\widetilde{W} = \Lambda^{-1/2} U^T W$$

אם נסמן $B = \Lambda^{-1/2} U^T$ אזי

$$\widetilde{W} = BW \sim N(0, BVB^T)$$

ומטריצת השונויות היא:

$$BVB^T = \Lambda^{-1/2} U^T V U \Lambda^{-1/2} = \Lambda^{-1/2} U^T U \Lambda U^T U \Lambda^{-1/2} = \Lambda^{-1/2} \Lambda \Lambda^{-1/2} = I$$

כלומר

$$\widetilde{W} \sim N(0, I)$$

ולכן

$$Q = \widetilde{W}^T \widetilde{W} = \sum_{i=1}^{d} \widetilde{W}_i^2 \sim \chi_d^2$$

וראינו כי ההתפלגות של סכום ריבועי משתנים מקריים מתפלגים נורמלי סטנדרטי היא התפלגות חי-בריבוע כנדרש.

משפט 2.13 – נניח $W \sim N(0, I)$ ו-P מטריצה סימטרית ואידמפוטנטית ו-$rank(P) = k$ אזי

$$||PW||^2 \sim \chi_k^2$$

הוכחה 2.13

$$||PW||^2 = W^T P^T P W = W^T P W$$

כיוון ש-P סימטרית יש פירוק ספקטראלי

$$P = U\Lambda U^T$$

מכיוון ש-$P^2 = P$ כל הערכים העצמיים של P שווים ל-1 או 0 ואז נקבל:

$$W^T P W = \sum_{i=1}^{k} W_i^2 \sim \chi_k^2$$

עד עתה עברנו על נושאים מוכרים, מעתה נתחיל את הנושא הראשון בקורס – הסקה סטטיסטית עבור וקטור התוחלות של וקטור מקרי שמפולג רב-נורמלית.

נושא 1.1 –התפלגות רב-נורמלית – אומד נראות מרבית

משפט 2.14 אומד נראות מרבית לווקטור התוחלות של וקטור מקרי מתפלג רב-נורמלית – נניח $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, V)$ ורוצים לעשות הסקה סטטיסטית. נניח כי V ידוע. נאמוד את μ לפי שיטת נראות מרבית.

$$L(\mu) = \prod_{i=1}^{n} f_{X_i}(X_i) = \prod_{i=1}^{n} 2\pi^{-d/2} \det(V)^{-1/2} \exp\left\{-\frac{1}{2}(X_i - \mu)^T V^{-1}(X_i - \mu)\right\}$$
$$= (2\pi)^{-nd/2} \det(V)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^T V^{-1}(X_i - \mu)\right\}$$

לוג-הנראות:

$$l(\mu) = \log\big(L(\mu)\big) = \log\left((2\pi)^{-nd/2}\, det(V)^{-n/2}\, exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^T V^{-1}(X_i - \mu)\right\}\right)$$
$$= \log\big((2\pi)^{-nd/2}\, det(V)^{-n/2}\big) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^T V^{-1}(X_i - \mu)$$

נסמן $B = V^{-1}$ ו-$H = \log\big((2\pi)^{-nd/2}\, det(V)^{-n/2}\big)$

$$= H - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^T B(X_i - \mu)$$
$$= H - \frac{1}{2}\sum_{i=1}^{n}\big(X_i^T B X_i - 2X_i^T B\mu + \mu^T B\mu\big)$$
$$= H - \frac{1}{2}\left[\sum_{i=1}^{n} X_i^T B X_i - 2\left(\sum_{i=1}^{n} X_i\right)^T B\mu + n\mu^T B\mu\right]$$
$$= H - \frac{1}{2}\sum_{i=1}^{n} X_i^T B X_i + n\bar{X}B\mu - \frac{n}{2}\mu^T B\mu$$

כאשר

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

כעת, על מנת למצוא את אומד הנראות המרבית של $\vec{\mu}$, נגזור את לוג-הנראות לפי כל אחת מהקואורדינטות של $\vec{\mu}$. לקואורדינטה $1 \le r \le d$ כלשהי נשווה ל-0 נקבל:

$$\frac{\partial l}{\partial \mu_r} = n\frac{\partial}{\partial \mu_r}\bar{X}^T B\mu - \frac{n}{2}\cdot\frac{\partial}{\partial \mu_r}\mu^T B\mu = 0$$

העמקה אלגברה 3 – נגזרות חלקיות של וקטורים מקריים

תזכורת 2.15 – נגזרות חלקיות של וקטורים מקריים -נזכור כי באופן כללי עבור וקטור מקרי כלשהו $w \in \mathbb{R}^d$ ווקטור קבוע a מתקיים:

$$\frac{\partial}{\partial w_r}a^T w = \frac{\partial}{\partial w_r}\sum_{j=1}^{d} a_j w_j = \sum_{j=1}^{d} a_j \cdot \frac{\partial w_j}{\partial w_r} = a^T\delta_{jr} = a^T\begin{cases}1, & j=r\\ 0, & j\neq r\end{cases}$$

$$\frac{\partial}{\partial w_r}w^T B w = \frac{\partial}{\partial w_r}\sum_{j=1}^{d}\sum_{k=1}^{d}(w^T)_j B_{jk} w_k = \frac{\partial}{\partial w_r}\sum_{j=1}^{d}\sum_{k=1}^{d} B_{jk} w_j w_k = \sum_{j=1}^{d}\sum_{k=1}^{d} B_{jk}\frac{\partial}{\partial w_r}\left[w_j w_k\right]$$
$$= \sum_{j=1}^{d}\sum_{k=1}^{d} B_{jk}\left[\frac{\partial w_k}{\partial w_r}w_j + \frac{\partial w_j}{\partial w_r}w_k\right]$$
$$= \sum_{j=1}^{d}\sum_{k=1}^{d} B_{jk}\left[\delta_{kr}w_j + \delta_{jr}w_k\right]$$
$$= \sum_{j=1}^{d}\sum_{k=1}^{d} B_{jk}\delta_{kr}w_j + \sum_{j=1}^{d}\sum_{k=1}^{d} B_{jk}\delta_{jr}w_k$$
$$= \sum_{k=1}^{d} B_{rk}w_k + \sum_{j=1}^{d} B_{jr}w_j$$
$$= (Bw)_r + (B^T w)_r$$
$$= [(B+B^T)w]_r$$

המשך הוכחה 2.14 – כזכור אנחנו רוצים להשוות ל-0 את לוג הנראות המרבית של וקטור התוחלות בהתפלגות רב-נורמלית. קיבלנו:

$$\frac{\partial l}{\partial \mu_r} = n\frac{\partial}{\partial \mu_r}\bar{X}^T B\mu - \frac{n}{2}\cdot\frac{\partial}{\partial \mu_r}\mu^T B\mu = 0$$

לפי התזכורת נקבל:

$$n\frac{\partial}{\partial \mu_r}\bar{X}^T B\mu - \frac{n}{2}\cdot\frac{\partial}{\partial \mu_r}\mu^T B\mu$$
$$= n(\bar{X}^T B)_r - \frac{n}{2}[(B+B^T)\mu]_r$$

אצלנו $B = V^{-1}$ מטריצה סימטרית לכן:

$$= n(\bar{X}^T B)_r - n(B\mu)_r$$

נשים לב כי $B \in \mathbb{R}^{d\times d}$ ו-$X, X^T, \tilde{X}, \tilde{X}^T, \mu \in \mathbb{R}^d$ כלומר יש לנו d משוואות ואנחנו צריכים לפתור את מערכת המשוואות:

$$\forall 1 \le r \le d \coloneqq n(\bar{X}^T B)_r - n(B\mu)_r = 0 \Rightarrow$$
$$\forall 1 \le r \le d \coloneqq [B\mu]_r = (\bar{X}^T B)_r \xRightarrow{vector-wise}$$
$$\forall 1 \le r \le d \coloneqq B\mu = B^T\bar{X} \xRightarrow{B^T=B}$$

$\forall 1 \le r \le d \coloneqq B\mu = B\bar{X} \xRightarrow{multiply\ left\ B^{-1}}$
$\forall 1 \le r \le d \coloneqq B^{-1}B\mu = B^{-1}B\bar{X} \Rightarrow$
$\hat{\mu} = \bar{X}$

שיעור 3

תזכורת 3.1 - נניח $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, V)$ ורוצים לעשות הסקה סטטיסטית. נניח כי V ידוע. נאמוד את μ לפי שיטת נראות מרבית. בפעם שעברה ראינו כי לפי שיטת נראות מרבית מגיעים למסקנה $\mu = \bar{X}$.

נושא 1.2 – התפלגות רב-נורמלית – מבחן יחס נראות

טענה 3.2 – מבחן יחס נראות עבור השערת אפס לווקטור תוחלות – אם נניח $H_0: \vec{\mu} = 0$ מול האלטרנטיבה $H_1: \vec{\mu} \neq 0$ אז נוכל לבדוק לפי מבחן יחס הנראות.

כמו בשיעור 2, נגדיר

$$\tilde{H} = H - \frac{1}{2}\sum_{i=1}^{n} X_i^T B X_i$$

כאשר

$$H = \log\left((2\pi)^{-nd/2}|V|^{-n/2}\right)$$
$$B = V^{-1}$$

לפיכך לפי מבחן יחס הנראות נקבל:

$$2[l(\hat{\mu}) - l(\mu_0)] = 2[l(\hat{\mu}) - l(0)]$$
$$= 2\left[\left(\tilde{H} + n\bar{X}^T V^{-1}\hat{\mu} - \frac{n}{2}\mu^T V^{-1}\mu\right)\right] - 2\tilde{H}$$
$$= 2\left[n\bar{X}^T V^{-1}\hat{\mu} - \frac{n}{2}\hat{\mu}^T V^{-1}\hat{\mu}\right]$$
$$\overset{\hat{\mu}=\bar{X}}{=} 2\left[n\bar{X}^T V^{-1}\bar{X} - \frac{n}{2}\bar{X}^T V^{-1}\bar{X}\right]$$
$$= \boldsymbol{n\bar{X}^T V^{-1}\bar{X}}$$

טענה 3.3 התפלגות הסטטיסטי של מבחן יחס הנראות – כזכור $X_i \sim N(\mu, V)$ ו-$X_1, \dots, X_n$ בלתי-תלויים. מתקיים:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\begin{pmatrix} I_d & \dots & I_d \end{pmatrix}\begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}$$

כאשר:

$$\begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \dots \\ \mu \end{pmatrix}, \begin{pmatrix} V & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V \end{pmatrix}\right)$$

כלומר וקטור התצפיות מתפלג רב-נורמלית עם וקטור תוחלות שבו כל רכיב הוא וקטור התוחלות של ההתפלגות של כל אחת מהתצפיות, ומטריצת השונויות היא מטריצת בלוקים שבה כל בלוק הוא מטריצת השונויות V. לכן באשר לממוצע מתקיים (לפי טענה 2.8):

$$\bar{X} \sim N_d\left(\mu, \frac{1}{n}V\right) \overset{H_0}{\sim} N\left(0, \frac{1}{n}V\right)$$

מסקנה 3.4 – סטטיסטי יחס הנראות של התפלגות רב-נורמלית

נגדיר

$$\mathbb{V}(aX) = a^2X$$

$$W = \sqrt{n}\bar{X}\overset{H_0}{\sim}\sqrt{n}N\left(0, \frac{1}{n}V\right) = N\left(0, \frac{1}{n}V \cdot \left(\sqrt{n}\right)^2\right) = N\left(0, \frac{1}{n} \cdot nV\right) = N(0, V)$$

סטטיסטי יחס הנראות הינו (לפי 3.2 ו-3.4 וממשפט 2.12):

$$\bar{X}^TV^{-1}\bar{X} = \frac{n}{n}\bar{X}^TV^{-1}\bar{X} = \frac{1}{n}\sqrt{n}\bar{X}^TV^{-1}\sqrt{n}\bar{X} = \frac{1}{n}W^TV^{-1}W \sim \chi_d^2$$

ראינו כי ההתפלגות במצב כזה היא התפלגות χ_d^2 עם d דרגות חופש.

מסקנה 3.5 – עבור רמת מובהקות כלשהי α, כלל הדחיה: לדחות את H_0 אם:

$$W^TV^{-1}W \geq \chi^2_{d;(1-\alpha)}$$

טענה 3.6 – במקרה כללי יותר $\mu = \mu^0$ כאשר μ^0 קבוע נרצה למצוא את סטטיסטי המבחן ואת ההתפלגות שלו. נגדיר:

$$\widetilde{X_\iota} = X_i - \mu^0$$

לכן

$$\widetilde{X_\iota}, \dots, \widetilde{X}_n \sim N(\mu - \mu^0, V)$$

במקרה כזה השערת האפס היא $H_0: \mathbb{E}[\widetilde{X}_i] = 0$ באופן שדומה למקרה הקודם שראינו.

לכן כלל הדחיה במקרה זה יהיה:

$$D \geq \chi^2_{d;(1-\alpha)}$$

כאשר

$$D = n\bar{\tilde{X}}^TV^{-1}\bar{\tilde{X}} = n(\bar{X} - \mu^0)^TV^{-1}(\bar{X} - \mu^0)$$

הערה 3.7 – **הכללה של משפט הגבול המרכזי למקרה הרב-ממדי** - נזכור כי לפי משפט הגבול המרכזי הקלאסי מתקיים כי אם $Y_1, \dots, Y_n$ בלתי תלויים ושווי התפלגות עם תוחלת $\mathbb{E}[Y_i] = \mu$ ושונות $\mathbb{V}(Y_i) = \sigma^2$, מתקיים:

$$\sqrt{n}(\bar{Y} - \mu) \overset{d}{\to} N(0, \sigma^2)$$

באופן דומה במקרה הרב-ממדי, אם $X_1, \dots, X_n$ בלתי-תלויים שווי-התפלגות בעלי תוחלת $\mathbb{E}[X_i] = \vec{\mu}$ ומטריצת שונויות $Cov[X_i] = V$ גם כאן מתקיים:

$$\sqrt{n}(\bar{X} - \mu) \overset{d}{\to} N(\vec{0}, V)$$

לכן עבור מדגם מספיק גדול נוכל להשתמש בסטטיסטי המבחן שראינו קודם כך שיתפלג גם כן חי-בריבוע. נציין כי במקרה של נתונים שאינם מפולגים רב-נורמלית, אין לסטטיסטי D אינטרפרטציה כסטטיסטי יחס נראות

נושא 1.3 – התפלגות רב-נורמלית - אזור סמך לאומד נראות מרבית

הערה 3.8 – **אזור סמך לווקטור התוחלות** - נוכל להפוך את המבחן לאזור סמך עבור μ המוגדר בתור קבוצת כל הערכים של μ^0 עבורם לא דוחים את $H_0: \mu = \mu^0$.

$$\left\{\mu^0 : (\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) \leq \frac{1}{n}\chi^2_{d;(1-\alpha)}\right\}$$

לפי הפירוק הספקטראלי מתקיים $V = U\Lambda U^T \Rightarrow V^{-1} = U\Lambda^{-1}U^T$ לכן:

$$(\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) = (\bar{X}-\mu^0)^T U\Lambda^{-1}U^T(\bar{X}-\mu^0) \overset{\eta=U^T\bar{X}-U^T\mu^0}{=} \eta^T\Lambda^{-1}\eta = \sum_{i=1}^{d}\frac{\eta_i^2}{\lambda_i^2}$$

ואזור הסמך הוא:

$$\left\{\eta : \eta^T\Lambda^{-1}\eta \leq \frac{1}{n}\chi^2_{d;(1-\alpha)}\right\}$$

במקרה שבו $d = 2$ אנחנו מסתכלים על אזור שתחום באי-השוויון:

$$\frac{\eta_1^2}{\lambda_1} + \frac{\eta_2^2}{\lambda_2} = a$$

זוהי למעשה אליפסה כפונקציה של $\vec{\eta} = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}$.

אם נגדיר באופן כללי

$$S(a) = \{\eta : \eta^T\Lambda^{-1}\eta \leq a\}$$

אזי כיוון שמתקיים $\eta = U\eta$ נשים לב גם כי מתקיים:

$$\{\mu^0 : (\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) \leq a\} = \{\mu^0 : U(\bar{X}-\mu^0) \in S(a)\} = \{\mu^0 : (U\bar{X}-U\mu^0) \in S(a)\}$$

הערה 3.9 – המהות של הפעולה $v \to Uv$. ניקח שני ווקטורים v_1, v_2. נסתכל על:

$$\widetilde{v_1} = Uv_1$$
$$\widetilde{v_2} = UV_2$$

ואז:

$$\|\tilde{v}_1\|^2 = \tilde{v}_1^T\tilde{v}_1 = (Uv_1)^T(Uv_1) = v_1^T U^T U v_1 = v_1^T I v_1 = v_1^T v_1 = \|v_1\|^2$$

כלומר הכפלה במטריצה U האורתוגונלית שומרת על הגודל (הנורמה) של הווקטור. באופן דומה:

$$\|\tilde{v}_2\|^2 = \|v_2\|^2$$

ולכן מתקיים גם השוויון:

$$\tilde{v}_1^T\tilde{v}_2 = v_1^T v_2$$

נזכור כי מתקיים

$$v_1^T v_2 = \|v_1\|\|v_2\|\cos(\theta)$$

ולכן:

$$\cos(\tilde{\theta}) = \frac{\tilde{v}_1^T\tilde{v}_2}{\|\tilde{v}_1\|\|\tilde{v}_2\|} = \tilde{v}_1^T\tilde{v}_2 = v_1^T v_2 = \|\tilde{v}_1\|\|\tilde{v}_2\|\cos(\tilde{\theta}) = \cos(\theta)$$

כלומר $\tilde{\theta} = \theta$, כלומר המכפלה שומרת גם על גודל וגם על זווית. זאת אומרת שהפעולה של מכפלה ב-U הינה סיבוב, שיקוף או שניהם.

טענה 3.10 – אם נסתכל על הביטוי שפיתחנו קודם

$(\bar{X}-\mu^0)^T U\Lambda^{-1}U^T(\bar{X}-\mu^0)$

נוכל לקבל את הפירוק הבא:

$$(\bar{X}-\mu^0)^T U\Lambda^{-1}U^T(\bar{X}-\mu^0) = \left(\Lambda^{-1/2}U\bar{X}-\Lambda^{-1/2}U\mu^0\right)^T\left(\Lambda^{-1/2}U\bar{X}-\Lambda^{-1/2}U\mu^0\right)$$

נציב

$$c_1 = \Lambda^{-1/2}U\bar{X}$$
$$c_2 = \Lambda^{-1/2}U\mu^0$$

ונקבל

$$= (c_1-c_2)^T(c_1-c_2)$$

לפיכך נוכל לחשוב על $(\bar{X}-\mu^0)V^{-1}(\bar{X}-\mu^0)$ כעל מרחק מוכלל המכונה **מרחק Mahalanobis.**

נושא 1.4.1 – התפלגות רב-נורמלית – עוצמת מבחן

הקדמה 3.11 – עוצמה סטטיסטית של מבחן סטטיסטי לווקטור תוחלות

$$P(\mu) = P_\mu(\text{דחיה}) = P_\mu\left((\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) \geq \frac{1}{n}\chi^2_{d;(1-\alpha)}\right)$$

אז צריך לחשב את ההתפלגות של $(\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0)$

כזכור ראינו כי $\bar{X}\sim N\left(\mu,\frac{1}{n}V\right)$ לכן אם נגדיר $\Delta = \mu-\mu^0$:

$$\bar{X}\sim N\left(\mu,\frac{1}{n}V\right) \Rightarrow (\bar{X}-\mu^0)\sim N\left(\mu-\mu^0,\frac{1}{n}V\right) \xRightarrow{\Delta=\mu-\mu^0}$$
$$(\bar{X}-\mu^0)\sim N\left(\Delta,\frac{1}{n}V\right)$$

לכן עבור $\Delta^* = \sqrt{n}\Delta = \sqrt{n}(\mu-\mu^0)$ נקבל:

$$(\bar{X}-\mu^0)\sim N\left(\Delta,\frac{1}{n}V\right) \Rightarrow \sqrt{n}(\bar{X}-\mu^0)\sim N\left(\sqrt{n}\Delta,V\right) \Rightarrow$$
$$\sqrt{n}(\bar{X}-\mu^0)\sim N(\Delta^*,V)$$

נתחיל עם המקרה $V=I$.

$$\sqrt{n}(\bar{X}-\mu^0)\sim N(\Delta^*,I) \xrightarrow{W=\sqrt{n}(\bar{X}-\mu^0)} W\sim N(\Delta^*,I) \xrightarrow{(\widetilde{W}=W-\Delta^*)} \widetilde{W}\sim N(0,I)$$

נסמן

$$H = \left(\widetilde{W}+\Delta^*\right)^T\left(\widetilde{W}+\Delta^*\right) = W^T W$$

טענה 3.11.1 - נטען כי ההתפלגות של H תלויה ב-Δ^* רק דרך $\|\Delta^*\|$

הוכחה 3.11.1 – נגדיר מטריצה U כך שהעמודה הראשונה של U הינה $u_1 = \frac{\Delta^*}{\|\Delta^*\|}$ ושאר העמודות $u_2,\dots,u_d$ מהווה בסיס אורתונורמלי עבור $\{Span(u_1)\}^\perp$ כך שבסך הכל המטריצה U היא מטריצה אורתונורמלית. אזי

$$H = \left(\widetilde{W}+\Delta^*\right)^T\left(\widetilde{W}+\Delta^*\right) = \left(\widetilde{W}+\Delta^*\right)^T UU^T\left(\widetilde{W}+\Delta^*\right) = \left(U\widetilde{W}+U\Delta^*\right)^T\left(U\widetilde{W}+U\Delta^*\right)$$

נזכור ש-$\widetilde{W}\sim N(0,I)$ לכן $U\widetilde{W}\sim N(0,UU^T) = N(0,I)$. למעשה:

$$U^T\Delta^* = \begin{pmatrix} -u_1^T - \\ \dots \\ -u_d^T - \end{pmatrix} \|\Delta^*\| u_1$$

שיעור 4

תזכורת 4.1 – התחלנו לעסוק במקרים של הסקה סטטיסטית עבור מדגם מקרי פשוט מהתפלגות רב-נורמלית כלשהי. בהינתן n תצפיות $X_1,\dots,X_n$ שכל אחת מהן היא וקטור $X_i \in \mathbb{R}^d$ אשר מקיימות $X_i \overset{iid}{\sim} N_d(\mu,V)$ לכל $1 \le i \le n$ ראינו כי אומד הנראות המרבית הינו $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$. במקרה של השערת אפס $H_0: \mu = \mu^0$ ראינו כי מתקיים $n(\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) \sim \chi_d^2$ ולכן כלל הדחיה להשערת האפס הוא שדוחים את השערת האפס אם מתקיים $n(\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) > \chi^2_{d;(1-\alpha)}$. לבסוף ראינו כי עוצמת המבחן היא $P(\mu) = P_\mu(\text{דחיה}) = P_\mu\left(n(\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) \ge \chi^2_{d;(1-\alpha)}\right)$.

הוכחת עוצמת המבחן 4.2 (המשך משיעור קודם) – בשיעור הקודם התחלנו להוכיח את עוצמת המבחן עבור המקרה $V = I$. הגדרנו עבור μ^* אלטרנטיבי:

$$Z = \sqrt{n}(\bar{X}-\mu^*) \overset{H_1}{\sim} N(0,I)$$
$$W = \sqrt{n}(\bar{X}-\mu^0) \overset{H_0}{\sim} N(0,I)$$

$$\Delta^* = \sqrt{n}\,(\mu^* - \mu^0)$$

טענו בשיעור הקודם כי ההתפלגות W^TW תלויה ב-Δ^* רק באמצעות הנורמה $\|\Delta^*\|$. בפעם שעברה הגדרנו גם:

$$\tilde{Z} = U^T Z \Rightarrow \tilde{Z} \sim N(0, U^T Cov(Z) U) \sim N(0, U^T U) \sim N(0,I)$$

כאשר U מוגדר כפי שראינו בשיעור הקודם: מטריצה U כך שהעמודה הראשונה של U הינה $u_1 = \frac{\Delta^*}{\|\Delta^*\|}$ ושאר העמודות $u_2,\dots,u_d$ מהווה בסיס אורתונורמלי עבור $\{Span(u_1)\}^\perp$ כך שבסך הכל המטריצה U היא מטריצה אורתונורמלית.

ואז מתקיים

$$\begin{aligned}
&W^TW = n(\bar{X}-\mu^0)^T(\bar{X}-\mu^0) = \\
&n(\bar{X}-\mu^*+\mu^*-\mu^0)^T(\bar{X}-\mu^*+\mu^*-\mu^0) = \\
&(Z+\Delta^*)^T(Z+\Delta^*) = \\
&(Z+\Delta^*)^T U U^T (Z+\Delta^*) = \\
&(U^TZ+U^T\Delta^*)^T(U^TZ+U^T\Delta^*) = \\
&\left(\tilde{Z}+U^T\Delta^*\right)^T\left(\tilde{Z}+U^T\Delta^*\right)
\end{aligned}$$

בשיעור הקודם ראינו כי

$$U^T\Delta^* = \begin{pmatrix} \frac{\Delta^{*T}}{\|\Delta^*\|} \\ \dots \\ u_d^T \end{pmatrix}\Delta^* = \begin{pmatrix} \frac{\Delta^{*T}\Delta^*}{\|\Delta^*\|} \\ \dots \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\|\Delta^*\|^2}{\|\Delta^*\|} \\ \dots \\ 0 \end{pmatrix} = \|\Delta^*\| \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

ולכן

$$W^TW = \left(\tilde{Z}+U^T\Delta^*\right)^T\left(\tilde{Z}+U^T\Delta^*\right) =$$

$$\left(\tilde{Z} + \|\Delta^*\| \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}\right)^T \left(\tilde{Z} + \|\Delta^*\| \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}\right) =$$

$$\begin{pmatrix} \tilde{Z}_1 + \|\Delta^*\| \\ \tilde{Z}_2 \\ \dots \\ \tilde{Z}_d \end{pmatrix}^T \begin{pmatrix} \tilde{Z}_1 + \|\Delta^*\| \\ \tilde{Z}_2 \\ \dots \\ \tilde{Z}_d \end{pmatrix}$$

$$(\tilde{Z}_1 + \|\Delta^*\| \quad \widetilde{Z_2} \quad \dots \quad \tilde{Z}_d) \begin{pmatrix} \tilde{Z}_1 + \|\Delta^*\| \\ \tilde{Z}_2 \\ \dots \\ \tilde{Z}_d \end{pmatrix} = \left(\tilde{Z}_1 + \|\Delta^*\|\right)^2 + \sum_{j=2}^{d} \tilde{Z}_j^2$$

כלומר W^TW הוא פונקציה של Δ^* רק מתוך $\|\Delta^*\|$ כנדרש.

העמקה הסתברות 2 – התפלגות חי-בריבוע לא מרכזית

הגדרה 4.3 – התפלגות חי-בריבוע לא-מרכזי - ההתפלגות של W^TW נקראת התפלגות חי-בריבוע לא מרכזי עם פרמטר אי-מרכזיות $\|\Delta^*\|^2$. ההתפלגות מסומנת על-ידי $\chi_d^2(\|\Delta^*\|^2)$.

עתה נתבונן בתכונות של התפלגות חי-בריבוע לא-מרכזי

טענה 4.4 – אם $Y\sim\chi_d^2(\lambda)$ אז נוכל להציג את Y בצורה $Y = Y_1 + Y_2$ כאשר $Y_1\sim\chi_1^2(\lambda)$ ו-$Y_2\sim\chi_{d-1}^2$ (רגיל) והמשתנים בלתי-תלויים $Y_1 \perp Y_2$. טענה זו נובעת ישירות מהפירוק שהגענו אליו בסעיף 4.2 שמראה את ההתפלגות כסכום של שני משתנים מקריים.

טענה 4.5 – פונקציית צפיפות של התפלגות חי-בריבוע לא מרכזית - כתוצאה מסעיף 4.4 נוכל לכתוב:

$$f_Y(y) = \int_0^\infty f_{Y_1}(y_1) f_{Y_2}(y - y_1) dy_1 = \int_0^\infty f_{\chi_1^2(\lambda)}(y_1) f_{\chi_{d-1}^2}(y - y_1) dy_1$$

ובדומה פונקציית הצפיפות המצטברת:

$$F_Y(y) = \int_0^\infty f_{Y_1}(y_1) \mathbb{P}(Y = y | Y_1 = y_1) dy_1 = \int_0^\infty f_{Y_1}(y_1) \mathbb{P}(Y_2 \le y - y_1) dy_1$$

טענה 4.6 – עבור $\lambda = 0$ ישנה זהות בין התפלגות חי-בריבוע לא-מרכזית להתפלגות חי-בריבוע רגילה:

$$\chi_d^2(0) = \chi_d^2$$

טענה 4.7 – **תוחלת התפלגות חי-בריבוע לא מרכזית**

$$\mathbb{E}\left[\chi_d^2(\lambda)\right] = d + \lambda$$

הוכחה 4.7 – נסתכל על הפירוק שראינו

$$Y = Y_1 + Y_2$$
$$\tilde{Z}\sim N(0, I)$$
$$Y_1 = \left(\tilde{Z}_1 + \lambda\right)^2 \sim \chi_d^2(\lambda)$$
$$Y_2 \sim \chi_{d-1}^2 \to \mathbb{E}[Y_2] = d - 1$$

לכן:

$$\mathbb{E}[Y_1] = \mathbb{E}\left[\left(\tilde{Z}_1 + \lambda\right)^2\right] = \mathbb{E}[\tilde{Z}_1^2] + 2\mathbb{E}\left[\tilde{Z}_1\sqrt{\lambda}\right] + \mathbb{E}[\lambda] = 1 + \sqrt{\lambda}\mathbb{E}[\tilde{Z}_1] + \lambda = 1 + 0 + \lambda = 1 + \lambda$$

$$\mathbb{E}[Y]=\mathbb{E}[Y_1+Y_2]=\mathbb{E}[Y_1]+\mathbb{E}[Y_2]=1+\lambda+d-1=\lambda+d$$

טענה 4.8 – נניח $\lambda_2>\lambda_1$

$$Y_1\sim\chi_d^2(\lambda_1)$$
$$Y_2\sim\chi_d^2(\lambda_2)$$

אזי מתקיים אי-שוויון סטוכסטי $Y_2\leq_s Y_1$ כלומר לכל y מתקיים

$$\mathbb{P}(Y_2>y)>\mathbb{P}(Y_1>y)$$

טענה 4.9 – נניח $Z\sim N(\theta,I)$ כאשר P מטריצת היטל בעלת דרגה $rank(P)=m$. אזי

$$\|PZ\|^2\sim\chi_m^2(\|\theta\|^2)$$

הוכחה דומה להוכחה על המשפט הרגיל – באמצעות הפירוק הספקטראלי.

טענה 4.10 – עבור $d=1$

$$\mathbb{P}(\chi_1^2(\lambda)\leq x)=\mathbb{P}\left(\left(\tilde{Z}_1+\sqrt{\lambda}\right)^2\leq x\right)=\mathbb{P}\left(\tilde{Z}_1+\sqrt{\lambda}\leq\sqrt{x}\right)=\mathbb{P}\left(-\sqrt{x}\leq\tilde{Z}_1+\sqrt{\lambda}\leq\sqrt{x}\right)$$
$$=\mathbb{P}\left(-\sqrt{x}-\sqrt{\lambda}\leq\tilde{Z}_1\leq\sqrt{x}+\sqrt{\lambda}\right)=\Phi\left(\sqrt{x}+\sqrt{\lambda}\right)-\Phi\left(-\sqrt{x}-\sqrt{\lambda}\right)$$

לכן:

$$f_{\chi_1^2(\lambda)}(x)=\frac{d}{dx}F_{\chi_1^2(\lambda)}(x)=\frac{d}{dx}\left[\Phi\left(\sqrt{x}-\sqrt{\lambda}\right)-\Phi\left(-\sqrt{x}-\sqrt{\lambda}\right)\right]$$
$$=\frac{1}{2}\cdot\frac{1}{\sqrt{x}}\Phi\left(\sqrt{x}-\sqrt{\lambda}\right)+\frac{1}{2}\cdot\frac{1}{\sqrt{x}}\Phi\left(-\sqrt{x}-\sqrt{\lambda}\right)$$
$$=\frac{1}{2\sqrt{x}}\cdot\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\sqrt{x}+\sqrt{\lambda}\right)^2\right\}+\frac{1}{2\sqrt{x}}\cdot\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(-\sqrt{x}-\sqrt{\lambda}\right)^2\right\}$$
$$=\frac{1}{2\sqrt{2\pi x}}\left[\exp\left\{-\frac{1}{2}\left(\sqrt{x}+\sqrt{\lambda}\right)^2\right\}+\exp\left\{-\frac{1}{2}\left(-\sqrt{x}-\sqrt{\lambda}\right)^2\right\}\right]$$

טענה 4.11 – צפיפות חי-בריבוע לא-מרכזי – כביטוי של התפלגות פואסון. עבור $D\sim Poisson\left(\frac{1}{2}\lambda\right)$.

$$f_{\chi_1^2(\lambda)}(x)=\sum_{j=0}^{\infty}\mathbb{P}(D=j)\cdot f_{\chi_{2d+1}^2}(x)$$

הוכחת הטענה – באמצעות פיתוח טיילור על הביטויים

$$\exp\left\{-\frac{1}{2}\left(\sqrt{x}-\sqrt{\lambda}\right)^2\right\}$$
$$\exp\left\{-\frac{1}{2}\left(-\sqrt{x}-\sqrt{\lambda}\right)^2\right\}$$

טענה 4.12 – עבור $Y\sim\chi_1^2(\lambda)$ ניתן לכתוב

$$Y=\sum_{j=1}^{2D+1}Z_j^2$$

כאשר $Z_1,\dots,Z_{2D+1}\overset{iid}{\sim}N(0,1)$.

טענה 4.13 – **פונקציה יוצרת מומנטים של התפלגות $\chi_1^2(\lambda)$** – עבור $t<2$ מתקיים

$$M(t) = \exp\left\{\frac{\lambda t}{1-2t}\right\}(1-2t)^{-1/2}$$

הוכחה באתר הקורס.

טענה 4.14 – עבור d כללי מתקיים עבור משתנה מקרי פואסוני $D \sim Poisson\left(\frac{1}{2}\lambda\right)$:

$$f_{\chi_d^2(\lambda)}(x) = \sum_{j=0}^{\infty} \mathbb{P}(D=j) f_{\chi_{2j+d}^2}(x)$$

עבור d כללי פונקציה יוצרת מומנטים היא

$$M(t) = \exp\left\{\frac{\lambda t}{1-2t}\right\}(1-2t)^{-d/2}$$

נושא 1.4.2 – התפלגות רב-נורמלית – עוצמת המבחן

נחזור לחישוב עוצמה סטטיסטית – הסיבה שבעטייה הגענו להתפלגות חי-בריבוע לא מרכזית.

טענה 4.15 – עוצמת המבחן לסטטיסטי $n(\bar{X}-\mu^0)^T(\bar{X}-\mu^0)$ הינה:

$$\mathbb{P}(\mu) = \mathbb{P}_\mu\left(n(\bar{X}-\mu^0)^T(\bar{X}-\mu^0) \geq \chi_{d;1-\alpha}^2\right)$$

ראינו כי עבור $\Delta^* = \sqrt{n}(\mu-\mu^0)$ מתקיים

$$(\bar{X}-\mu^0)^T(\bar{X}-\mu^0) \sim \chi_d^2(\|\Delta\|^*)$$

ולכן:

$$= \mathbb{P}_\mu\left(\chi_d^2(\|\Delta\|^*) \geq \chi_{d;(1-\alpha)}^2\right)$$

לכן נדרש חיפוש על n כדי להגיע לעוצמה הנדרשת כאשר $\|\Delta^*\|^2 = n\|\mu-\mu^0\|^2$

בדקנו עד עכשיו את המקרה הפרטי $V = I$. נתחיל לדבר על המקרה הכללי יותר עבור V כלשהו.

טענה 4.16 – המקרה הכללי הוא:

$$n(\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0)$$

נשתמש בפירוק חולסקי:

$$V = LL^T \Rightarrow V^{-1} = (LL^T)^{-1} = (L^T)^{-1}L^{-1} = (L^{-1})^T L^{-1}$$

לכן הסטטיסטי הוא:

$$\begin{aligned} &n(\bar{X}-\mu^0)^T V^{-1}(\bar{X}-\mu^0) \\ &= n(\bar{X}-\mu^0)^T (L^{-1})^T L^{-1}(\bar{X}-\mu^0) \end{aligned}$$

נגדיר

$$\begin{aligned} B &= L^{-1}\bar{X} \\ \gamma &= L^{-1}\mu \\ \gamma^0 &= L^{-1}\mu^0 \end{aligned}$$

ואז:

$$= n(B-\gamma^0)^T(B-\gamma^0)$$

ואז בדיקת ההשערה $H_0{:}\,\mu=\mu^0$ שקולה להשערה $H_0{:}\,\gamma=\gamma^0$.

נזכור כי $\bar{X}\sim N\left(\mu,\frac{1}{n}V\right)\sim N\left(L\gamma,\frac{1}{n}V\right)$. לכן

$$B=L^{-1}\bar{X}\sim N\left(L^{-1}(L\gamma),\frac{1}{n}L^{-1}\left[\frac{1}{n}V\right](L^{-1})^T\right)\sim N\left(\gamma,\frac{1}{n}L^{-1}\cdot\frac{1}{n}LL^T(L^T)^{-1}\right)\sim N\left(\gamma,\frac{1}{n}I\right)$$

ובמקרה זה ניתן להפעיל את מה שפיתחנו על מטריצת היחידה במקרה הפרטי $V=I$.

שיעור 5

נושא 1.5 - הסקה סטטיסטית עבור מדגם מקרי מהתפלגות רב-נורמלית במקרה של V לא ידועה

תזכורת 5.1 – ראינו כי פונקציית לוג-הנראות עבור התפלגות רב-נורמלית $X_1,\dots,X_n\sim N_d(\mu,V)$ היא:

$$l(\mu,V)=-\frac{nd}{2}\log(2\pi)-\frac{n}{2}\log(|V|)-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^TV^{-1}(X_i-\mu)$$

אנחנו רוצים למקסם את הפונקציה עבור μ ו-V. עבור וקטור התוחלות μ מתקיים מה שראינו במקרה שבו מטריצת השונויות ידועה:

$$\hat{\mu}=\bar{X}$$
$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

טענה 5.2 נציב את אומד הנראות המרבית של וקטור התוחלות בנוסחא שקיבלנו לפונקציית לוג הנראות ונקבל:

$$l(\hat{\mu},V)=-\frac{nd}{2}\log(2\pi)-\frac{1}{2}\log(|V|)-\frac{1}{2}\sum_{i=1}^{n}(X_i-\bar{X})^TV^{-1}(X_i-\bar{X})$$

עבור חישוב אומד הנראות המרבית של מטריצת השונויות נרצה למקסם את הביטוי כפונקציה של V. כדי לעשות זאת, רוצים לגזור לפי המרכיבים של V ולהשוות את הנגזרות ל-0. לפיכך נקדים ונעסוק בנגזרות של מטריצות באופן כללי ולאחר מכן נחזור לענייננו.

העמקה אלגברה 4 – נגזרות של מטריצות

הגדרה 5.3 – נגזרות של מטריצה – באופן כללי עבור מטריצה A שכל קואורדינטה בה ניתנת לייצוג כפונקציה של וקטור פרמטרים $\vec{\theta}$ ומסומנת $A(\vec{\theta})$, ניתן להגדיר את הנגזרת החלקית לפי הקואורדינטה ה-r של $\vec{\theta}$ להיות:

$$\frac{\partial A}{\partial\theta_r}=\left[\frac{\partial A_{ij}}{\partial\theta_r}\right]_{1\le i,j\le n}$$

טענה 5.4 – סכום נגזרות של מטריצות - עבור שתי מטריצות A,B עם פרמטריזציה $A(\theta),B(\theta)$ מתקיים

$$\frac{\partial}{\partial\theta_r}[A(\theta)+B(\theta)]=\frac{\partial A}{\partial\theta_r}+\frac{\partial B}{\partial\theta_r}$$

טענה 5.5 – מכפלת נגזרות של מטריצות - עבור שתי מטריצות A,B עם פרמטריזציה $A(\theta),B(\theta)$ מתקיים

$$\frac{\partial}{\partial\theta_r}A(\theta)B(\theta)=\left[\frac{\partial A}{\partial\theta_r}\right]\cdot[B(\theta)]+[A(\theta)]\cdot\left[\frac{\partial B}{\partial\theta_r}\right]$$

הוכחה 5.5 –

$$[A(\theta)B(\theta)]_{i,j}=\sum_{k=1}^{n}\big(A(\theta)\big)_{ik}\big(B(\theta)\big)_{jk}\Rightarrow$$

$$\frac{\partial}{\partial\theta_r}[A(\theta)B(\theta)]_{i,j}=\frac{\partial}{\partial\theta_r}\sum_{k=1}^{n}\big(A(\theta)\big)_{ik}\big(B(\theta)\big)_{kj}=\sum_{k=1}^{n}\frac{\partial}{\partial\theta_r}\left[\big(A(\theta)\big)_{ik}\big(B(\theta)\big)_{kj}\right]$$

$$=\sum_{k=1}^{n}\left[\frac{\partial A_{ik}}{\partial\theta_r}B(\theta)_{kj}+A(\theta)_{ik}\frac{\partial B_{kj}}{\partial\theta_r}\right]=\left[\frac{\partial A}{\partial\theta_r}B(\theta)\right]_{ij}+\left[A(\theta)\frac{\partial B}{\partial\theta_r}\right]_{ij}$$

טענה 5.6 – נניח כי $A\in\mathbb{R}^{n\times n}$ ריבועית והפיכה. אזי

$$\frac{\partial}{\partial\theta_r}A(\theta)^{-1}=-A(\theta)^{-1}\frac{\partial A}{\partial\theta_r}A(\theta)^{-1}$$

הוכחה 5.6 – כיוון ש-A הפיכה אזי $AA^{-1}=I$ ובפרט:

$$\frac{\partial}{\partial\theta_r}AA^{-1}=\frac{\partial}{\partial\theta_r}I=0$$

כיוון שמטריצת היחידה לא תלויה ב-θ הרי שהנגזרת שלה לפי θ_r היא מטריצת האפס.

נפעיל את הנוסחה שהתקבלה מטענה 5.5 כאשר $B=A^{-1}$ ונקבל

$$\frac{\partial}{\partial\theta_r}A(\theta)B(\theta)=\frac{\partial A}{\partial\theta_r}B(\theta)+A(\theta)\frac{\partial B}{\partial\theta_r}\Rightarrow\frac{\partial}{\partial\theta_r}A(\theta)A^{-1}(\theta)=\frac{\partial A(\theta)}{\partial\theta_r}A^{-1}(\theta)+\frac{\partial A^{-1}(\theta)}{\partial\theta_r}A\Rightarrow$$

$$\frac{\partial A}{\partial\theta_r}A^{-1}+A\frac{\partial A^{-1}}{\partial\theta_r}=0\Rightarrow A\frac{\partial A^{-1}}{\partial\theta_r}=-\frac{\partial A}{\partial\theta_r}A^{-1}\xRightarrow{\times A^{-1}\,left}A^{-1}A\frac{\partial A^{-1}}{\partial\theta_r}=-A^{-1}\frac{\partial A}{\partial\theta_r}A^{-1}\Rightarrow$$

$$\frac{\partial A^{-1}}{\partial\theta_r}=-A^{-1}\frac{\partial A}{\partial\theta_r}A^{-1}$$

הגדרה 5.7 – עבור מטריצה כלשהי $A=(A)_{ij}$ ניתן להציג **פרמטריזציה** שהיא למעשה וקטור הכולל את כל הערכים של A כך שיתקיים $\left[\frac{\partial A}{\partial\theta_{rs}}\right]_{ij}=E_{ij}^{(rs)}$ כאשר $E^{(rs)}$ היא מטריצה שכולה אפסים למעט בקואורדינטה rs שבה יש 1. ניתן להגדיר גם $E_{ij}^{(rs)}=\delta_{ir}\delta_{js}$ כאשר $\delta_{ab}=1\{a=b\}$ (Kronecker's Delta).
לכן באופן כללי $\frac{\partial A}{\partial A_{rs}}=E^{(rs)}$.

טענה 5.8 – במקרה שבו המטריצה A סימטרית נוכל לייצר פרמטריזציה שכוללת את כל האיברים על האלכסון ואת האיברים במשולש התחתון של המטריצה. במקרה כזה נקבל

$$\forall 1\le r\le n:\frac{\partial A}{\partial A_{rr}}=E^{(rr)}$$

$$\forall 1\le s<r\le n:\frac{\partial A}{\partial A_{rs}}=E^{(rs)}+E^{(sr)}$$

דוגמה 5.9 – עבור מטריצה $A\in\mathbb{R}^{3\times3}$ סימטרית

$$A=\begin{pmatrix}A_{11}&A_{12}&A_{13}\\A_{21}&A_{22}&A_{23}\\A_{31}&A_{32}&A_{33}\end{pmatrix}=\begin{pmatrix}A_{11}&A_{21}&A_{31}\\A_{21}&A_{22}&A_{32}\\A_{31}&A_{32}&A_{33}\end{pmatrix}$$

ניתן לייצג באמצעות הוקטור $\theta=(A_{11}\quad A_{21}\quad A_{31}\quad A_{22}\quad A_{32}\quad A_{33})^T$ שלא כולל את כל הקואורדינטות בגלל הסימטריות. לכן למשל:

$$\frac{\partial A}{\partial A_{32}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

ובגזירה על האלכסון

$$\frac{\partial A}{\partial A_{33}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

טענה 5.10 – תהי $A \in \mathbb{R}^{n\times n}$ מטריצה ריבועית (לא סימטרית) עם פרמטריזציה מלאה θ (אורך הווקטור θ הוא $n \times n$, כמספר האיברים במטריצה). אזי עבור קואורדינטה כלשהי A_{rs}, הנגזרת של הדטרמיננטה של A לפי הקואורדינטה הינה:

$$cof(A)_{rs} = \frac{\partial}{\partial A_{rs}}|A| = (-1)^{r+s}\left|A^{(rs)}\right|$$

כאשר $A^{(rs)}$ היא המטריצה המתקבלת כאשר מוחקים את השורה ה-r והעמודה ה-s של A.

דוגמה 5.11 – במקרה של מטריצה $A \in \mathbb{R}^{3\times 3}$

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}, \theta = (A_{11} \quad \dots \quad \dots \quad \dots \quad \dots \quad A_{33})^T$$

הנגזרת של הדטרמיננטה הינה לדוגמה, בגזירה לפי A_{12}:

$$\frac{\partial}{\partial A_{12}}\begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = \frac{\partial}{\partial A_{12}}\left\{A_{11}\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{12}\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{13}\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix}\right\}$$
$$= -\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix}$$

משסיימנו לסקור את ענין נגזרות של מטריצות, נקדים עוד שני עניינים ואז נחזור להסקה הסטטיסטית.

העמקה אלגברה 5 – $trace$ של מטריצה וכלל השרשרת הרב-משתני

הגדרה 5.12 – עקבה ($trace$) של מטריצה – תהי $A \in \mathbb{R}^{m\times m}$ מטריצה ריבועית, אזי העקבה של A מוגדרת על ידי:

$$trace(A) = tr(A) = \sum_{i=1}^{m} A_{ii}$$

טענה 5.13 – אם $A \in \mathbb{R}^{1\times 1}$ כלומר $A = [a]$ אזי $tr(A) = a = A$.

טענה 5.14 – תהיינה שתי מטריצות $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times m}$, אזי $tr(AB) = tr(BA)$.

תזכורת 5.13 – כלל השרשרת – עבור שתי פונקציות גזירות, $f: \mathbb{R} \to \mathbb{R}$ ו-$g: \mathbb{R} \to \mathbb{R}$, ופונקציה $h: \mathbb{R} \to \mathbb{R}$ המוגדרת על ידי

$$h(x) = g\big(f(x)\big)$$

אזי הנגזרת של h הינה:

$$h'(x) = g'\big(f(x)\big)f'(x)$$

הגדרה 5.14 – כלל השרשרת הרב-משתני – עבור שתי פונקציות

$$f: \mathbb{R}^p \to \mathbb{R}^q \coloneqq f\big(x_1, \dots, x_p\big) = \begin{bmatrix} f_1\big(x_1, \dots, x_p\big) \\ \dots \\ f_q\big(x_1, \dots, x_p\big) \end{bmatrix}$$

$$g: \mathbb{R}^q \to \mathbb{R}^s \coloneqq g(y_1, \dots, y_q) = \begin{bmatrix} g_1(y_1, \dots, y_q) \\ \dots \\ g_s(y_1, \dots, y_q) \end{bmatrix}$$

נסמן:

$$\dot{f}_{rj} = \frac{\partial f_r}{\partial x_j}$$

$$\dot{g}_{kl} = \frac{\partial q_k}{\partial y_l}$$

ואת הפונקציה $h: \mathbb{R}^p \to \mathbb{R}^s$ נגדיר על ידי:

$$h(x) = g(f(x)) \coloneqq h(x) = \begin{bmatrix} g_1(f(x)) \\ \dots \\ g_s(f(x)) \end{bmatrix}$$

אזי לפי כלל הנגזרת הרב-משתני מתקיים:

$$\frac{\partial h_u}{\partial x_j} = \sum_{t=1}^{q} \dot{g}_{ut}(f(x)) \dot{f}_{tj}(x)$$

הערה 5.15 – במקרה $s = 1$ מתקיים:

$$\dot{g}_l = \frac{\partial q}{\partial y_l}$$

$$\frac{\partial h}{\partial x_j} = \sum_{t=1}^{q} \dot{g}_t(f(x)) \dot{f}_{tj}(x)$$

טענה 5.16 – נחזור לנגזרות עם מטריצות. כאשר $A(\theta) \in \mathbb{R}^{m \times m}$ ריבועית מתקיים:

$$\frac{\partial}{\partial \theta_r} |A(\theta)| = tr\left([cof(A)]^T \cdot \frac{\partial A}{\partial \theta_r}\right)$$

הוכחה 5.16 – לכל $1 \le i \le m$ מתקיים שהדטרמיננטה של A היא סכום רץ על עמודות:

$$|A| = \sum_{j=1}^{m} A_{ij} cof(A_{ij})$$

לפי כלל השרשרת הרב-משתני מתקיים:

$$\frac{\partial}{\partial \theta_r} |A| = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial |A|}{\partial A_{ij}} \cdot \frac{\partial A_{ij}}{\partial \theta_r}$$

ראינו שמתקיים $\frac{\partial}{\partial A_{ij}} |A| = cof(A)_{ij}$. לכן:

$$\frac{\partial}{\partial \theta_r} |A| = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial |A|}{\partial A_{ij}} \cdot \frac{\partial A_{ij}}{\partial \theta_r} = \sum_{i=1}^{m} \sum_{j=1}^{m} cof(A)_{ij} \cdot \frac{\partial A_{ij}}{\partial \theta_r} = \sum_{i=1}^{m} \sum_{j=1}^{m} cof(A)_{ij} \cdot \frac{\partial A_{ij}}{\partial \theta_r}$$

$$\overset{change\ order\ of\ sums}{=} \sum_{j=1}^{m} \left[\sum_{i=1}^{m} cof(A)_{ij} \cdot \frac{\partial A_{ij}}{\partial \theta_r} \right] \overset{transpose\ matrix}{=} \sum_{j=1}^{m} \left[\sum_{i=1}^{m} [cof(A)]_{ji}^T \cdot \frac{\partial A_{ij}}{\partial \theta_r} \right]$$

$$= \sum_{j=1}^{m} \left[\left[[cof(A)]^T \frac{\partial A}{\partial \theta_r} \right]_{jj} \right] = tr\left([cof(A)]^T \cdot \frac{\partial A}{\partial \theta_r}\right)$$

שיעור 6

תזכורת 6.0 – ראינו בסוף השיעור הקודם כי עבור מטריצה ריבועית A מתקיים

$$\frac{\partial}{\partial\theta_r}|A(\theta)| = tr\left([cof(A)]^T\cdot\frac{\partial A}{\partial\theta_r}\right)$$

כאשר

$$[cof(A)]_{ab} = (-1)^{a+b}\left|A^{(ab)}\right|$$

ו-$A^{(ab)}$ מוגדרת להיות המטריצה A אחרי שהוסרו ממנה השורה ה-a והעמודה ה-b.

טענה 6.1 –

$$\frac{\partial}{\partial\theta_r}\log\left(\det\left(A(\theta)\right)\right) = tr\left(A^{-1}\frac{\partial A}{\partial\theta_r}\right)$$

הוכחה 6.1 – בגלל ש-$\frac{d}{du}(\log(u)) = \frac{1}{u}$ ולפי כלל השרשרת נשתמש בתזכורת נקבל:

$$\frac{\partial}{\partial\theta_r}\log\left(\det\left(A(\theta)\right)\right) = \frac{\frac{\partial}{\partial\theta_r}\det(A)}{\det(A)} = \frac{tr\left([cof(A)]^T\cdot\frac{\partial A}{\partial\theta_r}\right)}{\det(A)} = tr\left(\frac{[cof(A)]^T}{\det(A)}\cdot\frac{\partial A}{\partial\theta_r}\right)$$
$$= tr\left(A^{-1}\cdot\frac{\partial A}{\partial\theta_r}\right)$$

כאשר המעבר האחרון מתקיים בשל חוק קרמר, לפיו $A^{-1} = \frac{[cof(A)]^T}{\det(A)}$.

תזכורת 6.2 – חוק קרמר – עבור מערכת משוואות $Ax = b$ הפתרון $x = \begin{pmatrix} x_1 \\ \cdots \\ x_n \end{pmatrix}$ מתקבל כאשר $x_i = \frac{\det(\tilde{A}_i)}{\det(A)}$
ו-$\tilde{A}_i$ מוגדרת להיות המטריצה המתקבלת אחרי החלפת עמודה i ב-b.

נושא 1.5.1 – התפלגות רב-נורמלית - הסקה סטטיסטית כאשר V לא ידועה – מבחן יחס נראות

נחזור לנראות מרבית עבור מדגם מקרי מהתפלגות רב-נורמלית כאשר מטריצת השונויות V אינה ידועה.

תזכורת 6.3 – ראינו (טענה 5.2) כי עבור אומד נראות מרבית $\hat{\mu} = \bar{X}$.

$$l(\hat{\mu}, V) = -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(|V|) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \bar{X})^T V^{-1}(X_i - \bar{X})$$

נרצה לגזור לפי כל אחד מהרכיבים של מטריצת השונויות V כדי למצוא את המקסימום.

טענה 6.4 – כיוון ש-V מטריצת שונויות היא סימטרית, ניקח וקטור פרמטריזציה $\theta = \begin{pmatrix} V_{11} \\ \cdots \\ V_{dd} \end{pmatrix}$ שמייצג את כל הערכים על האלכסון ועל המשולש התחתון של המטריצה. מספר האיברים של הווקטור θ הוא $\frac{d(d+1)}{2}$
ולכן נקבל מערכת של $\frac{d(d+1)}{2}$ משוואות מהצורה:

$$\frac{\partial l(\hat{\mu}, V)}{\partial\theta_r} = 0$$

טענת עזר 6.4.1 – כיוון ש-$\sum_{i=1}^{n}(X_i - \bar{X})^T V^{-1}(X_i - \bar{X})$ הוא סקלר וסקלר שווה לעקבה של עצמו:

$$\sum_{i=1}^{n}(X_i-\bar{X})^T V^{-1}(X_i-\bar{X}) = \sum_{i=1}^{n} tr[(X_i-\bar{X})^T V^{-1}(X_i-\bar{X})]$$
$$\overset{tr(AB)=tr(BA)}{=} \sum_{i=1}^{n} tr[V^{-1}(X_i-\bar{X})(X_i-\bar{X})^T]$$
$$= tr\left[V^{-1}\left[\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T\right]\right]$$

טענת עזר 6.4.1.1 - המעבר האחרון נכון כי סכום עקבות שווה לעקבת הסכומים:

$$\sum_{i=1}^{n} tr(AB_i) = \sum_{i=1}^{n}\sum_{j=1}^{n}[(AB_i)]_{jj} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n} A_{jt}(B_i)_{tj} = \sum_{j=1}^{n}\sum_{t=1}^{n} A_{jt}\left(\sum_{i=1}^{n}(B_i)_{tj}\right) = tr\left(A\sum_{i=1}^{n} B_i\right)$$

המשך 6.4.1 - נסמן $\tilde{S} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T$. ונקבל:

$$= n \cdot tr\left(V^{-1}\tilde{S}\right)$$

וקואורדינה ספציפית הינה:

$$\left[\tilde{S}\right]_{tu} = \frac{1}{n}\sum_{i=1}^{n}[(X_i-\bar{X})(X_i-\bar{X})^T]_{tu} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})_t(X_i-\bar{X})_u^T$$

וזהו האומד $\hat{s}$ לשונות (עד כדי הבדל בחלוקה – כאן מחלקים ב-n ואילו באומד $\hat{s}$ מחלקים ב-$(n-1)$. לכן ניתן לרשום:

$$\boldsymbol{\tilde{S} = \frac{n-1}{n}\widehat{Cov}(X_i)}$$

כאשר $\widehat{Cov}(X_i)$ היא השונות המשותפת במדגם בין המרכיבים השונים של X_i.

המשך טענה 6.4 – את אומד הנראות המרבית נוכל לרשום לפי הפיתוח ב-6.4.1 באמצעות:

$$l(\hat{\mu},V) = -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(|V|) - \frac{1}{2}\sum_{i=1}^{n}(X_i-\bar{X})^T V^{-1}(X_i-\bar{X})$$
$$= -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(|V|) - \frac{n}{2}tr\left(V^{-1}\tilde{S}\right)$$

נגדיר $W = V^{-1}$ ונמקסם לפי W על פונקציית נראות מקבילה:

$$\tilde{l}(\hat{\mu},W) = -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(\det(W^{-1})) - \frac{n}{2}\cdot tr\left(W\tilde{S}\right)$$

נזכור ש-$\det(C^{-1}) = \frac{1}{\det(C)}$ לכן:

$$\tilde{l}(\hat{\mu},W) = -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{1}{\det(W)}\right) - \frac{n}{2}\cdot tr\left(W\tilde{S}\right)$$
$$= -\frac{nd}{2}\log(2\pi) + \frac{n}{2}\log(\det(W)) - \frac{n}{2}\cdot tr\left(W\tilde{S}\right)$$

כעת נגזור את הביטוי לפי W_{rs} עבור $1 \le r \le d$ ו-$r \le s \le d$. נסמן את הווקטור של כל האינדקסים ב-ϕ.

$$\frac{\partial\tilde{l}(\mu,W)}{\partial\phi_c} = 0 + \frac{n}{2}tr\left(W^{-1}\frac{\partial W}{\partial\phi_c}\right) - \frac{n}{2}\cdot tr\left(\frac{\partial W}{\partial\phi_c}\tilde{S}\right) = \frac{n}{2}tr\left(V\frac{\partial W}{\partial\phi_c}\right) - \frac{n}{2}\cdot tr\left(\frac{\partial W}{\partial\phi_c}\tilde{S}\right)$$

כאשר האיבר הראשון קבוע אז מתאפס, האיבר השני לפי טענה 6.1, והאיבר האחרון בגלל שנגזרת של עקבה היא עקבה של נגזרות $\frac{\partial}{\partial\theta_r}tr(A(\theta)B) = tr\left(\frac{\partial A(\theta)}{\theta_r}B\right)$ (הוכחה בתרגול).

טענת עזר 6.4.2 - נשים לב כי

$$tr\left(V\cdot\frac{\partial W}{\partial W_{rr}}\right)=tr\left(VE^{(rr)}\right)$$

דוגמה לטענת עזר 6.4.3 – נניח $d=3$

$$W=\begin{pmatrix}W_{11}&W_{12}&W_{13}\\W_{21}&W_{22}&W_{23}\\W_{31}&W_{32}&W_{33}\end{pmatrix}=\begin{pmatrix}W_{11}&W_{21}&W_{31}\\W_{21}&W_{22}&W_{32}\\W_{31}&W_{32}&W_{33}\end{pmatrix}\Rightarrow\frac{\partial W}{\partial W_{22}}=\begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix}=E^{(22)}$$

טענת עזר 6.4.3

$$tr\left(E^{(rr)}V\right)=V_{rr}$$
$$tr\left(E^{(rs)}V\right)=V_{rs}$$

הוכחת 6.4.3

$$tr\left(E^{(rr)}V\right)=\sum_{i=1}^{n}\left[E^{(rr)}V\right]_{ii}=\sum_{i=1}^{n}\sum_{j=1}^{n}E_{ij}^{(rr)}V_{ji}=V_{rr}$$

המשך טענה 6.4 – אם נסתכל על אינדקסים זהים (על האלכסון) נקבל:

$$\frac{\partial\tilde{l}(\mu,W)}{\partial W_{rr}}=\frac{n}{2}tr\left(V\frac{\partial W}{\partial W_{rr}}\right)-\frac{n}{2}\cdot tr\left(\frac{\partial W}{\partial W_{rr}}\tilde{S}\right)$$
$$\overset{6.4.2}{=}\frac{n}{2}tr\left(VE^{(rr)}\right)-\frac{n}{2}tr\left(E^{(rr)}\tilde{S}\right)$$
$$\overset{tr(AB)=tr(BA)}{=}\frac{n}{2}tr\left(E^{(rr)}V\right)-\frac{n}{2}tr\left(E^{(rr)}\tilde{S}\right)$$
$$\overset{6.4.3}{=}\frac{n}{2}V_{rr}-\frac{n}{2}\tilde{S}_{rr}$$

אם נשווה ל-0 נקבל:

$$V_{rr}=\tilde{S}_{rr}$$

עתה, אם נסתכל על אינדקסים שונים (על המשולש התחתון, לא על האלכסון) נקבל עבור $r\neq s$:

$$\frac{\partial\tilde{l}(\mu,W)}{\partial W_{rs}}=\frac{n}{2}tr\left(V\frac{\partial W}{\partial W_{rs}}\right)-\frac{n}{2}tr\left(\frac{\partial W}{\partial W_{rs}}\tilde{S}\right)$$

טענת עזר 6.4.4 – בגלל סימטריות

$$\frac{\partial W}{\partial W_{rs}}=E^{(rs)}+E^{(sr)}$$

המשך טענה 6.4

$$\overset{6.4.4}{=}\frac{n}{2}tr\left(V\left[E^{(rs)}+E^{(sr)}\right]\right)-\frac{n}{2}tr\left(\left[E^{(rs)}+E^{(sr)}\right]\tilde{S}\right)$$
$$\overset{tr(AB)=tr(BA)}{=}\frac{n}{2}tr\left(\left[E^{(rs)}+E^{(sr)}\right]V\right)-\frac{n}{2}tr\left(\left[E^{(rs)}+E^{(sr)}\right]\tilde{S}\right)$$
$$=\frac{n}{2}tr\left(E^{(rs)}V+E^{(sr)}V\right)-\frac{n}{2}tr\left(E^{(rs)}\tilde{S}+E^{(sr)}\tilde{S}\right)$$
$$\overset{6.4.3}{=}\frac{n}{2}\left[V_{rs}+V_{sr}\right]-\frac{n}{2}\left[\tilde{S}_{rs}+\tilde{S}_{sr}\right]$$
$$\overset{symmetry}{=}nV_{rs}-n\tilde{S}_{rs}$$

נשווה ל-0 ונקבל

$$\hat{V}_{rs}=\tilde{S}_{rs}$$

מסקנה 6.5 – בסך הכל קיבלנו כי אומדי הנראות המרבית עבור וקטור התוחלות ומטריצת השונויות הינם

$$\hat{\mu} = \bar{X}$$
$$\hat{V} = \tilde{S} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

טענה 6.6 סטטיסטי יחס הנראות עבור מקרה של מטריצת שונויות לא ידועה – נתחיל במקרה $H_0: \mu = 0$ ואז סטטיסטי יחס הנראות יהיה:

$$Statistic = 2\left[l\left(\hat{\mu}, \hat{V}\right) - l\left(0, \hat{V}^{(0)}\right)\right]$$

כאשר האומד של V תחת H_0 יהיה כפי שראינו $V^{(0)} \overset{H_0}{=} \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T$. נוכל לפיכך לחשב:

$$l\left(\hat{\mu}, \hat{V}\right) - l\left(0, \hat{V}^{(0)}\right) =$$
$$\left[-\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log\left(det\left(\hat{V}\right)\right) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \bar{X})^T\hat{V}^{-1}(X_i - \bar{X})\right]$$
$$-\left[-\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log\left(det\left(\hat{V}^{(0)}\right)\right) - \frac{1}{2}\sum_{i=1}^{n} X_i^T {\hat{V}^{(0)}}^{-1} X_i\right]$$
$$= \frac{n}{2}\log\left(det\left(\hat{V}^{(0)}\right)\right) - \frac{n}{2}\log\left(det\left(\hat{V}\right)\right) + \frac{1}{2}\sum_{i=1}^{n} X_i^T \hat{V}^{(0)} X_i - \frac{1}{2}\sum_{i=1}^{n}(X_i - \bar{X})^T\hat{V}^{-1}(X_i - \bar{X})$$

נסמן

$$A = \sum_{i=1}^{n}(X_i - \bar{X})^T\hat{V}^{-1}(X_i - \bar{X})$$
$$B = \sum_{i=1}^{n} X_i^T \hat{V}^{(0)} X_i$$

ונקבל:

$$= \frac{n}{2}\log\left(det\left(\hat{V}^{(0)}\right)\right) - \frac{n}{2}\log\left(det\left(\hat{V}\right)\right) + \frac{1}{2}(B - A)$$

טענת עזר 6.6.1 - כפי שראינו קודם סקלר שווה לעקבה של עצמו, ואז נוכל לפתח:

$$A = \sum_{i=1}^{n}(X_i - \bar{X})^T\hat{V}^{-1}(X_i - \bar{X}) = \sum_{i=1}^{n} tr\left[(X_i - \bar{X})^T\hat{V}^{-1}(X_i - \bar{X})\right]$$
$$\overset{tr(AB)=tr(BA)}{=} \sum_{i=1}^{n} tr\left[(X_i - \bar{X})(X_i - \bar{X})^T\hat{V}^{-1}\right]$$
$$\overset{6.4.1.1}{=} tr\left(\hat{V}^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T\right)$$
$$\overset{\tilde{S}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T}{=} tr\left(\hat{V}^{-1}n\tilde{S}\right)$$
$$\overset{\hat{V}=\tilde{S}}{=} n \cdot tr\left(\tilde{S}^{-1}\tilde{S}\right)$$
$$= n \cdot tr(I_d) = nd$$

באופן זהה נוכל לראות כי גם

$$B = nd$$

ולכן $\frac{1}{2}(B-A)=0$.

המשך טענה 6.6 - על כן הסטטיסטי המתקבל הוא:

$$2\left[l\left(\hat{\mu},\hat{V}\right)-l(0,\hat{V}^{(0)}\right] = n\log\left(det\left(\hat{V}^{(0)}\right)\right)-n\log\left(det\left(\hat{V}\right)\right)$$
$$= n\cdot\log\left(\frac{det\left(\hat{V}^{(0)}\right)}{det\left(\hat{V}\right)}\right)$$

טענת עזר 6.6.2 – נשים לב כי

$$\hat{V} = \tilde{S} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[X_iX_i^T - X_i\bar{X}^T - \bar{X}X_i^T + \bar{X}\bar{X}^T\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}X_iX_i^T - \sum_{i=1}^{n}X_i\bar{X}^T - \sum_{i=1}^{n}\bar{X}X_i^T + \sum_{i=1}^{n}\bar{X}\bar{X}^T$$
$$= \frac{1}{n}\sum_{i=1}^{n}X_iX_i^T - \frac{n}{n}\sum_{i=1}^{n}X_i\cdot\bar{X}^T - \bar{X}\frac{n}{n}\sum_{i=1}^{n}X_i^T + n\bar{X}\bar{X}^T$$
$$= \frac{1}{n}\sum_{i=1}^{n}X_iX_i^T - n\bar{X}\bar{X}^T - n\bar{X}\bar{X}^T + n\bar{X}\bar{X}^T$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[X_iX_i^T\right] - \bar{X}\bar{X}^T$$
$$= \hat{V}^{(0)} - \bar{X}\bar{X}^T$$

מפני ש-

$$\hat{V}^{(0)} = \tilde{S}^{(0)} = \frac{1}{n}\sum_{i=1}^{n}(X_i-0)(X_i-0)^T = \frac{1}{n}\sum_{i=1}^{n}X_iX_i^T$$

כלומר בסך הכל:

$$\hat{V} = \tilde{S} = \hat{V}^{(0)} - \bar{X}\bar{X}^T = \tilde{S}^{(0)} - \bar{X}\bar{X}^T \Rightarrow$$
$$\hat{V}^{(0)} = \tilde{S} + \bar{X}\bar{X}^T$$

המשך טענה 6.6

$$\frac{det(\hat{V}^{(0)})}{det(\hat{V})} \overset{6.6.2}{=} \frac{det(\tilde{S}+\bar{X}\bar{X}^T)}{det(\tilde{S})} = \frac{det(\tilde{S}+\bar{X}\bar{X}^T)}{det(\tilde{S}^{1/2}\tilde{S}^{1/2})} \overset{|AB|=|A|\cdot|B|}{=} \frac{det(\tilde{S}+\bar{X}\bar{X}^T)}{det(\tilde{S}^{1/2})\cdot det(\tilde{S}^{1/2})}$$
$$= \det(\tilde{S}^{1/2})^{-1} det(\tilde{S}+\bar{X}\bar{X}^T) \det(\tilde{S}^{1/2})^{-1} = \cdots = \det\left(I + (\tilde{S}^{-1/2}\bar{X})(\tilde{S}^{-1/2}\bar{X})^T\right)$$

שיעור 7

נושא 1.5.2 – התפלגות רב-נורמלית - הסקה סטטיסטית כאשר V לא ידועה – בדיקת השערות

תזכורת 7.1 – ראינו בשיעור שעבר פיתוח להסקה סטטיסטית עבור התפלגות רב-נורמלית $X_1,\dots,X_n \overset{iid}{\sim} N(\mu,V)$ במצב שבו μ ו-V אינם ידועים. ראינו כי אומדי הנראות המרבית הם:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\hat{V} = \tilde{S} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T$$

במצב של השערת אפס $H_0{:}\,\mu=0$ פיתחנו מבחן יחס נראות, כאשר הסטטיסטי של יחס הנראות שקול לסטטיסטי $\det(I+aa^T)$ כאשר $a=\tilde{S}^{-1/2}\bar{X}$, שכן:

$$\det(I+aa^T) = 1+\|a\|^2 = 1+a^Ta$$
$$a^Ta = (\tilde{S}^{-1/2}\bar{X})^T(\tilde{S}^{-1/2}\bar{X}) = \bar{X}^T\tilde{S}^{-1/2}\tilde{S}^{-1/2}\bar{X} = \bar{X}^T\tilde{S}^{-1}\bar{X}$$

כדי לבצע בדיקת השערות נידרש לפתח את ההתפלגות של $\tilde{S}$ ואת ההתפלגות המשותפת $(\bar{X},\tilde{S})$ כאשר כזכור ראינו כבר כי $\bar{X}\sim N\left(\mu,\frac{1}{n}V\right)$. כדי לעשות זאת נצטרך להכיר התפלגות חדשה – התפלגות Wishart.

העמקה הסתברות 2 – התפלגות Wishart

הגדרה 7.2 – התפלגות Wishart – נניח שקיימים אינדקסים $1\le i\le n$ כך שמתקיים $Y_i\sim N(\nu_i,C)$ ו-$Y_1,\dots,Y_n$ בלתי-תלויים. $Y_i\in\mathbb{R}^d$. נגדיר:

$$S = \sum_{i=1}^{n} Y_iY_i^T$$

נגדיר מטריצה M שעמודותיה הן $\nu_1,\dots,\nu_n$, ונגדיר $H=MM^T$.

אזי ההתפלגות של S היא התפלגות Wishart עם n דרגות חופש, מטריצת שונויות C ומטריצת אי-מרכזיות H. ההתפלגות מסומנת $W_d(n,C,H)$.

הערה 7.3 – מבחינת גדלים נשים לב:

$$S\in\mathbb{R}^{d\times d}$$
$$M\in\mathbb{R}^{d\times n}$$
$$M^T\in\mathbb{R}^{n\times d}$$
$$H=MM^T\in\mathbb{R}^{d\times d}$$

הערה 7.4 – נשים לב כי

$$H = MM^T = [\nu_1 \quad \cdots \quad \nu_n]\begin{bmatrix}\nu_1^T\\ \dots\\ \nu_n^T\end{bmatrix} \Rightarrow H_{rs} = \sum_{i=1}^{n}\nu_{ri}\nu_{si} = \left(\sum_{i=1}^{n}\nu_i\nu_i^T\right)_{rs}$$

זוהי בהכרח מטריצה חיובית, לפחות למחצה.

הערה 7.5 – כאשר $\vec{v_i}=\vec{0}$ לכל $1\le i\le n$ אנחנו מקבלים מטריצת אי-מרכזיות $H=0$ ובמצב כזה ההתפלגות היא התפלגות Wishart מרכזית ומסומנת $S\sim W_d(n,C)$.

הערה 7.6 - במקרים בהם לא נרצה לעסוק במטריצת אי-המרכזיות נסמן $S\sim W_d(n,C,\cdot)$.

נסקור עתה תכונות חשובות של התפלגות Wishart.

טענה 7.7 – **תכונה מספר 1** - נניח $S\sim W_d(k,C,H)$ ויהי וקטור קבוע (לא-מקרי) $l\in\mathbb{R}^d$. אזי $l^TSl\sim\chi_k^2(\lambda)$ כאשר $\lambda=\frac{1}{\sigma_l^2}l^THl$ ו-$\sigma_l^2=l^TCl$ כלומר

$$\lambda=\frac{l^THl}{l^TCl}$$

הוכחה 7.7 – נרשום $S=\sum_{i=1}^k Y_iY_i^T$ כאשר $Y_i\overset{iid}{\sim}N(v_i,C)$ כך שעבור $M=[v_1\quad\cdots\quad v_k]$ (כזכור מטריצה שבה כל עמודה היא וקטור התוחלות המקביל של Y_i, נקבל $H=MM^T=[v_1\quad\cdots\quad v_k]\begin{bmatrix}v_1^T\\ \ldots\\ v_k^T\end{bmatrix}$. תמיד ניתן למצוא מטריצה M כזאת – לדוגמה ניקח את השורש הסימטרי של H (שקיים תמיד כי H חיובית, לפחות למחצה).

נשים לב כי:

$$Y_i\sim N(v_i,C)\Rightarrow$$
$$Z_i=l^TY_i\sim N(l^Tv_i,l^TCl)\overset{\theta_i=l^Tv_i,\sigma_l^2=l^TCl}{=}N(\theta_i,\sigma_l^2)\Rightarrow$$
$$Z_i'=\frac{1}{\sigma_l}Z_i\sim N\left(\frac{\theta_l}{\sigma_l},1\right)=N(\theta',1)$$

בהתאם לזאת נוכל לחשב:

$$l^TSl=l^T\left[\sum_{i=1}^k Y_iY_i^T\right]l=\sum_{i=1}^k(l^TY_i)(Y_i^Tl)\overset{Z_i=l^TY_i}{=}\sum_{i=1}^k Z_i^2=\sigma_l^2\sum_{i=1}^k Z_i'^2\sim\sigma_l^2\chi_k^2(\lambda)$$

וזה מתפלג חי-בריבוע. נזכיר כי לפי ההגדרה של התפלגות חי-בריבוע לא מרכזית מתקיים $\sum_{i=1}^k Z_i'^2\sim\chi_k^2(\lambda)\Leftrightarrow\lambda=\sum_{i=1}^k\theta_i'^2$.

טענה 7.8 – **תכונה מספר 2** - נניח $U_i\sim N_d(\xi,G)$ ונגדיר את המטריצה $\mathcal{U}\in\mathbb{R}^{k\times d}$ על ידי:

$$\mathcal{U}=\begin{bmatrix}U_{11}&\ldots&U_{1d}\\ \ldots&\ldots&\ldots\\ U_{k1}&\ldots&U_{kd}\end{bmatrix}$$

כמו כן נניח שקיימת מטריצה קבוצה $A\in\mathbb{R}^{k\times k}$. אזי שני התנאים הבאים שקולים:

תנאי א': $U^T AU \sim W_d(k, G, \cdot)$.

תנאי ב': $l^T U^T AUl \sim \sigma_l^2 \chi_k^2(\cdot)$.

עבור כל $l \in \mathbb{R}^d$ עם $\sigma_l^2 = l^T Gl$.

טענה 7.9 – תכונה מספר 3 - נגדיר $\mathcal{U}$ כמו שמוגדר בטענה 7.8.

א. תהיינה מטריצה A_1, A_2 קבועות (לא-מקריות). נגדיר $Q_1 = \mathcal{U}^T A_1 \mathcal{U}$ ו-$Q_2 = \mathcal{U}^T A_2 \mathcal{U}$. אזי Q_1, Q_2 בלתי-תלויים עם התפלגות Wishart אם ורק אם לכל $\ell \in \mathbb{R}^d$ המשתנים המקריים $T_1 = \ell^T U^T A_1 U\ell$ ו-$T_2 = \ell^T U^T A_2 U\ell$ מתפלגים $\chi_d^2(\cdot)$ ו-T_1, T_2 בלתי-תלויים.

ב. נניח A מטריצה ריבועית קבועה (לא-מקרית) ו-$b \in \mathbb{R}^k$ וקטור קבוע. אזי:

$$\mathcal{U}^T b \sim N(\cdot,\cdot)$$
$$\mathcal{U}^T A\mathcal{U} \sim W_d(k, G, \cdot)$$

באופן בלתי תלוי אם ורק אם לכל $\ell \in \mathbb{R}^d$

$$\ell^T \mathcal{U}^T b \sim N(\cdot,\cdot)$$
$$\ell^T \mathcal{U}^T A U\ell \sim \chi_k^2(\cdot)$$

כאשר הדבר מתקיים באופן בלתי תלוי לכל $\ell \in \mathbb{R}^d$

טענה 7.10 – תכונה מספר 4 - נניח $S_1 \sim W_d(k_1, C)$ ו-$S_2 \sim W_d(k_2, C)$ בלתי תלויים. אזי $S_1 + S_2 \sim W_d(k_1 + k_2, C)$.

טענה 7.11 – תכונה מספר 5 - נניח $S \sim W_d(k, C, \cdot)$ ו-$B \in \mathbb{R}^{q \times d}$ מטריצה קבועה. אזי $BSB^T \sim W_q(k, BCB^T, \cdot)$.

טענה 7.12 – תכונה מספר 6 - נניח $S \sim W_d(k, \Gamma)$ מרכזי, אזי לכל $1 \leq \mathrm{r} \leq \mathrm{d}$ מתקיים:

$$\Omega = (\Gamma^{-1})_{rr} [(S^{-1})_{rr}]^{-1} = \frac{(\Gamma^{-1})_{rr}}{(S^{-1})_{rr}} \sim \chi_{k-(d-1)}^2 = \chi_{k-d+1}^2$$

ו-Ω בלתי תלוי מהמשתנים המקריים S_{ij} (לא כולל r). [הוכחה לטענה בשיעור 8]

טענה 7.13 – תכונה מספר 7 נניח $S \sim W_d(k, C)$ מרכזי ו-$l \in \mathbb{R}^d$ לא מקרי. אזי

$$\frac{l^T C^{-1} l}{l^T S^{-1} l} \sim \chi_{k-d+1}^2$$

טענה 7.14 – תכונה מספר 8 - נניח $S \sim W_d(k, C)$ ו-C לא סינגולרית. אזי

$$\frac{\det(S)}{\det(C)} = \prod_{r=1}^{d} V_r$$

כאשר $V_r \sim \chi^2_{k-d+r}$ ו-$V_1, \dots, V_d$ בלתי-תלויים.

שיעור 8

נמשיך לראות תכונות של התפלגות Wishart.

תזכורת 8.1 – ראינו בשבוע שעבר את טענה 7.12 – נניח $S \sim W_d(k, \Gamma)$ מרכזי, אזי לכל $1 \leq \mathrm{r} \leq \mathrm{d}$ מתקיים:

$$\Omega = (\Gamma^{-1})_{rr}[(S^{-1})_{rr}]^{-1} = \frac{(\Gamma^{-1})_{rr}}{(S^{-1})_{rr}} \sim \chi^2_{k-(d-1)} = \chi^2_{k-d+1}$$

ו-Ω בלתי תלוי מהמשתנים המקריים S_{ij} (לא כולל r).

הוכחה 7.12 \ 8.1 - נוכל לכתוב את S בצורה $S = \sum_{i=1}^k U_i U_i^T$ כאשר $U_i \sim N(0, \Gamma)$ לכל $1 \leq i \leq k$. נניח בלי הגבלת כלליות ש-$r = d$ (אחרת ניתן לעשות חילופי שורות במטריצה כך שנגיע למצב כזה).

טענת עזר 8.1.1 – הופכי של מטריצת בלוקים - עבור מטריצת בלוקים $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ ריבועית שבה A, B, C, D הן מטריצות בפני עצמן, מוגדר $E = D - CA^{-1}B$ וההופכי של המטריצה מוגדר על ידי:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BE^{-1}CA^{-1} & -A^{-1}BE^{-1} \\ -E^{-1}CA^{-1} & E^{-1} \end{bmatrix} = \begin{bmatrix} G & H \\ K & L \end{bmatrix}$$

במקרה הסימטרי $C = B^T$ ומתקבל:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BE^{-1}B^TA^{-1} & -A^{-1}BE^{-1} \\ -E^{-1}B^TA^{-1} & E^{-1} \end{bmatrix} = \begin{bmatrix} G & H \\ K & L \end{bmatrix}$$

המשך הוכחה 8.1 – נכתוב את המטריצה Γ בצורה של מטריצת בלוקים:

$$\Gamma = \begin{pmatrix} A & b \\ b^T & \gamma \end{pmatrix}$$

כאשר

$$A \in \mathbb{R}^{(d-1)\times(d-1)}$$

$$b \in \mathbb{R}^{(d-1)\times 1}$$

$$b^T \in \mathbb{R}^{1\times(d-1)}$$

$$\gamma \in \mathbb{R}^{1\times 1}$$

נבחר i כלשהו, נסתכל על $U_i \sim N(0, \Gamma)$ וגם כאן נפרק:

$$U_i = \begin{pmatrix} U_{i1} \\ U_{i2} \\ \dots \\ U_{i,d-1} \\ U_{id} \end{pmatrix} = \begin{pmatrix} | \\ \widetilde{U}_i \\ | \\ | \\ U_{id} \end{pmatrix}$$

כאשר

$$\widetilde{U}_i \in \mathbb{R}^{(d-1)\times 1} = \begin{pmatrix} U_{i1} \\ \dots \\ U_{i,(d-1)} \end{pmatrix}$$

נסתכל על ההתפלגות המותנית $U_{id}|\widetilde{U}_i$.

תזכורת 8.1.2 - נזכור כי עבור התפלגות רב-נורמלית, אם $T \sim N(\mu, V)$, ונגדיר $T^{(1)} = \begin{pmatrix} T_1 \\ \dots \\ T_m \end{pmatrix}, T^{(2)} = \begin{pmatrix} T_{m+1} \\ \dots \\ T_n \end{pmatrix}$, החלוקה $\vec{\mu} = \begin{pmatrix} \mu^{(0)} \\ \mu^{(1)} \end{pmatrix}$ באותו אופן ומטריצת השונויות הינה מטריצת בלוקים $V = \begin{pmatrix} V^{(11)} & V^{(12)} \\ V^{(21)} & V^{(22)} \end{pmatrix}$
אזי ההתפלגות המותנית הינה:

$$T^{(2)}|T^{(1)} \sim N_d\left(\mu^{(2)} - V^{(21)}{V^{(11)}}^{-1}\left(T^{(1)} - \mu^{(1)}\right), V^{(22)} - V^{(21)}{V^{(11)}}^{-1}V^{(12)}\right)$$

המשך הוכחה 8.1 – ניעזר בתזכורת 8.1.2 ונוכל להסיק:

$$U_{id}|\widetilde{U}_i \sim N\left(-b^T A^{-1}\widetilde{U}_i, \gamma - b^T A^{-1} b\right)$$

לפי טענת עזר 8.1.1 (נציב $A = A, B = b, C = b^T, D = \gamma$) נקבל:

$$(\Gamma^{-1})_{dd} = E^{-1} = [D - CA^{-1}B]^{-1} = [\gamma - b^T A^{-1} b]^{-1}$$

ונסמן

$$\sigma^2 = \gamma - b^T A^{-1} b = \frac{1}{(\Gamma)^{-1}_{dd}} \in \mathbb{R}$$

$$\beta^T = -b^T A^{-1} \Rightarrow \beta = (-b^T A^{-1})^T = -{A^{-1}}^T b \overset{A\ symmetric}{=} -A^{-1} b \in \mathbb{R}^{(d-1)\times 1}$$

כלומר

$$U_{id}|\widetilde{U}_i \sim N\left(-b^T A^{-1}\widetilde{U}_i, \gamma - b^T A^{-1} b\right) = N\left(\beta^T \widetilde{U}_i, \sigma^2\right) = N\left(\widetilde{U}_i^T \beta, \sigma^2\right)$$

כעת נגדיר:

$$Y = \begin{pmatrix} U_{1d} \\ \dots \\ U_{kd} \end{pmatrix} \in \mathbb{R}^{k\times 1}$$

$$X = \begin{pmatrix} U_{11} & \cdots & U_{1,d-1} \\ \vdots & \ddots & \vdots \\ U_{k1} & \cdots & U_{k,d-1} \end{pmatrix} = \begin{pmatrix} - \widetilde{U}_1^T - \\ \dots \\ - \widetilde{U}_k^T - \end{pmatrix} \in \mathbb{R}^{k\times(d-1)}$$

נשים לב כי:

$$X\beta = \begin{pmatrix} U_{11} & \cdots & U_{1,d-1} \\ \vdots & \ddots & \vdots \\ U_{k1} & \cdots & U_{k,d-1} \end{pmatrix}\begin{pmatrix} \beta_1 \\ \dots \\ \beta_{d-1} \end{pmatrix} = \begin{pmatrix} - \widetilde{U}_1^T - \\ \dots \\ - \widetilde{U}_k^T - \end{pmatrix}\begin{pmatrix} \beta_1 \\ \dots \\ \beta_{d-1} \end{pmatrix} = \begin{pmatrix} \widetilde{U}_1^T \beta \\ \dots \\ \widetilde{U}_k^T \beta \end{pmatrix}$$

והווקטור המתקבל הוא בדיוק התוחלת של ההתפלגות המותנית $U_{id}|\widetilde{U}_i$. קיבלנו אם כן:

$$Y|X \sim N(X\beta, \sigma^2 I)$$

נגדיר

$$\varepsilon = Y - X\beta \sim N(0, \sigma^2 I)$$

לכן:

$$X\beta + \varepsilon = X\beta + Y - X\beta = Y$$

ולמסקנה

$$Y = X\beta + \varepsilon$$

כלומר יש לנו k תצפיות ו-$(d-1)$ משתנים מסבירים.

עתה נתסכל על מטריצת ההטלה

$$P_X = X(X^TX)^{-1}X^T$$

וכיוון ש-$P_XX = X$ נקבל

$$(I - P_X)Y = (I - P_X)(X\beta + \varepsilon) = X\beta + \varepsilon - P_XX\beta + P_X\varepsilon = X\beta + \varepsilon - X\beta + P_X\varepsilon = (I - P_X)\varepsilon = Q\varepsilon$$

כאשר $Q = I - P_X$ וכאמור $\varepsilon \sim N(0, \sigma^2 I)$.

נזכור כי $Q^T = Q$ וממשפט שלמדנו קודם (טענה 4.9) מתקבל (בהתפלגות המותנית בתנאי X):

$$\frac{1}{\sigma^2}\|Q\varepsilon\|^2 \mid X \sim \chi^2_{k-(d-1)}$$

עתה נפרק את S בצורה דומה למה שעשינו עד כה.

$$S = \sum_{i=1}^{k} U_iU_i^T = \sum_{i=1}^{k}\begin{pmatrix}\widetilde{U}_i \\ U_{id}\end{pmatrix}\begin{pmatrix}\widetilde{U}_i^T & U_{id}\end{pmatrix} = \sum_{i=1}^{k}\begin{pmatrix}\widetilde{U}_i\widetilde{U}_i^T & \widetilde{U}_iU_{id} \\ U_{id}\widetilde{U}_i^T & U_{id}U_{id}\end{pmatrix} = \begin{pmatrix}X^TX & X^TY \\ Y^TX & Y^TY\end{pmatrix}$$

כאשר ה-X, Y בביטוי האחרון הם כפי שהוגדר לעיל.

קיבלנו בסך הכל כי בקואורדינטה ה-dd של המטריצה S נמצא Y^TY ובקואורדינטה ה-dd של Γ^{-1} נמצא $\sigma^2 = \frac{1}{\Gamma_{dd}^{-1}}$.

עתה, על ידי שימוש נוסף בטענת עזר 8.1.1 נציב $\begin{pmatrix}X^TX & X^TY \\ Y^TX & Y^TY\end{pmatrix} = \begin{pmatrix}A & B \\ B^T & D\end{pmatrix}$ ונקבל:

$$\begin{aligned}
\left[S_{dd}^{-1}\right]^{-1} &= \frac{1}{(S^{-1})_{dd}} = D - CA^{-1}B = D - B^TA^{-1}B \\
&= Y^TY - (X^TY)^T(X^TX)^{-1}X^TY \\
&= Y^TY - Y^TX(X^TX)^{-1}X^TY \\
&= Y^T(I - X(X^TX)^{-1}X^T) \\
&= [(I - P_X)Y]^T[(I - P_X)Y] \\
&= \|Q\varepsilon\|^2
\end{aligned}$$

בסך הכל קיבלנו

$$[(S^{-1})_{dd}]^{-1} = \|Q\varepsilon\|^2 = \sigma^2\chi^2_{k-d+1}$$

ומכאן המסקנה:

$$\frac{[(S^{-1})_{dd}]^{-1}}{[(\Gamma^{-1})_{dd}]^{-1}}$$

כפי שראינו בהתחלה, בחרנו d בלי הגבלת כלליות ומכאן שהדבר נכון לכל $1 \le i \le r$ ולכן סיימנו את ההוכחה.

עתה, נמשיך לסקור תכונות של התפלגות Wishart.

טענה 8.2 – נניח $S \sim W_d(k, C)$ כאשר C הפיכה. אזי

$$\frac{\det(S)}{\det(C)} = \prod_{r=1}^{d} V_r$$
$$\forall r: V_r \overset{indep}{\sim} \chi^2_{k-d+r}$$

ההוכחה מבוססת על הוכחת 8.1 \ 7.12.

טענה 8.3 – נניח $S_1 \sim W_d(k_1, C)$ ו-$S_2 \sim W_d(k_2, C)$ בלתי-תלויים. כאשר $k_1 \ge d$ אזי

$$\frac{\det(S_1)}{\det(S_1+S_2)} = \prod_{r=1}^{d} H_r$$
$$\forall r: H_r \overset{indep}{\sim} Beta\left(\frac{k_1-r+1}{2}, \frac{k_2}{2}\right)$$

נושא 1.5.2 – המשך בדיקת השערות עבור מגדם מקרי פשוט מהתפלגות רב-נורמלית

כזכור, הסיבה לפיתוח של התפלגות Wishart הייתה בגלל שקיבלנו עבור הסקה ממדגמם מקרי פשוט מהתפלגות $N_d(\mu, V)$ את אומדי הנראות המרבית:

$$\hat{\mu} = \bar{X}$$
$$\hat{V} = \hat{S} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

ראינו כי במצב של בדיקת השערות $H_0: \mu = 0$ מתקבל סטטיסטי המבחן $\bar{X}^T \tilde{S}^{-1} \bar{X}$. הרחבנו בענין התפלגות Wishart כיוון שהיא מאפשרת להגדיר את ההתפלגות של סטטיסטי המבחן הנ"ל.

טענה 8.4 –

$$\bar{X} \sim N_d\left(\mu, \frac{1}{n}V\right)$$
$$(n-1)\tilde{S} \sim W_d(n-1, V)$$

והסטטיסטיים בלתי-תלויים.

הוכחה 8.4 – נזכור כי ראינו (טענה 7.9 חלק ב') כי עבור $U_i \sim N_d(\xi, G)$ כאשר המטריצה $\mathcal{U} \in \mathbb{R}^{k \times d}$ מוגדרת על ידי: $\mathcal{U} = \begin{bmatrix} U_{11} & \dots & U_{1d} \\ \dots & \dots & \dots \\ U_{k1} & \dots & U_{kd} \end{bmatrix}$, ועבור $A \in \mathbb{R}^{k \times k}$ ו-$b \in \mathbb{R}^k$ מטריצה ווקטור קבועים בהתאמה, מתקיים

$$\mathcal{U}^T b \sim N(\cdot, \cdot)$$

$$\mathcal{U}^T A \mathcal{U} \sim W_d(k, G, \cdot)$$

אם ורק אם לכל $l \in \mathbb{R}^d$ באופן בלתי תלוי

$$l^T \mathcal{U}^T b \sim N(\cdot,\cdot)$$
$$l^T \mathcal{U}^T A \mathcal{U} l \sim \chi_k^2(\cdot)$$

המשך ההוכחה בשיעור הבא.

שיעור 9

נמשיך את ההוכחה משבוע שעבר. נזכור שראינו את הטענה

תזכורת 9.0 – נניח $X_1, \dots, X_n \overset{iid}{\sim} N_d(\mu, V)$ כאשר

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

אזי

$$\bar{X} \sim N_d\left(\mu, \frac{1}{n}V\right)$$
$$(n-1)S \sim W_d(n-1, V)$$

כאשר $(n-1)S \perp \bar{X}$. ראינו כי כדי להוכיח טענה זאת ניעזר בטענה 7.9 חלק ב' אשר לפיה: עבור $U_i \sim N_d(\xi, G)$ כאשר המטריצה $\mathcal{U} \in \mathbb{R}^{k\times d}$ מוגדרת על ידי: $\mathcal{U} = \begin{bmatrix} U_{11} & \dots & U_{1d} \\ \dots & \dots & \dots \\ U_{k1} & \dots & U_{kd} \end{bmatrix}$, ועבור $A \in \mathbb{R}^{k\times k}$ ו-$b \in \mathbb{R}^k$ מטריצה ווקטור קבועים בהתאמה, מתקיים

$$\mathcal{U}^T b \sim N(\cdot,\cdot)$$
$$\mathcal{U}^T A \mathcal{U} \sim W_d(k, G, \cdot)$$

אם ורק אם לכל $l \in \mathbb{R}^d$ באופן בלתי בלתי

$$l^T \mathcal{U}^T b \sim N(\cdot,\cdot)$$
$$l^T \mathcal{U}^T A \mathcal{U} l \sim \chi_k^2(\cdot)$$

עתה להוכחה. ניקח $l \in \mathbb{R}^d$, נגדיר $Q_i = l^T X_i$ לכל $1 \le i \le n$. אזי

$$Q_i = l^T X_i \overset{iid}{\sim} N(l^T\mu, l^T V l) \overset{\sigma_l^2 = l^T V l}{=} N(l^T\mu, \sigma_l^2)$$

אנחנו יודעים (קורסים קודמים) כי מתקיים (הרחבה של תוצאה עבור מדגם מקרי פשוט מהתפלגות נורמלית רגילה)

$$\bar{Q} \sim N\left(l^T\mu, \frac{\sigma_l^2}{n}\right)$$

$$\sum_{i=1}^{n}(Q_i-\bar{Q})\sim\sigma_l^2\chi_{n-1}^2$$

נגדיר

$$\mathcal{X}=\begin{bmatrix}X_{11}&\cdots&X_{1d}\\\vdots&\ddots&\vdots\\X_{n1}&\cdots&X_{nd}\end{bmatrix}\Rightarrow$$

$$\mathcal{X}l=\begin{bmatrix}X_{11}&\cdots&X_{1d}\\\vdots&\ddots&\vdots\\X_{n1}&\cdots&X_{nd}\end{bmatrix}\begin{bmatrix}l_1\\\ldots\\l_d\end{bmatrix}=\begin{bmatrix}X_{11}l_1+\cdots+X_{1d}l_d\\\ldots\\X_{n1}l_1+\cdots+X_{nd}l_d\end{bmatrix}=\begin{bmatrix}l^TX_1\\\ldots\\l^TX_n\end{bmatrix}=\begin{bmatrix}Q_1\\\ldots\\Q_n\end{bmatrix}$$

ניתן לרשום בהמשך לכך כאשר $\vec{1}_n=\begin{bmatrix}1\\\ldots\\1\end{bmatrix}\in\mathbb{R}^n$:

$$\bar{Q}=\frac{1}{n}\sum_{i=1}^{n}Q_i=\frac{1}{n}\vec{1}_n\begin{bmatrix}Q_1\\\ldots\\Q_n\end{bmatrix}=\frac{1}{n}\vec{1}_n^T\mathcal{X}l\overset{1_n\mathcal{X}l\ is\ scalar}{=}\frac{1}{n}l^T\mathcal{X}^T\vec{1}_n\overset{b=\frac{1}{n}\vec{1}_n}{=}l^T\mathcal{X}^Tb$$

כמו כן

$$\sum_{i=1}^{n}(Q_i-\bar{Q})^2=\sum_{i=1}^{n}Q_i^2-n\bar{Q}^2$$

כאשר

$$\sum_{i=1}^{n}Q_i^2=[Q_1\quad\ldots\quad Q_n]\begin{bmatrix}Q_1\\\ldots\\Q_n\end{bmatrix}=(\mathcal{X}l)^T\mathcal{X}l=l^T\mathcal{X}^T\mathcal{X}l$$

$$\bar{Q}=\frac{1}{n}l^T\mathcal{X}^T\vec{1}_n$$

$$\Rightarrow\bar{Q}^2=\left(\frac{1}{n}l^T\mathcal{X}^T\vec{1}_n\right)\left(\frac{1}{n}l^T\mathcal{X}^T\vec{1}_n\right)^T=\left(\frac{1}{n}l^T\mathcal{X}^T\vec{1}_n\right)\left(\frac{1}{n}\vec{1}_n\mathcal{X}l\right)=\left(\frac{1}{n}\right)^2l^T\mathcal{X}^T\vec{1}_n\vec{1}^T{}_n\mathcal{X}l$$

$$\Rightarrow n\bar{Q}^2=n\left(\frac{1}{n}\right)^2l^T\mathcal{X}^T\vec{1}_n\vec{1}^T{}_n\mathcal{X}l=\frac{1}{n}l^T\mathcal{X}^T\vec{1}_n\vec{1}^T{}_n\mathcal{X}l=l^T\mathcal{X}^T\left[\frac{1}{n}\vec{1}_n\vec{1}_n^T\right]\mathcal{X}l$$

ולבסוף:

$$\sum_{i=1}^{n}(Q_i-\bar{Q})^2=\sum_{i=1}^{n}Q_i^2-n\bar{Q}^2=l^T\mathcal{X}^T\mathcal{X}l-l^T\mathcal{X}^T\left[\frac{1}{n}\vec{1}_n\vec{1}_n^T\right]\mathcal{X}l$$

$$=l^T\mathcal{X}^T\left[I-\frac{1}{n}\vec{1}_n\vec{1}_n^T\right]\mathcal{X}l\overset{A=I-\frac{1}{n}\vec{1}_n\vec{1}_n^T}{=}l^T\mathcal{X}^TA\mathcal{X}l$$

ולכן הראינו הוכחה לתנאי ב' של טענה 7.9ב ומכאן שתנאי א' גם מתקיים. במקרה שלנו

$$\mathcal{U}^T=\mathcal{X}^T$$

$b = \frac{1}{n}\vec{1}_n$

$\mathcal{U}^T b = \begin{bmatrix} X_{11} & \cdots & X_{1d} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nd} \end{bmatrix} \cdot \frac{1}{n} \cdot \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} = \begin{bmatrix} \bar{X}_1 \\ \dots \\ \bar{X}_d \end{bmatrix} = \bar{X} \Rightarrow \bar{X} \sim N_d\left(\mu, \frac{1}{n}V\right)$

$\mathcal{U}^T A \mathcal{U} = \cdots = (n-1)S \sim W_d(n-1, V)$

הערה 9.2 - נתבונן עתה על המשתנה המקרי $\Delta = \bar{X}^T S^{-1} \bar{X}$ שהוא כפי שראינו סטטיסטי יחס הנראות וננסה למצוא את ההתפלגות של הסטטיסטי הזה. לשם כך ניזכר בתכונה נוספת של התפלגות Wishart (תכונה 7 \ טענה 7.13): נניח $S \sim W_d(k, C)$ מרכזי ו-$l \in \mathbb{R}^d$ לא מקרי. אזי $\frac{l^T C^{-1} l}{l^T S^{-1} l} \sim \chi^2_{k-d+1}$.

טענה 9.3 – נתבונן על המשתנה המקרי

$$\Delta = \bar{X}^T S^{-1} \bar{X} = (\bar{X}^T V^{-1} \bar{X}) \frac{\bar{X}^T S^{-1} \bar{X}}{\bar{X}^T V^{-1} \bar{X}}$$

לפי מה שראינו עתה בהוכחה מתקיים

$$(n-1)S \sim \frac{1}{n-1} W_d(n-1, V) \perp \bar{X} \Rightarrow$$
$$(n-1)S|\bar{X} \sim \frac{1}{n-1} W_d(n-1, V) \overset{7.13}{\Longrightarrow}$$
$$\frac{1}{n-1} \cdot \frac{\bar{X}^T V^{-1} \bar{X}}{\bar{X}^T\left((n-1)S^{-1}\right)\bar{X}} |\bar{X} \sim \chi^2_{(n-1)-d+1} = \chi^2_{n-d}$$

נסמן $R = \frac{\bar{X}^T V^{-1} \bar{X}}{\bar{X}^T S^{-1} \bar{X}}$. אזי

$$\Delta = \frac{\bar{X}^T V^{-1} \bar{X}}{R} = \bar{X}^T S^{-1} \bar{X} \cdot \frac{\bar{X}^T V^{-1} \bar{X}}{\bar{X}^T V^{-1} \bar{X}} = \frac{\bar{X}^T V^{-1} \bar{X}}{\bar{X}^T V^{-1} \bar{X} / \bar{X}^T S^{-1} \bar{X}} =$$

הראינו קודם כי

$$n\bar{X}^T V^{-1} \bar{X} \overset{H_0}{\sim} \chi^2_d$$
$$n\bar{X}^T V^{-1} \bar{X} \overset{H_1}{\sim} \chi^2_d(\lambda) \overset{\lambda = n\mu^T V \mu}{=} \chi^2_d(n\mu^T V \mu)$$

כלומר אנחנו מקבלים כי Δ הינו משתנה מקרי שבו ההתפלגות של המונה היא $\bar{X}^T V^{-1} \bar{X} \sim \frac{1}{n}\chi^2_d$ וההתפלגות של המכנה היא $R \sim \chi^2_{n-d}$.

נגדיר

$$T^2 = n\bar{X}^T S^{-1} \bar{X} = n\Delta \sim (n-1)\frac{\chi^2_d(\lambda)}{\chi^2_{n-d}} \Rightarrow$$
$$\frac{T_2/d}{1/(n-d)} \sim (n-1)\frac{\chi^2_d(\lambda)/d}{\chi^2_{n-d}/(n-d)} \Rightarrow$$
$$\frac{n-d}{d(n-1)} T^2 \sim \frac{\chi^2_d(\lambda)/d}{\chi^2_{n-d}/(n-d)} = F_{d,n-d}(\lambda)$$

זו התפלגות F עם פרמטר אי-מרכזיות λ.

נושא 1.5.3 – בדיקת השערות עבור שני מדגמים בלתי תלויים מהתפלגות רב-נורמלית

הגדרה 9.4 – נניח כי

$$\left\{X^{(1)}\right\} = X_1^{(1)}, \dots, X_{n_1}^{(1)} \overset{iid}{\sim} N(\mu_1, V)$$
$$\left\{X^{(2)}\right\} = X_1^{(2)}, \dots, X_{n_2}^{(2)} \sim N(\mu_2, V)$$

מתקיימת אי-תלות בין שתי הקבוצות, ומוגדר $N = n_1 + n_2$.

ראינו בתרגיל כי אומדי הנראות המרבית הם:

$$\hat{\mu}_1 = \bar{X}^{(1)}$$
$$\hat{\mu}_2 = \bar{X}^{(2)}$$
$$\hat{V} = \frac{1}{n_1 + n_2} \sum_{m=1}^{2} \sum_{i=1}^{n_m} \left(X_i^{(m)} - \bar{X}^{(m)}\right)\left(X_i^{(m)} - \bar{X}^{(m)}\right)^T$$

כאשר אם רוצים אומד חסר הטיה מוסיפים במכנה $n_1 + n_2 - 2$.

$$\hat{S} = \frac{1}{n_1 + n_2 - 2} \sum_{m=1}^{2} \sum_{i=1}^{n_m} \left(X_i^{(m)} - \bar{X}^{(m)}\right)\left(X_i^{(m)} - \bar{X}^{(m)}\right)^T$$

ו-$\hat{S}$ אומד חסר הטיה ל-$\hat{V}$.

טענה 9.5 - נרצה לבדוק מה המצב עבור השערת האפס $H_0: \mu_1 = \mu_2$ או באופן שקול עבור הפרש אומדים $\Delta = \mu_1 - \mu_2$ השערת האפס היא $H_0: \Delta = 0$.

ניתן להוכיח כי:

$$\widehat{\Delta} = \bar{X}^{(1)} - \bar{X}^{(2)} \sim N\left(\Delta, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)V\right)$$

וכפי שראינו עתה:

$$(n_1 + n_2 - 2)S \sim W_d(n_1 + n_2 - 2, V)$$

כאשר שני הסטטיסטיים הם בלתי-תלויים, בדומה למה שראינו עבור מדגם יחיד.

בהמשך לכך, אנחנו מקבלים את הסטטיסטי

$$T^2 = \frac{n_1 n_2}{N} \widehat{\Delta}^T S^{-1} \widehat{\Delta}$$

ומתקיים

$$\frac{N - d - 1}{(N-2)d} T^2 \sim F_{d, N-d-1}(\lambda) = F_{d, N-d-1}\left(\frac{n_1 n_2}{N} \Delta^T V^{-1} \Delta\right)$$

טענה 9.6 אזור סמך – עבור הסטטיסטי

$$\left(\widehat{\Delta} - \Delta\right)^T S^{-1} \left(\widehat{\Delta} - \Delta\right)$$

הסטטיסטי

$$\frac{N - d - 1}{(N-2)d} \cdot \frac{n_1 n_2}{N} \left(\widehat{\Delta} - \Delta\right)^T S^{-1} \left(\widehat{\Delta} - \Delta\right) \sim F_{d, N-d-1}$$

ואם כך ניתן לבנות אזור סמך עבור Δ מהצורה:

$$\left\{\delta:\left(\widehat{\Delta}-\delta\right)^T V^{-1}\left(\widehat{\Delta}-\delta\right)\le \frac{N}{n_1 n_2}\cdot\frac{(N-2)d}{N-d-1}F_{d,N-d-1;(1-\alpha)}\right\}$$

ההסבר:

$$\widehat{\Delta}\sim N\left(\Delta,\left(\frac{1}{n_1}+\frac{1}{n_2}\right)V\right)\Rightarrow$$
$$\widehat{\Delta}-\Delta\sim N\left(0,\left(\frac{1}{n_1}+\frac{1}{n_2}\right)V\right)\Rightarrow$$
$$c\left(\widehat{\Delta}-\Delta\right)^T S^{-1}\left(\widehat{\Delta}-\Delta\right)\sim F$$

טענה 9.7 - **מדגמים בלתי-תלויים מההתפלגות רב-נורמלית עם מטריצות שונויות שונות -** נניח כי

$$\left\{X^{(1)}\right\}=X_1^{(1)},\dots,X_{n_1}^{(1)}\overset{iid}{\sim}N(\mu_1,V_1)$$
$$\left\{X^{(2)}\right\}=X_1^{(2)},\dots,X_{n_2}^{(2)}\sim N(\mu_2,V_2)$$

מתקיימת אי-תלות בין שתי הקבוצות, ומוגדר $N=n_1+n_2$.

ראינו בתרגיל כי אומדי הנראות המרבית הם:

$$\hat{\mu}_1=\bar{X}^{(1)}$$
$$\hat{\mu}_2=\bar{X}^{(2)}$$

נגדיר

$$\Delta=\mu_1-\mu_2$$
$$\widehat{\Delta}=\bar{X}^{(1)}-\bar{X}^{(2)}\sim N\left(\Delta,\frac{1}{n_1}V_1+\frac{1}{n_2}V_2\right)$$

אם V_1,V_2 ידועים, אזי

$$\widehat{\Delta}^T\left(\frac{1}{n_1}V_1+\frac{1}{n_2}V_2\right)^{-1}\widehat{\Delta}\sim\chi_d^2$$

שיעור 10

תזכורת 10.1 – התחלנו לנתח מקרה שבו ישנם שני מדגמים בלתי-תלויים שכל אחד מהם מתפלג רב-נורמלית, בפרט במקרה של בדיקת השערות כאשר מטריצות השונויות של שני המדגמים הן שונות. כזכור, המדגמים שלנו מוגדרים על-ידי:

$$\left\{X^{(1)}\right\}=X_1^{(1)},\dots,X_{n_1}^{(1)}\overset{iid}{\sim}N(\mu_1,V_1)$$
$$\left\{X^{(2)}\right\}=X_1^{(2)},\dots,X_{n_2}^{(2)}\sim N(\mu_2,V_2)$$

מתקיימת אי-תלות בין שתי הקבוצות, ומוגדר $N=n_1+n_2$. נגדיר

$$\Delta=\mu_1-\mu_2$$

והשערת האפס הינה:

$$H_0:\Delta=0$$
$$H_1:\Delta\neq 0$$

ראינו בתרגיל כי אומדי הנראות המרבית הם:

$$\hat{\mu}_1=\bar{X}^{(1)}$$

$$\hat{\mu}_2 = \bar{X}^{(2)}$$

$$\hat{\Delta} = \bar{X}^{(1)} - \bar{X}^{(2)} \sim N\left(\Delta, \frac{1}{n_1}V_1 + \frac{1}{n_2}V_2\right)$$

טענה 10.2 - אם V_1, V_2 **ידועים**, אזי באופן כללי

$$Q = \left(\hat{\Delta}^T - \Delta\right)^T \left(\frac{1}{n_1}V_1 + \frac{1}{n_2}V_2\right)^{-1} \left(\hat{\Delta} - \Delta\right) \sim \chi^2_d$$

ותחת השערת האפס:

$$Q^{(0)} = \hat{\Delta}^T \left(\frac{1}{n_1}V_1 + \frac{1}{n_2}V_2\right)^{-1} \hat{\Delta} \overset{H_0}{\sim} \chi^2_d$$

טענה 10.3 – אם V_1, V_2 **לא ידועים ו-n_1, n_2 גדולים מספיק**, אזי באופן כללי:

$$\hat{V}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(X_{1i} - \bar{X}^{(1)}\right)\left(X_{1i} - \bar{X}^{(1)}\right)^T$$
$$\hat{V}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\left(X_{2i} - \bar{X}^{(2)}\right)\left(X_{2i} - \bar{X}^{(2)}\right)^T$$

לפי חוק המספרים הגדולים

$$\hat{V}_1 \overset{p}{\to} V_1$$
$$\hat{V}_2 \overset{p}{\to} V_2$$

ואז במצב זה באופן כללי:

$$Q = \left(\hat{\Delta}^T - \Delta\right)^T \left(\frac{1}{n_1}\hat{V}_1 + \frac{1}{n_2}\hat{V}_2\right)^{-1} \left(\hat{\Delta} - \Delta\right) \overset{d}{\to} \chi^2_d$$

ותחת השערת האפס:

$$Q^{(0)} = \hat{\Delta}^T \left(\frac{1}{n_1}\hat{V}_1 + \frac{1}{n_2}\hat{V}_2\right)^{-1} \hat{\Delta} \overset{d}{\to} \chi^2_d$$

טענה 10.4 – עבור מדגמים קטנים ומטריצות שונויות לא ידועות אין ביטוי מדויק להתפלגות של Q ו-$Q^{(0)}$ אלא קיימים רק קירובים:

$$Q \overset{approximate}{\sim} \frac{\nu d}{\nu - d + 1} F_{d, \nu - d + 1}$$

יש בספרות כמה אופציות של ν. נוסחה עבור האופציה המובילה זמינה באתר הקורס.

סיכום 10.5

- מטריצות שונויות זהות בשני המדגמים – ההתפלגות המדויקת קיימת לכל גודל מדגם, הן עבור מטריצת שונויות ידועה והן עבור מטריצת שונויות לא ידועה (טענה 9.5)
- מטריצות שונויות שונות בשני המדגמים – אם המטריצות ידועות יש התפלגות מדויקת (חי-בריבוע, טענה 10.2), אם המטריצות לא ידועות והמדגם גדול מספיק יש התפלגות מדויקת לפי חוק המספרים הגדולים (טענה 10.3), אם המדגם לא מספיק גדול ישנם מספר קירובים (10.4)

טענה 10.6 – בדיקת השערות על שוויון בין שתי מטריצות שונויות כאשר התוחלות והשונויות לא ידועות – עבור $X^{(1)} \perp X^{(2)}$

$$X^{(1)} = X_{11}, \dots, X_{1n_1} \overset{iid}{\sim} N_d(\mu_1, V_1)$$
$$X^{(2)} = X_{21}, \dots, X_{2n_2} \overset{iid}{\sim} N_d(\mu_2, V_2)$$

נרצה לבדוק את השערת האפס הבוחנת שוויון בין מטריצות השונויות כלומר

$$H_0: V_1 = V_2$$
$$H_1: V_1 \neq V_2$$

האומד תחת השערת האפס הינו:

$$V^* = \frac{1}{n_1 + n_2} \sum_{k=1}^{2} \sum_{i=1}^{n_k} \left(X_{ki} - \bar{X}^{(k)}\right)\left(X_{ki} - \bar{X}^{(k)}\right)$$

ותחת ההשערה האלטרנטיבית כפי שראינו:

$$\hat{V}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \left(X_{1i} - \bar{X}^{(1)}\right)\left(X_{1i} - \bar{X}^{(1)}\right)^T$$
$$\hat{V}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \left(X_{2i} - \bar{X}^{(2)}\right)\left(X_{2i} - \bar{X}^{(2)}\right)^T$$

לפי מבחן יחס הנראות מתקיים:

$$W = 2\left(\ell^{(full)} - \ell^{(H_0)}\right)$$

כאשר

$$\ell^{(full)} = \left[-\frac{n_1 d}{2}\log(2\pi) - \frac{n_1}{2}\log\left(det\left(\hat{V}_1\right)\right) - \frac{1}{2}\sum_{i=1}^{n_1}\left(X_{1i} - \bar{X}^{(1)}\right)^T \hat{V}_1^{-1}\left(X_{1i} - \bar{X}^{(1)}\right)\right] + \left[-\frac{n_2 d}{2}\log(2\pi) - \frac{n_2}{2}\log\left(det\left(\hat{V}_2\right)\right) - \frac{1}{2}\sum_{i=1}^{n_2}\left(X_{2i} - \bar{X}^{(2)}\right)^T \hat{V}_2^{-1}\left(X_{2i} - \bar{X}^{(2)}\right)\right]$$

ואילו תחת השערת האפס נקבל (נסמן כמו בעבר $N = n_1 + n_2$):

$$\ell^{(H_0)} = \left[-\frac{Nd}{2}\log(2\pi) - \frac{N}{2}\log(det(V^*)) - \frac{1}{2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}\left(X_{ki} - \bar{X}^{(k)}\right)^T V^{*-1}\left(X_{ki} - \bar{X}^{(k)}\right)\right]$$

טענת עזר 10.6.1 – בסכום הפנימי של פונקציית לוג-הנראות במקרה של המודל המלא יש סקלר ששווה לפיכך לעקבה של עצמו ואז ניתן להפוך בו את סדר המכפלה:

$$\sum_{i=1}^{n_1}\left(X_{1i} - \bar{X}^{(1)}\right)^T \hat{V}_1^{-1}\left(X_{1i} - \bar{X}^{(1)}\right)$$
$$= tr\left(\sum_{i=1}^{n_1}\left(X_{1i} - \bar{X}^{(1)}\right)^T \hat{V}_1^{-1}\left(X_{1i} - \bar{X}^{(1)}\right)\right)$$
$$= tr\left(\hat{V}_1^{-1}\left[\sum_{i=1}^{n_1}\left(X_{1i} - \bar{X}^{(1)}\right)\left(X_{1i} - \bar{X}^{(1)}\right)^T\right]\right)$$
$$= tr\left(\hat{V}_1^{-1} n_1 \hat{V}_1\right)$$

$$= tr(n_1 I_d)$$
$$= n_1 d$$

באופן דומה כמובן:

$$\sum_{i=1}^{n_2}\left(X_{2i}-\bar{X}^{(2)}\right)^T \hat{V}_2^{\,-1}\left(X_{2i}-\bar{X}^{(2)}\right)=n_2 d$$

ועבור השערת האפס:

$$\sum_{k=1}^{2}\sum_{i=1}^{n_k}\left(X_{ki}-\bar{X}^{(k)}\right)^T V^{*-1}\left(X_{ki}-\bar{X}^{(k)}\right)=Nd$$

המשך טענה 10.6 מכאן מתקבל:

$$\ell^{(full)}=\left[-\frac{n_1 d}{2}\log(2\pi)-\frac{n_1}{2}\log\left(det\left(\hat{V}_1\right)\right)-\frac{n_1 d}{2}\right]+\left[-\frac{n_2 d}{2}\log(2\pi)-\frac{n_2}{2}\log\left(det\left(\hat{V}_2\right)\right)-\frac{n_2 d}{2}\right]$$
$$\ell^{(H_0)}=\left[-\frac{Nd}{2}\log(2\pi)-\frac{N}{2}\log(det(V^*))-\frac{Nd}{2}\right]$$
$$W=2\left(\ell^{(full)}-\ell^{(H_0)}\right)=N\cdot\log(det(V^*))-\left[n_1\log\left(det\left(\hat{V}_1\right)+n_2\, log\left(det\left(\hat{V}_2\right)\right)\right]\right.$$

טענה 10.7 - לא קיים ביטוי מדויק להתפלגות של סטטיסטי מבחן יחס הנראות W, אולם קיים קירוב אסימפטוטי לפי משפט Wilks:

$$W=2\left(\ell^{(full)}-\ell^{(H_0)}\right)\overset{\overset{H_0}{approximate}}{\sim}\chi_m^2$$

כאשר $m=\frac{d(d+1)}{2}$.

הערה 10.8 – משפט Wilks- נניח $Y_i\overset{iid}{\sim}f_i(y|\theta)$ משתנים מקריים בלתי-תלויים המתפלגים תחת התפלגות פרמטרית כלשהי המבוטאת בידי הווקטור $\vec{\theta}$, כאשר $\vec{\theta}\in\mathbb{R}^p$. נניח כי הווקטור $\vec{\theta}$ מורכב משני תתי-וקטורים כאשר $\theta_1\in\mathbb{R}^{p_1},\theta_2\in\mathbb{R}^{p_2}$ וכמובן $p=p_1+p_2$ ובסך הכל:

$$\vec{\theta}=\begin{bmatrix}\vec{\theta}^{(1)}\\ \vec{\theta}^{(2)}\end{bmatrix}$$

משפט Wilks עוסק בבדיקת השערת האפס $H_0:\theta^{(2)}=\vec{0}$ לעומת ההשערה האלטרנטיבית $H_1:\theta^{(2)}\neq 0$.
כלומר אם אומד הנראות המרבית מסומן ב-MLE אזי:

$$\ell^{(full)}\Rightarrow\hat{\theta}=MLE$$
$$\ell^{(H_0)}\Rightarrow\begin{matrix}\theta^{(2)}=0\\ \tilde{\theta}^{(1)}=MLE\ of\ \theta^{(1)}\ under\ H_0\end{matrix}$$

לפי המשפט, סטטיסטי מבחן יחס הנראות מתפלג חי-בריבוע עם מספר דרגות חופש שהוא ההפרש בין מספר הפרמטרים במודל המלא למספר הפרמטרים תחת השערת האפס כלומר $p-p_1=p_1+p_2-p_1=$ p_2. כלומר:

$$W=2\left(\ell^{(full)}-\ell^{H_0}\right)\overset{H_0}{\sim}\chi_{p_2}^2$$

דוגמה 10.9 – יישום עבור בדיקת שוויון בין שתי מטריצות שונויות בגודל 2×2 – במקרה הפשוט ביותר $d=2$ (נזכור כי מדובר במטריצות שונויות לכן הן בהכרח סימטריות):

$$V_1 = \begin{pmatrix} V_{11}^{(1)} & V_{12}^{(1)} \\ V_{21}^{(1)} & V_{22}^{(1)} \end{pmatrix} = \begin{pmatrix} V_{11}^{(1)} & \boldsymbol{V_{12}^{(1)}} \\ \boldsymbol{V_{12}^{(1)}} & V_{22}^{(1)} \end{pmatrix} \overset{parametrization}{=} \begin{pmatrix} \phi_1 & \phi_2 \\ \phi_2 & \phi_3 \end{pmatrix}$$

$$V_2 = \begin{pmatrix} V_{11}^{(2)} & V_{12}^{(2)} \\ V_{21}^{(2)} & V_{22}^{(2)} \end{pmatrix} = \begin{pmatrix} V_{11}^{(2)} & \boldsymbol{V_{12}^{(2)}} \\ \boldsymbol{V_{12}^{(2)}} & V_{22}^{(2)} \end{pmatrix} \overset{parametrization}{=} \begin{pmatrix} \phi_4 & \phi_5 \\ \phi_5 & \phi_6 \end{pmatrix}$$

כאשר:

$$\vec{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \phi_5 \\ \phi_6 \end{pmatrix} = \begin{pmatrix} V_{11}^{(1)} \\ V_{12}^{(1)} \\ V_{22}^{(1)} \\ V_{11}^{(2)} \\ V_{12}^{(2)} \\ V_{22}^{(2)} \end{pmatrix}$$

נגדיר

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 - \phi_1 \\ \phi_5 - \phi_2 \\ \phi_6 - \phi_3 \end{pmatrix}$$

ובמצב זה השערת האפס שאנחנו מעוניינים לבדוק $H_0\colon V_1 = V_2$ שקולה להשערת האפס:

$$H_0\colon \begin{pmatrix} \theta_4 \\ \theta_5 \\ \theta_6 \end{pmatrix} = \vec{0}$$

כלומר מספר דרגות החופש שאנחנו דורשים בהשערת האפס הוא $3 = \frac{2\cdot 3}{2} \overset{d=2}{=} \frac{d(d+1)}{2}$. כמובן תחת השערה אלטרנטיבית אנחנו נדרשים למספר פרמטרים כמספר הפרמטרים בשתי המטריצות (הסימטריות) גם יחד כלומר $2\left(\frac{d(d+1)}{2}\right) = d(d+1) \overset{d=2}{=} 2\cdot 3 = 6$.

עתה לפי מה שראינו סטטיסטי יחס הנראות הינו:

$$W = N \cdot \log(det(V^*)) - [n_1 \log\left(det\left(\hat{V}_1\right) + n_2\, log\left(det\left(\hat{V}_2\right)\right)\right]$$

ותחת השערת האפס לפי משפט $Wilks$ מתקיים:

$$W = N \cdot \log(det(V^*)) - [n_1 \log\left(det\left(\hat{V}_1\right) + n_2\, log\left(det\left(\hat{V}_2\right)\right)\right] \overset{H_0}{\sim} \chi_m^2$$

כאשר לפי משפט $Wilks$ מספר דרגות החופש הוא ההפרש בין מספר דרגות החופש במודל המלא למספר דרגות החופש במודל תחת השערת האפס כלומר:

$$m = 2 \cdot \frac{d(d+1)}{2} - \frac{d(d+1)}{2} = \frac{d(d+1)}{2} \overset{d=2 \text{ in our case}}{=} 3$$

הגדרה 10.10 – תיקון Bartlett לשתי קבוצות – הוצע תיקון שמקרב את ההתפלגות של סטטיסטי המבחן להתפלגות חי-בריבוע יותר מאשר הסטטיסטי המקורי:

$$W_{bartlett} = (N-2)\cdot \log(det(V^*)) - \left[(n_1-1)\log\left(det\left(\hat{V}_1\right) + (n_2-1)\,log\left(det\left(\hat{V}_2\right)\right)\right]\right.$$

הגדרה 10.11 – תיקון Bartlett עבור מספר כללי של קבוצות – עבור g קבוצות:

$$W_{bartlett} = (N-g)\cdot \log(det(V^*)) - \sum_{r=1}^{g}(n_r-1)\log\left(\det\left(\hat{V}^{(r)}\right)\right.$$

במצב כזה מספר דרגות החופש לפי משפט Wilks יהיה $m = (g-1)\cdot\frac{d(d+1)}{2}$.

הערה 10.12 – Box M-test – קיים גם קירוב עם התפלגות F שהוצע בידי Box. נוסחה באתר.

52801 מודלים סטטיסטיים למתקדמים 2023א
מרצה: פרופ' דוד צוקר
סיכום: אריאל וישנה

נושא 2 – מודלים של עקומות גדילה

נושא 2.1 – עקומות גדילה לינאריות – המקרה הפשוט

הקדמה 10.13 – נניח קיים מדגם עם n תצפיות $Y_1,\dots,Y_n$ וקיימות מספר תצפיות על פני זמן $t_1,\dots,t_J$ כאשר בכל נקודת זמן אנחנו מודדים את הנתונים, כלומר בכל נקודת זמן t_j אנחנו מקבלים תצפיות $Y_{1j},\dots,Y_{nj}$ ובסופו של דבר עבור כל מושא דגימה Y_i אנחנו מקבלים אוסף תצפיות $Y_{i1},\dots,Y_{iJ}$.

באופן כללי: Y_{ij} הוא התצפית בנקודת זמן t_j של מושא דגימה Y_i.

דוגמה 10.14 – מדידות של גבהי ילדים $1,\dots,n$ לאורך נקודות זמן $t_1,\dots,t_J$.

הקדמה 10.15 – עקומת גדילה לינארית - במקרה הפשוט הראשוני נניח:

1. כל מושאי הדגימה נמדדים באותם זמנים $t_1,\dots,t_J$.
2. עקומת הגדילה של כל מושא דגימה (ילד) היא לינארית

במצב זה המודל הלינארי המייצג הוא:

$$\boldsymbol{Y_{ij}=\beta_{1i}+\beta_{2i}t_{ij}+\varepsilon_{ij}}$$

התפלגות המשתנים הינה (באופן בלתי תלוי זה בזה):

$$\forall 1\le i\le n := \begin{pmatrix}\beta_{1i}\\ \beta_{2i}\end{pmatrix}\overset{iid}{\sim}N_2\left(\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix},A\right)$$
$$\varepsilon_{ij}\overset{iid}{\sim}N(0,\sigma^2)$$

כאשר:

3. β_1 – נקודת חתך כללית כלשהי (ממוצע עבור כלל הילדים)
4. β_2 – מקדם גדילה ממוצע כלשהו
5. β_{1i} – נקודת החתך עבור ילד i
6. β_{2i} – מקדם הגדילה עבור ילד i

כלומר מתקיים $\mathbb{E}\left[Y_{ij}\right]=\beta_1+\beta_2t_j$.

מבחינה זו ניתן גם להציג זאת כל ידי:

$$\beta_{1i}=\beta_1+b_{i1}$$
$$\beta_{2i}=\beta_2+b_{2i}$$

כלומר אנחנו מייצגים את ההפרש של כל ילד מהממוצע הכללי באמצעות b_{i1},b_{i2} ואז ההתפלגות היא:

$$\forall 1\le i\le n := \begin{pmatrix}\beta_{1i}\\ \beta_{2i}\end{pmatrix}\overset{iid}{\sim}N_2\left(\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix},A\right)\Rightarrow\begin{pmatrix}b_{i1}\\ b_{i2}\end{pmatrix}=\begin{pmatrix}\beta_{1i}-\beta_1\\ \beta_{2i}-\beta_2\end{pmatrix}\overset{iid}{\sim}N_2\left(\begin{pmatrix}0\\ 0\end{pmatrix},A\right)$$

הגדרה 10.16 – ניסוח מטריציוני של עקומות גדילה לינאריות:

$$X=\begin{bmatrix}1 & t_1\\ \dots & \dots\\ 1 & t_J\end{bmatrix}\in\mathbb{R}^{J\times 2}$$
$$Y_i=\begin{bmatrix}Y_{i1}\\ \dots\\ Y_{iJ}\end{bmatrix}\in\mathbb{R}^{J\times 1}$$
$$\vec{\beta}=\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}\in\mathbb{R}^{2\times 1}$$

ומתקיים:

$$\mathbb{E}[Y_i] = X\vec{\beta}$$

בדומה כפי שראינו ב-10.15

$$b_i = \begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} \sim N\left(\vec{0}, A\right)$$

ואז:

$$Y_i = X\beta + Xb_i + \varepsilon_i$$

כאשר וקטור השגיאות הוא:

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \cdots \\ \varepsilon_{iJ} \end{bmatrix} \sim N_J(0, \sigma^2 I)$$

ומניחים $b_i \perp \varepsilon_i$. בהתאם לכך ניתן לחשב את השונות המשותפת של התצפיות Y_i על ידי:

$$\begin{aligned} &Cov(Y_i) = \\ &Cov(X\beta + Xb_i + \varepsilon_i) \overset{X\beta\ non-random}{=} \\ &Cov(Xb_i + \varepsilon_i) = \\ &Cov(Xb_i) + Cov(\varepsilon_i) = \\ &XCov(b_i)X^T + \sigma^2 I = \\ &XAX^T + \sigma^2 I \end{aligned}$$

ועל כן בסך הכל:

$$Y_i \sim N(X\beta, XAX^T + \sigma^2 I) \overset{V = XAX^T + \sigma^2 I}{=} N(X\beta, V)$$

שיעור 11

הערה 11.0 – עבור שיעורים 11-12 ראו את רשימות המרצה באתר הקורס (Balanced Linear Growth Curve Model)

תזכורת 11.1 – התחלנו ללמוד על מודלים של דגימות בזמן עם קשרים לינאריים:

$$Y_{ij} = \beta_{1j} + \beta_{2i} t_j + \varepsilon_{ij}$$
$$Y_i = X\beta_i + \varepsilon_i \sim N(X\beta, XAX^T + \sigma^2 I)$$

רשימת הפרמטרים במודל הינה – 2 פרמטרים הקשורים לתוחלת ו-4 פרמטרים הקשורים לשונות. שני הפרמטרים הקשורים לתוחלת יסומנו בווקטור $\vec{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$. ארבעת הפרמטרים הקשורים לשונות יסומנו בווקטור

$$\vec{\theta} = \begin{pmatrix} A_{11} \\ A_{12} \\ A_{22} \\ \sigma^2 \end{pmatrix}$$

את הפרמטרים הללו נרצה לאמוד באמצעות שיטת הנראות המרבית.

נושא 2.1.1 אומד נראות מרבית לפרמטרים של מודל עקומת גדילה לינארית

טענה 11.2 – פונקציית נראות ופונקציית לוג-נראות: נסמן $V = XAX^T + \sigma^2 I$ ונחשב את פונקציית הנראות:

$$L(\vec{\beta},\vec{\theta}) = \prod_{i=1}^{n} f_{Y_i}(Y_i|\vec{\beta},\vec{\theta}) = \prod_{i=1}^{n}\left[\left(\frac{1}{\sqrt{2\pi}}\right)^J \cdot \det(V)^{-1/2} \cdot \ \exp\{-\frac{1}{2}(Y_i - X\beta)^T V^{-1}(Y_i - X\beta)\right]$$

ובהמשך לכך פונקציית לוג-הנראות

$$l(\vec{\beta},\vec{\theta}) = \log\left(L(\vec{\beta},\vec{\theta})\right) = -\frac{nJ}{2}\log(2\pi) - \frac{n}{2}\log(det(V)) - \frac{1}{2}\sum_{i=1}^{n}(Y_i - X\beta)^T V^{-1}(Y_i - X\beta)$$

טענה 11.3 – נגזרת לוג-הנראות של וקטור התוחלות $\vec{\beta}$ - עתה כדי למצוא את אומד הנראות המרבית נגזור ונשווה את הנגזרות ל-0.

$$\frac{\partial l}{\partial \beta_r} = -\frac{1}{2}\frac{\partial}{\partial \beta_r}\sum_{i=1}^{n}(Y_i - X\beta)^T V^{-1}(Y_i - X\beta)$$

טענת עזר 11.3.1 – באמצעות פיתוח אלגברי

$$\begin{aligned}(Y_i - X\beta)^T V^{-1}(Y_i - X\beta) &= Y_i^T V^{-1} Y_i - Y_i^T V^{-1} X\beta - (X\beta)^T V^{-1} Y_i + (X\beta)^T V^{-1}(X\beta)\\ &= Y_i^T V^{-1} Y_i - Y_i^T V^{-1} X\beta - \beta^T X^T V^{-1} Y_i + \beta^T X^T V^{-1} X\beta\\ &= Y_i^T V^{-1} Y_i - Y_i^T V^{-1} X\beta - \left(Y_i^T V^{-1} X\beta\right)^T + \beta^T X^T V^{-1} X\beta\\ &\overset{all\ scalars}{=} Y_i^T V^{-1} Y_i - 2Y_i^T V^{-1} X\beta + \beta^T (X^T V^{-1} X)\beta\end{aligned}$$

כאשר המעבר האחרון אפשרי בגלל שכל האיברים המסוכמים הם סקלרים.

נסמן

$$d^T = Y_i^T V^{-1} X$$
$$\Omega = X^T V^{-1} X$$

נשים לב כי Ω מטריצה סימטרית.

ואז בהמשך לפיתוח נקבל:

$$= Y_i^T V^{-1} Y_i - 2d^T\beta + \beta^T \Omega\beta$$

המשך טענה 11.3 – נגזור את הביטוי החדש שקיבלנו לפי β_r:

$$\begin{aligned}&\frac{\partial}{\partial \beta_r}(Y_i - X\beta)^T V^{-1}(Y_i - X\beta) =\\ &\overset{11.3.1}{=} \frac{\partial}{\partial \beta_r}\left[Y_i^T V^{-1} Y_i - 2d^T\beta + \beta^T \Omega\beta\right]\\ &= 0 - 2d_r + [(\Omega + \Omega^T)\beta]_r\end{aligned}$$

כאשר האיבר הראשון מתאפס שכן איננו תלוי ב-β, האיבר השני ישירות, האיבר השלישי בדומה לטענה שפיתחנו בעבר.

כעת בגלל ש-Ω מטריצה סימטרית מתקבל:

$$= -2dr + 2(\Omega\beta)_r$$

לכן בסך הכל הנגזרת של לוג-הנראות לפי $\vec{\beta}$ היא:

$$\frac{\partial l}{\partial \beta_r} = -\frac{1}{2}\frac{\partial}{\partial \beta_r}\sum_{i=1}^{n}(Y_i - X\beta)^T V^{-1}(Y_i - X\beta)$$

$$\overset{(\Omega\beta)_r = d_r \Rightarrow \Omega\beta = d \Rightarrow \beta = \Omega^{-1}d}{=} -\frac{1}{2}\frac{\partial}{\partial \beta_r}\sum_{i=1}^{n}\{-2d_{ir} + 2(\Omega\beta)_r\}$$

נשווה לאפס ונקבל:

$$\hat{\beta} = (n\Omega)^{-1}\sum_{i=1}^{n} d_i \Rightarrow$$

$$\boldsymbol{\hat{\beta} = \left(\sum_{i=1}^{n} X^T V^{-1} X\right)^{-1} \sum_{i=1}^{n} X^T V^{-1} Y_i}$$

משפט עזר 11.3.2 – עבור מטריצות P, Q הפיכות אזי

$$(P + ZQZ^T)^{-1} = P^{-1} + P^{-1}Z(Z^T P^{-1} Z + Q^{-1})^{-1} Z^T P^{-1}$$

טענת עזר 11.3.3 – נציב בהתבסס על משפט עזר 11.3.2

$$P = \sigma^2 I \Rightarrow P^{-1} = \frac{1}{\sigma^2} I$$
$$Z = X$$
$$Q = A$$

ונקבל:

$$V^{-1} = (XAX^T + \sigma^2 I) = \frac{1}{\sigma^2} I - \left(\frac{1}{\sigma^2}\right)^2 X\left[A^{-1} + \frac{1}{\sigma^2}(X^T X)\right]^{-1} X^T$$

נסמן

$$\Gamma = \left[A^{-1} + \frac{1}{\sigma^2}(X^T X)\right]^{-1}$$

ונקבל:

$$\boldsymbol{V^{-1} = \frac{1}{\sigma^2} I - \left(\frac{1}{\sigma^2}\right) X\Gamma X^T}$$

טענת עזר 11.3.4 - נגדיר

$$\hat{\beta}_i = (X^T X)^{-1} X^T Y_i$$

ונפתח לפי הגדרת Y_i:

$$\hat{\beta}_i = (X^T X)^{-1} X^T [X\beta + \varepsilon_i] = \beta + (X^T X)^{-1} X^T \varepsilon_i$$

כאשר

$$\hat{Y}_i = X\hat{\beta}_i$$
$$e_i = Y_i - \hat{Y}_i$$
$$Y_i = X\hat{\beta}_i + e_i$$
$$X^T e_i = 0$$

המשך טענה 11.3 - לפיכך נקבל בנוסחה שקיבלנו קודם ל-$\hat{\beta}$ (עם ההצבה של V ושל Y_i שהראינו):

$$\hat{\beta}=\left(\sum_{i=1}^{n}X^TV^{-1}X\right)^{-1}\sum_{i=1}^{n}X^TV^{-1}Y_i$$

$$\overset{input\ V\ by\ 11.3.3}{=}\left(\sum_{i=1}^{n}X^T\left(\frac{1}{\sigma^2}I-\left(\frac{1}{\sigma^2}\right)^2X\Gamma X^T\right)^{-1}X\right)^{-1}\sum_{i=1}^{n}X^T\left(\frac{1}{\sigma^2}I-\left(\frac{1}{\sigma^2}\right)^2X\Gamma X^T\right)^{-1}Y_i$$

$$\overset{input\ Y_i\ by\ 11.3.4}{=}\left(\sum_{i=1}^{n}X^T\left(\frac{1}{\sigma^2}I-\left(\frac{1}{\sigma^2}\right)^2X\Gamma X^T\right)X\right)^{-1}\sum_{i=1}^{n}X^T\left(\frac{1}{\sigma^2}I-\left(\frac{1}{\sigma^2}\right)^2X\Gamma X^T\right)^{-1}\left[X\hat{\beta}_i+e_i\right]$$

$$=\left(\sum_{i=1}^{n}X^T\left(\frac{1}{\sigma^2}I-\left(\frac{1}{\sigma^2}\right)X\Gamma X^T\right)^{-1}X\right)^{-1}\sum_{i=1}^{n}\left(\frac{1}{\sigma^2}X^T\left[X\hat{\beta}_i+e_i\right]-\left(\frac{1}{\sigma^2}\right)^2X^TX\Gamma X^T\left[X\hat{\beta}_i+e_i\right]\right)$$

עבור החלק השני נשים לב כי מתקיים:

$$X^T\left(X\hat{\beta}_i+e_i\right)=X^TX\hat{\beta}_i+X^Te_i=X^TX\hat{\beta}_i+0=X^TX\hat{\beta}_i$$

נציב ונקבל:

$$\hat{\beta}=\cdots=\left(\sum_{i=1}^{n}X^T\left(\frac{1}{\sigma^2}I-\left(\frac{1}{\sigma^2}\right)X\Gamma X^T\right)^{-1}X\right)^{-1}\sum_{i=1}^{n}\left(\frac{1}{\sigma^2}X^TX\hat{\beta}_i-\left(\frac{1}{\sigma^2}\right)^2X^TX\Gamma X^TX\hat{\beta}_i\right)$$

$$=\left(\sum_{i=1}^{n}X^T\left(\frac{1}{\sigma^2}I-\left(\frac{1}{\sigma^2}\right)X\Gamma X^T\right)^{-1}X\right)^{-1}X^T\left[\sum_{i=1}^{n}\frac{1}{\sigma^2}\left[I-\frac{1}{\sigma^2}X\Gamma \mathrm{X}^{\mathrm{T}}\right]X\right]\hat{\beta}_i=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i$$

טענה 11.4 – נשים לב כי

$$\hat{\beta}_i=(X^TX)^{-1}X^TY_i\Rightarrow$$
$$\hat{\beta}_i=(X^TX)^{-1}X^T[X\beta_i+\varepsilon_i]=\beta_i+(X^TX)^{-1}X^T\varepsilon_i$$

ולכן השונות המשותפת היא:

$$Cov\left(\hat{\beta}_i\right)=Cov(\beta_i+(X^TX)^{-1}X^T\varepsilon_i)=Cov(\beta_i)+Cov((X^TX)^{-1}X^T\varepsilon_i)=A+\sigma^2(X^TX)^{-1}$$

טענה 11.5 - נגזרת לוג-הנראות לפי וקטור השונויות $\vec{\theta}$ – כפי שראינו פונקציית לוג-הנראות היא:

$$l\left(\hat{\beta},\theta\right)=-\frac{nJ}{2}\log(2\pi)-\frac{n}{2}\log(det(V))-\frac{1}{2}\sum_{i=1}^{n}(Y_i-X\beta)^TV^{-1}(Y_i-X\beta)=$$

$$-\frac{nJ}{2}\log(2\pi)-\frac{n}{2}\log(det(V))-\frac{1}{2}tr\left[\sum_{i=1}^{n}(Y_i-X\beta)^TV^{-1}(Y_i-X\beta)\right]=$$

$$-\frac{nJ}{2}\log(2\pi)-\frac{n}{2}\log(det(V))-\frac{1}{2}tr\left[V^{-1}\sum_{i=1}^{n}(Y_i-X\beta)^T(Y_i-X\beta)\right]$$

כאשר

$$\vec{\theta}=\begin{pmatrix}A_{11}\\A_{12}\\A_{22}\\\sigma^2\end{pmatrix}$$

נסמן $Cov(\hat{\beta}_i) = C = A + \sigma^2(X^TX)^{-1}$ וקיימת טרנספומרציה הפיכה המביאה את הווקטור θ לווקטור

$$\tilde{\theta} = \begin{pmatrix} C_{11} \\ C_{12} \\ C_{22} \\ \sigma^2 \end{pmatrix}$$

נגזור לפי הרכיבים של $\tilde{\theta}$ תוך שימוש בטענות שלמדנו על נגזרות עם מטריצות:

$$\frac{\partial l}{\partial \tilde{\theta}_r} = -\frac{n}{2} tr\left(V^{-1}\frac{\partial V}{\partial \tilde{\theta}_r}\right) + \frac{1}{2} tr\left(V^{-1}\frac{\partial V}{\partial \tilde{\theta}_r} V^{-1} \sum_{i=1}^{n} (Y_i - X\beta)^T V^{-1} (Y_i - X\beta)\right)$$

טענת עזר 11.5.1

$$C = A + \sigma^2(X^TX)^{-1} \Rightarrow$$
$$A = C - \sigma^2(X^TX)^{-1} \Rightarrow$$
$$XAX^T = X[C - \sigma^2(X^TX)^{-1}]X^T = XCX^T = \sigma^2 X(X^TX)^{-1}X^T$$

ולכן

$$V = XAX^T + \sigma^2 I = XCX^T + \sigma^2(I - \sigma^2 X(X^TX)^{-1}X^T)$$

שיעור 12

הערה 12.0 – השיעור לא עבר הגהת המרצה, ישנם סיכומים של המרצה באתר הקורס

30.11.22

מודלים א' - שבוע 6 הרצאה 2

מודלים של עקומת גדילה - המשך

$$Y_{ij} = \beta_{1i} + \beta_{2i} t_j + \varepsilon_{ij} \qquad i=1,\dots,n \quad , j=1,\dots,J$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \qquad \theta = \begin{bmatrix} A_{11} \\ A_{12} \\ A_{22} \\ \sigma^2 \end{bmatrix} \qquad , V = XAX^T + \sigma^2 I \qquad , Y_i \sim N(\beta, V)$$

לוג הנראות:

$$l(\theta,\beta) = -\frac{nJ}{2}\log 2\pi - \frac{n}{2}\log\det(V) - \frac{1}{2}\sum_{i=1}^{n}(Y_i - X\beta)^T V^{-1}(Y_i - X\beta)$$

אומדים את β,θ לפי נראות מרבית. כפי שסברה קיבלנו:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i \quad \Rightarrow \quad \hat{\beta}_i = (X^TX)Y_i$$

רואים כי $Y_i = X\hat{\beta}_i + e_i \qquad X^Te_i = 0$

ה Cov: $Cov(\hat{\beta}_i) = A + \sigma^2(X^TX)^{-1} = C$

ה שונות:

$$\begin{aligned} V &= XAX^T + \sigma^2 I = X(C - \sigma^2(X^TX)^{-1})X^T + \sigma^2 I \\ &= XCX^T + \sigma^2(I - X(X^TX)^{-1}X^T) \\ &= XCX^T + \sigma^2 Q \end{aligned}$$

$\left(Q = I - X(X^TX)^{-1}X^T \rightarrow\right.$ מטריצת הטלה על תת מרחב $\left. Im(X)^\perp \ , \ X^TQ = 0\right)$

ולכן ניתן לסמן את $\tilde{\theta} = \begin{bmatrix} C_{11} \\ C_{12} \\ C_{22} \\ \sigma^2 \end{bmatrix}$ ורצה למצוא את האומדים עבור $\tilde{\theta}$

$$\frac{\partial l}{\partial \theta_r} = -\frac{n}{2}\log trace\left(V^{-1}\frac{\partial V}{\partial \theta_r}\right) + \frac{1}{2}trace\left(V^{-1}\frac{\partial V}{\partial \theta_r}V^{-1}\sum_{i=1}^{n}(Y_i - X\hat{\beta})(Y_i - X\hat{\beta})^T\right)$$

$$\frac{\partial V}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}XCX^T + \sigma^2 Q = Q$$

$$\frac{\partial V}{\partial C_{11}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = E_{11} \qquad \frac{\partial V}{\partial C_{12}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = E_{12} + E_{21} \qquad \frac{\partial V}{\partial C_{22}} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = E_{22}$$

נשתמש בפיתוח למטריצת ההופכית לאסס. נרצה להכין את מה קורה עבור $\Sigma(Y_i - X\beta)(Y_i - X\beta)$

$$V^{-1} = \frac{1}{\sigma^2}I - \left(\frac{1}{\sigma^2}\right)^2 X\Gamma X^T$$

$$\begin{aligned} \Gamma &= (A^{-1} + \sigma^2(X^TX)^{-1})^{-1} \\ &= \left[(C + \sigma^2(X^TX)^{-1})^{-1} + \sigma^2 Q\right]^{-1} \end{aligned}$$

$$\sum_{i=1}^{n}(Y_i - X\hat{\beta})(Y_i - X\hat{\beta})^T = \sum_{i=1}^{n}(\overbrace{X\hat{\beta}_i + e_i}^{Y_i} - X\hat{\beta})(X\hat{\beta}_i + e_i - X\hat{\beta})^T$$

$$= \sum_{i=1}^{n}(X(\hat{\beta}_i - \hat{\beta}) + e_i)(X(\hat{\beta}_i - \hat{\beta}) + e_i)^T$$

$$\mathrm{trace}\left(V^{-1}\frac{\partial V}{\partial\sigma^2}\right) =$$

$$= \mathrm{trace}(V^{-1}Q)$$

$$= \mathrm{trace}\left[\left(\frac{1}{\sigma^2}I - \left(\frac{1}{\sigma^2}\right)^2 X\Gamma X^T\right)Q\right]$$

$$= \mathrm{trace}\left[\frac{1}{\sigma^2}Q - \underbrace{\left(\frac{1}{\sigma^2}\right)^2 X\Gamma \underbrace{X^T Q}_{0}}_{0}\right]$$

$$= \frac{1}{\sigma^2}\mathrm{trace}(Q)$$

מה שראינו

$$\overset{*}{=} \frac{1}{\sigma^2}(J-2)$$

$$\mathrm{trace}(Q) = \mathrm{trace}\left(I - X(X^TX)^{-1}X\right) = \mathrm{trace}(I_J) - \mathrm{trace}\left(\underbrace{X(X^TX)^{-1}X^T}\right)$$

$$= \mathrm{trace}(I_J) - \mathrm{trace}(I_2) = J-2$$

שבמטריצה X יש 2 עמודות

ועכשיו נרצה לחשב את השני

פיתוח של זה בהמשך.

$$\mathrm{trace}\left(\overbrace{V^{-1}\frac{\partial V}{\partial\sigma^2}V^{-1}}\sum_{i=1}^{n}(y_i - X\hat{\beta})^T(y_i - X\hat{\beta})^T\right) =$$

$$= \mathrm{trace}\left(\left(\frac{1}{\sigma^2}\right)^2 \cancel{V^{-1}}Q\sum_{i=1}^{n}(y_i - X\hat{\beta})(y_i - X\hat{\beta})^T\right)$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \mathrm{trace}\left[\cancel{V^{-1}}Q\sum_{i=1}^{n}\left(X(\hat{\beta}_i - \hat{\beta}) + e_i\right)\left(X(\hat{\beta}_i - \hat{\beta}) + e_i\right)^T\right]$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \mathrm{trace}\left(\cancel{V^{-1}}\sum_{i=1}^{n}\left(\underbrace{QX}_{0}(\hat{\beta}_i - \hat{\beta}) + Qe_i\right)\left(X(\hat{\beta}_i - \hat{\beta}) + e_i\right)^T\right)$$

$$Q = I - X(X^TX)^{-1}X^T$$

$$QX = 0$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \mathrm{trace}\left(\cancel{V^{-1}}\sum_{i=1}^{n}Qe_i\left(X(\hat{\beta}_i - \hat{\beta}) + e_i\right)^T\right)$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \mathrm{trace}\left(\cancel{V^{-1}}\sum_{i=1}^{n}Qe_i\left((\hat{\beta}_i - \hat{\beta})^T X^T + e_i^T\right)\right)$$

$$\cancel{= \left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{n}\mathrm{trace}\left(Qe_i(\hat{\beta}_i - \hat{\beta})^T X^T + Qe_ie_i^T\right)}$$

נרצה לפתוח את הביטוי $V^{-1}\frac{\partial V}{\partial \sigma^2}V^{-1}$.

$$\frac{\partial V}{\partial \sigma^2} = Q \quad , \quad V^{-1}Q = \frac{1}{\sigma^2}Q$$

$$V^{-1}\frac{\partial V}{\partial \sigma^2}V^{-1} = V^{-1}\frac{\partial V}{\partial \sigma^2}V^{-1} = \left[V^{-1}\frac{\partial V}{\partial \sigma^2}\right]V^{-1} = \frac{1}{\sigma^2}QV^{-1} =$$

$$\cancel{= \frac{1}{\sigma^2}Q\left[\frac{1}{\sigma^2}Q - \left(\frac{1}{\sigma^2}\right)^2 QX\Gamma X\right]}$$

$$= \frac{1}{\sigma^2}Q\left[\frac{1}{\sigma^2}I - \left(\frac{1}{\sigma^2}\right)^2 X\Gamma X^T\right]$$

$$= \frac{1}{\sigma^2}\left[\frac{1}{\sigma^2}Q - \left(\frac{1}{\sigma^2}\right)^2 QX\Gamma X^T\right)$$

$$= \left(\frac{1}{\sigma^2}\right)^2$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \sum_{i=1}^{n} \text{trace}\left(\left((\hat{\beta}_i - \hat{\beta})^T X^T + e_i^T\right)Qe_i\right)$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \sum_{i=1}^{n} \text{trace}\left(\underbrace{(\hat{\beta}_i - \hat{\beta})\cancel{Qe}\overbrace{X^T Q}^{0}e_i}_{0} + \underbrace{e_i^T Q e_i}_{e_i^T e_i}\right)$$

$X^T Q = (Q^T X)^T = QX^T = 0$

Q סימטרית

$$e_i = y_i - \hat{y}_i = y_i - X(X^TX)^{-1}X^T y_i = (I - X(X^TX)^{-1}X^T)y_i = Qy_i$$

$\cancel{e_i^T Q e_i}$

$$\cancel{e_i = 0} \quad Qe_i = QQy_i = Qy_i = e_i$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \sum_{i=1}^{n} \text{trace}(e_i^T e_i)$$

$$= \left(\frac{1}{\sigma^2}\right)^2 \sum_{i=1}^{n} e_i^T e_i$$

לכן קיבלנו שלב השני הוא:

$$\frac{\partial \ell}{\partial \theta_r} = -\frac{n}{2}\frac{1}{\sigma^2}(J-2) + \frac{1}{2}\left(\frac{1}{\sigma^2}\right)^2 \sum_{i=1}^{n} e_i^T e_i$$

נשווה את הנגזרת לאפס, ונקבל:

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \Downarrow \quad \boxed{\hat{\sigma}^2 = \frac{1}{n(J-2)} \sum_{i=1}^{n} e_i^T e_i}$$

גוזרים לפי C_{11}, C_{21}, C_{22} ומקבלים את האומד הבא:

$$\boxed{\hat{C} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\beta}_i - \hat{\beta})(\hat{\beta}_i - \hat{\beta})^T}$$

<u>סיכום:</u>

קיבלנו

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_i$$

$$\hat{C} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\beta}_i - \hat{\beta})(\hat{\beta}_i - \hat{\beta})^T$$

ראינו ש:

$$\hat{\beta}_i \overset{\text{במית}}{\approx} N(\beta, C)$$

$$\hat{\beta} \overset{\text{במית}}{\approx} N(\beta, \tfrac{1}{n} C)$$

ב"ת

$$n(\hat{C}) = \sum_{i=1}^{n} (\hat{\beta}_i - \hat{\beta})(\hat{\beta}_i - \hat{\beta})^T \sim W_2 (n-1) C$$

נובע מדברים שהוכחו בהרצאות קודמות.

שיעור 13

תזכורת 13.1 – מודל של עקומות גדילה לינאריות – כפי שראינו בשיעורים הקודמים אנחנו מסתכלים על מודל מסוג

$$Y_{ij} = \beta_{1i} + \beta_{2i} t_j + \varepsilon_{ij}$$

עבור אינדקסים רצים $1 \le i \le n$ ו-$1 \le j \le J$, כאשר

$$\beta_{1i} = \beta_1 + b_{1i}$$
$$\beta_{2i} = \beta_2 + b_{2i}$$
$$\overrightarrow{b_\iota} = \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \sim N(0, A)$$
$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

כאשר $\overrightarrow{b_\iota} \perp \varepsilon_{ij}$.

תזכורת 13.2 – אומד נראות מרבית של וקטור התוחלות - ראינו את האמידה לפי שיטת הנראות המרבית של וקטורי הפרמטרים:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

$$\theta = \begin{bmatrix} A_{11} \\ A_{12} \\ A_{22} \\ \sigma^2 \end{bmatrix}$$

הנאמדת לפי

$$\boldsymbol{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\hat{\beta}_i}$$

כאשר לכל i:

$$\boldsymbol{\hat{\beta}_i} = \left(\boldsymbol{X^T X}\right)^{-1}\boldsymbol{X^T Y_i} \sim \boldsymbol{N}(\boldsymbol{\beta}, \boldsymbol{C})$$

והמטריצה C מוגדרת על ידי:

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{\sigma^2}\left(\boldsymbol{X^T X}\right)^{-1}$$

תזכורת 13.3 – אומד נראות מרבית של וקטור מטריצת השונויות - בהמשך לכך על מנת לאמוד את הפרמטרים של מטריצת השונויות, ביצענו החלפת משתנים על המטריצה C והגדרנו את הווקטור:

$$\tilde{\theta} = \begin{bmatrix} C_{11} \\ C_{12} \\ C_{22} \\ \sigma^2 \end{bmatrix}$$

וראינו בהתאם לכך כי האומדים הם:

$$\boldsymbol{\hat{\sigma}^2} = \frac{1}{\boldsymbol{n(J-2)}}\sum_{i=1}^{n}\boldsymbol{e_i^T e_i}$$

כאשר השארית e_i מוגדרת על ידי

$$e_i = Y_i - X\hat{\beta}_i$$

ולבסוף מטריצת השונויות נאמדת על ידי:

$$\hat{C} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_i - \hat{\beta}\right)\left(\hat{\beta}_i - \hat{\beta}\right)^T$$

בהמשך נעבוד עם האומד הקרוב לכך (שהוא גם אומד חסר-הטיה):

$$\boldsymbol{\hat{C}} = \frac{1}{\boldsymbol{n-1}}\sum_{i=1}^{n}\left(\boldsymbol{\hat{\beta}_i} - \boldsymbol{\hat{\beta}}\right)\left(\boldsymbol{\hat{\beta}_i} - \boldsymbol{\hat{\beta}}\right)^{\boldsymbol{T}}$$

תזכורת 13.4 – התפלגות אומדי נראות מרבית במודל עקומת גדילה לינארית - מבחינת ההתפלגויות של האומדים הללו ראינו כי:

$$\hat{\beta}_i \overset{indep}{\sim} N(\beta, C)$$

ומכאן התוצאות

$$\boldsymbol{\hat{\beta}} \sim \boldsymbol{N}\left(\boldsymbol{\beta}, \frac{1}{\boldsymbol{n}}\boldsymbol{C}\right)$$

$$(\boldsymbol{n-1})\boldsymbol{\hat{C}} = \sum_{i=1}^{n}\left(\boldsymbol{\hat{\beta}_i} - \boldsymbol{\hat{\beta}}\right)\left(\boldsymbol{\hat{\beta}_i} - \boldsymbol{\hat{\beta}}\right)^{\boldsymbol{T}} \sim \boldsymbol{W_2}(\boldsymbol{n-1}, \boldsymbol{C})$$

כאשר $\hat{\beta} \perp (n-1)\hat{C}$ בלתי-תלויים.

הערה 13.5 – יוצא מהחישובים לעיל כי:

$$Cov(\hat{\beta}) = \frac{1}{n}C$$
$$\widehat{Cov}(\hat{\beta}) = \frac{1}{n}\hat{C}$$

נושא 2.1.2 – בדיקת השערת על וקטור התוחלות בעקומות גדילה

הקדמה 13.6 – במקרה של עקומות גדילה, נרצה לבדוק את השערת האפס

$$H_0: \beta = \beta^{(0)}$$

ניעזר לשם כך במבחן וולד ($Wald$).

תזכורת 13.7 – מבחן וולד – מבחן וולד מוגדר על ידי:

$$W = \left(\hat{\beta} - \beta^{(0)}\right)^T \widehat{Cov}\left(\hat{\beta}\right)^{-1}(\hat{\beta} - \hat{\beta}_0)$$

הערה 13.8 – במקרה שלפנינו נראה כי מבחן יחס הנראות ומבחן וולד שקולים זה לזה, אך זה איננו במקרה הכללי (במקרה הכללי הם שקולים אסימפטוטית בלבד).

טענה 13.9 – סטטיסטי מבחן וולד במקרה שלנו הינו:

$$\left(\hat{\beta} - \beta^{(0)}\right)^T \widehat{Cov}\left(\hat{\beta}\right)^{-1}\left(\hat{\beta} - \beta^{(0)}\right) \overset{13.5}{=}$$
$$n\left(\hat{\beta} - \beta^{(0)}\right)^T \hat{C}^{-1}\left(\hat{\beta} - \beta^{(0)}\right) =$$
$$n\left(\hat{\beta} - \beta^{(0)}\right)^T \hat{C}^{-1}\left(\hat{\beta} - \beta^{(0)}\right)\frac{\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right)}{\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right)} =$$
$$\frac{1}{1/\left(\hat{\beta} - \beta^{(0)}\right)^T \hat{C}^{-1}\left(\hat{\beta} - \beta^{(0)}\right)}\frac{n\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right)}{\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right)} =$$
$$\frac{n\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right)}{\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right)/\left(\hat{\beta} - \beta^{(0)}\right)^T \hat{C}^{-1}\left(\hat{\beta} - \beta^{(0)}\right)}$$

כאשר (בדומה למקרים שראינו על סטטיסטי הוטלינג) במונה:

$$n\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right) \overset{H_0}{\sim} \chi^2_2$$

ובמכנה:

$$\left(\hat{\beta} - \beta^{(0)}\right)^T C^{-1}\left(\hat{\beta} - \beta^{(0)}\right)/\left(\hat{\beta} - \beta^{(0)}\right)^T \hat{C}^{-1}\left(\hat{\beta} - \beta^{(0)}\right) \sim \frac{1}{n-1}\chi^2_{n-2}$$
$$\hat{C} \sim W_2(n, C, \cdot)$$

וההתפלגויות שלהם בלתי-תלויות (ראו בשיעור על התפלגות Wishart). אם נגדיר

$$\boldsymbol{T^2 = n\left(\hat{\beta} - \beta^{(0)}\right)^T \hat{C}^{-1}\left(\hat{\beta} - \beta^{(0)}\right)}$$

אזי

$$\boldsymbol{\frac{n-1}{2(n-2)}T^2 \sim F_{2,n-2}}$$

מסקנה 13.10 – כדי לבדוק את השערת האפס $\beta = \beta^0$ יש לנו מבחן סטטיסטי שניתן להשוות בו את הערך הקריטי להתפלגות F.

נושא 2.1.3 – אזור סמך לווקטור התוחלות בעקומות גדילה לינאריות

טענה 13.11 – עבור $\alpha \in (0,1)$, אזור סמך ברמת ביטחון $1-\alpha$ לוקטור התוחלות $\vec{\beta}$ יהיה מוגדר על ידי:

$$\left\{\beta : \left(\hat{\beta} - \beta\right)^T \hat{C}^{-1}\left(\hat{\beta} - \beta\right) \le \frac{1}{n} \cdot \frac{2(n-1)}{n-2} F_{2,n-2;(1-\alpha)}\right\}$$

נושא 2.2 – עקומות גדילה לינאריות – הרחבות אפשריות

הקדמה 13.12 – מודל הבסיס שלנו הוא כזכור:

$$Y_{ij} = \beta_{1i} + \beta_{2i}t_j + \varepsilon_{ij}$$

עבור אינדקסים רצים $1 \le i \le n$ ו-$1 \le j \le J$, כאשר

$$\begin{aligned}
\beta_{1i} &= \beta_1 + b_{1i} \\
\beta_{2i} &= \beta_2 + b_{2i} \\
\overrightarrow{b_i} &= \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \sim N(0, A) \\
\varepsilon_{ij} &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}$$

למודל זה קיימות מספר הרחבות אפשריות

הגדרה 13.13 – הרחבה 1 - מודל ריבועי – במקום המודל הלינארי הפשוט ניתן להרחיב למקרה ריבועי על ידי:

$$Y_{ij} = \beta_{1i} + \beta_{2i}t_j + \beta_{3i}t_j^2 + \varepsilon_{ij}$$

הגדרה 13.14 – הרחבה 2 - זמני מדידות משתנים – הנחנו במודל הפשוט כי המדידות קבועות בזמן ובלתי-תלויות במושא התצפיות. ניתן להרחיב זאת למקרה שבו זמני המדידה תלויים בתצפית:

$$Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \varepsilon_{ij}$$

כאשר t_{ij} הינו נקודת המדידה ה-j של דגימה i (ומתקיים t_{ij} איננו בהכרח שווה ל-$t_{i'j}$ עבור $i \ne i'$)

הגדרה 13.15 – הרחבה 3 – הוספת משתנים מסבירים – מלבד התצפיות עצמן ניתן להוסיף משתני רקע מסבירים:

$$\beta_{i1} = \beta_1 + \sum_{k=1}^{p_1} \gamma_{ik}X_{ik}$$
$$\beta_{2i} = \beta_2 + \sum_{k=1}^{p_2} \gamma_{2k}X_{ik}$$

הגדרה 13.16 – הרחבה 4 – לאפשר תלות בין ε_{ij} – הנחנו במודל הפשוט כי השגיאות ε_{ij} מתפלגות באופן בלתי-תלוי ושווה התפלגות.

דוגמה 13.6.1 ניתן להרחיב זאת למודל **m-dependent** שבו עבור פרמטר m כלשהו (למשל $m = 2$) מטריצת השונויות בין ε_{ij} היא מהצורה:

$$Corr(\varepsilon_{ij}) = \begin{pmatrix} 1 & \phi_1 & \phi_2 & 0 & 0 \\ \phi_1 & 1 & \phi_1 & \phi_2 & 0 \\ \phi_2 & \phi_1 & 1 & \phi_1 & \phi_2 \\ 0 & \phi_2 & \phi_1 & 1 & \phi_1 \\ 0 & 0 & \phi_2 & \phi_1 & 1 \end{pmatrix}$$

דוגמה 13.6.2 – בדומה ניתן לייצר מודל אוטו-רגרסיבי $AR(k)$ כאשר למשל עבור $k = 1$ ופרמטר כלשהו ρ נקבל שונות משותפת בין השגיאות מהצורות:

$$Corr(\varepsilon_{ij}, \varepsilon_{il}) = \rho^{|j-l|}$$
$$Corr(\varepsilon_{ij}, \varepsilon_{il}) = \rho^{|t_{ij}-t_{il}|}$$

נושא 2.3 – המודל הלינארי המעורב הכללי

הגדרה 13.7 – המודל הלינארי המעורב הכללי מנוסח באופן הבא:

$$\boldsymbol{Y_{ij}} = \boldsymbol{X_{ij}^T\beta} + \boldsymbol{Z_{ij}^T b_i} + \boldsymbol{\varepsilon_{ij}}$$

כאשר עבור וקטורי פרמטרים כלשהם $\overrightarrow{\theta^{(1)}}, \overrightarrow{\theta^{(2)}}$

$$\overrightarrow{b_i} \sim N\left(0, G\left(\theta^{(1)}\right)\right)$$
$$\varepsilon_i \sim N\left(0, R_i\left(\theta^{(2)}\right)\right)$$

כאשר

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \cdots \\ \varepsilon_{iJ_i} \end{bmatrix}$$

הערה 13.8 – נשים לב שהגדרנו פונקציה יחידה $G\left(\theta^{(1)}\right)$ המגדירה את מטריצת השונויות המשותפות של $\vec{b_i}$ לכל i (מספר האפקטים המקריים הוא זהה לכל הדגימות). לעומת זאת עבור ε_i אנחנו יכולים לקבל מטריצות שונויות שונות עבור יחידות שונות (עם אותה פרמטריזציה), ולכן מטריצות השונויות המסומנות הן הן R_i.

הגדרה 13.9 – מבחינה מטריציונית, נגדיר:

$$\forall 1 \le i \le n \coloneqq Y_i = \begin{bmatrix} Y_{i1} \\ \cdots \\ Y_{iJ_i} \end{bmatrix}$$
$$\forall 1 \le i \le n \coloneqq X_i = \begin{bmatrix} -X_{i1}^T - \\ \cdots \\ -X_{iJ_i}^T - \end{bmatrix} \in \mathbb{R}^{J_i \times p}, \qquad X_{ij} \in \mathbb{R}^p$$
$$\vec{\beta} = \begin{bmatrix} \beta_1 \\ \cdots \\ \beta_p \end{bmatrix} \in \mathbb{R}^p$$
$$\forall 1 \le i \le n \coloneqq b_i = \begin{bmatrix} b_{i1} \\ \cdots \\ b_{iq} \end{bmatrix} \in \mathbb{R}^q$$
$$\forall 1 \le i \le n \coloneqq Z_i = \begin{bmatrix} -Z_{i1}^T - \\ \cdots \\ -Z_{iJ_i}^T - \end{bmatrix} \mathbb{R}^{J_i \times q}, \qquad Z_{ij} \in \mathbb{R}^q$$

הערה 13.10 – יש לנו n תצפיות, לכל תצפית בכל זמן דגימה אנחנו דוגמים p משתנים מסבירים הקשורים לתצפית, וישנם במקביל q אפקטים מקריים. כל דגימה i נצפית ב-J_i זמנים התלויים בדגימה. הגדלים הם בהתאם.

הגדרה 13.11 – המודל הלינארי המעורב הכללי – אם מסתכלים עבור תצפית כלשהי i ועל כל זמני המדידות שלה $1,\dots,J_i$ מקבלים:

$$\mathbb{E}[\boldsymbol{Y_i}] = \boldsymbol{X_i\beta}$$
$$\boldsymbol{Y_i} = \boldsymbol{X_i\beta} + \boldsymbol{Z_i b} + \boldsymbol{\varepsilon_i}$$

הערה 13.12 – נשים לב להכללה שאנחנו עושים במקרה המורחב למקרה הפשוט. נשים לב כי במקרה הפשוט נוכל לכתוב:

$$Y_i = \begin{bmatrix} Y_{i1} \\ \dots \\ Y_{iJ} \end{bmatrix} \in \mathbb{R}^{J\times 2}$$
$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \in \mathbb{R}^{2\times 1}$$
$$X_i = \begin{bmatrix} 1 & t_1 \\ \dots & \dots \\ 1 & t_J \end{bmatrix} \in \mathbb{R}^{J\times 2}$$
$$Z_i = \begin{bmatrix} 1 & t_1 \\ \dots & \dots \\ 1 & t_J \end{bmatrix} \in \mathbb{R}^{J\times 2}$$
$$b_i = \begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix}$$

אם נכפיל נקבל:

$$X_i\beta = \begin{bmatrix} 1 & t_1 \\ \dots & \dots \\ 1 & t_J \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_1 + \beta_2 t_1 \\ \dots \\ \beta_1 + \beta_2 t_J \end{bmatrix}$$

ובדומה:

$$Z_i b_i = \begin{bmatrix} 1 & t_1 \\ \dots & \dots \\ 1 & t_J \end{bmatrix}\begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} = \begin{bmatrix} b_1 + b_2 t_1 \\ \dots \\ b_1 + b_2 t_J \end{bmatrix}$$

ובמובן זה אנחנו משחזרים את המודל הפשוט, כאשר ההרחבה מבחינת הנוסחה היא:

$$\begin{aligned} Y_{ij} &= [\beta_1 + b_{i1}] + [\beta_2 + b_{i2}]t_j + \varepsilon_{ij} \\ &= \begin{bmatrix} 1 & t_j \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 & t_j \end{bmatrix}\begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} + \varepsilon_{ij} \\ &= \begin{bmatrix} 1 & t_j \end{bmatrix}\vec{\beta} + \begin{bmatrix} 1 & t_j \end{bmatrix}\vec{b}_i + \varepsilon_{ij} \end{aligned}$$

כמובן במקרה המורחב אנחנו גם לא מוגבלים בכך שמספר האפקטים המקריים שווה ל-2 או שהדגימות יהיו באותם קבועי זמנים.

הגדרה 13.13 – וקטוריציה מלאה של המודל המורחב נחזור למודל הכללי שבו פתחנו:

$$\boldsymbol{Y_i} = \boldsymbol{X_i\beta} + \boldsymbol{Z_i b_i} + \boldsymbol{\varepsilon_i}$$
$$b_i \sim N\left(0, G\left(\theta^{(1)}\right)\right)$$
$$\varepsilon_i \sim N\left(0, R_i\left(\theta^{(2)}\right)\right)$$

נרצה להגדיר וקטוריזציה מלאה למודל. נגדיר

$$N = \sum_{i=1}^{n} J_i$$

ניקח את כל הווקטורים $Y_1, \dots, Y_n$ שבהם כזכור לכל $1 \le i \le n$ מתקיים $Y_i \in \mathbb{R}^{J_i}$ ונגדיר את הווקטור $\vec{Y}$ להיות שרשור הווקטורים:

$$\vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \dots \\ \vec{Y}_n \end{bmatrix} \in \mathbb{R}^N$$

עבור מטריצות האפקטים הקבועים $X_1, \dots, X_n$ שבהן מתקיים $X_i \in \mathbb{R}^{J_i \times p}$ נגדיר מטריצת בלוקים X על ידי:

$$X = \begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix} \in \mathbb{R}^{N \times p}$$

עבור האפקטים המקריים $Z_1, \dots, Z_n$ שבהם מתקיים $Z_i \in \mathbb{R}^{J_i \times q}$ נגדיר את המטריצה Z להיות מטריצת בלוקים אלכסונית:

$$Z = \begin{bmatrix} Z_1 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & Z_n \end{bmatrix} \in \mathbb{R}^{N \times nq}$$

לבסוף, את הווקטורים של המקדמים $b_1, \dots, b_n$ שבהם $b_i \in \mathbb{R}^q$ נגדיר על ידי:

$$b = \begin{bmatrix} \vec{b}_1 \\ \dots \\ \vec{b}_n \end{bmatrix} \in \mathbb{R}^{nq}$$

נשים לב כי

$$Zb = \begin{bmatrix} Z_1 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & Z_n \end{bmatrix} \begin{bmatrix} \vec{b}_1 \\ \dots \\ \vec{b}_n \end{bmatrix} = \begin{bmatrix} Z_1 \vec{b}_1 \\ \dots \\ Z_n \vec{b}_n \end{bmatrix} \in \mathbb{R}^{N \times 1}$$

באופן דומה $\varepsilon_i \in \mathbb{R}^{J_i}$ והווקטור המשורשר יהיה:

$$\vec{\varepsilon} = \begin{bmatrix} \vec{\varepsilon}_1 \\ \dots \\ \vec{\varepsilon}_n \end{bmatrix} \in \mathbb{R}^N$$

ואז ננסח את המודל בצורה

$$\boldsymbol{Y = X\beta + Zb_i + \varepsilon}$$

טענה 13.14 – בחזרה למודל נשים לב כי:

$$Y_i = X_i \beta + Z_i b_i + \varepsilon_i$$

$$\mathbb{E}[Y_i] = X_i \beta$$
$$Cov[Y_i] = Cov[X_i \beta + Z_i b_i + \varepsilon_i]$$

$$
\begin{aligned}
&= Z_i Cov[b_i] Z_i^T + Cov[\varepsilon_i] \\
&\overset{by\ definition}{=} Z_i G\left(\theta^{(1)}\right) Z_i^T + R_i\left(\theta^{(2)}\right) \\
&= V_i
\end{aligned}
$$

כאשר

$$Y_i \sim N(X_i\beta, V_i)$$

ובאופן כללי

$$Y \sim N(X\beta, V)$$

כאשר V מטריצת בלוקים אלכסונית מהצורה

$$V = diag\{V_1, \dots, V_n\}$$

שיעור 14

תזכורת 14.1 – אנחנו עוסקים במודל הלינארי המעורב הכללי. המודל הינו כזכור מהצורה הבאה:

$$\forall_{i\in\{1,\dots,n\}, j\in\{1,\dots,J_i\}}: Y_{ij} = X_{ij}^T\beta + Z_{ij}^T b_i + \varepsilon_{ij}$$

כאשר

7. וקטור האפקטים הקבועים $\vec{\beta}$ הוא בגודל p
8. וקטור האפקטים המקריים b_i הוא בגודל q

ומבחינת התפלגויות ראינו כי:

$$
\begin{aligned}
b_i &\sim N\left(0, G\left(\theta^{(1)}\right)\right) \\
\varepsilon_i &\sim N\left(0, R_i\left(\theta^{(2)}\right)\right)
\end{aligned}
$$

הפרמטרים שיש לנו במודל (שאותם נרצה לאמוד) הם $\vec{\beta}, \theta^{(1)}, \theta^{(2)}$.

הערה 14.2 הנחות לגבי הפרמטרים של $\boldsymbol{\theta^{(2)}}$ – ראינו בשיעור הקודם כמה אפשרויות שונות למבנה של הפרמטרים של השגירות $R_i\left(\theta^{(2)}\right)$. האפשרויות למבנים שראינו:

1. אי-תלות בין השגיאות $R_i = \sigma^2 I$
2. מבנה m-dependent
3. מבנה של $AR(1)$.

כל אחד מהמבנים הללו מאפשרים פרמטריזציה של $\theta^{(2)}$ עם מספר יחסית מצומצם של פרמטרים

הערה 14.3 הנחות לגבי הפרמטרים של $\boldsymbol{\theta^{(1)}}$ - לגבי $\theta^{(1)}$ אנחנו מניחים מבנה כללי של G ללא מבנה מסוים. במקרה כזה הפרמטריזציה היא פרמטריזציה מלאה למעט האילוץ שהמטריצה סימטרית חיובית. לכן כפי שראינו בעבר, עבור q בגודל כלשהו, מספר הפרמטרים יהיה $\frac{q(q+1)}{2}$. לדוגמה עבור $q = 3$ וקטור הפרמטרים יהיה $\theta^{(1)} = [G_{11} \quad G_{21} \quad G_{31} \quad G_{22} \quad G_{32} \quad G_{33}]^T$.

אולם במקרה כללי זה, האילוצים על הפרמטרים הם מסובכים בגלל שנדרש המטריצה תהיה סימטריות חיובית.

הגדרה 14.4 פרמטריזציה של $\boldsymbol{\theta^{(1)}}$ באמצעות פירוק לוג חולסקי - קיימת אפשרות אחרת לעבוד עם פרמטריזציה במונחי פירוק חולסקי של המטריצה G, כך שבאמצעות פירוק חולסקי יתקיים $G = LL^T$, כאשר

כזכור L מטריצה משולשית תחתונה (כל מה שמעל האלכסון הוא 0) והאיברים באלכסון חייבים להיות גדולים מ-0. במצב כזה אנחנו עוברים (בדוגמה $q = 3$) לפרמטריזציה מסוג:

$$\theta^{(1)} = \begin{bmatrix} L_{11} \\ L_{21} \\ L_{31} \\ L_{22} \\ L_{32} \\ L_{33} \end{bmatrix}$$

במצב זה עדין נותרים אילוצים על הפרמטרים המחייבים $L_{kk} > 0$.

לכן ישנה אפשרות לבצע פרמטריזציית לוג-חולסקי מהצורה הבאה (שוב עבור $q = 3$):

$$\theta^{(1)} = \begin{bmatrix} \widetilde{L_{11}} \\ \widetilde{L_{21}} \\ \widetilde{L_{31}} \\ \widetilde{L_{22}} \\ \widetilde{L_{32}} \\ \widetilde{L_{33}} \end{bmatrix} = \begin{bmatrix} \log(L_{11}) \\ L_{21} \\ L_{31} \\ \log(L_{22}) \\ L_{32} \\ \log(L_{33}) \end{bmatrix}$$

באופן כזה אנחנו מבטלים את האילוצים ומאפשרים לכל הפרמטרים לקבל כל ערך ממשי וכך פתרון המשוואות בהמשך יתאפשר יותר בקלות וללא אילוצים על הפתרון.

תזכורת 14.5 – עם X_i, Z_i כמו שהגדרנו בשיעור הקודם אפשר להסתכל על Y_i כווקטור של תצפיות בכל נקודת זמן:

$$Y_i = \begin{bmatrix} Y_{i1} \\ \dots \\ Y_{iJ_i} \end{bmatrix}$$

תחת המשוואה שאנחנו יכולים להציג עתה בצורה ווקטורית:

$$\boldsymbol{Y_i = X_i\beta + Z_i b_i + \varepsilon_i}$$

כאשר גדלי המטריצות הינם (לכל $i \in \{1, \dots, n\}$):

$$Y_i \in \mathbb{R}^{J_i}$$
$$X \in \mathbb{R}^{J_i \times p}$$
$$Z \in \mathbb{R}^{J_i \times q}$$
$$b_i \in \mathbb{R}^q$$
$$\varepsilon_i \in \mathbb{R}^{J_i}$$

במצב זה

$$Y_i \sim N(X_i\beta, V_i)$$
$$V_i = Z_i G Z_i^T + R_i$$

תזכורת 14.6 מודל מטריציוני מלא של המודל המעורב הכללי - ראינו שניתן להציג את כל התצפיות באופן הבא. עבור $N = \sum_{i=1}^{n} J_i$ נגדיר את הווקטור $\vec{Y}$ כשרשור של כל התצפיות:

$$Y = \begin{bmatrix} Y_1 \in \mathbb{R}^{J_1} \\ Y_2 \in \mathbb{R}^{J_2} \\ \dots \\ Y_n \in \mathbb{R}^{J_n} \end{bmatrix} \in \mathbb{R}^N$$

את וקטור השגיאות $\vec{\varepsilon}$ נגדיר באופן זהה:

$$\vec{\varepsilon} = \begin{bmatrix} \vec{\varepsilon}_1 \in \mathbb{R}^{J_1} \\ \vec{\varepsilon}_2 \in \mathbb{R}^{J_2} \\ \dots \\ \vec{\varepsilon}_n \in \mathbb{R}^{J_n} \end{bmatrix} \in \mathbb{R}^N$$

באופן דומה:

$$\vec{b} = \begin{bmatrix} b_1 \in \mathbb{R}^q \\ \dots \\ b_n \in \mathbb{R}^q \end{bmatrix} \in \mathbb{R}^{nq}$$

המטריצה X היא מטריצת בלוקים שהיא שרשור אנכי של כל המטריצות $X_1, \dots, X_n$:

$$X = \begin{bmatrix} X_1 \in \mathbb{R}^{J_1 \times p} \\ X_2 \in \mathbb{R}^{J_2 \times p} \\ \dots \\ X_n \in \mathbb{R}^{J_n \times p} \end{bmatrix} \in \mathbb{R}^{N \times p}$$

את המטריצה Z נגדיר בתור מטריצת בלוקים אלכסונית מהצורה:

$$Z = \begin{bmatrix} Z_1 \in \mathbb{R}^{J_1 \times q} & 0 & 0 & 0 \\ 0 & Z_2 \in \mathbb{R}^{J_2 \times q} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & Z_n \in \mathbb{R}^{J_n \times q} \end{bmatrix} \in \mathbb{R}^{N \times nq}$$

את המטריצה V נגדיר בתור מטריצת בלוקים אלכסונית מהצורה:

$$V = \begin{bmatrix} V_1 \in \mathbb{R}^{J_1 \times J_1} & 0 & 0 & 0 \\ 0 & V_2 \in \mathbb{R}^{J_1 \times J_1} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & V_n \in \mathbb{R}^{J_n \times J_n} \end{bmatrix} \in \mathbb{R}^{N \times N}$$

הערה 14.7 – את המודל כולו ניתן עכשיו לנסח מחדש תחת המשוואה:

$$\vec{Y} = X\beta + Z\vec{b} + \vec{\varepsilon}$$

כאשר

$$\vec{Y} \in \mathbb{R}^N$$
$$X \in \mathbb{R}^{N \times p}$$
$$\beta \in \mathbb{R}^{p \times 1}$$
$$Z \in \mathbb{R}^{N \times nq}$$
$$b \in \mathbb{R}^{q \times 1}$$
$$\varepsilon \in \mathbb{R}^{N \times 1}$$

ולכן בסך הכל אנחנו מקבלים

$$X\beta \in \mathbb{R}^N$$
$$Z\vec{b} \in \mathbb{R}^N$$

ואכן מקבלים N חיזויים כנדרש.

טענה 14.8 התפלגות הפרמטרים בווקטוריזציה מלאה

$$\vec{Y} = X\beta + Zb + \varepsilon$$

מבחינת התפלגויות אנחנו מקבלים לאור ההנחה של אי-תלות בין היחידות את ההתפלגויות הבאות:

$$b\sim N\left(0,\begin{pmatrix} G & 0 & 0\\ 0 & \dots & 0\\ 0 & 0 & G\end{pmatrix}\right)=N(0,\mathcal{G})$$

כאשר בגלל ש-$G\in\mathbb{R}^{q\times q}$ הרי שמטריצת הבלוקים האלכסונית $\mathcal{G}\in\mathbb{R}^{nq\times nq}$.

ובאופן דומה:

$$\vec{\varepsilon}\sim N\left(0,\begin{pmatrix} R_1 & 0 & 0\\ 0 & \dots & 0\\ 0 & 0 & R_n\end{pmatrix}\right)=N(0,R)$$

כאשר המטריצה R היא מטריצת בלוקים אלכסונית, ובגלל שכל $R_i\in\mathbb{R}^{J_i\times J_i}$ הרי ש-$R\in\mathbb{R}^{N\times N}$.

לאור זאת אנחנו מסיקים את ההתפלגות של הוקטור הכללי:

$$Y\sim N(X\beta,V)$$

כאשר V היא מטריצת הבלוקים האלכסונית $diag\{V_1,\dots,V_n\}$ שראינו לעיל והיא מוגדרת על ידי:

$$V=ZGZ^T+R=\begin{bmatrix} V_1 & 0 & 0 & 0\\ 0 & V_2 & 0 & 0\\ 0 & 0 & \dots & 0\\ 0 & 0 & 0 & V_n\end{bmatrix}=\begin{bmatrix} Z_1GZ_1^T+R_1 & 0 & 0 & 0\\ 0 & Z_2GZ_2^T+R_2 & 0 & 0\\ 0 & 0 & \dots & 0\\ 0 & 0 & 0 & Z_nGZ_n^T+R_n\end{bmatrix}\in\mathbb{R}^{N\times N}$$

הגדרה 14.9 – פרמטריזציה של הצורה הווקטורית המלאה המודל הלינארי המעורב הכללי – גם כאן הפרמטרים הם $\beta,\theta^{(1)},\theta^{(2)}$. לשם הנוחות נגדיר

$$\vec{\theta}=\begin{bmatrix}\vec{\theta}^{(1)}\\ \vec{\theta}^{(2)}\end{bmatrix}$$

נושא 2.3.1 – המודל הלינארי המעורב הכללי – אומדי נראות מרבית

טענה 14.10 – פונקציית הנראות של המודל הווקטורי המלא – תחת המודל הווקטורי המלא והפרמטריזציה שהגדרנו פונקציית הנראות היא:

$$L(\vec{\beta},\vec{\theta})=(2\pi)^{-N/2}\cdot\det(V)^{-1/2}\cdot\exp\left\{-\frac{1}{2}(Y-X\beta)^TV^{-1}(Y-X\beta)\right\}$$

טענה 14.11 – פונקציית לוג נראות של המודל הווקטורי המלא

$$\ell(\vec{\beta},\vec{\theta})=\log\left(L(\vec{\beta},\vec{\theta})\right)=-\frac{N}{2}\log(2\pi)-\frac{1}{2}\log(det(V))-\frac{1}{2}(Y-X\beta)^TV^{-1}(Y-X\beta)$$

עתה כדי למצוא את אומדי הנראות המרבית של המודל הווקטורי המלא נרצה לגזור את פונקציית לוג-הנראות ולהשוות ל-0. נזכור כי ראינו את הבסיס לכך בשיעורים הקודמים.

טענה 14.12 – על בסיס החישוב שעשינו כבר על המקרה הפשוט (המאוזן) אנחנו מקבלים:

$$\frac{\partial\ell}{\partial\beta_r}=[X^TV^{-1}(Y-X\beta)]_r$$

$$\frac{\partial \ell}{\partial \theta_s} = -\frac{1}{2} tr\left(V^{-1}\dot{V}_s\right) + \frac{1}{2}(Y_i - X\beta)^T V^{-1}\dot{V}_s V^{-1}(Y_i - X\beta)$$

בניגוד למקרה הפשוט (המאזן), במקרה הכללי אין שיטה אנליטית לפתרון אלא צריכים לפנות לשיטות נומריות.

הגדרה 14.13 – שיטות נומריות אפשריות לפתרון:

1. Newton-Raphson (NR)
2. Fisher Scoring (FS)
3. Expectation-Maximization (EM)

הגדרה 14.14 – שיטת ניוטון-רפסון – אם נגדיר וקטור הכולל את כל הפרמטרים שלנו:

$$\vec{\phi} = \begin{bmatrix} \vec{\beta} \\ \vec{\theta}^{(1)} \\ \vec{\theta}^{(2)} \end{bmatrix} \in \mathbb{R}^k$$

כאשר נניח $\theta^{(1)} \in \mathbb{R}^{u_1}, \theta^{(2)} \in \mathbb{R}^{u_2}$ וכפי שראינו כבר $\beta \in \mathbb{R}^p$ ולכן $k = u_1 + u_2 + p$.

אזי שיטת ניוטון-רפסון מגדירה את התהליך האיטרטיבי הבא:

$$\vec{\phi}^{(m+1)} = \phi^{(m)} - \nabla^2\ell\left(\vec{\phi}^{(m)}\right)^{-1}\nabla\ell\left(\phi^{(m)}\right)$$

כאשר האתחול $\vec{\phi}^{(0)}$ נקבע על ידינו. $\nabla\ell$ הוא הגרדיאנט ו-$\nabla^2\ell$ הוא ההסיאן:

$$\nabla\ell = \begin{bmatrix} \frac{\partial \ell}{\partial \phi_1} \\ \cdots \\ \frac{\partial \ell}{\partial \phi_k} \end{bmatrix}$$

$$\forall_{s,t\in\{1,\dots,k\}} := [\nabla^2\ell]_{st} = \frac{\partial^2 \ell}{\partial \phi_s \partial \phi_t}$$

הערה 14.15 – לרוב שיטת ניוטון מתכנסת, אם היא לא מתכנסת זה הרבה פעמים נובע ממספר קטן מדי של תצפיות ביחס למספר הפרמטרים.

הגדרה 14.16 – שיטת פישר - Fisher Scoring – השיטה דומה לשיטת ניוטון-רפסון והיא גם שיטה איטרטיבית המוגדרת באופן הבא:

$$\vec{\phi}^{(m+1)} = \phi^{(m)} - \left(\mathbb{E}\left[\nabla^2\ell\left(\vec{\phi}^{(m)}\right)^{-1}\right]\right)\nabla\ell\left(\phi^{(m)}\right)$$

כלומר במקום ההסיאן מסתכלים על התוחלת של ההסיאן. היתרון הוא שהתוחלת של ההסיאן תמיד חיובית לחלוטין וגם הנוסחאות יותר פשוטות.

טענה 14.17 – כפי שראינו ב-14.12,

$$\frac{\partial \ell}{\partial \beta_r} = [X^T V^{-1}(Y - X\beta)]_r$$
$$\frac{\partial \ell}{\partial \theta_r} = -\frac{1}{2} tr\left(V^{-1}\dot{V}_s\right) + \frac{1}{2}(Y_i - X\beta)^T V^{-1}\dot{V}_s V^{-1}(Y_i - X\beta)$$

נרצה למצוא את הנגזרות השניות

טענה 14.17.1 – גזירה פעמיים לפי $\vec{\beta}$

$$\frac{\partial \ell}{\partial \beta_r \beta_s} \overset{14.12}{=} \frac{\partial \ell}{\partial \beta_s}\left(X^T V^{-1}(Y - X\beta)\right)_r$$

נסמן $A = X^T V^{-1}$:

$$= \frac{\partial \ell}{\partial \beta_s}\left(A(Y - X\beta)\right)_r = \frac{\partial \ell}{\partial \beta_s}(AY - AX\beta)_r \overset{AY\ constant}{=} \frac{\partial \ell}{\partial \beta_s}(-AX\beta)_r = -\frac{\partial \ell}{\partial \beta_s}(AX\beta)_r$$

נשים לב:

$$(AX\beta)_r = \sum_{v=1}^{p}\sum_{w=1}^{p} A_{rv}X_{vw}\beta_w$$

ולכן הנגזרת במקרה זה היא

$$\frac{\partial \ell}{\partial \beta_s}(AX\beta)_r = \frac{\partial \ell}{\partial \beta_s}\sum_{v=1}^{p}\sum_{w=1}^{p} A_{rv}X_{vw}\beta_w = \sum_{v=1}^{p}\sum_{w=1}^{p} A_{rv}X_{vw}\frac{\partial \beta}{\partial \beta_s} = \sum_{v=1}^{p} A_{rv}X_{vs} = (AX)_{rs}$$

ומכאן

$$\frac{\partial \ell}{\partial \beta_r \beta_s} = -(AX)_{rs} = -(X^T V^{-1} X)_{rs}$$

ובסך הכל

$$\left(\boldsymbol{\ell^{\beta\beta}}\right)_{rs} = \frac{\boldsymbol{\partial^2 \ell}}{\boldsymbol{\partial \beta_r \partial \beta_s}} \Rightarrow \boldsymbol{\ell^{\beta\beta}} = -\left(\boldsymbol{X^T V^{-1} X}\right)$$

טענה 14.17.2 – גזירה לפי β ואז לפי θ

$$\frac{\partial \ell}{\partial \beta_r \theta_s} = \frac{\partial \ell}{\partial \theta_s}\left[\frac{\partial \ell}{\partial \beta_r}\right] \overset{14.12}{=} \frac{\partial \ell}{\partial \theta_s}[X^T V^{-1}(Y - X\beta)]_r = -\left[(X^T V^{-1}\dot{V}_S V^{-1}(Y - X\beta)\right]_r$$

כאשר

$$\frac{\partial}{\partial \theta_s}V^{-1} = -V^{-1}\dot{V}_s V^{-1}$$
$$\dot{V}_s = \frac{\partial V}{\partial \theta_s}$$

ובסך הכל

$$\frac{\ell}{\partial \beta_r \partial \theta_s} = -\left[(X^T V^{-1}\dot{V}_S V^{-1}(Y - X\beta)\right]_r \Rightarrow \boldsymbol{\ell^{\beta\theta}} = -(\boldsymbol{X^T V^{-1}\dot{V}_S V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta})$$

טענה 14.17.3 – גזירה לפי θ ואז לפי β - בגלל שמדובר בהסיאן אז המטריצה סימטרית ולכן גם

$$\boldsymbol{\ell^{\theta\beta}} = -[\left(\boldsymbol{X^T V^{-1}\dot{V}_S V^{-1}}(\boldsymbol{Y} - \boldsymbol{X\beta})\right]^{\boldsymbol{T}}$$

טענה 14.17.4 – גזירה פעמיים לפי θ

$$\frac{\ell}{\partial \theta_r \partial \theta_s} \overset{14.12}{=} \frac{\ell}{\partial \theta_s}\left[-\frac{1}{2}tr\left(V^{-1}\dot{V}_r\right) + \frac{1}{2}(Y - X\beta)^T V^{-1}\dot{V}_r V^{-1}(Y - X\beta)\right]$$

$$= \frac{\partial \ell}{\partial \theta_s}\left[-\frac{1}{2}tr\left(V^{-1}\dot{V}_r\right)\right] + \frac{\partial \ell}{\partial \theta_s}\left[\frac{1}{2}(Y - X\beta)^T V^{-1}\dot{V}_r V^{-1}(Y - X\beta)\right]$$

בשני המקרים נשתמש בכלל שראינו לנגזרת על מכפלת מטריצות

$$\frac{\partial}{\partial \theta}AB = \frac{\partial A}{\partial \theta}B + A\frac{\partial B}{\partial \theta}$$

עבור הביטוי הראשון נקבל:

$$\frac{\partial \ell}{\partial \theta_s}\left[-\frac{1}{2}tr\left(V^{-1}\dot{V}_r\right)\right] = -\frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - \frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right)$$

עבור הביטוי השני נקבל:

$$\frac{\partial \ell}{\partial \theta_s}\left[\frac{1}{2}(Y - X\beta)^T V^{-1}\dot{V}_r V^{-1}(Y - X\beta)\right] = \frac{1}{2}(Y - X\beta)^T\left[\frac{\partial}{\partial \theta_s}V^{-1}\dot{V}_r V^{-1}\right](Y - X\beta)$$

נשתמש פעמים חוזרות בפיתוח שראינו ב-14.17.2. הנגזרת הפנימית היא:

$$\begin{aligned}
&Q = \frac{\partial}{\partial \theta_s}V^{-1}\dot{V}_r V^{-1} \overset{A=V^{-1},B=\dot{V}_r V^{-1}}{=} \\
&\left(\frac{\partial}{\partial \theta_s}V^{-1}\right)\dot{V}_r V^{-1} + V^{-1}\left(\frac{\partial}{\partial \theta_s}\dot{V}_r V^{-1}\right) \overset{A=\dot{V}_r,B=V^{-1}}{=} \\
&\left(\frac{\partial}{\partial \theta_s}V^{-1}\right)\dot{V}_r V^{-1} + V^{-1}\left[\left(\frac{\partial}{\partial \theta_s}\dot{V}_r\right)V^{-1} + \dot{V}_r\left(\frac{\partial}{\partial \theta_s}V^{-1}\right)\right] = \\
&\left(\frac{\partial}{\partial \theta_s}V^{-1}\right)\dot{V}_r V^{-1} + V^{-1}\left(\frac{\partial}{\partial \theta_s}\dot{V}_r\right)V^{-1} + V^{-1}\dot{V}_r\left(\frac{\partial}{\partial \theta_s}V^{-1}\right) \overset{\frac{\partial}{\partial \theta_s}V^{-1}=V^{-1}\frac{\partial V}{\partial \theta_s}V^{-1}}{=} \\
&\boldsymbol{\left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r V^{-1} + V^{-1}\ddot{V}_{rs}V^{-1} - V^{-1}\dot{V}_r V^{-1}\dot{V}_s V^{-1}}
\end{aligned}$$

כלומר:

$$\frac{\partial \ell}{\partial \theta_s}\left[\frac{1}{2}(Y - X\beta)^T V^{-1}\dot{V}_r V^{-1}(Y - X\beta)\right] = \frac{1}{2}(Y - X\beta)^T Q(Y - X\beta)$$

ובסך הכל

$$\boldsymbol{\begin{aligned}
\frac{\ell(\beta,\theta)}{\partial \theta_r \partial \theta_s} &= -\frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - \frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}(Y - X\beta)^T\left[\frac{\partial}{\partial \theta_s}V^{-1}\dot{V}_r V^{-1}\right](Y - X\beta) \\
&= -\frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - \frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}(Y - X\beta)^T Q(Y - X\beta)
\end{aligned}}$$

מסקנת ביניים 14.18 – חישבנו אם כן את כל הנגזרות הראשונות והשניות של המודל הכללי. הדבר מאפשר לנו לפי הצורך להשתמש בשיטת ניוטון-רפסון על מנת למצוא את אומד הנראות המרבית.

טענה 14.19 – כדי להשתמש בשיטת פישר נצטרך לחשב גם את התוחלות (לפי Y).

$$\mathbb{E}\left[\ell^{\beta\beta}\right] = -\left(X^T V^{-1} X\right)$$

$$\mathbb{E}\left[\frac{\ell}{\partial \beta_r \partial \theta_s}\right] = \mathbb{E}\left[-\left[\left(X^T V^{-1}\dot{V}_s V^{-1}(Y - X\beta)\right]_r\right] = -X^T V^{-1}\dot{V}_s V^{-1}\mathbb{E}[Y - X\beta] = 0$$

בגלל ש-$\mathbb{E}[Y] = X\beta$ ולכן $\mathbb{E}[Y - X\beta] = \mathbb{E}[Y] - \mathbb{E}[X\beta] = X\beta - X\beta = 0$.

לבסוף

$$\mathbb{E}\left[(Y - X\beta)^T Q(Y - X\beta)\right]$$

$$= \mathbb{E}\left[tr\left((Y - X\beta)^T Q(Y - X\beta)\right)\right]$$
$$= \mathbb{E}[tr(Q(Y - X\beta)(Y - X\beta)^T)]$$
$$= tr(Q \cdot \mathbb{E}[(Y - X\beta)(Y - X\beta)^T])$$
$$= tr(Q \cdot V)$$
$$= tr\left(\left(\left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r V^{-1} + V^{-1}\ddot{V}_{rs}V^{-1} - V^{-1}\dot{V}_r V^{-1}\dot{V}_s V^{-1}\right) V\right)$$
$$= tr\left(\left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r V^{-1}V + V^{-1}\ddot{V}_{rs}V^{-1}V - V^{-1}\dot{V}_r V^{-1}\dot{V}_s V^{-1}V\right)$$
$$= tr\left(\left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r + V^{-1}\ddot{V}_{rs} - V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right)$$

לכן

$$\mathbb{E}\left[\frac{\ell}{\partial\theta_r\partial\theta_s}\right] = \mathbb{E}\left[-\frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - \frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}(Y - X\beta)^T Q(Y - X\beta)\right]$$
$$= -\frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - -\frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}tr\left(\left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r + V^{-1}\ddot{V}_{rs} - V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right)$$
$$= -\frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - \frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) + \frac{1}{2}tr\left[V^{-1}\ddot{V}_{rs}\right]$$
$$-\frac{1}{2}tr\left[V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right]$$
$$= -\frac{1}{2}tr\left[V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right]$$

סיכום 14.20 נגזרות ונגזרות שניות של פונקציית לוג הנראות במודל הווקטורי המלא – להלן סיכום קצר של וקטורי הנגזרות ומטריצת הנגזרות השניות של פונקציית לוג הנראות במודל הווקטורי המלא של המודל הלינארי המעורב הכללי.

נגזרות ראשונות

$$\boldsymbol{\frac{\partial\ell}{\partial\beta_r} = \left[X^T V^{-1}(Y - X\beta)\right]_r}$$
$$\boldsymbol{\frac{\partial\ell}{\partial\theta_r} = -\frac{1}{2}tr\left(V^{-1}\dot{V}_s\right) + \frac{1}{2}(Y - X\beta)^T V^{-1}\dot{V}_s V^{-1}(Y - X\beta)}$$

וקטור הנגזרות הראשונות הוא:

$$\nabla\ell = \begin{bmatrix} \partial\ell/\partial\beta_1 \\ \dots \\ \partial\ell/\partial\beta_k \\ \partial\ell/\partial\theta_1 \\ \dots \\ \partial\ell/\partial\theta_{u_1+u_2} \end{bmatrix} = \begin{bmatrix} [X^T V^{-1}(Y - X\beta)]_1 \\ \dots \\ [X^T V^{-1}(Y - X\beta)]_k \\ -\frac{1}{2}tr\left(V^{-1}\dot{V}_1\right) + \frac{1}{2}(Y - X\beta)^T V^{-1}\dot{V}_1 V^{-1}(Y - X\beta) \\ \dots \\ -\frac{1}{2}tr\left(V^{-1}\dot{V}_{V_{u_1+u_2}}\right) + \frac{1}{2}(Y - X\beta)^T V^{-1}\dot{V}_{V_{u_1+u_2}} V^{-1}(Y - X\beta) \end{bmatrix}$$

נגזרות שניות:

$$\boldsymbol{\left(\ell^{\beta\beta}\right)_{rs} = \frac{\partial^2\ell}{\partial\beta_r\partial\beta_s} \Rightarrow \ell^{\beta\beta} = -\left(X^T V^{-1}X\right)}$$
$$\boldsymbol{\ell^{\beta\theta} = -(X^T V^{-1}\dot{V}_s V^{-1}(Y - X\beta)}$$
$$\boldsymbol{\ell^{\theta\beta} = -[\left(X^T V^{-1}\dot{V}_s V^{-1}(Y - X\beta)\right]^T}$$
$$\boldsymbol{\frac{\ell}{\partial\theta_r\partial\theta_s} = -\frac{1}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - \frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}(Y - X\beta)^T\left[\frac{\partial}{\partial\theta_s}V^{-1}\dot{V}_r V^{-1}\right](Y - X\beta)}$$

$$= -\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{tr}\left(-\boldsymbol{V}^{-\mathbf{1}}\dot{\boldsymbol{V}}_{\boldsymbol{s}}\boldsymbol{V}^{-\mathbf{1}}\dot{\boldsymbol{V}}_{\boldsymbol{r}}\right) - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{tr}\left(\boldsymbol{V}^{-\mathbf{1}}\ddot{\boldsymbol{V}}_{\boldsymbol{rs}}\right) + \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{Y}-\boldsymbol{X\beta})^{\boldsymbol{T}}\boldsymbol{Q}(\boldsymbol{Y}-\boldsymbol{X\beta})$$

נסמן

$$\left[\ell^{\theta\theta}\right]_{rs} = \frac{\ell}{\partial\theta_r\partial\theta_s}$$

מטריצת ההסיאן היא:

$$\nabla^2\ell = \begin{bmatrix} \ell^{\beta\beta} & \ell^{\beta\theta} \\ \ell^{\theta\beta} & \ell^{\theta\theta} \end{bmatrix}$$
$$= \begin{bmatrix} -(X^TV^{-1}X) & -(X^TV^{-1}\dot{V}_SV^{-1}(Y-X\beta) \\ -\left[\left(X^TV^{-1}\dot{V}_SV^{-1}(Y-X\beta)\right]^T\right. & -\frac{1}{2}tr\left(-V^{-1}\dot{V}_sV^{-1}\dot{V}_r\right) - \frac{1}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}(Y-X\beta)^TQ(Y-X\beta) \end{bmatrix}$$

תוחלת מטריצת ההסיאן:

$$\mathbb{E}\left[\ell^{\beta\beta}\right] = -(X^TV^{-1}X)$$
$$\mathbb{E}\left[\ell^{\beta\theta}\right] = \mathbb{E}\left[\ell^{\theta\beta}\right] = 0$$
$$\mathbb{E}\left[\ell^{\theta\theta}\right] = -\frac{1}{2}tr\left[V^{-1}\dot{V}_rV^{-1}\dot{V}_s\right]$$

ובסך הכל:

$$\mathbb{E}[\nabla^2\ell] = \mathbb{E}\begin{bmatrix} \ell^{\beta\beta} & \ell^{\beta\theta} \\ \ell^{\theta\beta} & \ell^{\theta\theta} \end{bmatrix} = \begin{bmatrix} \mathbb{E}\left[\ell^{\beta\beta}\right] & \mathbb{E}\left[\ell^{\beta\theta}\right] \\ \mathbb{E}\left[\ell^{\theta\beta}\right] & \mathbb{E}\left[\ell^{\theta\theta}\right] \end{bmatrix} = \begin{bmatrix} -(X^TV^{-1}X) & 0 \\ 0 & -\frac{1}{2}tr\left[V^{-1}\dot{V}_rV^{-1}\dot{V}_s\right] \end{bmatrix}$$

52801 מודלים סטטיסטיים למתקדמים 2023א
מרצה: פרופ' דוד צוקר
סיכום: אריאל וישנה

סיכום 14.21 – השוואה בין שלבים שונים של פיתוח המשוואות במקרה הפשוט (טיוטה)

		תצפית בודדת בנקודת זמן בודדת	תצפית בודדת בכל נקודות הזמן (מספר הנקודות קבוע לכל הדגימות - $(t_1, \dots, t_J$)	כלל התצפיות בכל נקודות הזמן
מודל	משוואת המודל	$Y_{ij} = \beta_{i1} + \beta_{2i} t_j + \varepsilon_{ij}$ $Y_{ij} = X_j^T \vec{\beta}_i + \varepsilon_{ij}$	$\vec{Y}_i = X\vec{\beta}_i + \vec{\varepsilon}_i$	$\vec{Y} = X\vec{\beta} + Z\vec{b} + \varepsilon$
הגדרות וגדלים	משתנה מוסבר	$Y_{ij} \in \mathbb{R}$	$\vec{Y}_i = \begin{bmatrix} Y_{i1} \\ \dots \\ Y_{iJ} \end{bmatrix} \in \mathbb{R}^J$	$\vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \dots \\ \vec{Y}_n \end{bmatrix} \in \mathbb{R}^{J \times n}$
		$X_j^T = [1 \quad t_J]$	$X = \begin{bmatrix} 1 & t_1 \\ \dots & \dots \\ 1 & t_J \end{bmatrix}$	
	אפקטים קבועים	$\beta_i = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \in \mathbb{R}^{2 \times 1}$	$\beta_i = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \in \mathbb{R}^{2 \times 1}$	
	שגיאות	$\varepsilon_{ij} \in \mathbb{R}$	$\varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \dots \\ \varepsilon_{iJ} \end{bmatrix} \in \mathbb{R}^J$	$\vec{\varepsilon} = \begin{bmatrix} \vec{\varepsilon}_1 \\ \dots \\ \vec{\varepsilon}_n \end{bmatrix} \in \mathbb{R}^N$
	שונויות		$V = XAX^T + \sigma^2 I$	
התפלגויות			$\vec{\beta}_i \sim N(\beta, A) = N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{pmatrix}\right)$	
		$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$	$\vec{\varepsilon}_i \sim N(0, \sigma^2 I)$	
		$Y_{ij} \sim N(X_{ij}^T \beta, \sigma^2)$	$Y_i \sim N(0, V)$	$Y \sim N(X\beta, V)$
אמידה	פרמטריזציה	$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \theta = \begin{bmatrix} A_{11} \\ A_{12} \\ A_{21} \\ A_{22} \end{bmatrix}$		
	נראות מרבית	$\hat{\beta}_i = (X^T X)^{-1} X^T Y_i$ $\vec{\beta} = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_i$ $\hat{\sigma}^2 = \frac{1}{n(J-2)} \sum_{i=1}^n e_i^T e_i$ $\hat{C} = \frac{1}{n-1} \sum_{i=1}^n (\hat{\beta}_i - \hat{\beta})(\hat{\beta}_i - \hat{\beta})^T$		
	התפלגויות	$\hat{\beta}_i \sim N(\beta, A + \sigma^2 (X^T X)^{-1})$ $\hat{\beta} \sim N_2\left(\beta, \frac{1}{n} C\right)$ $n(\hat{\beta} - \beta^{(0)})^T C^{-1} (\hat{\beta} - \beta^{(0)}) \overset{H_0}{\sim} \chi_2^2$ $(n-1)\hat{C} \sim W_2(n-1, C)$		
משתני עזר		$C = Cov(\hat{\beta}_i) = A + \sigma^2 (X^T X)^{-1}$ $\Omega = X^T V^{-1} X$ $V^{-1} = \left(\frac{1}{\sigma^2}\right) I - \left(\frac{1}{\sigma^2}\right)^2 X \Gamma X^T$ $\Gamma = (A^{-1} + (\sigma^2)^{-1} X^T X)^{-1}$		

סיכום 14.22 – השוואה בין שלבים שונים של פיתוח המשוואות במקרה המעורב הכללי

סיכום של המשוואות והגדלים במקרים של המקרה המעורב הכללי:

1. תצפית בודדת i בנקודת זמן בודדת j
2. תצפית בודדת i בכלל נקודות הזמן. מספר נקודות הזמן במודל הכללי הוא J_i כלומר תלוי בתצפית עליה אנחנו מסתכלים. בתוך המדידות השונות על התצפית אנחנו לא מניחים אי-תלויות אלא מניחים תלות בין השגיאות ותלות בין האפקטים המקריים.
3. כלל התצפיות בכלל נקודות הזמן. מספר התצפיות בכל נקודות הזמן הוא $N = \sum_{i=1}^{n} J_i$. אנחנו כן מניחים אי-תלות בין תצפיות שונות ולכן מטריצות השונויות הן מטריצות בלוקים אלכסוניות, שמבטאות את התלות בתוך אותה תצפית i ואת אי-התלות בין כל תצפיות שונות $i \neq j$.

		תצפית בודדת בנקודת זמן בודדת	תצפית בודדת בכל נקודות הזמן	כלל התצפיות בכל נקודות הזמן
מודל	משוואת המודל	$Y_{ij} = X_{ij}^T\vec{\beta} + Z_{ij}^T\vec{b}_i + \varepsilon_{ij}$	$\vec{Y}_i = X_i\vec{\beta} + Z_i\vec{b}_i + \vec{\varepsilon}_i$	$\vec{Y} = X\vec{\beta} + Z\vec{b} + \varepsilon$
הגדרות וגדלים	משתנה מוסבר	$Y_{ij} \in \mathbb{R}$	$\vec{Y}_i = \begin{bmatrix} Y_{i1} \\ \dots \\ Y_{iJ_i} \end{bmatrix} \in \mathbb{R}^{J_i}$	$\vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \dots \\ \vec{Y}_n \end{bmatrix} \in \mathbb{R}^{N}$
		$X_{ij} = \begin{bmatrix} X_{i1} \\ \dots \\ X_{ip} \end{bmatrix} \in \mathbb{R}^{p\times 1}$	$X_i = \begin{bmatrix} -X_{i1}^T- \\ \dots \\ -X_{iJ_i}^T- \end{bmatrix} \in \mathbb{R}^{J_i\times p}$	$X = \begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix} \in \mathbb{R}^{N\times p}$
	אפקטים קבועים	$\beta = \begin{bmatrix} \beta_1 \\ \dots \\ \beta_p \end{bmatrix} \in \mathbb{R}^{p\times 1}$	$\beta = \begin{bmatrix} \beta_1 \\ \dots \\ \beta_p \end{bmatrix} \in \mathbb{R}^{p\times 1}$	$\beta = \begin{bmatrix} \beta_1 \\ \dots \\ \beta_p \end{bmatrix} \in \mathbb{R}^{p\times 1}$
		$Z_{ij} = \begin{bmatrix} Z_{i1} \\ \dots \\ Z_{iq} \end{bmatrix} \in \mathbb{R}^{q\times 1}$	$Z_i = \begin{bmatrix} -Z_{i1}^T- \\ \dots \\ -Z_{iJ_i}^T- \end{bmatrix} \in \mathbb{R}^{J_i\times q}$	$Z = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Z_n \end{bmatrix} \in \mathbb{R}^{N\times nq}$
	אפקטים משתנים	$b_i = \begin{bmatrix} b_{i1} \\ \dots \\ b_{iq} \end{bmatrix} \in \mathbb{R}^{q\times 1}$	$\vec{b}_i = \begin{bmatrix} b_{i1} \\ \dots \\ b_{iq} \end{bmatrix} \in \mathbb{R}^{q\times 1}$	$\vec{b} = \begin{bmatrix} \vec{b}_1 \\ \dots \\ \vec{b}_n \end{bmatrix} \in \mathbb{R}^{nq}$
	שגיאות	$\varepsilon_{ij} \in \mathbb{R}$	$\varepsilon_i \in \mathbb{R}^{J_i}$	$\vec{\varepsilon} = \begin{bmatrix} \vec{\varepsilon}_1 \\ \dots \\ \vec{\varepsilon}_n \end{bmatrix} \in \mathbb{R}^{N}$
	שונויות		$V_i = Z_i G\left(\theta^{(1)}\right) Z_i^T + R_i\left(\theta^{(2)}\right) \in \mathbb{R}^{J_i\times J_i}$	$V = \begin{bmatrix} V_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & V_n \end{bmatrix} \in \mathbb{R}^{N\times N}$
התפלגויות		$\vec{b}_i \sim N\left(0, G\left(\theta^{(1)}\right)\right)$	$\vec{b}_i \sim N\left(0, G\left(\theta^{(1)}\right)\right)$	$\vec{b}$
		$\varepsilon_{ij} \sim N(0, \sigma^2)$	$\varepsilon_i \sim N\left(0, R_i\left(\theta^{(2)}\right)\right)$	$\vec{\varepsilon}$
			$Y_i \sim N(X_i\beta, V_i)$	$Y \sim N(X\beta, V)$

שיעור 15

תזכורת 15.1 – אנחנו עוסקים כזכור במודל הלינארי המעורב הכללי, שבו מתקיים כפי שראינו:

$$Y_i = X_i\vec{\beta} + Z_i\vec{b}_i + \varepsilon_i$$

ובכתיב וקטורי מלא:

$$Y = X\vec{\beta} + Z\vec{b} + \vec{\varepsilon}$$

ומתקיים

$$Y \sim N(X\beta, V)$$

אשר V מטריצת בלוקים אלכסונית $V = diag\{V_1, \dots, V_n\}$ ולכל $1 \le i \le n$ המטריצה V_i מוגדרת על ידי:

$$V_i = Z_i G Z_i^T + R_i$$

כפי שראינו הפרמטרים מתפלגים רב-נורמלית עם הפרמטריזציה:

$$b_i \sim N\left(0, G\left(\theta^{(1)}\right)\right)$$
$$\varepsilon_i \sim N\left(0, R_i\left(\theta^{(2)}\right)\right)$$

תוך ההגדרות הללו ראינו כי הפרמטריזציה היא באמצעות הפרמטרים של $\beta, \theta^{(1)}, \theta^{(2)}$ וראינו את האומדים לפי הנראות המרבית של פונקציית לוג-הנראות:

$$\ell\left(\vec{\beta}, \vec{\theta}\right) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(det(V)) - \frac{1}{2}(Y - X\beta)^T V^{-1}(Y - X\beta)$$

האמידה נעשית באמצעות גזירת פונקצית לוג-הנראות והשוואה לאפס. כפי שראינו אין פתרון סגור לאומדים וצריך לפתור באמצעות שיטות נומריות לפתרון כמו ניוטון-רפסון או Fisher-scoring.

הערה 15.2 – Generalized Least Squares (GLS) - אם V היה מטריצת היחידה היינו מקבלים את אומד הריבועים הפחותים:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

במקרה הכללי שלנו, ראינו כי הנגזרת של הנראות לפי אחת מהקואורדינטות של β היא

$$\frac{\partial \ell}{\partial \beta_r} = [X^T V^{-1}(Y - X\beta)]$$

כלומר מערכת המשוואות שצריך לפתור היא מערכת שבה כל משוואה $1 \le r \le p$ היא מהצורה:

$$[X^T V^{-1}(Y - X\beta)]_r = 0$$

ובסך הכל נדרש אם כן:

$$[X^T V^{-1}(Y - X\beta)] = \vec{0} \Rightarrow$$
$$X^T V^{-1} X\beta = X^T V^{-1} Y \Rightarrow$$
$$\hat{\beta} = (X^T V X)^{-1} X^T V^{-1} Y$$

ניתן לראות כאן את הדמיון בין אומד הריבועים הפחותים הרגיל לבין האומד שיצא במקרה שלנו. בגלל ההכללה הזאת האומד ל-β נקרא אומד ריבועים פחותים מוכלל (Generalized Least Squares). כמובן שאין בהכרח עדין פתרון סגור ל-$\hat{\beta}$ כי המטריצה V תלויה בפרמטריזציה שלנו לא רק של $\vec{\beta}$ אלא גם של $\theta^{(1)}, \theta^{(2)}$.

נושא 2.3.2 – המודל הלינארי המעורב הכללי – התפלגות אומדי נראות מרבית

משפט 15.3 התפלגות אומד נראות מרבית ומטריצת האינפורמציה של פישר – יהיו $W_1, \dots, W_n$ וקטורי תצפיות בלתי-תלויים המתפלגים בהתפלגות כלשהי שניתנת לפרמטריזציה של $\vec{\phi} \in \mathbb{R}^k$. יהי $\hat{\phi}$ אומד נראות מרבית ל-ϕ. מטריצת האינפורמציה של פישר (Fisher Information Matrix) מוגדרת על ידי:

$$\mathfrak{I}(\phi) = -\mathbb{E}[\nabla^2 \ell]$$

אזי

$$\hat{\phi} \dot{\sim} N(\phi, \mathfrak{I}(\phi)^{-1})$$

טענה 15.4 – במקרה שלנו, ראינו כי:

$$\mathbb{E}\left[\ell^{\beta\beta}\right] = -(X^T V^{-1} X)$$
$$\mathbb{E}\left[\ell^{\beta\theta}\right] = \mathbb{E}\left[\ell^{\theta\beta}\right] = 0$$

$$\mathbb{E}\left[\left(\ell^{\theta\theta}\right)_{rs}\right] = -\frac{1}{2} tr\left(\left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r + V^{-1}\ddot{V}_{rs} - V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right)$$

וניתן להראות (שיעורי בית תרגיל 7)

$$\mathbb{E}\left[\left(\ell^{\theta\theta}\right)_{rs}\right] = -\frac{1}{2} tr\left(V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right)$$

נסמן

$$\mathbb{E}\left[\left(\ell^{\theta\theta}\right)_{rs}\right] = \mathfrak{I}^{\theta\theta}$$

בסך הכל:

$$\mathfrak{I}(\beta, \theta) = \begin{bmatrix} (X^T V^{-1} X) & 0 \\ 0 & \mathfrak{I}^{\theta\theta} \end{bmatrix}$$

אזי לפי משפט 15.3 נקבל:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\theta} \end{bmatrix} \sim N\left(\begin{bmatrix} \beta \\ \theta \end{bmatrix}, \begin{bmatrix} (X^T V^{-1} X)^{-1} & 0 \\ 0 & \left(\mathfrak{I}^{\theta\theta}\right)^{-1} \end{bmatrix}\right)$$

נושא 2.3.3 – המודל הלינארי המעורב הכללי – רווחי סמך, אזורי סמך, ובדיקת השערות לאומדי נראות מרבית

מסקנה 15.5 – רווחי סמך לאומדי הנראות המרבית – קואורדינטה יחידה – על בסיס ההתפלגות שהראינו ניתן לבנות רווחי סמך ל-β_r על ידי:

$$\boldsymbol{\hat{\beta}_r \pm z_{1-\alpha/2}\sqrt{([X^T V^{-1} X]^{-1})_{rr}}}$$

מסקנה 15.6 – אזור סמך לאומדי הנראות המרבית – וקטור מלא - נשים לב כי ניתן לבנות אזור סמך עבור וקטור הפרמטרים כולו באמצעות משפט שראינו בשיעורים קודמים:

$$\left(\hat{\beta} - \beta\right)^T \left[\widehat{Cov}\left(\hat{\beta} - \beta\right)\right]^{-1} \left(\hat{\beta} - \beta\right) \dot{\sim} \chi^2_p$$

זאת בשל המשפט שראינו שאם $W \sim N_d(0, V)$ אזי $W^T V^{-1} W \sim \chi^2_d$. לפיכך אזור סמך לכלל הווקטור הוא:

$$\left\{\boldsymbol{\beta} \in \mathbb{R}^{\boldsymbol{p}}: \left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)^{\boldsymbol{T}}\left[\widehat{\boldsymbol{Cov}}\left(\widehat{\boldsymbol{\beta}}\right)\right]^{-\mathbf{1}}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right) \leq \boldsymbol{\chi}^{\mathbf{2}}_{\boldsymbol{p};(\mathbf{1}-\boldsymbol{\alpha})}\right\}$$

באופן אחר ניתן לרשום זאת באמצעות פיתוח של מטריצת השונויות:

$$Cov(\hat{\beta}) = (X^T V^{-1} X)^{-1} \Rightarrow$$
$$\left[Cov(\hat{\beta})\right]^{-1} = X^T V^{-1} X$$

ולכן ביטוי נוסף לרווח הסמך הינו

$$\left\{\boldsymbol{\beta} \in \mathbb{R}^{\boldsymbol{p}}: \left(\boldsymbol{X\beta}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\left(\boldsymbol{X\beta}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \leq \boldsymbol{\chi}^{\mathbf{2}}_{\boldsymbol{p};(\mathbf{1}-\boldsymbol{\alpha})}\right\}$$

הגדרה 15.7 רווחי סמך לאומדי נראות מרבית באמצעות התפלגות $\boldsymbol{t}$ – במקרה של גודל מדגם קטן עד בינוני, קיימות שיטות שאומדות את ההתפלגות לא כהתפלגות נורמלית אלא לפי התפלגות t. זאת על ידי:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}} \pm \boldsymbol{t}_{\boldsymbol{d};(\mathbf{1}-\boldsymbol{\alpha})}\sqrt{\left(\left[\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\boldsymbol{X}\right]^{-\mathbf{1}}\right)_{\boldsymbol{rr}}}$$

קיימות כמה שיטות להגדרה של מספר דרגות החופש d.

הגדרה 15.8. – שיטת Satterthwaite – אנחנו מסתכלים על $\hat{\beta}_r$, אומד הנראות המרבית. נניח כקירוב כי:

$$\widehat{\mathbb{V}}(\hat{\beta}_r) \dot{\sim} \mathbb{V}\left(\frac{\chi^2_d}{d}\right)$$

ונמצא ערך של d מתאים לפי מומנטים של $\widehat{\mathbb{V}}(\hat{\beta}_r)$.

התוחלת של האומד לשונות היא:

$$\mathbb{E}\left[\widehat{\mathbb{V}}(\hat{\beta}_r)\right] \doteq \mathbb{V}(\beta_r) \cdot \mathbb{E}\left[\frac{\chi^2_d}{d}\right] = \mathbb{V}(\beta_r) \cdot 1 = \mathbb{V}(\beta_r)$$

והשונות של האומד לשונות היא:

$$\mathbb{V}\left[\widehat{\mathbb{V}}(\hat{\beta}_r)\right] \doteq \left(\mathbb{V}(\beta_r)\right)^2 \cdot \frac{1}{d_2}\mathbb{V}(\chi^2_d) = \left(\mathbb{V}(\beta_r)\right)^2 \cdot \frac{1}{d_2} 2d = \frac{2}{d}\left[\mathbb{V}(\hat{\beta}_r)\right]^2 \Rightarrow$$
$$\mathbb{V}\left[\widehat{\mathbb{V}}(\hat{\beta}_r)\right] = \frac{2\left[\mathbb{V}(\hat{\beta}_r)\right]^2}{d} \Rightarrow d = \frac{2\left[\mathbb{V}(\hat{\beta}_r)\right]^2}{\mathbb{V}\left[\widehat{\mathbb{V}}(\hat{\beta}_r)\right]}$$

מכאן ה-d המוצע הוא:

$$d = \frac{2\left(\widehat{\mathbb{V}}(\hat{\beta}_r)\right)^2}{\widehat{\mathbb{V}}\left(\widehat{\mathbb{V}}(\hat{\beta}_r)\right)}$$

כאשר בפועל משתמשים בקירובים שפותחו עבור $\mathbb{V}\left(\widehat{\mathbb{V}}(\hat{\beta}_r)\right)$.

הערה 15.8.1 – כזכור נסמכנו בחישובים על $\mathbb{E}[\chi^2_d] = d, \mathbb{V}[\chi^2_d] = 2d$.

הערה 15.8.2 – נשים לב שאומד לפרמטרים הוא משתנה מקרי ואותו אנחנו מנסים לאמוד, ולכן יש לו גם שונות ותוחלת

$$\hat{\beta} \sim N(\beta, \Omega^{\beta\beta})$$
$$\Omega^{\beta\beta} = (X^T V^{-1}(\theta) X)^{-1}$$
$$\mathbb{V}(\hat{\beta}_r) = [(X^T V^{-1}(\theta) X)^{-1}]_{rr} \Rightarrow$$

$$\widehat{\mathbb{V}}(\hat{\beta}_r) = \left[\left(X^T V^{-1}(\hat{\theta})X\right)^{-1}\right]_{rr} \Rightarrow$$
$$\mathbb{V}\left(\widehat{\mathbb{V}}(\beta_r)\right) = \mathbb{V}\left(\left[\left(X^T V^{-1}(\hat{\theta})X\right)^{-1}\right]_{rr}\right)$$

מסקנה 15.9 – באופן דומה, אם נרצה לבצע בדיקת השערות

$$H_0: \beta_r = 0$$
$$H_1: \beta_r \neq 0$$

נוכל לעשות זאת על ידי:

$$\boldsymbol{\zeta_r} = \frac{\boldsymbol{\widehat{\beta}_r}}{\sqrt{([\boldsymbol{X^T V^{-1} X}]^{-1})_{\boldsymbol{rr}}}} \overset{\boldsymbol{H_0}}{\sim} \boldsymbol{N(0,1)}$$

אפשר גם להשתמש בהתפלגות t כמתואר לעיל.

טענה 15.10 רווח סמך ל-$\boldsymbol{X_{ij}^T \beta}$ – כזכור

$$Y_{ij} = X_{ij}^T \beta + Z_{ij}^T b_i + \varepsilon_{ij}$$

ולכן התוחלת של Y_{ij} היא

$$\mathbb{E}[Y_{ij}] = X_{ij}^T \beta$$

כפי שראינו

$$\hat{\beta} \dot{\sim} N(\beta, \Omega^{\beta\beta}) \Rightarrow$$
$$X_{ij}^T \hat{\beta} \dot{\sim} N(X_{ij}^T \beta, X_{ij}^T \Omega^{\beta\beta} X_{ij})$$

לכן ניתן לייצר את רווחי הסמך עבור $X_{ij}^T \beta$:

$$\boldsymbol{X_{ij}^T \widehat{\beta} \pm t_{d;(1-\alpha)} \sqrt{X_{ij}^T \widehat{\Omega}^{\beta\beta} X_{ij}}}$$

דוגמה 15.10.1 – האומד הנ"ל שימושי במקרה שבו למשל אנחנו מנסים לאמוד את התצפית הבאה בסדרה על בסיס התצפיות הקודמות. למשל עבור $t_{i1}, \dots, t_{i,J_i}$ נניח שאנחנו מנסים לאמוד את התצפית t_{i,J_i+1}. במצב כזה אנחנו מעוניינים למשל להעריך מה התוחלת:

$$\mathbb{E}[Y_{i,J_i+1}] = [1 \quad t_{i,J_i+1}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

נושא 2.3.4 – המודל הלינארי המעורב הכללי – אמידה של אפקטים מקריים

טענה 15.11 – רווח סמך ל-$\boldsymbol{\begin{bmatrix} Y_i \\ b_i \end{bmatrix}}$ – כזכור $Y_i = X_i \beta + Z_i b_i + \varepsilon_i$. ניתן לפיכך להסתכל על ההתפלגות:

$$\begin{bmatrix} Y_i \\ b_i \end{bmatrix} \sim N\left(\begin{bmatrix} X_i \beta \\ 0 \end{bmatrix}, \begin{bmatrix} V_i & Cov(Y_i, b_i) \\ [Cov(Y_i, b_i)]^T & G \end{bmatrix}\right) = N\left(\begin{bmatrix} X_i \beta \\ 0 \end{bmatrix}, \begin{bmatrix} V_i & Z_i G \\ (Z_i G)^T & G \end{bmatrix}\right)$$

שכן

$$Cov(Y_i, b_i) = Cov(X\beta + Z_i b_i + \varepsilon_i, b_i) = Cov(Z_i b_i, b_i) = Z_i Cov(b_i, b_i) = Z_i \mathbb{V}(b_i) = Z_i G$$

תזכורת 15.11.1 – כזכור ראינו בשיעורים קודמים התפלגות מותנית של וקטור תוחלות המתפלג רב-נורמלית. ראינו כי אם:

$$\begin{bmatrix} W^{(1)} \\ W^{(2)} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \begin{bmatrix} V^{(11)} & V^{(12)} \\ V^{(21)} & V^{(22)} \end{bmatrix}\right)$$

אזי

$$W^{(2)}|W^{(1)} \sim N\left(\mu^{(2)} - V^{(21)}\left(V^{(11)}\right)^{-1}\left(W^{(1)} - \mu^{(1)}\right), V^{(22)} - V^{(21)}{V^{(11)}}^{-1}V^{(12)}\right)$$

המשך טענה 15.11 – במקרה שלנו נוכל ליישם:

$$\begin{bmatrix} \boldsymbol{Y_i} \\ \boldsymbol{b_i} \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} \boldsymbol{X_i\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{V_i} & \boldsymbol{Z_iG} \\ (\boldsymbol{Z_iG})^{\boldsymbol{T}} & \boldsymbol{G} \end{bmatrix}\right)$$

ולכן

$$\boldsymbol{b_i}|\boldsymbol{Y_i} \sim \boldsymbol{N}\left(\boldsymbol{GZ_i^TV_i^{-1}}(\boldsymbol{Y_i} - \boldsymbol{X\beta}), \boldsymbol{G} - \boldsymbol{G^TZ_i^TV_i^{-1}Z_iG}\right)$$

דוגמה 15.12 – נניח שיש נקודת זמן עתידית כלשהי t^* שבה אנחנו מעוניינים להעריך את גודלי האפקט. אזי

$$\beta_{1i} + \beta_{2i}t^* = (\beta_1 + b_{i1}) + (\beta_2 + b_{2i})t^*$$

ומכאן החיזוי שלנו יהיה

$$prediction = \left(\hat{\beta}_1 + \hat{b}_{i1}\right) + \left(\hat{\beta}_2 + \hat{b}_{2i}\right)t^*$$

כאשר

$$\widehat{\boldsymbol{b}}_{\boldsymbol{i}} = \widehat{\mathbb{E}}[\boldsymbol{b_i}|\boldsymbol{Y_i}] = \widehat{\boldsymbol{G}}\boldsymbol{Z_i}\widehat{\boldsymbol{V}}_{\boldsymbol{i}}^{-1}\left(\boldsymbol{Y_i} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$$

דוגמה 15.13 - $\boldsymbol{b_i|Y_i}$ **במקרה הפשוט** – ננסה להדגים באמצעות המקרה הפשוט ביותר של מקרה חד-ממדי

$$Y_{ij} = \beta + b_i + \varepsilon_{ij}$$

כאשר $\beta, b_i \in \mathbb{R}$, ההתפלגויות הן $b_i \sim N\left(0, \sigma_b^2\right), \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$. כאשר אנחנו משתמשים בוקטור אחדות בודד כמקרה הפשוט ביותר של X_i, Z_i כלומר:

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i = \vec{1}_{J_i}\beta + \vec{1}_{J_i}b_i + \varepsilon_i$$

במקרה הזה

$$G = \left[\sigma_b^2\right] \in \mathbb{R}^{1\times 1}$$
$$Z_i = \vec{1}_{J_i}$$

בדוגמה זו

$$b_i|Y_i \sim GZ_i^TV^{-1}(Y_i - X\beta) = \sigma_b^2\vec{1}_{J_i}^TV^{-1}\left(Y_i - \vec{1}_{J_i}\beta\right)$$

במקרה זה:

$$V_i = Z_iGZ_i^T + R_i = \vec{1}_{J_i}\left[\sigma_b^2\right]\vec{1}_{J_i}^T + \sigma_\varepsilon^2 I_{J_i}$$

את החישוב של V_i^{-1} נבצע בשיעור הבא.

שיעור 16

תזכורת 16.1 – כזכור אנחנו עוסקים במודל הלינארי המעורב הכללי ובשיעור הקודם התחלנו לעסוק באמידה של אפקטים מקריים בתוך המודל. המודל בכללותו הוא כזכור עבור תצפית כלשהי i:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$
$$V_i = Z_i G Z_i^T + R_i$$

כאשר

$$Y_i \sim N(X_i\beta, V_i)$$
$$b_i \sim N(0, G)$$
$$G = cov(b_i)$$

ראינו בשיעור הקודם כי ניתן לבטא את ההתפלגות המשותפת באופן הבא:

$$\begin{bmatrix} Y_i \\ b_i \end{bmatrix} \sim N\left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, \begin{bmatrix} V_i & Z_i G \\ (Z_i G)^T & G \end{bmatrix}\right)$$

ואת ההתפלגות המותנית היא:

$$b_i | Y_i \sim N\left(G Z_i^T V_i^{-1}(Y_i - X\beta), G - G^T Z_i^T V_i^{-1} Z_i G\right)$$

ומכאן הגענו לאומד לאפקט המקרי באמצעות:

$$\hat{b}_i = \widehat{\mathbb{E}}[b_i | Y_i] = \hat{G} Z_i \hat{V}_i^{-1}\left(Y_i - X\hat{\beta}\right)$$

תזכורת 16.2 – עוד ראינו בשיעור הקודם את המודל הפשוט ביותר מהצורה:

$$Y_i = \vec{1}_{J_i}\mu + \vec{1}_{J_i} b_i + \varepsilon_i$$

כלומר כאשר האפקט המקרי b_i הוא סקלר. ההתפלגויות הן:

$$b_i \sim N\left(0, \sigma_b^2\right)$$
$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

כאשר במקרה זה:

$$G = \left[\sigma_b^2\right]$$
$$R_i = \sigma_\varepsilon^2 I$$

טענה 16.3 – מטריצת השונויות V_i במקרה הפשוט

$$w = \frac{\sigma_b^2}{\sigma_\varepsilon^2}$$

ואז נוכל לבטא את המטריצה V_i ואת המטריצה ההופכית באופן הבא:

$$V_i = Z_i G Z_i^T + R_i =$$
$$\vec{1}_{J_i}\left[\sigma_b^2\right]\vec{1}_{J_i}^T + \sigma_\varepsilon^2 I_{J_i} =$$
$$\sigma_b^2 \vec{1}_{J_i}\vec{1}_{J_i}^T + \sigma_\varepsilon^2 I_{J_i} =$$
$$\sigma_\varepsilon^2\left(I_{J_i} + w 1_{J_i} 1_{J_i}^T\right)$$

וההופכית:

$$V_i^{-1} = \frac{1}{\sigma_\varepsilon^2}\left(I_{J_i} + w 1_{J_i} 1_{J_i}^T\right)^{-1}$$

נעזר בשוויון Woodbury שראינו בעבר המוגדר על ידי:

$$(P+ZQZ^T)^{-1}=P^{-1}-P^{-1}Z(Q^{-1}+Z^TPZ)^{-1}Z^TP^{-1}$$

נייש את השוויון עבור

$$P=I_{J_i}\in\mathbb{R}^{J_i\times J_i}$$
$$Q=[w]\in\mathbb{R}^{1\times 1}$$
$$Z=\vec{1}_{J_i}\in\mathbb{R}^{J_i\times 1}$$

ונקבל

$$\left(I_{J_i}+w\vec{1}_{J_i}\vec{1}_{J_i}^T\right)^{-1}=I_{J_i}^{-1}-I^{-1}\vec{1}_{J_i}\left(\frac{1}{w}+\vec{1}_{J_i}^TI\vec{1}_{J_i}\right)^{-1}\vec{1}_{J_i}^TI^{-1}=I-I1_{J_i}^T\left(\frac{1}{w}+J_i\right)^{-1}1_{J_i}^TI$$
$$=I-I1_{J_i}^T\left(\left(\frac{1}{w}+J_i\right)\right)^{-1}1_{J_i}^TI=I-I1_{J_i}^T\left(\left(\frac{1+wJ_i}{w}\right)\right)^{-1}1_{J_i}^TI=I-I1_{J_i}^T\frac{w}{1+wJ_i}1_{J_i}^TI=$$
$$I_{J_i}-\frac{w}{1+wJ_i}1_{J_i}1_{J_i}^T=I_{J_i}-\xi 1_{J_i}1_{J_i}^T$$

כאשר

$$\xi=\frac{w}{1+wJ_i}$$

ומכאן נוכל לקבל:

$$V_i^{-1}=\frac{1}{\sigma_\varepsilon^2}\left(I_{J_i}-\xi 1_{J_i}1_{J_i}^T\right)$$

טענה 16.4 – אומד לאפקטים המקריים במקרה הפשוט – בהתאם לחישובים שביצענו נוכל להגדיר:

$$\hat{b}_i=\hat{G}Z_i^T\hat{V}_i^{-1}\left(Y_i-X\hat{\beta}\right)$$

כאשר V_i^{-1} כפי שהגדרנו מעלה, $\hat{G}=\sigma_b^2$ ו-$Z_i=1_{J_i}$. כמו כן במקרה שלנו המשוואה הכללית היא

$$Y_i=\vec{1}_{J_i}\mu+\vec{1}_{J_i}b_i+\varepsilon_i$$

כלומר $\beta=\mu$ ומכאן $\hat{\beta}=\hat{\mu}$.
אם כך נוכל להציב ולקבל:

$$\hat{b}_i=\sigma_b^2\vec{1}_{J_i}^T\left[\frac{1}{\sigma_\varepsilon^2}\left(I_{J_i}-\xi\vec{1}_{J_i}\vec{1}_{J_i}^T\right)\right]\left(Y_i-\vec{1}_{J_i}\hat{\mu}\right)$$

נסתכל בביטוי הפנימי $Z_i^T\hat{V}_i^{-1}$:

$$\vec{1}_{J_i}^T\left[\frac{I_{J_i}-\xi\vec{1}_{J_i}\vec{1}_{J_i}^T}{\sigma_\varepsilon^2}\right]=\frac{1}{\sigma_\varepsilon^2}\left[\vec{1}_{J_i}^T-\xi\vec{1}^T\vec{1}\vec{1}^T\right]=\frac{1}{\sigma_\varepsilon^2}(1-J_i\xi)\vec{1}_{J_i}^T$$

ובסך הכל:

$$\hat{b}_i=\sigma_b^2\vec{1}_{J_i}^T\left[\frac{I_{J_i}-\xi\vec{1}_{J_i}\vec{1}_{J_i}^T}{\sigma_\varepsilon^2}\right]\left(Y_i-\vec{1}_{J_i}\hat{\mu}\right)$$
$$=\frac{\sigma_b^2}{\sigma_\varepsilon^2}(1-J_i\xi)\left(\vec{1}^TY_i-\vec{1}^T\vec{1}\hat{\mu}\right)$$

בביטוי השלישי נוכל לפתח

$$\vec{1}_{J_i}Y_i = \sum_{j=1}^{J_i} Y_{ij} = J_i\bar{Y}_{i\cdot}$$
$$\vec{1}_{J_i}^{\;T}\vec{1}_{J_i} = J_i$$

ולכן:

$$\hat{b}_i = \frac{\sigma_b^2}{\sigma_\varepsilon^2}(1 - J_i\xi)(J_i\bar{Y}_{i\cdot} - J_i\hat{\mu}) =$$
$$\frac{\sigma_b^2}{\sigma_\varepsilon^2}\left(1 - \frac{J_i w}{1 + J_i w}\right)(J_i\bar{Y}_{i\cdot} - J_i\hat{\mu}) =$$
$$\frac{\sigma_b^2}{\sigma_\varepsilon^2}\left(\frac{(1 + J_i w) - J_i w}{1 + J_i w}\right)(J_i\bar{Y}_{i\cdot} - J_i\hat{\mu}) =$$
$$\frac{\sigma_b^2}{\sigma_\varepsilon^2}\left(\frac{1}{1 + J_i w}\right)(J_i\bar{Y}_{i\cdot} - J_i\hat{\mu}) =$$
$$\left(\frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + \sigma_\varepsilon^2 J_i w}\right)(\bar{Y}_{i\cdot} - \hat{\mu}) \overset{w=\frac{\sigma_\varepsilon^2}{\sigma_b^2}}{=}$$
$$\left(\frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + J_i\sigma_b^2}\right)(\bar{Y}_{i\cdot} - \hat{\mu})$$

ובסך הכל:

$$\boldsymbol{\hat{b}_i = \frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + J_i\sigma_b^2}(\bar{Y}_{i\cdot} - \hat{\mu})}$$

טענה 16.5 – נסתכל על

$$\mu_i = \mathbb{E}\left[Y_{ij}\middle|b_i\right]$$

כאשר

$$\mu_i = \mu + b_i$$

והאומד הוא

$$\hat{\mu}_i = \hat{\mu} + \hat{b}_i = \hat{\mu} + \frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + J_i\sigma_b^2}(\bar{Y}_{i\cdot} - \hat{\mu}) = \boldsymbol{\frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + J_i\sigma_b^2}\bar{Y}_{i\cdot} + \left(1 - \frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + J_i\sigma_b^2}\right)\hat{\mu}}$$

כלומר אנחנו מקבלים שהאומד הכללי עבור כל יחידה i הוא ממוצע משוקלל שבו מצד אחד:

1. $\bar{Y}_{i\cdot}$ - ממוצע התצפיות עבור יחידה i בלבד
2. $\hat{\mu} = \mathbb{E}\left[Y_{ij}\right]$ - האומד הכללי של Y_{ij} בהתחשב בכל המדגם (ללא התניה על b_i).

הערה 16.6 - נשים לב:

1. אם J_i מאוד גדול אז החלק של $\bar{Y}$ גובר והחלק של $\hat{\mu}$ מתאפס. כלומר האומד של $\hat{\mu}_i$ ייצא קרוב לממוצע של התצפיות ביחידה i (כי יש מספיק תצפיות על היחידה עצמה).
2. אם σ_b^2 גדול אזי יש שונות גדולה בין תצפיות שונות ולכן עדיף להשתמש בתצפיות על דוגמה i בלבד ולכן בקירוב $\hat{\mu}_i \approx \bar{Y}_{i\cdot}$.
3. אם σ_ε^2 גדול (יש הרבה רעש) אז עדיף כבר להשתמש בכלל הנתונים כלומר $\hat{\mu}_i \approx \hat{\mu}$

טענה 16.7 – אם נחזור למודל הווקטורי המלא:

$$Y=\begin{bmatrix}Y_1\\ \dots\\ Y_n\end{bmatrix}=X\beta+Zb+\varepsilon$$

ובהתאם למה שראינו עתה נכליל:

$$\hat{b}=\begin{bmatrix}\hat{b}_1\\ \dots\\ \hat{b}_n\end{bmatrix}=\mathcal{G}Z^T\hat{V}^{-1}\left(Y-X\hat{\beta}\right)$$

כאשר $\mathcal{G}$ מטריצת בלוקים אלכסונית שבה בכל בלוק המטריצה G.

נניח V_i ידוע. ניתן לפתח באמצעות הצבת Y:

$$\begin{aligned}&\hat{b}-b=\mathcal{G}Z^TV^{-1}\left[X\beta+Zb+\varepsilon-X\hat{\beta}\right]-b=\\&\mathcal{G}Z^TV^{-1}\left[X\left(\beta-\hat{\beta}\right)+Zb+\varepsilon\right]-b=\\&\mathcal{G}Z^TV^{-1}X\left(\beta-\hat{\beta}\right)+GZ^TV^{-1}Zb+GZ^TV^{-1}\varepsilon-b=\\&D\left(\hat{\beta}-\beta\right)+Fb+H\varepsilon\end{aligned}$$

כאשר

$$\begin{aligned}&D=-\mathcal{G}Z^TV^{-1}X\\&F=-(I-\mathcal{G}Z^TV^{-1}Z)\\&H=\mathcal{G}Z^TV^{-1}\end{aligned}$$

כלומר

$$\hat{b}-b=D\left(\hat{\beta}-\beta\right)+Fb+H\varepsilon$$

כזכור האומד לנראות של β הוא:

$$\begin{aligned}&\hat{\beta}=(X^TV^{-1}X)^{-1}X^TV^{-1}Y=\\&(X^TV^{-1}X)^{-1}X^TV^{-1}(X\beta+Zb+\varepsilon)=\\&(X^TV^{-1}X)^{-1}(X^TV^{-1}X)\beta+(X^TV^{-1}X)^{-1}X^TV^{-1}Zb+(X^TV^{-1}X)^{-1}X^TV^{-1}\varepsilon=\\&\beta+(X^TV^{-1}X)^{-1}X^TV^{-1}Zb+(X^TV^{-1}X)^{-1}X^TV^{-1}\varepsilon=\\&\beta+KZb+K\varepsilon\end{aligned}$$

כאשר

$$K=(X^TV^{-1}X)^{-1}X^TV^{-1}$$

מכאן

$$\hat{\beta}-\beta=\beta+KZb+K\varepsilon-\beta=KZb+K\varepsilon$$

ולכן אם נמשיך לפתח את $\hat{b}$ נקבל

$$\begin{aligned}&\hat{b}-b=D\left(\hat{\beta}-\beta\right)+Fb+H\varepsilon=\\&D[KZb+K\varepsilon]+Fb+H\varepsilon=\\&(DKZ+F)b+(H+DK)\varepsilon\end{aligned}$$

מסקנה 16.8 – בפיתוח שעשינו הצלחנו לבטא את הביטויים $\hat{\beta}-\beta,\hat{b}-b$ באמצעות המשתנים המקריים שיש לנו – b ו-ε. בסך הכל הצלחנו להגיע לביטוי:

$$\begin{bmatrix}\boldsymbol{\widehat{\beta}-\beta}\\ \boldsymbol{\widehat{b}-b}\end{bmatrix}\sim\boldsymbol{N(0,C)}$$

כאשר

$$\boldsymbol{C=Cov\left(\begin{bmatrix}\widehat{\beta}-\beta\\ \widehat{b}-b\end{bmatrix}\right)=Cov\left(\begin{bmatrix}KZb+K\varepsilon\\ (DKZ+F)b+(H+DK)\varepsilon\end{bmatrix}\right)}$$

ואת הביטויים הללו ניתן כאמור לחשב מתוך התצפיות.

דוגמה 16.9 – נניח ש-i הם בתי חולים שונים, Y_{ij} זמני אשפוז אחרי ניתוח מסוים. הנוסחה הכללית היא:

$$Y_{ij}=\mu+b_i+\varepsilon_{ij}$$

הביטוי $\mu_i=\mu+b_i$ מתאר את זמן האשפוז הצפוי בבית חולים i.

את זמן האשפוז הנ"ל ניתן לאמוד באמצעות

$$\hat{\mu}_i=\hat{\mu}+\hat{b}_i=\hat{\beta}+\hat{b}_i=\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{b}_i\end{bmatrix}$$

כלומר אנחנו אומדים את זמן האשפוז הממוצע בבית החולים i כסכום של האומד הכללי של זמן האשפוז בכל בתי החולים ושל אפקט האשפוז בבית החולים i באופן ספציפי.

ניתן לייצר רווח סמך עבור $\hat{\mu}_i$. שכן

$$\hat{\mu}_i-\mu_i=\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{b}_i\end{bmatrix}-\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}\beta\\ b_i\end{bmatrix}=\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}\hat{\beta}-\beta\\ \hat{b}_i-b_i\end{bmatrix}=\vec{1}_2^T\begin{bmatrix}\hat{\beta}-\beta\\ \hat{b}_i-b_i\end{bmatrix}$$

נוכל לכתוב זאת גם בצורה:

$$\hat{\mu}_i-\mu_i=\begin{bmatrix}1 & e_i^T\end{bmatrix}\begin{bmatrix}\hat{\beta}-\beta\\ \hat{b}_i-b_i\end{bmatrix}$$

כאשר $e_i\in\mathbb{R}^n$ הוא וקטור שבו 1 בקואורדינטה ה-i ו-0 בכל קואורדינטה אחרת.

עבור C שראינו קודם מתקיים

$$\hat{\mu}_i-\mu_i\sim N\left(0,a_i^TCa_i\right)$$

עם $a_i^T=\begin{bmatrix}1 & e_i^T\end{bmatrix}$.

לכן רווח הסמך הוא:

$$\boldsymbol{\widehat{\mu_i}\pm z_{1-\alpha/2}\sqrt{a_i^TCa_i}}$$

שיעור 17

תזכורת 17.1 – ראינו בשיעור הקודם מצב של אמידה של $\hat{b}=\mathbb{E}[b|Y]$ תחת המודל הלינארי המעורב. ראינו כי:

$$\begin{bmatrix}\hat{\beta}-\beta\\ \hat{b}-b\end{bmatrix}\sim N(0,C)$$

כאשר באופן כללי

$$\beta\in\mathbb{R}^p$$

$b \in \mathbb{R}^{nq}$
$b_i \in \mathbb{R}^q$

וראינו כי לכל $a \in \mathbb{R}^{p+nq}$ מתקיים

$$a^T \begin{bmatrix} \hat{\beta} - \beta \\ \hat{b} - b \end{bmatrix} \sim N(0, a^T C a)$$

ראינו דוגמה 1 עם משתנה מסביר יחיד בשיעור הקודם. עתה נבחן דוגמה נוספת.

דוגמה 17.2 – נניח

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + b_{i1} + b_{i2} t_{ij} + \varepsilon_{ij}$$

נניח שאנחנו מעוניינים לאמוד את Y_{ij} בזמן חדש כלשהו t^*.

נסמן Y_{ij}^* להיות הערך של Y עבור פרט i בזמן t^*.

נסמן:

$$\Psi_{ij}^* = \mathbb{E}\left[Y_{ij}^* \middle| b_i\right] = \begin{bmatrix} 1 & t^* \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 & t^* \end{bmatrix} \begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix}$$

וכן:

$$\hat{\Psi}_{ij}^* = \mathbb{E}\left[\widehat{Y_{ij}^* \middle| b_i}\right] = \begin{bmatrix} 1 & t^* \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} + \begin{bmatrix} 1 & t^* \end{bmatrix} \begin{bmatrix} \hat{b}_{i1} \\ \hat{b}_{i2} \end{bmatrix}$$

ולכן:

$$\hat{\Psi}_{ij}^* - \Psi_{ij}^* = \begin{bmatrix} 1 & t^* \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{bmatrix} + \begin{bmatrix} 1 & t^* \end{bmatrix} \begin{bmatrix} \hat{b}_{i1} - b_{i1} \\ \hat{b}_{i2} - b_{i2} \end{bmatrix}$$

עבור $i = 1$ נוכל לרשום באופן הבא:

$$\hat{\Psi}_{ij}^* - \Psi_{ij}^* = \begin{bmatrix} 1 & t^* & 1 & t^* & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \\ \hat{b}_{12} - b_{12} \\ \hat{b}_{21} - b_{21} \\ \hat{b}_{22} - b_{22} \\ \dots \\ \hat{b}_{n2} - b_{n2} \end{bmatrix}$$

ונסמן $a_1^T = \begin{bmatrix} 1 & t^* & 1 & t^* & 0 & \dots & 0 \end{bmatrix}$ ונוכל לרשום:

$$\hat{\Psi}_{1j}^* - \Psi_{1j}^* = a_1^T \begin{bmatrix} \hat{\beta} - \beta \\ \hat{b} - b \end{bmatrix}$$

עבור $i = 2$ נוכל לרשום באופן דומה:

$$a_2^T = \begin{bmatrix} 1 & t^* & 0 & 0 & 1 & t^* & 0 & \dots & 0 \end{bmatrix}$$

ואז:

$$\widehat{\Psi}^*_{2j} - \Psi^*_{2j} = a_2^T \begin{bmatrix} \hat{\beta} - \beta \\ \hat{b} - b \end{bmatrix}$$

באופן זה נוכל לבנות מטריצה ששורותיה $a_1^T, \dots, a_n^T$.

דוגמה 17.2.1 – רווח סמך ל-Ψ^*_{ij} בזמן t^*

$$\widehat{\Psi}^*_{ij} \pm z_{1-\alpha/2}\sqrt{a_i^T \hat{C} a_i^T}$$

אם נסמן

$$d = z_{1-\alpha/2}\sqrt{a_i^T \hat{C} a_i^T}$$

נקבל

$$\begin{aligned}
&\mathbb{P}\left(\Psi^*_{ij} \in \text{רווח}\right) \\
&= \mathbb{P}\left(\Psi^*_{ij} - d \le \widehat{\Psi}^*_{ij} \le \Psi^*_{ij} + d\right) \\
&= \mathbb{P}\left(-d \le \widehat{\Psi}^*_{ij} - \Psi^*_{ij} \le d\right) \\
&\doteq \mathbb{P}\left(-d \le N\left(0, \mathbb{V}\left(\widehat{\Psi}^*_{ij}\right)\right) \le d\right) \\
&= \mathbb{P}\left(-z_{1-\alpha/2} \le N(0,1) \le z_{1-\alpha/2}\right)
\end{aligned}$$

הערה 17.2.2 נשים לב שהמעבר השלישי ברווח הסמך הוא בקירוב, כיוון שלא הבאנו בחשבון את ההבדל בין C ל-$\hat{C}$. הוצעו תיקונים בספרות אולם בפועל השינויים הם לא גדולים.

דוגמה 17.2.3 – רווח חיזוי עבור Y^*_{ij} – נוכל גם לחזות

$$\widehat{\Psi}^*{}_{ij} \pm \tilde{d}$$

כאשר

$$\mathbb{P}\left(Y^*_{ij} \in \text{רווח}\right) = 1 - \alpha$$

יש לנו

$$Y^*_{ij} = \Psi^*_{ij} + \varepsilon^*_{ij}$$

נניח שאנחנו במצב שבו ε_{ij} בלתי תלויים. אזי

$$\begin{aligned}
&Y_{ij} - \widehat{\Psi}^*_{ij} = \left(Y_{ij} - \Psi^*_{ij}\right) + \left(\Psi^*_{ij} - \widehat{\Psi}^*{}_{ij}\right) \\
&= \varepsilon^*_{ij} + \left(\Psi^*_{ij} - \widehat{\Psi}_{ij}\right)
\end{aligned}$$

כיוון שההתפלגות הנורמלית היא סימטרית ראינו $\left(\widehat{\Psi}_{ij}{}^* - \Psi^*_{ij}\right) \sim N\left(0, a_i^T C a_i\right)$ ולכן גם

$$\left(\Psi^*_{ij} - \widehat{\Psi}^*_{ij}\right) \dot{\sim} N\left(0, a_i^T C a_i\right)$$

כמו כן הנחנו אי-תלות ולכן

$$\varepsilon^*_{ij} \dot{\sim} N(0, \sigma^2_\varepsilon)$$

כאשר ההתפלגות של שני המחוברים היא בלתי-תלויה בגלל הנחת אי-התלות על השגיאות. לכן בסך הכל אנחנו מקבלים:

$$Y^*_{ij} - \widehat{\Psi}^*_{ij} \dot{\sim} N\left(0, a_i^T C a_i + \varepsilon^*_{ij}\right)$$

וכמו שראינו קודם נקבל באופן דומה **רווח סמך**:

$$\hat{\Psi}_{ij}^* \pm z_{1-\alpha/2}\sqrt{a_i^T \hat{C} a_i^T + \hat{\sigma}_\varepsilon^2}$$

נושא סטטיסטיקה 3 – ניתוח מרכיבים ראשיים – Principal Component Analysis PCA

הגדרה 17.3 – מרכיבים ראשיים – יהיו n וקטורים $X_1,\dots,X_n$ כאשר $X_i = \begin{bmatrix} X_{i1} \\ \cdots \\ X_{ip} \end{bmatrix} \in \mathbb{R}^p$ לכל i. נניח כי הוקטוורים בלתי-תלויים ושווי התפלגות $X_1,\dots,X_n$ כאשר $\mathbb{E}[X_i] = \mu \in \mathbb{R}^p$ ו-$Cov(X_i) = V \in \mathbb{R}^{p\times p}$. נרצה לייצר וקטורים חלופיים $\tilde{X}_1,\dots,\tilde{X}_n$ אשר שומרים את כמות האינפורמציה הגבוהה ביותר על הווקטורים המקוריים. פורמלית, נרצה שיתקיים לכל i ועבור מטריצה אורתנורמלית כלשהי U:

$$\tilde{X}_i = U\tilde{Y}_i \in \mathbb{R}^q$$

כאשר $\tilde{Y}_i \in \mathbb{R}^q$ ו-$q \ll p$ כלומר אנחנו רוצים להוריד את ממד הפיצ'רים של X מממד כלשהו p לממד קטן יותר q ("דחיסה" \ "הורדת ממד").

הערה 17.4 – נניח μ, V ידועים. באופן מעשי משתמשים לרוב ב-$\hat{\mu}, \hat{V}$ האומדים המוכרים שראינו בעבר בהתפלגות רב-נורמלית:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$$

$$\hat{V} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$$

הערה 17.5 – אנחנו מניחים ש-$n > p$ כלומר מספר הדוגמאות גדול ממספר המשתנים המסבירים (הפיצ'רים).

הערה 17.6 – נניח $\mu = 0$ (אחרת נוכל לתקנן ולייצר וקטורים חדשים שבהם הדבר מתקיים). ללא הנחות נוספות על V.

טענה 17.7 – יהי $a_1,\dots,a_p$ בסיס אורתונורמלי ל-$\mathbb{R}^p$ ווקטור כלשהו X שהוא אחד מבין וקטורי התצפיות הנתונים. אם $a_1,\dots,a_p$ בסיס אז לכל וקטור $X' \in \mathbb{R}^p$ קיימים סקלרים $\alpha_1,\dots,\alpha_p$ כך שמתקיים:

$$X' = \sum_{k=1}^p \alpha_k \cdot a_k$$

חוץ מזה $a_1,\dots,a_p$ הוא בסיס אורתורנומלי לכן:

$$\|a_k\|^2 = a_k^T a_k = 1$$

אזי במונחי הבסיס הנתון נוכל להציג את X באמצעות:

$$X = \sum_{k=1}^p \left(a_k^T X\right) a_k$$

$$\tilde{X} = \sum_{k=1}^{q} \left(a_k^T X\right) a_k$$

אם נסמן

$$H = \begin{bmatrix} | & | & | & | \\ a_1 & \dots & \dots & a_q \\ | & | & | & | \end{bmatrix}$$

אזי

$$\tilde{X} = P_H X$$

כאשר

$$P_H = H(H^T H)^{-1} H^T = HIH^T = HH^T$$

ולכן

$$\tilde{X} = P_H X = HH^T X$$

טענה 17.8 – **צמצום למינימום של השגיאה בין X ל-$\tilde{X}$** - נמדוד את המרחק בין $\tilde{X}$ ל-X לפי תוחלת שגיאה ריבועית (MSE):

$$MSE = \mathbb{E}\left[\left\|X - \tilde{X}\right\|^2\right]$$

נוכל לפתח את המרחק באמצעות פיתוח מתוך X:

$$X = P_H X + (I - P_H)X = \tilde{X} + (I - P_H)X$$

כאשר $P_H, I - P_H$ שני וקטורים ניצבים (P_H הוא היטל על תת-המרחב של q הוקטורים $a_1, \dots, a_q$ ו-$I - P_H$ הוא היטל על יתר הווקטורים). אם כן נסתכל על הנורמה (הכללה של משפט פיתגורס לנורמות):

$$\|X\|^2 = \left\|\tilde{X}\right\|^2 + \left\|X - \tilde{X}\right\|^2$$

והתוחלת:

$$\mathbb{E}[\|X\|^2] = \mathbb{E}\left[\left\|\tilde{X}\right\|^2 + \left\|X - \tilde{X}\right\|^2\right] = \mathbb{E}\left[\left\|\tilde{X}\right\|^2\right] + \mathbb{E}\left[\left\|\tilde{X} - X\right\|^2\right] = \mathbb{E}\left[\left\|\tilde{X}\right\|^2\right] + MSE\left(X, \tilde{X}\right) \Rightarrow$$
$$MSE\left(X, \tilde{X}\right) = \mathbb{E}[\|X\|^2] - \mathbb{E}\left[\left\|\tilde{X}\right\|^2\right] = constant - \mathbb{E}\left[\left\|\tilde{X}\right\|^2\right]$$

נשים לב כי התוחלת $\mathbb{E}[\|X\|^2]$ נקבעת על ידי ההתפלגות של X, ולכן היא קבועה. מכאן שכדי להביא את ה-MSE למינימום צריך להביא למקסימום את $\mathbb{E}\left[\left\|\tilde{X}\right\|^2\right]$. נפתח אם כן את $\tilde{X}$:

$$\tilde{X} = \sum_{k=1}^{q} \left(a_k^T X\right) a_k$$

נשים לב גם כאן כי $a_1, \dots, a_q$ הוא בסיס אורתונורמלי. הדבר מאפשר לנו לפשט את הנורמה כיוון שרוב המחוברים יתאפסו בשל האורתונורמליות:

$$\left\|\tilde{X}\right\|^2 = \tilde{X}^T \tilde{X} = \left(\sum_{k_1=1}^{q} \left(a_{k_1}^T X\right) a_{k_1}\right)\left(\sum_{k_2=1}^{q} \left(a_{k_2}^T X\right) a_{k_2}\right) = \sum_{k_1=1}^{q} \sum_{k_2=2}^{q} \left(a_{k_1}^T X\right)\left(a_{k_2}^T X\right) a_{k_1}^T a_{k_2}$$

$$= \sum_{k=1}^{q} \left(a_k^T X\right)^2$$

נזכור כי $\mathbb{V}\left(a_k^T X\right) = a_k^T V a_k$ ולכן התוחלת של הנורמה היא:

$$\mathbb{E}\left[\left\|\tilde{X}\right\|^2\right] = \mathbb{E}\left[\sum_{k=1}^{q}\left(a_k^T X\right)^2\right] = \sum_{k=1}^{q}\mathbb{E}\left[\left(a_k^T X\right)^2\right] = \sum_{k=1}^{q}\mathbb{V}\left(a_k^T X\right) = \sum_{k=1}^{q} a_k^T V a_k$$

כאשר המעבר הלפני אחרון בגלל ש-$\mathbb{E}[X] = 0$ (הערה 17.6).

נפרק את V באמצעות פירוק ספקטראלי:

$$V = U\Lambda U^T$$

כאשר U מטריצה אורתונורמלית ו-Λ מטריצה אלכסונית עם הערכים העצמיים של V $\lambda_1, \dots, \lambda_p$.

נשים לב כי

$$\|U^T a_k\|^2 = (U^T a_k)^T (U^T a_k) = a_k^T U U^T a_k = a_k^T I a_k = a_k^T a_k$$

כלומר ההכפלה במטריצה אורתונרמלית עם דרגה מלאה לא משנה את הנורמה. מכאן אם נמשיך את הפיתוח הקודם:

$$\mathbb{E}\left[\left\|\tilde{X}\right\|^2\right] = \sum_{k=1}^{q} a_k^T V a_k = \sum_{k=1}^{q} (U^T a_k)^T \Lambda (U^T a_k) \overset{z_k = U^T a_k}{=} \sum_{k=1}^{q} z_k^T \Lambda z_k$$

ניקח $z_k = e_k$, כלומר נבחר את המטריצה האורתנורמלית U באופן הזה:

$$z_k = e_k \rightarrow$$
$$U^T a_k = e_k$$
$$a_k = U e_k = u_k$$

הגדרה 17.8 - נגדיר

$$Y_r = u_r^T X$$

אזי Y_r נקרא המרכיב הראשי של X מספר r.

$$\tilde{X} = \sum_{k=1}^{q}\left(a_k^T X\right)a_k = \sum_{k=1}^{q}\left(u_k^T X\right)a_k = \sum_{k=1}^{q} Y_k u_k$$
$$X = \sum_{k=1}^{p} Y_k u_k$$

כלומר $Y = U^T X, X = UY$.

לכן

$$Cov(Y) = Cov(U^T X) = U^T Cov(X) U = U^T V U = U^T U \Lambda U^T U = \Lambda$$

כלומר הכניסות של Y בלתי מתואמים ול-Y_k יש שונות λ_k.

טענה 17.9

$$X = UY$$

$$X_r = \sum_{s=1}^{p} U_{rs} Y_s = \sum_{s=1}^{p} (U_s)_r Y_s$$

כאשר U_s הוא וקטור עצמי מספר s, ו-$(U_s)_r$ הוא הכניסה ה-r של הווקטור U_S.

הערה 17.10 – אם נסמן

$$W_{rs} = (U_s)_r Y_s$$

אז השונות היא:

$$\mathbb{V}[W_{rs}] = [(u_s)_r]^2 \mathbb{V}[Y_s] = (u_s)_r^2 \lambda_s$$

והשונות המשותפת היא (כאשר $s_1 \neq s_2$):

$$Cov\left(W_{rs_1}, W_{rs_2}\right) \\ = Cov\left(\left(u_{s_1}\right)_r Y_{s_1}, \left(u_{s_2}\right)_r Y_{s_2}\right) = \left(u_{s_1}\right)_r \left(u_{s_2}\right)_r Cov\left(Y_{s_1}, Y_{s_2}\right) = 0$$

שיעור 18

תזכורת 18.1 – התחלנו לעסוק בנושא של ניתוח מרכיבים ראשיים (Principal Component Analysis - PCA).

תזכורת 18.1.1 - כזכור, אנחנו עוסקים במצב של וקטור מקרי $X \in \mathbb{R}^p$ כאשר

$$\mathbb{E}[X] = \vec{\mu} \in \mathbb{R}^p$$
$$Cov(X) = V \in \mathbb{R}^{p \times p}$$

נניח בלי הגבלת כלליות כי $\vec{\mu} = 0$.

יהיו וקטורים $a_1, \dots, a_p$ המהווים בסיס אורתונורמלי עבור $\mathbb{R}^p$. מכאן שנוכל לייצג את הווקטור X באמצעות קומבינציה לינארית של הווקטורים בבסיס, כלומר:

$$X = \sum_{r=1}^{p} (a_r^T X) a_r$$

נרצה להוריד ממד בלי לאבד הרבה אינפורמציה. לכן מגדירים עבור $q \ll p$ וקטור חלופי:

$$\tilde{X} = \sum_{r=1}^{q} (a_r^T X) a_r$$

נשים לב כי עדיין $\tilde{X} \in \mathbb{R}^p$, פשוט הוא נוצר כתוצאה של סכימה של קומבינציה לינארית עם כמות איברים קטנה מהממד המלא של X.

תזכורת 18.1.2 - מטרתנו למזער את ה-MSE בין הוקטור המקורי X לווקטור בממד המוקטן $\tilde{X}$, כאשר כזכור ה-MSE מוגדר כתוחלת ריבועי השגיאות:

$$MSE\left(X, \tilde{X}\right) = \mathbb{E}\left[\left\|X - \tilde{X}\right\|^2\right]$$

ראינו כי על מנת למזער את ה-MSE הפתרון האופטימלי הינו לקחת וקטורים $a_1, \dots, a_p$ השווים ל-$u_1, \dots, u_p$, כאשר $u_1, \dots, u_p$ הם הווקטורים העצמיים של מטריצת השונויות V בסדר יורד, כלומר אם λ_i מתאים ל-u_i אזי הסידור $u_1, \dots, u_p$ תואם לסדר הערכים העצמיים $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

במקרה זה מגדירים את המטריצה U להיות מטריצה שעמודותיה הם הווקטורים $u_1, \dots, u_p$.

$$U = \begin{bmatrix} | & | & | & | \\ u_1 & \dots & \dots & u_p \\ | & | & | & | \end{bmatrix} \in \mathbb{R}^{p\times p}$$

בגלל שכל עמודה במטריצה היא וקטור כלשהו מבין $u_1, \dots, u_p$ אזי:

$$[U]_{rs} = (u_s)_r$$

כלומר הרכיב ה-r בווקטור u_s, המסומן $(u_s)_r$, שווה לקואורדינטה ה-rs של המטריצה U.

תזכורת 18.1.3 - כמו כן מגדירים וקטור חדש Y על ידי:

$$Y = U^T X \in \mathbb{R}^{p\times 1}$$
$$Y_r = u_r^T X \in \mathbb{R}$$

כאשר Y_r נקרא המרכיב הראשי (Principal Component) ה-r של X. ראינו כי עבור הווקטור החדש Y מתקיים כי מטריצת השונויות $\Lambda \in \mathbb{R}^{p\times p}$ היא:

$$\Lambda = Cov(Y) = diag\{\lambda_1, \dots, \lambda_p\} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix}$$

ובפרט לכל $1 \le t \le p$ מתקיים $\mathbb{V}[Y_t] = \lambda_t$.

טענה 18.2 – באשר ליחס בין Y ל-X, ניתן לראות כי מתקיים:

$$Y = U^T X \Rightarrow$$
$$X = UY$$

ולכן עבור קואורדינטה ספציפית מתקיים

$$X_r = (UY)_r = \sum_{t=1}^{p} [U]_{rt} Y_t \overset{18.1.2}{=} \sum_{t=1}^{p} (u_t)_r Y_t \overset{w_{rt}=(u_t)_r Y_t}{=} \sum_{t=1}^{p} w_{rt}$$

כאשר $\mathbb{R}^p \ni w_r = \begin{bmatrix} w_{r1} \\ \cdots \\ w_{rp} \end{bmatrix}$ מוגדר באמצעות $w_{rt} = [U]_{rt} Y_t = (u_t)_r Y_t$.

ומכאן ניתן להגיע למסקנה כי את הווקטור בממד המוקטן נוכל להציג באמצעות:

$$\boldsymbol{\tilde{X}_r = \sum_{t=1}^{q} w_{rt}}$$

הערה 18.3 – הערכים $(u_r)_s$ נקראים **loadings**.

טענה 18.4 – שונות של $\boldsymbol{w_{rt}}$ - נגדיר $G \in \mathbb{R}^{p\times p}$ מטריצה שבה השורה ה-r מייצגת את השונויות המשותפות עבור הווקטור w_r. כאשר:

$$[G]_{rt} = \mathbb{V}[w_{rt}] = [U]_{rt}^2 \mathbb{V}(Y_t) = [U]_{rt}^2 \lambda_t$$

ואילו עבור $t_1 \neq t_2$ מתקיים:

$$Cov\left(w_{rt_1}, w_{rt_2}\right) = [U]_{rt_1}[U]_{rt_2} Cov\left(Y_{t_1}, Y_{t_2}\right) = 0$$

שכן $Cov\left(Y_{t_1}, Y_{t_2}\right) = 0$ (ראינו כי השונות המשותפת של Y היא מטריצה אלכסונית).

בסך הכל אם נסמן את השורה ה-r של G על ידי G_r נקבל:

$$G_r = \left[[U]_{r1}^2\lambda_1 \quad \dots \quad [U]_{rp}^2\lambda_p\right]$$

טענה 18.5 – מתוך השונות שמצאנו עבור w_r ניתן גם לבטא את השונות של X_r. נשים לב כי הראינו (18.4) שלכל $t_1 \neq t_2$ מתקיים $Cov\left(w_{rt_1}, w_{rt_2}\right) = 0$ כלומר $w_{r1}, \dots, w_{rp}$ בלתי מתואמים. לפיכך נוכל לפתח:

$$\mathbb{V}[X_r] = \mathbb{V}\left[\sum_{t=1}^{p} w_{rt}\right] = \sum_{t=1}^{p} \mathbb{V}[w_{rt}] \overset{18.4}{=} \sum_{t=1}^{p} [U]_{rt}^2 \lambda_t$$

טענה 18.6 – בין X_r ל-w_{rt} –

מתוך החישובים לעיל נוכל לחשב את הפרופורציה של השונות של X_r המוסברת על ידי Y_t על ידי:

$$\frac{\mathbb{V}[w_{rt}]}{\mathbb{V}[X_r]} = \frac{\lambda_t [U]_{rt}^2}{\sum_{t'=1}^{p} [U]_{rt'}^2 \lambda_{t'}}$$

כמו כל נוכל לחשב את השונות המשותפת על ידי:

$$\begin{aligned}
&Cov(X_r, w_{rt}) \\
&= Cov\left(\sum_{t'=1}^{p} [U]_{rt'}^2 \lambda_{t'}, w_{rt}\right) = \sum_{t'=1}^{p} Cov(w_{rt'}, w_{rt}) = \sum_{\substack{t'=1 \\ t' \neq t}}^{p} Cov(w_{rt'}, w_{rt}) + Cov(w_{rt}, w_{rt}) \\
&= \sum_{\substack{t'=1 \\ t' \neq t}}^{p} 0 + \mathbb{V}(w_{rt}) = \mathbb{V}(w_{rt})
\end{aligned}$$

אם נרצה לחשב את הקורלציה ביניהם נקבל:

$$Corr(X_r, w_{rt}) = \frac{Cov(X_r, w_{rt})}{\sqrt{\mathbb{V}(X_r)\mathbb{V}(w_{rt})}} = \frac{\mathbb{V}(w_{rt})}{\sqrt{\mathbb{V}(X_r)\mathbb{V}(w_{rt})}} = \sqrt{\frac{\mathbb{V}(w_{rt})}{\mathbb{V}(X_r)}}$$

טענה 18.7 – בין X_r ל-$\tilde{X}_r$ – מתוך 18.6 וחישובים נוספים לעיל נוכל לחשב את השונות המשותפת בין הווקטור המקורי X_r והווקטור בממד המוקטן $\tilde{X}_r$:

$$Cov\left(X_r, \tilde{X}_r\right) = Cov\left(\sum_{t_1=1}^{p} w_{rt_1}, \sum_{t_2=1}^{q} w_{rt_2}\right) = \sum_{t_1=1}^{p}\sum_{t_2=1}^{q} Cov\left(w_{rt_1}, w_{rt_2}\right) = \sum_{t=1}^{q} \mathbb{V}(w_{rt})$$

כאשר המעבר האחרון בגלל שכפי שראינו $w_{r1}, \dots, w_{rp}$ בלתי מתואמים כלומר:

$$Cov\left(w_{rt_1}, w_{rt_2}\right) = \begin{cases} \mathbb{V}(w_{rt}), & t_1 = t_2 = t \\ 0, & otherwise \end{cases}$$

ניתן גם כאן לחשב את הקורלציה על ידי:

$$Corr\left(\tilde{X}_r, X_r\right) = \frac{Cov\left(\tilde{X}_r, X_r\right)}{\sqrt{\mathbb{V}\left(\tilde{X}_r\right)\mathbb{V}(X_r)}} = \frac{\mathbb{V}\left(\tilde{X}_r\right)}{\sqrt{\mathbb{V}\left(\tilde{X}_r\right)\mathbb{V}(X_r)}} = \sqrt{\frac{\mathbb{V}\left(\tilde{X}_r\right)}{\mathbb{V}(X_r)}}$$

כלומר אנחנו מגיעים כאן לביטוי המאפשר לנו להעריך את מידת הקורלציה בין הווקטור המקורי לווקטור המוקטן.

הערה 18.8 – נשים לב כי $V = Cov(X)$ סימטרית חיובית למחצה לכן קיים פירוק ספקטראלי $V = U\Lambda U^T$ ומכאן גם $\Lambda = U^T VU$ כאשר Λ מטריצה אלכסונית שבה הערכים העצמיים של V. כיוון שהמטריצה אלכסונית ומכילה ערכים עצמיים מתקיים:

$$\sum_{t=1}^{p} \lambda_t = tr(\Lambda) = tr(U^T VU) \overset{tr(AB)=tr(BA)}{=} tr(VUU^T) \overset{UU^T=I}{=} tr(VI) = tr(V) = \sum_{r=1}^{p} \mathbb{V}[X_r]$$

כזכור $\lambda_t = \mathbb{V}[Y_t]$. מכאן ניתן להגדיר:

$$\pi_t = \frac{\lambda_t}{\sum_{t'=1}^{p} \lambda_{t'}} = \frac{\mathbb{V}[Y_t]}{\sum_{r=1}^{p} \mathbb{V}[X_r]}$$

זוהי **פרופורציית השונות של $X_1, \dots, X_p$ שמוסברת על ידי Y_t**.

נשים לב כמובן כי סכום הפרופורציות הללו מסתכם ל-1 (כל המרכיבים הראשיים מאפשרים 100 אחוז מן השונות המוסברת):

$$\sum_{t=1}^{p} \pi_t = \sum_{t=1}^{p} \frac{\lambda_t}{\sum_{t'=1}^{p} \lambda_{t'}} = \frac{\sum_{t=1}^{p} \lambda_t}{\sum_{t'=1}^{p} \lambda_{t'}} = 1$$

הערה 18.9 – אם נחזור למטריצה G שמוגדרת על ידי $G_{rt} = \mathbb{V}[w_{rt}] = [U]_{rt}^2 \lambda_t$ אזי:

$$\mathbb{V}[X_r] = \sum_{t=1}^{p} \mathbb{V}[w_{rt}] = \sum_{t=1}^{p} [U]_{rt}^2 \lambda_t = \sum_{t=1}^{p} G_{rt}$$

בכיוון ההפוך אם מסתכלים על עמודה כלשהי t ב-G, נשים לב שכאשר סוכמים על העמודה אנחנו בעצם סוכמים על וקטור אורתונורמלי כלשהו. לכן:

$$\sum_{r=1}^{p} G_{rt} = \sum_{r=1}^{p} [U]_{rt}^2 \lambda_t = \lambda_t \sum_{r=1}^{p} [U]_{rt}^2 = \lambda_t \sum_{r=1}^{p} [(u_t)_r]^2 = \lambda_t \|u_t\|^2 \overset{u_t\ orthonormal}{=} \lambda_t \cdot 1 = \lambda_t$$

סיכום ביניים 18.10 – להלן עיקרי נוסחאות בעניין PCA:

$\mathbb{E}[X] = \vec{\mu} = 0$ $\mathbb{V}[X] = V = U\Lambda U^T$	$X \in \mathbb{R}^p$
	$u_1, \dots, u_p\ orthonormal\ base$
$\mathbb{V}[X_r] = V_{rr}$	$X_r \in \mathbb{R}$
$\mathbb{V}[Y_r] = \mathbb{V}[u_r^T X] = u_r^T U\Lambda U^T u_r = \lambda_r \in \mathbb{R}$	$Y_r = u_r^T X \in \mathbb{R}$
$\mathbb{V}[Y] = \Lambda = diag\{\lambda_1, \dots, \lambda_p\} \in \mathbb{R}^{p\times p}$	$Y = U^T X \in \mathbb{R}^{p\times 1}$
$G_{rt} = \mathbb{V}[w_{rt}] = [U]_{rt}^2 \lambda_t$	$w_{rt} = [U]_{rt} Y_t = (u_t)_r Y_t = (u_t)_r u_t^T X$
$\mathbb{V}[X_r] = \sum_{t=1}^{p} G_{rt}$	$X_r = \sum_{t=1}^{p} [U]_{rt} Y_t = \sum_{t=1}^{p} w_{rt}$
$Corr(X_r, w_{rt}) = \sqrt{\frac{\mathbb{V}(w_{rt})}{\mathbb{V}(X_r)}}$	$Cov(X_r, w_{rt}) = \mathbb{V}(w_{rt})$
	$\tilde{X}_r = \sum_{t=1}^{q} w_{rt}$

$Corr(\tilde{X}_r, X_r) = \frac{Cov(\tilde{X}_r, X_r)}{\sqrt{\mathbb{V}(\tilde{X}_r)\mathbb{V}(X_r)}} = \sqrt{\frac{\mathbb{V}(\tilde{X}_r)}{\mathbb{V}(X_r)}}$	$Cov(X_r, \tilde{X}_r) = \sum_{t=1}^{q} \mathbb{V}(w_{rt})$

מסקנה 18.11 – הבנייה שיצרנו של $\tilde{X}$ ממזערת את ה-MSE ביחס ל-X באמצעות סכימה על q המרכיבים הראשיים התואמים ל-q הערכים העצמיים הגדולים ביותר של המטריצה V.

כל מרכיב הוא צירוף לינארי כלשהו של התצפיות המקוריות, וכל מרכיב אורתוגונלי למרכיבים הקודמים. המרכיבים כאמור מסודרים לפי כמות השונות המוסברת שהם מספקים לגבי התצפיות.

בסך הכל התהליך שביצענו הוא כדלקמן:

1. מתוך הווקטור המקורי $X \in \mathbb{R}^p$ יצרנו וקטור חדש $Y = U^T X \in \mathbb{R}^p$.
2. מתוך הווקטור החדש אנחנו לוקחים תת-וקטור בגודל q: $Y^{(q)} = \begin{bmatrix} Y_1 \\ \dots \\ Y_q \end{bmatrix} \in \mathbb{R}^q$.
3. כמו ש-$X = UY$ אזי $\tilde{X} = U^{(q)} Y^{(q)}$ כאשר $U^{(q)}$ הינה המטריצה הבנויה מ-q העמודות הראשונות של U (סוכמים רק את q המרכיבים הראשיים).

הגדרה 18.12 – **Scree Plot** – דרך נוחה להתבונן על היחס בין $\tilde{X}$ ל-X היא באמצעות Scree Plot שבו בציר ה-x נשים את האינדקסים $1, \dots, p$ של הערכים העצמיים לפי הסדר ובציר ה-y את π_t את השונות המוסברת על ידי כל אחד מהמרכיבים הראשיים. נקבל פונקציה מונוטונית יורדת ומכאן נוכל להבין כמה כל משתנה Y תורם לשונות המוסברת (הגרף נסכם ל-1, כלל השונות המוסברת).

שיעור 19

נושא 4 – השוואות מרובות Multiple Comparisons

הקדמה 19.1 – השוואות מרובות נצרכות בכמה מקרים:

1. כאשר משווים בין 2 קבוצות לפי מספר מדדים שונים
2. כאשר משווים בין 2 קבוצות במספר תת-אוכלוסיות
3. להשוות בין מספר קבוצות גדול מ-2

נושא השוואות מרובות הוא אוסף של שיטות סטטיסטיות שאמורות לשלוט על בעיה של גילוי תופעות שווא.

הגדרה 19.2 – נניח שישנן m השערות שאנחנו מעוניינים לבדוק, כאשר כל השערה הינה מהצורה:

$$H_{0_i}: \mu_i = \mu_i^0$$
$$H_{1_i}: \mu_i \neq \mu_i^0$$

בכל מבחן אנחנו מקבלים סטטיסטי T_i ומעוניינים לבדוק את ההשערה i באמצעות כלל הדחיה:

$$|T_i| \geq c_i$$

כאשר ה-$p.value$ להשערה ה-i מסומן p_i ומוגדר על ידי:

$$p_i = \mathbb{P}\left(|T_i| \geq t_i^{obs}\right)$$

תחת H_{0_i} אם T_i הוא מ"מ רציף אזי:

$$p_i \overset{H_{0_i}}{\sim} U(0,1)$$

הגדרה 19.3 - השערת האפס הגלובלית – בהתאם ל-19.2, השערת האפס הגלובלית מוגדרת על ידי:

$$H_0^{Global} = H_0^G = \bigcap_{i=1}^{m} H_{0_i}$$

כלומר זוהי ההשערה לפיה כל השערות האפס מתקיימות.

המשלים של ההשערה הגלובלית $\left(\bigcap_{i=1}^{m} H_{0_i}\right)^c$ הוא ההשערה לפיה ישנה לפחות השערה אחת כלשהי i עבורה H_{0_i} לא נכונה.

טענה 19.4 – את סטטוס ההשערות ניתן להבנות בטבלת אמת מהצורה הבאה:

	דוחים	לא דוחים	
m_0	V	U	H_{0_i} נכון
$m-m_0$	S	T	H_{0_i} לא נכון
$\boldsymbol{m}$	$\boldsymbol{R} = V+S$	$\boldsymbol{m-R}$	

עבור כל השערה יש אפשרות שהיא נכונה או לא נכונה, ועבור כל השערה ישנה אפשרות שדחינו או לא דחינו את ההשערה.

נשים לב שמבין הנתונים בטבלה, אנחנו יודעים את m (כמה השערות יש) ואת R (כמה השערות דחינו ולא דחינו). לעומת זאת אנחנו לא יודעים את m_0, שהוא מספר ההשערות שבאמת נכון, ולא את V, שהוא מספר ההשערות שדחינו למרות שלא היה צריך לדחות אותן.

נשים לב לסימונים:

- V – מספר דחיות השווא (איננו יודעים מה הערך של V למעט כך שהוא חסום על ידי $V \le R \le m$)
- S – מספר דחיות האמת
- U – מספר אי-הדחיות הנכון
- T – מספר אי-הדחיות הלא-נכון (כאשר היה ראוי לדחות אך לא נדחה)

ישנן מספר שיטות לבדיקת השערות במצבים אלו. נתחיל לדון בשיטות הללו עתה.

הגדרה 19.5 – Familywise Error Rate (FWER) – ההסתברות שאנחנו דוחים לשווא השערה כלשהי, כלומר מצב שבו קיימת לפחות השערת אפס נכונה אחת אשר נדחית.

$$FWER = \mathbb{P}(V \ge 1) = \mathbb{P}\left(at\ least\ one\ H_{0_i}\ wrongly\ rejected\right)$$

הגדרה 19.6 – שיטה שולטת על ה-$\boldsymbol{FWER}$ במובן החלש – עבור רמה כלשהי α ניתן להגדיר שליטה על ה-$FWER$ במובן החלש באמצעות:

$$FWER = \mathbb{P}_{H_0^G}(V \ge 1) \le \alpha$$

כלומר ההסתברות תחת השערת האפס הגלובלית שדחינו פעם אחת בטעות השערת אפס כלשהי למרות שהיתה נכונה.

הגדרה 19.7 – שיטת שולטת על ה-$\boldsymbol{FWER}$ במובן החזק – עבור רמה כלשהי α נגדיר שליטה במובן החזק באמצעות:

$$FWER = \mathbb{P}(V \ge 1) \le \alpha$$

עבור כל קומבינציה של H_{0_i}.

טענה 19.8 – נניח שאנחנו עושים m מבחנים בלתי-תלויים ברמה $\alpha = 0.05$. נסמן

$$\pi(m, 0.05) = \mathbb{P}_{H_0^G}(V \geq 1)$$

בגלל אי-תלות בין המבחנים מתקיים:

$$\mathbb{P}_{H_0^G}(V \geq 1) = 1 - \mathbb{P}_{H_0^G}(no\ test\ is\ rejected) \overset{tests\ independent}{=} 1 - (1-\alpha)^m$$

הערה 19.9 – נשים לב לגידול האקספוננציאלי של הסיכוי לטעות (להגיע לפחות לדחיה אחת שגויה, כלומר דחיה של השערת אפס שהיא בפועל נכונה) כאשר מספר ההשערות גדל:

π - הסיכוי לדחיית שווא אחת לפחות	m – מספר ההשערות
0.05	1
0.098	2
0.143	3
0.185	4
0.226	5
0.401	10

נושא 4.1 – שיטות תיקון להשוואות מרובות

נושא 4.1.1 – שיטות תיקון בלי הנחת אי-תלות בין ה-$p.values$

הגדרה 19.10 – שיטת בונפרוני (Bonferroni) לתיקון השוואות מרובות – נניח שישנן m השערות, לכל בדיקת השערות i, דוחים את H_{0_i} אם:

$$p_i \leq \frac{\alpha}{m}$$

טענה 19.11 – בשיטת בונפרוני יש לנו

$$FWER = \mathbb{P}(V \geq 1) \leq \mathbb{E}[V] = \frac{m_0}{m}\alpha \leq \alpha$$

כלומר רמת המובהקות של ה-$FWER$ תחת שיטת בונפרוני נותרת מתחת לרמת המובהקות הנדרשת α.

הערה 19.11.1 – נשים לב כי אי-השוויון $\mathbb{P}(V \geq 1) \leq \mathbb{E}[V]$ מתקיים עבור כל שיטה לתיקון השוואות מרובות. שכן:

$$V \geq \mathbb{I}(V \geq 1) \rightarrow$$
$$\mathbb{E}[V] \geq \mathbb{E}[\mathbb{I}(V \geq 1)] = \mathbb{P}(V \geq 1)$$

הוכחת טענה 19.11 – נסמן

$$N = \{i : H_{0_i}\ is\ correct\}$$
$$R_i = \begin{cases} 1, & H_{0_i}\ was\ rejected \\ 0, & H_{0_i}\ was\ not\ rejected \end{cases}$$

אם כן:

$$V = \sum_{i \in N} R_i$$

כלומר סך הדחיות מתוך השערות האפס שהן דווקא נכונות.

ניתן לחשב את התוחלת:

$$\mathbb{E}[V]=\mathbb{E}\left[\sum_{i\in N}R_i\right]\overset{linearity\ of\ expectation}{=}\sum_{i\in N}\mathbb{E}[R_i]=\sum_{i\in N}\mathbb{P}(R_i=1)$$

$$=\sum_{i\in N}\mathbb{P}\left(p_i\le\frac{\alpha}{m}\right)\overset{p_i\overset{H_{0_i}}{\sim}U(0,1)}{=}\sum_{i\in N}\frac{\alpha}{m}=\frac{m_0\alpha}{m}\le\alpha$$

כאשר m_0 הוא מספר ה-H_{0_i} הנכונות וכמובן $m_0\le m$.

הערה 19.13 – נשים לב כי בשיטת בונפרוני יש לנו שליטה במובן החזק, כי אנחנו שולטים בנפרד על כל השערה.

הערה 19.14 – יש הרואים בשיטת בונפרוני שיטה שמרנית ("מחמירה"). עם זאת יש לשים לב כי אם המבחנים בלתי תלויים אזי:

$$\mathbb{P}^{bonferroni}(V\ge1)=1-\left(1-\frac{\alpha}{m}\right)^m\approx1-e^{-\alpha}$$

כאשר המעבר האחרון בגלל ש-

$$\lim_{m\to\infty}\left(1-\frac{\alpha}{m}\right)^m=e^{-\alpha}$$

לדוגמה אם $\alpha=0.05$ אזי

$$e^{-\alpha}=0.951$$

כלומר עבור מספר גדול מאוד של השערות בלתי-תלויות, שיטת בונפרוני איננה שמרנית יתר על המידה (מבחינת קפדנות על דחיית השערות אפס).

הגדרה 19.15 – שיטת פישר Fisher Least Significant Difference (LSD) Procedure (פישר הוא אותו פישר שעל שמו מוגדרת מטריצת האינפורמציה של פישר) - השיטה כוללת שני שלבים:

1. בדיקת H_0^G ברמה α באמצעות מבחן גלובלי כמו T^2 של $Hoteling$. אם לא דוחים את ההשערה אז עוצרים. אם דוחים עוברים לשלב 2.
2. בודקים כל השערה ברמת מובהקות α.

תזכורת 19.16 – סטטיסטי הוטלינג מוגדר על ידי:

$$T^2=n(\bar{X}-\mu^0)^TS^{-1}(\bar{X}-\mu^0)$$

עבור $X_i\in\mathbb{R}^d$ המתפלג $X_i\sim N(\mu,V)$, השערת אפס גלובלית $H_0^G:\mu=\mu^0$ כך ש- $\mu_i=\mu_i^0$ לכל i.

הערה 19.17 – תכונות הפרוצדורה של פישר:

1. קיימת שליטה על ה-$FWER$ במובן החלש.
2. אין שליטה במובן החזק.

הערה 19.18 – שיטת פישר מבטיחה "מחסום" עבור מקרים בהם השערת האפס הגלובלית לא נדחית כלל, אולם במקרים בהם ההשערה נדחית אין יותר "מעצורים" ברגע שנפרץ המחסום הראשוני. נניח למשל $d=5$, כאשר:

$$\mu^0=\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix},\qquad\mu=\begin{bmatrix}1000000\\0\\0\\0\\0\end{bmatrix}$$

אזי השערת האפס הגלובלית תידחה בגלל הפער בהשערה הראשונה. אולם לאחר מכן (שלב 2 בתהליך) אנחנו נמשיך ל-4 מבחנים שאותם נעשה בלי שום תיקון להשוואות מרובות.

הגדרה 19.19 – **שיטת הולם (Holm) לתיקון השוואות מרובות** – נסדר את ה-p_i שיצאו במבחנים בסדר עולה ובהתאם נסדר מחדש את ההשערות:

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$$
$$H_0^{(1)}, H_0^{(2)}, \ldots \ldots \ldots, H_0^{(m)}$$

שיטת הולם מוגדרת באמצעות השלבים הבאים:

1. נבדוק אם $p_{(1)} > \frac{\alpha}{m}$.
1.1 אם כן, עוצרים (מקבלים את השערת האפס הגלובלית).
1.2 אם לא, דוחים את $H_0^{(1)}$ וממשיכים הלאה.
2. בכל שלב j, נבדוק האם $p_{(j)} > \frac{\alpha}{m+1-j}$:
1.3 אם כן, עוצרים עם דחיה של $H_0^{(1)}, \ldots, H_0^{(j-1)}$ ואי-דחיה של $H_0^{(j)}, \ldots, H_0^{(m)}$.
1.4 אם לא, דוחים גם את $H_0^{(j)}$ וממשיכים הלאה.

משפט 19.20 – שיטת $Holm$ שולטת ב-$FWER$ במובן החזק

הוכחה 19.20 – עבור מספר השערות $m, \ldots, 1$ נבצע את פרוצדורת $Holm$ ונסמן ב-j^* את ה-H_{0_j} הראשונה שדחינו בטעות. ברור כי $j^* \leq m - m_0 + 1$ (מצב של שוויון הוא במצב הקיצוני שבו דחינו קודם את ה-H_{0_j}-שהן אכן לא נכונות ולאחר מכן דחינו השערת אפס נכונה כלשהי בטעות). אם כן:

$$j^* \leq m - m_0 + 1 \Leftrightarrow$$
$$m_0 \leq m - j^* + 1$$

אזי $Holm$ דוחה את $H_0^{(j^*)}$ אם ורק אם התנאי הבא מתקיים:

$$p_{(1)} \leq \frac{\alpha}{m}, p_{(2)} \leq \frac{\alpha}{m-1}, p_{(3)} \leq \frac{\alpha}{m-2}, \ldots, p_{(j^*)} \leq \frac{\alpha}{m - j^* + 1}$$

כלומר:

$$p_{(j^*)} \leq \frac{\alpha}{m_0}$$

נסמן כמו מוקדם:

$$N = \{i: H_{0_i}\ is\ correct\}$$

אזי

$$\mathbb{P}(V \geq 1) \leq \mathbb{P}\left(\min_{i \in N} p_i \leq \frac{\alpha}{m_0}\right) = \mathbb{P}\left(\bigcup_{i \in N}\left\{p_j \leq \frac{\alpha}{m_0}\right\}\right) \leq \mathbb{P}\left(\sum_{j \in N} \mathbb{P}\left(p_j \leq \frac{\alpha}{m_0}\right)\right) \overset{p_j \overset{H_{j_0}}{\sim} U(0,1)}{=} \sum_{j \in N} \frac{\alpha}{m_0}$$

כלומר קיבלנו

$$\mathbb{P}(V \geq 1) \leq \sum_{j \in N} \frac{\alpha}{m_0} = m_0\left(\frac{\alpha}{m_0}\right) = \alpha$$

דוגמה 19.21 – נניח $m = 4$, כאשר

$$p_1 = 0.1$$

$p_2 = 0.5$
$p_3 = 0.01$
$p_4 = 0.03$

רוצים לשלוט ב-$FWER$ ברמה 0.1.

לפי שיטת בונפרוני, נקבל

$$p_i \leq \frac{\alpha}{m} = \frac{0.1}{4} = 0.025 = \begin{pmatrix} not\ reject \\ not\ reject \\ reject \\ not\ reject \end{pmatrix}$$

לפי שיטת הולם:

$$p_{(1)} = 0.01 \leq \frac{\alpha}{m} = 0.025 \to Reject$$
$$p_{(2)} = 0.03 \leq \frac{\alpha}{m-1} = \frac{0.1}{3} = 0.0333 \to Reject$$

מכאן והלאה כבר לא נדחה. כלומר בסך הכל:

$$\begin{pmatrix} not\ reject \\ not\ reject \\ reject \\ reject \end{pmatrix}$$

שיעור 20

תזכורת 20.1 – אנחנו עוסקים בנושא של השערות מרובות, מצב שבו יש לנו m השערות אפס שונות $H_{01}, \dots, H_{0m}$ ורוצים לפתח שיטה שתמנע דחיות שווא. כזכור, סימנו את ערכי ה-$p.value$ השונים שיצאו בבדיקות השערות האפס בתור $p_1, \dots, p_m$ כאשר כזכור $p_i \overset{H_{0i}}{\sim} U(0,1)$. נסמן $D_i = \mathbb{I}\{H_{0i}\ is\ rejected\}$. ואז:

$$R = \sum_{i=1}^{m} D_i = total\ number\ of\ rejections$$

אם בפועל קבוצת השערות האפס הנכונות היא

$$N = \{i: H_{0i}\ is\ correct\}$$

אזי נוכל לסמן את מספר הדחיות השגויות בסימון

$$V = \sum_{i \in N} D_i$$

הגדרנו את ה-$FWER$ (Family-Wise Error Rate) בתור ההסתברות לדחות לפחות השערת אפס נכונה אחת:

$$FWER = \mathbb{P}(V \geq 1)$$

ראינו עד כה שלוש שיטות לתיקון השוואות מרובות:

1. שיטת בונפרוני – לדחות את H_{0i} אם $p_i \leq \frac{\alpha}{m}$.
2. שיטת הולם – המוגדרת בשלבים, מסדרים את ערכי ה-p-value בסידור $p_{(1)}, \dots, p_{(m)}$ מהקטן לגדול, כאשר בשלב j בודקים האם $p_{(j)} > \frac{\alpha}{m+1-j}$. אם כן עוצרים ודוחים את $H_0^{(1)}, \dots, H_0^{(j-1)}$. אם לא ממשיכים את התהליך.

3. שיטת פישר

לבסוף הגדרנו את השערת האפס הגלובלית H_0^G בתור $H_0^G = \cap_{i=1}^m H_{0i}$.

הגדרה 20.2 – ערך p מתוקן – הערך α הקטן ביותר עבורו ניתן לדחות את H_{0i}, נסמנו p'.

טענה 20.3 – בשיטות השונות ערכי ה-$p.value$ המתוקנים ניתנים להגדרה באופן הבא:

<u>שיטת בונפרוני</u> - $mp_i \le \alpha \Leftrightarrow p_i \le \frac{\alpha}{m}$ כלומר ערך ה-p המתוקן (שבו דוחים את H_{0i}) בשיטת בונפרוני הוא $p_i' = mp_i$.

<u>שיטת הולם</u> – דוחים את H_{0_i} אם ורק אם $p_{(j)} \le \frac{\alpha}{m-j+1}$ עבור $j = 1, \dots, i$:

$$
\begin{aligned}
& p_{(j)} \le \frac{\alpha}{m-j+1}, \quad j = 1, \dots, i \\
& \Leftrightarrow (m-j+1)p_{(j)} \le \alpha, \qquad j = 1, \dots, i \\
& \Leftrightarrow \max_{j=1,\dots i}\left((m-j+1)p_{(j)}\right) \le \alpha
\end{aligned}
$$

כלומר הערך p המתוקן עבור שיטת הולם הוא

$$p'_{(i)} = \max_{j=1,\dots i}(m-j+1)p_{(j)}$$

סיכום ביניים 20.4 – שיטות תיקון שונות

שיטה	שליטה במובן החלש	שליטה במובן החזק	ערך p מתוקן
בונפרוני		V	mp_i
פישר	V	X	
הולם			$\max_{j=1,\dots i}(m-j+1)p_{(j)}$

נושא 4.1.2 – שיטות תיקון כאשר יש תלות בין ה-$p.values$

הערה 20.5 עד כה הצגנו שיטות לתיקון שאינן מניחות הנחות על מבנה התלות בין ערכי ה-$p.value$ השונים. עתה נציג שיטה חדשה שיש בה הנחה כזאת, המבוססת על התפלגות רב-נורמלית.

דוגמה 20.6 – תלות בין ערכי $p.value$ בניתוח שונות חד-כיווני – במודל ניתוח שונות חד-כיווני אנחנו מניחים $Y_{ij} = \mu_i + \varepsilon_i$, ההפרש הוא $\Delta_{ij} = \mu_i - \mu_j$ והאומד להפרש הוא $\hat{\Delta}_{ij} = \bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}$. מכאן ניתן לחשב את השונות המשותפת על ידי:

$$Cov\left(\bar{Y}_{j\cdot} - \bar{Y}_{i\cdot}, \bar{Y}_{j\cdot} - \bar{Y}_{i\cdot}\right) = Cov\left(\bar{Y}_{j\cdot}, \bar{Y}_{j\cdot}\right) - Cov\left(\bar{Y}_{j\cdot}, \bar{Y}_{i\cdot}\right) - Cov\left(\bar{Y}_{i\cdot}, \bar{Y}_{j\cdot}\right) + Cov\left(\bar{Y}_{i\cdot}, \bar{Y}_{i\cdot}\right)$$

כאשר הביטוי האחרון הוא $\mathbb{V}(\bar{Y}_{i\cdot}) = \frac{\sigma_\varepsilon^2}{n_i} > 0$. כלומר יש לנו שני הפרשים שאנחנו רוצים לבדוק ויש לנו תלות בין האומדים האלה.

הגדרה 20.7 – שיטת תיקון שגיאות מבוססת התפלגות רב-נורמלית

נניח שקיים וקטור פרמטרים כלשהו שרוצים לאמוד $\vec{\phi} \in \mathbb{R}^m$ באומד $\hat{\phi} \dot{\sim} N(\phi, \Omega)$. בהקשר שלנו, נניח שווקטור השערות האפס שלנו הוא:

$$H_{0i}: \phi_i = \phi_i^0 \Leftrightarrow \overrightarrow{H_0}: \vec{\phi} = \vec{\phi}^0$$

נגדיר את הערך המתוקנן:

$$Z_i = \frac{\hat{\phi}_i - \phi_i^0}{\sqrt{\Omega_{ii}}} \overset{H_{0i}}{\dot{\sim}} N(0,1)$$

יחד כל הערכים המתוקנים ניתנים להגדרה בווקטור $\vec{Z} \in \mathbb{R}^m$ על ידי:

$$\vec{Z} = \begin{bmatrix} Z_1 \\ \dots \\ Z_m \end{bmatrix} \overset{H_0}{\dot{\sim}} N(\vec{0}, \Gamma)$$

כאשר $\Gamma = I_m$ במצב של אי-תלות בין הערכים המתוקננים השונים. כל קואורדינטה במטריצה Γ אם כן מבטאת את:

$$\Gamma_{rs} = Corr(\phi_r, \phi_s) = \frac{\Omega_{rs}}{\sqrt{\Omega_{rr}\Omega_{ss}}}$$

נוכל להשתמש בכלל הדחיה הבא: **לכל i, נדחה את H_{0i} אם $|Z_i| \geq Z^*$**. כאשר Z^* מוגדר להיות הערך של Z כך ש-

$$\mathbb{P}\left(\max_{1\leq i\leq m} |\vec{\zeta_i}| \leq Z^*\right) = 1 - \alpha$$

והווקטור $\vec{\zeta} \in \mathbb{R}^m$ הוא וקטור שמקיים:

$$\zeta \sim N(0, \Gamma)$$
$$\mathbb{V}(\zeta_i) = 1$$
$$Corr(\zeta_i, \zeta_j) = \Gamma_{ij}$$

דוגמה 20.8 – עבור $m = 2$ הווקטור Z מגדיר מרחב מלבני במישור שבו מתקבלות שתי השערות האפס, המתוחם בנקודות $(-z_1^*, z_1^*, -z_2^*, z_2^*)$.

$$\vec{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \qquad \vec{\phi}^0 = \begin{pmatrix} \phi_1^0 \\ \phi_2^0 \end{pmatrix}, \qquad \hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \sim N_2(\phi, \Omega)$$

$$\vec{Z} = \begin{pmatrix} \frac{\hat{\phi}_1 - \phi_1^0}{\sqrt{\Omega_{11}}} \\ \frac{\hat{\phi}_2 - \phi_2^0}{\sqrt{\Omega_{22}}} \end{pmatrix} \overset{H_0}{\sim} N(0, \Gamma)$$

$$\Gamma = \begin{pmatrix} 1 & \frac{\Omega_{12}}{\sqrt{\Omega_{11}\Omega_{22}}} \\ \frac{\Omega_{12}}{\sqrt{\Omega_{11}\Omega_{22}}} & 1 \end{pmatrix}$$

$$Z^* = \begin{pmatrix} Z_1^* \\ Z_2^* \end{pmatrix}$$

נושא 4.1.3 – False Discovery Rate - FDR

הערה 20.9 – הקריטרון של $FWER$ הוא קריטריון מחמיר שאנחנו רוצים לשמור עליו כאשר מספר ההשערות קטן עד בינוני ומחיר הטעות (אפילו עבור טעות אחת) הוא גבוה. לעומת זאת יש מקרים שבהם יש לנו אלפי השערות ושליטה על ה-$FWER$ היא מחמירה מדי. קריטריון אחר שניתן להשתמש בו ה-FDR, קריטריון שלא חזק כמו ה-$FWER$ אבל עדין מאפשר שליטה בכמות השגיאות. הקריטריון הומצא על ידי בנימיני והוכברג.

הגדרה 20.10 – False Discovery Proportion – כזכור, R הוא מספר הדחיות הכולל, ו-V הוא מספר הדחיות השגויות. אזי ה-False Discovery Proportion (FDP) הוא אחוז הדחיות שבו היתה טעות, ומוגדר על ידי:

$$\boldsymbol{FDP} = \begin{cases} \boldsymbol{V/R}, & \boldsymbol{R \geq 1} \\ \boldsymbol{0}, & \boldsymbol{R = 0} \end{cases}$$

הגדרה 20.11 – False Discovery Rate (FDR) מוגדר על ידי

$$\boldsymbol{FDR} = \mathbb{E}[\boldsymbol{FDP}]$$

עבור אחוז טעות שאנחנו מוכנים לסבול $q \in (0,1)$ אנחנו רוצים לוודא שיתקיים:

$$FDR \leq q$$

הגדרה 20.12 – הפרוצדורה של בנימיני-הוכברג – שיטת BH – כמו בשיטת הולם, מסדרים את ערכי ה-$p.value$ בסדר עולה ואת ההשערות באופן מקביל:

$$\begin{gathered} p_{(1)} \leq \cdots \leq p_{(m)} \\ H_0^{(1)}, \dots, \dots, H_0^{(m)} \end{gathered}$$

נגדיר i^* להיות הערך של i הכי גדול כך ש-$p_{(i)} \leq \frac{i \cdot q}{m}$. אז דוחים את $H_0^{(1)}, \dots, H_0^{(i^*)}$.

$$i^* = \max_{i=1,\dots,m} p_{(i)} \leq \frac{i \cdot q}{m}$$

משפט 20.13 – אם $p_1, \dots, p_m$ בלתי תלויים אזי עבור שיטת BH מתקיים

$$FDR(BH) = \frac{m_0 q}{m}$$

כאשר m_0 הוא מספר השערות האפס הנכונות.

הוכחה 20.13 – אם $m_0 = 0$ אז התנאי מתקיים ממילא שכן כל הדחיות נכונות, ואז $FDR = 0$ ו-$\frac{m_0 q}{m} = 0$, לכן נניח $1 \leq m_0$.

סימוני עזר 20.13.1 –

$$\begin{aligned} &D_i = \mathbb{I}(H_{0i}\ was\ rejected) \\ &N = \{i : H_{0i}\ is\ correct\}, \\ &|N| = m_0 \\ &R = \sum_{i=1}^{m} D_i = total\ number\ of\ rejections \\ &V = \sum_{i \in N} D_i = number\ of\ false\ rejections \\ &FDP = \begin{cases} V/R, & R \geq 1 \\ 0, & R = 0 \end{cases} \end{aligned}$$

נגדיר $R' = \max(R, 1)$ ואז נוכל להציג את ה-FDP באמצעות:

$$FDP = \frac{V}{R'} = \frac{\sum_{i \in N} D_i}{R'} = \sum_{i \in N} \frac{D_i}{R'} = \sum_{i \in N} X_i$$

כאשר נגדיר

$$X_i = \frac{D_i}{R'}$$

טענת עזר 20.3.2 - נוכיח בהמשך כי מתקיים

$$\mathbb{E}[X_i] = \frac{q}{m} \forall_{i \in N}$$

המשך הוכחה 20.3 - נניח כי טענת העזר מתקיימת לעת עתה. אזי:

$$FDR = \mathbb{E}[FDP] = \mathbb{E}\left[\sum_{i \in N} X_i\right] \overset{linearity}{=} \sum_{i \in N} \mathbb{E}[X_i] \overset{20.3.2}{=} \sum_{i \in N} \frac{q}{m} \overset{|N|=m_0}{=} \frac{m_0 q}{m}$$

הוכחת 20.3.2 – מספר הדחיות R חסום בין 0 ל-m.

$$X_i = \frac{D_i}{R'} = D_i \sum_{k=1}^{m} \frac{\mathbb{I}\{R = k\}}{R} = D_i \sum_{k=1}^{m} \frac{\mathbb{I}(R = k)}{k}$$

כאשר אם $R = 0$ אזי $D_i = 0$ לכן לא צריך לכלול את $k = 0$ בתוך הסכום.

נגדיר $R(p_i \to 0)$ להיות מספר הדחיות שהיינו מקבלים אם היינו מחליפים את p_i (לפני סידור הערכים) ב-0.

טענת עזר 20.3.2.1 -

$$D_i \cdot \mathbb{I}\{R = k\} = D_i \cdot \mathbb{I}(R(p_i \to 0) = k)$$

הוכחת טענת עזר 20.3.2.1 – אם $D_i = 0$ השוויון מתקיים באופן טריוויאלי.

אם לעומת זאת $D_i = 1$ אזי הפעולה $p_i \to 0$ לא גורמת לשום שינוי כי כבר דחינו את H_{0i}. לכן השוויון מתקיים.

טענת עזר 20.3.2.2 –

$$D_i \cdot \mathbb{I}(R(p_i \to 0) = k) = \mathbb{I}\left(p_i \le \frac{kq}{m}\right) \cdot \mathbb{I}(R(p_i \to 0) = k)$$

הוכחת טענת עזר 20.3.2.3 – גם כאן נוכיח באמצעות חלוקה למקרים. ראשית נשים לב כי הפעולה $p_i \to 0$ לא יכולה להוריד את מספר הדחיות הכולל. אלא, או שלא נגרם כל שינוי מבחינת מספר הדחיות, או שמספר הדחיות הכולל (R) גדל ב-1.

במקרה א', לא נגרם שינוי, בפרט הפעולה $p_i \to 0$ לא הופך מצב של $D_i = 0$ למצב $D_i = 1$.

המשך ההוכה בשיעור הבא.

שיעור 21

הערה 21.1 – נשים לב כי שליטה ב-FDR במובן החלש ברמה כלשהי q פירושה

$$\mathbb{E}_{H_0^G}[FDP] \le q$$

ושליטה ב-FDR במובן החזק פירושה עבור כל מערך אפשרי של $H_{01}, \ldots, H_{0m}$:

$$\mathbb{E}[FDP] \le q$$

הערה 21.2 - תחת H_0^G, יש לנו $FDR = FWER$ (הוכחה בתרגיל בית)

הערה 21.3 – נשים לב כי:

$$FDP = \begin{cases} V/R, & R \geq 1 \\ 0, & R = 0 \end{cases} = \begin{cases} V/R, & V \geq 1 \\ 0, & otherwise \end{cases}$$

אם $R \geq 1$ ו-$V = 0$ אז

$$FDP = \frac{V}{R} = 0$$

טענה 21.4 - $FWER \geq FDR$.

הוכחת טענה 21.4 – נשים לב כי $V/R \in [0,1]$. ואז:

$$FDP = \begin{cases} V/R, & R \geq 1 \\ 0, & R = 0 \end{cases} \leq \mathbb{I}\{V \geq 1\} = \begin{cases} 1, & V \geq 1 \\ 0, & otherwise \end{cases}$$

ובתוחלת:

$$FDR = \mathbb{E}[FDP] \leq \mathbb{E}[\mathbb{I}\{V \geq 1\}] \overset{expec.of\ indicator\ is\ prob.}{=} \mathbb{P}(V \geq 1) \overset{by\ definition}{=} FWER$$

תזכורת 21.5 – הפרודצורה של בנימיני-הוכברג - מסדרים את ערכי ה-$p.value$ בסדר עולה ואת ההשערות באופן מקביל:

$$p_{(1)} \leq \cdots \leq p_{(m)}$$
$$H_0^{(1)}, \dots, \dots, H_0^{(m)}$$

נגדיר i^* להיות הערך של i הכי גדול כך ש-$p_{(i)} \leq \frac{i^* \cdot q}{m}$. אז דוחים את $H_0^{(1)}, \dots, H_0^{(i^*)}$.

הערה 21.6 – הצגה גרפית של BH – מייצרים גרף שבו

- בציר ה-x אנחנו ממקמים את האינדקס i (לאחר הסידור מחדש מהקטן לגדול)
- בציר ה-y ערכי ה-$p_{(i)}$
- על גבי הגרף אנחנו לוקחים קו עם שיפוע $y = \frac{q}{m}$.
- מסתכלים על כל ערכי ה-$p.value$, כאשר חלק מהערכים $p_{(i)}$ מתחת לקו וחלק מעל הקו.
- i^* מוגדר להיות האינדקס הימני ביותר (הגדול ביותר) כך ש-$p_{(i)}$ מתחת לקו.

נשים לב – ערכי ה-$p_{(i)}$ הם מונוטוניים-עולים ביחס ל-i, אבל ביחס לשיפוע יכול להיות שיהיו ערכים מעל הקו אבל אז הם ימשיכו באופן מונונטוני עולה נמוך כך שיהיה ערך בהמשך שיהיה מתחת לקו השיפוע.

תזכורת 21.7 – עצרנו באמצע הוכחה של משפט 20.13 – אם $p_1, \dots, p_m$ בלתי תלויים אזי עבור שיטת BH מתקיים $FDR(BH) = \frac{m_0 q}{m}$, כאשר m_0 הוא מספר ההשערות השגויות. נרצה לסיים עתה את ההוכחה.

נזכיר את ההגדרות שבהן השתמשנו בהוכחה:

$$D_i = \mathbb{I}(H_{0i}\ was\ rejected)$$
$$N = \{i : H_{0i}\ is\ correct\}$$
$$R = \sum_{i=1}^{m} D_i = total\ number\ of\ rejections$$
$$V = \sum_{i \in N} D_i = number\ of\ false\ rejections$$
$$FDP = \begin{cases} V/R, & R \geq 1 \\ 0, & R = 0 \end{cases}$$

$$R' = \max(R, 1)$$

$$FDP = \frac{V}{R'} = \frac{\sum_{i\in N} D_i}{R'} = \sum_{i\in N} \frac{D_i}{R'} = \sum_{i\in N} X_i$$

ראינו את טענת עזר 20.3.2 -

$$\forall_{i\in N} \coloneqq \mathbb{E}[X_i] = \frac{q}{m}$$

אשר בשביל הוכחתה ראינו שתי טענות עזר נוספות:

1. טענת עזר 20.3.2.1

$$D_i \cdot \mathbb{I}\{R = k\} = D_i \cdot \mathbb{I}(R(p_i \to 0) = k)$$

2. טענת עזר 20.3.2.2

$$D_i \cdot \mathbb{I}(R(p_i \to 0) = k) = \mathbb{I}\left(p_i \le \frac{kq}{m}\right) \cdot \mathbb{I}(R(p_i \to 0) = k)$$

נשים לב:

$$X_i = \frac{D_i}{R'} = \sum_{k=1}^{m} D_i \cdot \frac{\mathbb{I}\{R = k\}}{k} = D_i \sum_{k=1}^{m} \mathbb{I}\{R = k\} = D_i$$

המשך הוכחת טענת עזר 20.3.2.2 –

נגדיר $R(p_i \to 0)$ להיות מספר הדחיות שהיינו מקבלים אם היינו מחליפים את p_i (לפני סידור הערכים) ב-0. נשים לב – הזזה של $p_{(i)}$ כלשהו ל-0 באופן גרפי פירושה שמאפסים ערך כלשהו ומסיטים ימינה את כל הערכים האחרים (שחייבים להיות גדולים או שווים לאפס).

שוב, נשים לב כי הפעולה $p_i \to 0$ לא יכולה להוריד את מספר הדחיות הכולל. אלא רק המצבים הבאים אפשריים:

מקרה א' - לא נגרם כל שינוי מבחינת מספר הדחיות $\boldsymbol{R = R(p_i \to 0)}$ - במצב זה בהכרח $D_i = 1$, כי אחרת הפעולה $p_i \to 0$ היתה מעבירה את D_i מ-0 ל-1. בנוסף יש לנו $R = k$ ולכן

$$p_{(i)} \le p_{(k)} \overset{BH\ procedure}{\le} \frac{kq}{m}$$

מקרה ב' - מספר הדחיות הכולל $\boldsymbol{R}$ גדל (עקרונית יכול לגדול ביותר מ-1). כלומר כאשר איפסנו את p_i באמצעות הפעולה $p_i \to 0$ הגדלנו את מספר הדחיות הכולל. אזי חייב להיות $D_i = 0$.

אם $R(p_i \to 0) \neq k$ אז השוויון מתקיים באופן טריוויאלי $0 = 0$.

אחרת, נניח $R(p_i \to 0) = k$, ואז אנחנו במצב הבא:

1. $p_i \to 0$ גרם לעליה של R.
2. $R < k$.

שני הדברים האלה יכולים לקרות יחד רק אם

$$p_i = p_{(k)} > \frac{kq}{m}$$

עתה, משסיימנו את הוכחת 20.3.2.1, 20.3.2.2, נחזור להוכחת 20.3.2, כזכור הטענה היא:

$$\forall_{i\in N} \coloneqq \mathbb{E}[X_i] = \frac{q}{m}$$

נגדיר F_i להיות הווקטור המקרי שכולל את כל ה-$p.values$ חוץ מ-p_i כלומר:

$$F_i^T = [p_1 \quad \cdots \quad p_{i-1} \quad p_{i+1} \quad \cdots \quad p_m]$$

מתוך הטענות שהוכחנו אנחנו יודעים:

$$D_i \cdot \mathbb{I}\{R = k\} = D_i \cdot \mathbb{I}(R(p_i \to 0) = k)$$

$$D_i \cdot \mathbb{I}(R(p_i \to 0) = k) = \mathbb{I}\left(p_i \le \frac{kq}{m}\right) \cdot \mathbb{I}(R(p_i \to 0) = k)$$

משתי הטענות הנ"ל אנחנו יכולים להסיק:

$$D_i \cdot \mathbb{I}\{R = k\} = \mathbb{I}\left(p_i \le \frac{kq}{m}\right) \cdot \mathbb{I}(R(p_i \to 0) = k)$$

ואז:

$$X_i = \frac{D_i}{R'} = D_i \sum_{k=1}^{m} \frac{\mathbb{I}\{R = k\}}{R} = \sum_{k=1}^{m} \frac{D_i \cdot \mathbb{I}(R = k)}{k} = \sum_{k=1}^{m} \frac{\mathbb{I}\left(p_i \le \frac{kq}{m}\right) \cdot \mathbb{I}(R(p_i \to 0) = k)}{k}$$

ניתן אם כן להוציא תוחלת מותנית של X_i בהינתן F_i ולקבל:

$$\mathbb{E}[X_i | F_i] =$$
$$\mathbb{E}\left[\sum_{k=1}^{m} \frac{\mathbb{I}\left(p_i \le \frac{kq}{m}\right) \cdot \mathbb{I}(R(p_i \to 0) = k)}{k} | F_i\right] =$$
$$\sum_{k=1}^{m} \frac{\mathbb{E}\left[\mathbb{I}\left(p_i \le \frac{kq}{m}\right) \cdot \mathbb{I}(R(p_i \to 0) = k) | F_i\right]}{k} \overset{\mathbb{I}(R(p_i \to 0)=k) \perp p_i \Rightarrow \mathbb{I}(R(p_i \to 0)=k) \perp F_i}{=}$$
$$\sum_{k=1}^{m} \frac{\mathbb{I}(R(p_i \to 0) = k) \cdot \mathbb{E}\left[\mathbb{I}\left(p_i \le \frac{kq}{m}\right) | F_i\right]}{k} \overset{cond.\ expectation\ \ to\ prob. because\ of\ ind.\ \ across\ p_1,\dots,p_m}{=}$$
$$\sum_{k=1}^{m} \frac{\mathbb{I}(R(p_i \to 0) = k) \cdot \mathbb{E}\left[\mathbb{I}\left(p_i \le \frac{kq}{m}\right)\right]}{k} \overset{expectation\ of\ indicator\ is\ prob.}{=}$$
$$\sum_{k=1}^{m} \frac{\mathbb{I}(R(p_i \to 0) = k) \cdot \mathbb{P}\left(p_i \le \frac{kq}{m}\right)}{k} \overset{p_i \sim U(0,1)}{=}$$
$$\sum_{k=1}^{m} \frac{\mathbb{I}(R(p_i \to 0) = k) \frac{kq}{m}}{k} =$$
$$\frac{q}{m} \sum_{k=1}^{m} \mathbb{I}(R(p_i \to 0) = k) \overset{running\ on\ all\ options\ 1,\dots,m\ exactly\ one\ is\ true}{=}$$
$$\frac{q}{m} \cdot 1 = \frac{q}{m}$$

הוכחנו עבור $\forall i \in N$:

$$\mathbb{E}[X_i | F_i] = \frac{q}{m}$$

לכן התוחלת הבלתי-מותנית:

$$\mathbb{E}[X_i] = \mathbb{E}\big[\mathbb{E}[X_i|F_i]\big] = \mathbb{E}\left[\frac{q}{m}\right] = \frac{q}{m}$$

כנדרש.

סיכום 21.8 – מה שראינו בסך הכל:

$$FDP = \frac{V}{R'} = \frac{\sum_{i\in N} D_i}{R'} \overset{X_i=\frac{D_i}{R'}}{=} \sum_{i\in N} X_i$$

הוכחנו כי עבור $i \in N$:

$$\mathbb{E}[X_i] = \frac{q}{m}$$

ואז:

$$FDR = \mathbb{E}[FDP] = \mathbb{E}\left[\sum_{i\in N} X_i\right] = \sum_{i\in N} \mathbb{E}[X_i] = \sum_{i\in N} \frac{q}{m} = \frac{m_0 q}{m}$$

כמובן מתקיים $m_0 < m$ ולכן גם $\frac{m_0}{m} q < q$ כלומר פרוצדורת BH משמרת רמת מובהקות q.

שיעור 22

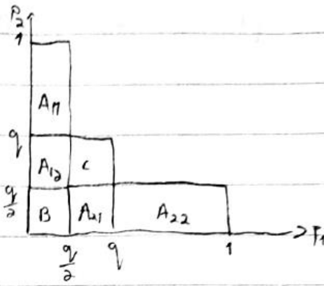
המשך - השוואות מרובות

<u>BH תחת תלות:</u>

דוגמה - נניח $X_1,\dots,X_n \overset{iid}{\sim} N_m(\mu,V)$. רוצים לבדוק $H_{0i}:\mu_i=0$ עבור $i=1,\dots,m$.

כאשר $m=2$:

נניח ש- H_{01}, H_{02} שניהם נכונים כך ש- $P_i \overset{H_{0i}}{\sim} U(0,1)$ אך יכולה להיות תלות בין P_1 ו- P_2.



- אם $P_1,P_2 \in A_{12}\cup A_{21}\cup C\cup B$ אז נדחה את שתי ההשערות H_{01},H_{02}
- ב- A_{11} דוחים את H_{01} ולא את H_{02}.
- ב- A_{22} דוחים את H_{02} ולא את H_{01}.

מכיוון ששתי ההשערות נכונות, אז כל דחייה היא שגיאה ו- $V=R$.

לכן $FDR = \begin{cases} 1 & R\ge 1 \\ 0 & R=0 \end{cases}$.

נניח בה"כ $P_1 \le \frac{3}{4}q$, $P_2 \le q$, $P_1 \le P_2$.

$FDR = E(FDP) = P(R \ge 1) = P(A_{11}) + P(A_{22}) + P(A_{12}) + P(A_{21}) + P(B) + P(C)$

דחיית $H_{02} = \{p: p_2 \le \frac{q}{2}\} = B \cup A_{21} \cup A_{22} \Leftarrow P(B) + P(A_{21}) + P(A_{22}) = \frac{q}{2}$ (כיוון ש $P_i \sim U(0,1)$)

דחיית $H_{01} = \{p: p_1 \le \frac{q}{2}\} = B \cup A_{12} \cup A_{11} \Leftarrow P(B) + P(A_{12}) + P(A_{11}) = \frac{q}{2}$

לכן- $q = 2P(B) + P(A)$ (כאשר A הוא איחוד A_{ij}). $\Leftarrow P(A) = q - 2P(A)$

נציב ב- FDR - $P(R \ge 1) = q + P(C) - P(B)$

רוצים כעת למצוא את $P(C)$ ו- $P(B)$ בשביל לקבל את ה- FDR הגרוע ביותר.

$P(C) = P(\frac{q}{2} \le P_1 \le q, \frac{q}{2} \le P_2 \le q) \le P(\frac{q}{2} \le P_1 \le q) = \frac{q}{2}$

לכן המקרה הגרוע יהיה כאשר $P(B) = 0$, $P(C) = \frac{q}{2}$. $\Leftarrow P(R \ge 1) = \frac{3}{2}q$

<u>מסקנה</u>- ההסתברות המקסימלית לטעות אחת לפחות היא $\frac{3}{2}q$. זאת לעומת מצב של אי-תלות בו ההסתברות המקסימלית לטעות היא q. במקרה של $m=2$.

$$q = 2\mathbb{P}(B) + \mathbb{P}(A) \to \mathbb{P}(A) = q - 2\mathbb{P}(B)$$
$$\mathbb{P}(C) \le \frac{q}{2}$$
$$\mathbb{P}(R \ge 1) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) = q + \mathbb{P}(C) - \mathbb{P}(B)$$

<u>משפטים:</u>

1. תחת m כלשהו ותלות כלשהי, מתקיים $FDR_{BH} \le S(m)\frac{m_0}{m} \cdot q$, כאשר $S(m) = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{m}$ עבור התפלגות כללית.
2. עבור כל m אפשר למצוא התפלגות משותפת של $(P_1, \dots, P_m)$ עם $P_i \sim U(0,1)$ לכל i.
ומתקיים- $FDR_{BH} = \min\{q \cdot S(m), 1\}$

הערה- ניתן לראות ש- BH חוסם בצורה פחות חזקה מבונפרוני.

נושא 4.1.4 – PDRS

<u>תנאי PDRS</u>: Positive regression dependence on a subset

1. עבור $z, w \in \mathbb{R}^m$ נגיד ש- $z \ge w$ אם $z_i \ge w_i$ לכל i.
2. עבור $D \subseteq \mathbb{R}^m$, נגיד ש- D הוא סט עולה אם עבור כל $w \in D$, מתקיים $z \ge w \Rightarrow z \in D$.
הגדרה שקולה- $\{z: z \ge w\} \subset D$ לכל $w \in D$.
3. נניח $X_1, \dots, X_m$ וקטור מקרי. ההתפלגות המשותפת של הוקטור נקראת PDRS על I_0 (סט אינדקסים) אם עבור כל $i \in I_0$, ועבור כל D עולה, הפונקציה $h(x) = P((X_1, \dots, X_m) \in D / X_i = x)$ היא פונקציה עולה ממש של x.

<u>טענה:</u>

אם $X_1,\dots,X_m$ מקיימים PDRS, והפונקציות $f_1,f_2,\dots,f_m$ הן פונקציות עולות ממש, אזי הוקטור $(f_1(X_1),f_2(X_2),\dots,f_m(X_m))$, מקיים PDRS גם כן.

<u>הערה:</u>

אם יש לנו השערות $H_{0i}:\mu_i\le\mu_i^0$ מול $H_{1i}:\mu_i>\mu_i^0$, ויש לנו וקטור סטטיסטים $(X_1,\dots,X_m)$, נניח כי הוקטור מקיים PDRS על N, כאשר $N=\{i: H_{0i} \text{ נכונה}\}$.
דוחים את H_{0i} כאשר $X_i\ge C$.
אזי אם נגדיר $P_i=F_{H_{0i}}(X_i)\sim U(0,1)$, נקבל שגם $P_1,\dots,P_m$ הינו PDRS על N.

שיעור 23

תזכורת 23.1 – הגדרנו בשיעור הקודם את המושג PDRS.

טענה 23.2 – נניח כי $X\sim N(\mu,V)$. אזי X מקיים את תכונת PDRS על I_0 אם ורק אם לכל $i\in I_0$ ולכל $j\in\{1,\dots,m\}$ מתקיים $V_{ij}\ge 0$.

הוכחת טענה 23.2 – לא תינתן בכיתה, ניתן למצוא את ההוכחה בהרצאה מספר 7 של Candes באתר הקורס.

משפט 23.3 – נניח שקיימות m השערות $H_{01},\dots,H_{0m}$, כאשר p_i הוא ה-$p.value$ עבור H_{0i} וקבוצת השערות האפס הנכונות היא: $N=\{i: H_{0i}\ is\ correct\}$ ו-m_0 הוא מספר ההשערות הנכונות. אם הווקטור $(p_1,\dots,p_m)$ הוא PDRS על N אזי:

$$FDR\big(BH(q)\big)\le\frac{q\cdot m_0}{m}$$

כלומר אם מתקיים PDRS אזי יש שמירה ברמה q של ה-False Discovery Rate על תהליך BH.

נושא 4.1.5 – הסקה אחרי בחירה ו-False Discovery Rate

דוגמה 23.4 – נתבונן בדוגמה הבאה:

1. מריצים רגרסיה רב-משתנית על קובץ נתונים כלשהו.
2. מחליטים לפי התוצאות שהתקבלו אילו משתנים להשאיר במודל. לדוגמה נניח שמחליטים לזרוק מהמודל את כל המשתנים המסבירים עבורם ערך ה-$p.value$ גדול או שווה לרמת מובהקות כלשהי, או לחילופין מבצעים stepwise regression באמצעות AIC או BIC.
3. רוצים לחשב רווחי סמך ל-β_i (המקדמים) של המשתנים שנותרו במודל לאחר שלב ה- feature selection.
4. אם לוקחים את רווחי הסמך ללא התייחסות לתהליך הבחירה ולערכי ה-$p.value$ אנחנו מקבלים תוצאות מוטות.

הערה 23.5 - AIC ו-BIC – תחת מודל לינארי מלא מתקיים:

$$Y_i=\sum_{j=0}^{p}\beta_jX_{ij}+\varepsilon_i$$
$$\varepsilon_i\sim N(0,\sigma^2)$$

פונקציית הנראות היא:

$$\mathcal{L}(\vec{\beta},\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left(Y_i-\sum_{j=0}^{p}\beta_jX_{ij}\right)^2\right\}$$

ולוג הנראות היא:

$$\ell(\vec{\beta},\sigma^2)=\log\left(\mathcal{L}(\vec{\beta},\sigma^2)\right)$$

כדי 'לקנוס' על מספר גדול של משתנים ישנן שיטות רגולריזציה. אם p' הוא מספר המשתנים במודל הרגרסיה ו-n הוא מספר הדוגמאות אזי:

$$\tilde{\ell}_{AIC}(\vec{\beta},\sigma^2)=2p'-2\ell(\vec{\beta},\sigma^2)$$
$$\tilde{\ell}_{BIC}(\vec{\beta},\sigma^2)=p'\cdot\log(n)-2\ell(\vec{\beta},\sigma^2)$$

שני הקריטריונים הם לינאריים במספר המשתנים במודל, כאשר ה-AIC הוא גם פונקציה לינארית בגודל המדגם ואילו ה-BIC הוא פונקציה לוגריתמית ביחס לגודל המדגם.

דוגמה 23.6 – ניתן שתי דוגמאות למקרה שתיארנו בדוגמה 23.4.

דוגמה 23.6.1 (Buja) – נניח מודל רגרסיה:

$$Y_i=\beta_0X_{i0}+\sum_{j=1}^{p}\beta_jX_{ij}+\varepsilon_i$$

כאשר X_0 הוא המשתנה שאנחנו מעוניינים בו במיוחד.

נניח

$$n=250$$
$$p=10$$
$$\alpha=0.05$$
$$\varepsilon_i\overset{iid}{\sim}N(0,1)$$

כיוון שאנחנו מעוניינים ב-X_0 (ובפרט מעוניינים לאמוד רווחי סמך עבור X_0), אנחנו מותירים אותו תמיד במודל, ועושים stepwise selection על יתר המשתנים המסבירים במודל $X_1,\dots,X_p$ (בדוגמה של Buja בוצעה אופטימיזציה באמצעות BIC). בסוף התהליך מקבלים תת-קבוצה כלשהי של המשתנים המסבירים אשר נותרים במודל הסופי.

מחשבים רווח סמך עבור β_0 לפי הנוסחה הסטנדרטית (ללא תיקונים) באמצעות:

$$\hat{\beta}_0\pm t_{n-(p+1),1-\frac{\alpha}{2}}\sqrt{\widehat{\mathbb{V}}(\hat{\beta}_0)}\overset{given\ numbers}{=}t_{239,0.975}\sqrt{\widehat{\mathbb{V}}(\hat{\beta}_0)}$$

אפשר להציג זאת באופן שקול על ידי רווח הסמך הקבוצה הבאה:

$$CI(\beta_0)=\left\{\tilde{\beta}_0:\left|T(\tilde{\beta}_0)\right|\le t_{239,0.975}\right\}=\left\{\tilde{\beta}_0:\left|\frac{\hat{\beta}_0-\tilde{\beta}_0}{\sqrt{\widehat{\mathbb{V}}(\hat{\beta}_0)}}\right|\le t_{239,0.975}\right\}$$

כאשר

$$T(\tilde{\beta}_0)=\frac{\hat{\beta}_0-\tilde{\beta}_0}{\sqrt{\widehat{\mathbb{V}}(\hat{\beta}_0)}}$$

תחת $H_0:\beta_0=0$ מתקיים $T(0)\overset{H_0}{\sim}t_{239}$.

הבעיה שניתן לראות אם נבצע סימולציה למול הנתונים הוא שההתפלגות תחת השערת האפס ללא תיקונים היא (ככל הנראה) משמעותית גדולה יותר מההתפלגות האמיתית של $T(0)$ תחת בחירת משתנים המחושבת על ידי הסימולציה. כלומר אנחנו חושבים שרמת הכיסוי שווה של רווח הסמך לערך האמיתי של ההתפלגות הוא 95% אולם בפועל רמת הכיסוי האמתית שווה ל-83.5%.

דוגמה 23.6.2 – נניח

$$\theta_i \overset{iid}{\sim} N(0,0.04)$$
$$Z_i \sim N(\theta_i, 1)$$

אזי רווח הסמך הסטנדרטי (ללא תיקונים) עבור θ_i הוא ברמת מובהקות של 90%:

$$\Im_i = [Z_i - 1.64, Z_i + 1.64]$$

בלי בחירה מתקיים:

$$\mathbb{P}(\theta_i \in \Im_i) = 0.9$$

נניח כעת שביצענו את הניסוי מספר גדול של פעמים ויש לנו עתה n רווחי סמך עבור $\theta_1, \dots, \theta_n$. נניח שאנחנו לוקחים רק את רווחי הסמך עבורם $0 \notin \Im$, כלומר אנחנו משאירים רווחי סמך שבהם אנחנו באופן מדויק אומדים שהערכים של θ_i חיוביים או שליליים. במצב כזה אנחנו מקבלים:

$$\mathbb{P}(\theta_i \in \Im_i | 0 \notin \Im_i) = 0.043$$

מצב זה נפוץ מאוד בניסויים מדעיים שבהם אנחנו מסתכלים על מספר גדול של פרמטרים ומפרסמים את רווחי הסמך המקוריים ללא התחשבות בבחירה שביצענו.

תזכורת 23.7 – נזכיר כי Family-Wise Error Rate מוגדר על ידי:

$$FWER_{test} = \mathbb{P}(V \geq 1)$$

כאשר V הוא מספר המקרים שבהם דחינו השערת אפס למרות שהיתה נכונה.

הגדרה 23.8 –

$$V_{CI}^{tot} = \{i \in \{1, \dots, m\} : \theta_i \notin CI_i\}$$

עבור שיטת בחירה כלשהי שבוחרת $S \subseteq \{1, \dots, m\}$ נגדיר:

$$V_{CI}^{S} = \{i \in S : \theta_i \notin CI_i\}$$

עוד נגדיר:

$$FWER_{CI} = \mathbb{P}\left(V_{CI}^{S} \geq 1\right)$$
$$FWER_{CI}^{*} = \mathbb{P}(V_{CI}^{tot} \geq 1)$$

ברור כי

$$V_{CI}^{S} \leq V_{CI}^{tot}$$
$$FWER_{CI} \leq FWER_{CI}^{*}$$

הערה 23.8.1 – אם יש לנו שיטה ששומרת על $FWER_{test}$ במסגרת של בדיקת השערות, נוכל לבנות ממנה שיטה לחישוב רווחי סמך ששומרת על $FWER^{*}$.

דוגמה 23.9 – בשיטת בונפרוני עבור m השערות וערכי $p.value$ מסומנים $p_1, \dots, p_m$ דוחים את H_{0i} אם $p_i \leq \frac{\alpha}{m}$ באופן ששומר על ה-$FWER$. במסגרת רווחי סמך זה אומר שיש לנו $\theta_1, \dots, \theta_m$ ו-$CI_i(\alpha)$ הוא רווח הסמך עבור θ_i שיש לו רמת כיסוי אם הסתכלנו על θ_i בלבד ללא תיקונים. במצב זה נוכל לייצר רווח סמך בשיטת בונפרוני על ידי:

$CI_i\left(\frac{\alpha}{m}\right)$

אזי במצב כזה:

$FWER^*_{CI}(Bonferroni) \leq \alpha$

ההוכחה תהיה זהה להוכחה שהיתה לנו במקרה של בדיקת השערות.

$FWER^*_{CI}(Bonferroni) = \mathbb{P}(V_{CI}^{tot} \geq 1)$

הוכחה:

$V_{CI}^{tot} \geq \mathbb{I}\{V_{CI}^{tot} \geq 1\} \rightarrow \mathbb{E}[V_{CI}^{tot}] \geq \mathbb{E}[\mathbb{I}\{V_{CI}^{tot} \geq 1\}] = \mathbb{P}(V_{CI}^{tot} \geq 1)$

$$FWER^*_{CI}(Bonferroni) = \mathbb{P}(V_{CI}^{tot} \geq 1) \leq \mathbb{E}[V_{CI}^{tot}] = \mathbb{E}\left[\sum_{i\in N} V_{CI_i}\right] = \sum_{i\in N}\mathbb{E}\left[R_{CI_i}\right] = \sum_{i\in N}\mathbb{P}\left(R_{CI_i}\right)$$
$$= \sum_{i\in N}\frac{\alpha}{m} = \frac{m_o\alpha}{m}$$

רווח סמך רגיל:

$$\hat{\beta}_j \pm t_{n-p-1,1-\frac{\alpha}{2}} \cdot s\sqrt{(X^TX)^{-1}_{jj}}$$

רווח סמך של בונפרוני יהיה

$$\hat{\beta}_j \pm t_{n-p-1,1-\frac{\alpha/m}{2}} \cdot s\sqrt{(X^TX)^{-1}_{jj}}$$

דוגמה 23.10 – נניח $\theta_1, \dots, \theta_m$ ונסמן

$\vec{\theta}\dot{\sim}N(\theta, \Omega)$

רווח סמך של θ_i במקרה ש-θ_i הוא הפרמטר היחיד נתון על ידי:

$CI_i(\alpha) = \hat{\theta}_i \pm z_{1-\alpha/2}\sqrt{\Omega_{ii}}$

עבור השערת אפס $H_{0i}: \theta_i = \theta_i^0$ מגדירים את הסטטיסטי המתוקנן

$$Z_i = \frac{\hat{\theta}_i - \theta_i^0}{\sqrt{\Omega_{ii}}}$$

דוחים את H_{0i} אם $|Z_i| \geq Z^*$, כאשר Z^* מקיים את התכונה הבאה:

$\mathbb{P}\left(\max_{1\leq i\leq n}|\zeta| \leq Z^*\right) = 1 - \alpha$

כאשר

$\zeta \sim N(0, \Gamma)$

והמטריצה Γ מוגדרת על ידי:

$$\Gamma_{rs} = Corr(\zeta_r, \zeta_s) = \frac{\Omega_{rs}}{\sqrt{\Omega_{rr}\Omega_{rs}}}$$

את זה ראינו במקרה של בדיקת השערות. גם כאן אפשר להפוך זאת לשיטה של חישוב רווחי סמך על ידי כך שבמקום שנשתמש בערך הכללי $z_{1-\alpha/2}$ נשתמש בערך z^*. אזי במצב כזה:

$$\left\{\theta_i \in \hat{\theta}_i \pm z^*\sqrt{\Omega_{ii}}\right\} = \left\{\left|\frac{\hat{\theta}_i - \theta_i^0}{\sqrt{\Omega_{ii}}}\right| \le Z^*\right\}$$

שיעור 24

תזכורת 24.1 – כזכור אנחנו בנושא הסקה אחרי בחירה. נניח שישנם

- פרמטרים התחלתיים $\theta_1, \dots, \theta_m$.
- פרוצדורה לבחירה של פרמטרים שבהם אנחנו מעוניינים
- S הוא אוסף האינדקסים של הפרמטרים שאכן נבחרו
- פרוצדורה באמצעותה מייצרים CI_i רווח סמך עבור הפרמטר θ_i

הגדרנו בשיעור הקודם את $FWER_{CI}$ ואת $FWER_{CI}^*$, כאשר:

$$V_{CI}^{tot} = \{i \in \{i, \dots, m\}: \theta_i \notin CI_i\}$$
$$V_{CI}^{S} = \{i \in S: \theta_i \notin CI_i\}$$

ובהתאם

$$FWER_{CI} = \mathbb{P}\left(V_{CI}^S \ge 1\right)$$
$$FWER_{CI}^* = \mathbb{P}(V_{CI}^{tot} \ge 1)$$

הצגנו פרוצדורות ששומרות על $FWER_{CI}^*$ מפני מצב שבו יש לי ביטחון יתר ברווחי הסמך. השיטות שראינו עד כה:

1. שיטת בונפרוני
2. שיטה מבוססת התפלגות רב-נורמלית

הערה 24.2 – נשים לב כי המאורע שבו כל רווחי הסמך מכסים את הפרמטרים שלהם מוגדר באמצעות:

$$1 - FWER_{CI} = \mathbb{P}\left(V_{CI}^S = 0\right) = \mathbb{P}(\forall_{i\in S}: \theta_i \in CI_i)$$
$$1 - FWER_{CI}^* = \mathbb{P}(V_{CI}^{tot} = 0) = \mathbb{P}\left(\forall_{i\in\{1,\dots,m\}}: \theta_i \in CI_i\right)$$

הערה 24.3 – פרוצדורה ששומרת על $FWER_{CI}^*$ נקראת פרוצדורה של כיסוי בו-זמני.

נושא 4.1.5 – False Coverage Rate (FCR)

הגדרה 24.4 – False Coverage Rate (FCR) – נגדיר V_{CI}^S כמו קודם, ונגדיר $R_{CI}^S = |S|$. אזי אחוז שגיאת הכיסוי מוגדר על ידי:

$$FCP = \begin{cases} V_{CI}^S / R_{CI}^S, & R_{CI}^S \ge 1 \\ 0, & R_{CI}^S = 0 \end{cases}$$
$$FCR = \mathbb{E}[FCP] = \mathbb{E}\left[\frac{V_{CI}^S}{\max\{R_{CI}^S, 1\}}\right]$$

כלומר ה-FCP הוא אחוז הטעויות שלנו, כאשר טעות מוגדרת להיות מצב שבו יצרנו רווח סמך לפרמטר כלשהו אבל הפרמטר איננו בתוך רווח הסמך, כלומר מצב $\theta_i \notin CI_i(\alpha)$ הוא מצב טעות.

טענה 24.5 – *FCR* במצב ללא בחירת פרמטרים - אם יש לנו מערך של רווחי סמך $CI_1(\alpha), \dots, CI_m(\alpha)$, כך שלכל רווח סמך $CI_i(\alpha)$ מתקיים $\mathbb{P}\left(\theta_i \in CI_i(\alpha)\right) = 1 - \alpha$, כאשר לא ביצענו בחירה, כלומר מסתכלים על θ_i כפרמטר יחיד, אזי:

$R_{CI}^{S}=R_{CI}^{tot}=m$

$FCP=\frac{1}{m}\sum_{i=1}^{m}\mathbb{I}\{\theta_i\notin CI_i(\alpha)\}$

$FCR=\mathbb{E}[FCP]=\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[\mathbb{I}\{\theta_i\notin CI_i(\alpha)\}]=\frac{1}{m}\sum_{i=1}^{m}\mathbb{P}\left(\theta_i\notin CI_i(\alpha)\right)=\frac{1}{m}\sum_{i=1}^{m}\left[1-\mathbb{P}\left(\theta_i\in CI_i(\alpha)\right)\right]$

$=\frac{1}{m}\sum_{i=1}^{m}[1-(1-\alpha)]=\frac{1}{m}\sum_{i=1}^{m}\alpha=\frac{1}{m}m\alpha=\alpha$

טענה 24.6 – אם יש פרודצדורה ששומרת על $FWER_{CI}$, אזי היא גם שומרת על ה-FCR.

הוכחה 24.6 – נשים לב כי תמיד מתקיים

$$0\le V_{CI}^{S}\le R_{CI}^{S}\Rightarrow 0\le V_{CI}^{S}/R_{CI}^{S}\le 1$$

לכן:

$$FCP=\begin{cases}V_{CI}^{S}/R_{CI}^{S}, & R_{CI}^{S}\ge 1\\ 0, & R_{CI}^{S}=0\end{cases}\le\mathbb{I}\{V_{CI}^{S}\ge 1\}$$

ואז

$$FCR=\mathbb{E}[FCP]\le\mathbb{E}\left[\mathbb{I}\left(V_{CI}^{S}\ge 1\right)\right]=FWER_{CI}$$

הגדרה 24.7 – פרוצדורת בנימיני ויקותיאלי BY לשמירת ה-FCR – נניח קיימים:

- פרמטרים $\theta_1,\dots,\theta_m$
- סטטיסטיים $T_1,\dots,T_m$
- פרוצדורה לבחור פרמטרים שאנחנו רוצים להתעניין בהם

נגדיר $S(T_1,\dots,T_m)$ להיות אוסף האינדקסים של הפרמטרים שנבחרו.

לכל פרמטר נמצא את נקודת הגבול עבורה בוחרים או לא בוחרים את הפרמטר. כלומר t הוא הערך המינימלי כך שעדיין ייבחר הפרמטר θ_i. לכל פרמטר θ_i:

$$R^{(i)}=\min_t\{\#S(T_1,\dots,T_{i-1},t,T_{i+1},\dots,T_m):i\in S(T_1,\dots,T_{i-1},t,T_{i+1},\dots,T_m)\}$$

בסך הכל ניתן לייצר וקטור שהוא כאורך מספר הפרמטרים שבו בכל קואורדינטה אנחנו מקבלים מספר כלשהו של ערכים שנדחו, כלומר:

$$\vec{R}=\begin{bmatrix}R^{(1)}\\ \dots\\ R^{(m)}\end{bmatrix}\in\{0,\dots,m\}^m$$

הרווח הסמך המתוקן של θ_i יהיה:

$$CI_i\left(\frac{\alpha}{m}R^{(i)}\right)$$

דוגמה 24.8 –

נניח

- פרמטרים $\theta_1,\dots,\theta_m$
- סטטיסטיים $T_1,\dots,T_m$
- ערכי $p_1,\dots,p_m$ כך ש-$p_i=\mathbb{P}\left(T_i\ge t_i^{obs}\right)$

- נניח שבוצע שיטת בונפרוני לבחירת פרמטרים

נתונים בדוגמה:

- $m = 4$
- נניח רמת המובהקות $\alpha = 0.1$. כלומר דרוש $p_i \leq 0.025$.
- נניח שה-$p.values$ שהתקבלו הם $p_1 = 0.01, p_2 = 0.1, p_3 = 0.02, p_4 = 0.3$.
- באמצעות שיטת בונפרוני נבחרים רק שני ערכים - p_1, p_3.
- במצב זה $S = \{1,3\}$. הגודל של S הוא $R_{CI}^S = \#S = |S| = 2$.
- לכל אחד הערכים שהתקבלו אנחנו יכולים לבצע את התהליך שהורדנו קודם. לדוגמה עבור p_1
    - אנחנו יכולים להוריד את T_1 כך ש-p_1 יעלה עד לערך הקבלה הגבוה ביותר כלומר 0.025.
    - יוצרים סדרת ערכים חדשה לאחר ההחלפה: $p_1' = 0.025, p_2 = 0.1, p_3 = 0.02, p_4 = 0.3$.
    - במצב כזה מסתכלים שוב על שיטת הבחירה שלנו ובוחנים אילו ערכים מתקבלים ואילו לא מתקבלים. במצב שלנו עדין מתקבלים הערכים p_1', p_3 כלומר $R^{(1)} = 2$ גם כן.

הערה 24.9 – מספר הערות על שיטת בנימיני-יקותיאלי

1. במקרים רבים $R^{(i)} = R_{CI}^S$ (ראו למשל בדוגמה 24.8)
2. תמיד מתקיים $R^{(i)} \leq R_{CI}^S$, שכן בבניה של $R^{(i)}$ אנחנו מורידים את מספר הפרמטרים שנבחרו ולכן הוא חייב להיות קטן מהמדגם המקורי.
3. אם $R_{CI}^S = m$, משתמשים ב-$CI_i(\alpha)$ ללא תיקון (כי במקרה זה $R^{(i)} = m$ לכל i)
4. אם $R^{(i)} = R_{CI}^S = m$, אזי הרווח שמשתמשים בו עבור הפרמטר היחיד שנבחר הינו $CI_i\left(\frac{\alpha}{m}\right)$ כמו בבונפרוני.

משפט 24.10 – אם $T_1, \dots, T_m$ אזי שיטת BY שומרת את ה-FCR ברמה α.

הוכחה 24.10 – נגדיר

$$Q_i = \mathbb{I}\left\{i \in S, \theta_i \notin CI_i\left(\frac{\alpha}{m} R^{(i)}\right)\right\}$$
$$V_{CI}^S = \sum_{i=1}^{m} Q_i$$

נגדיר (בדומה להוכחה של BH) את X_i באמצעות: $X_i = \frac{Q_i}{\max\{R_{CI}^S, 1\}}$. אזי $FCR = \mathbb{E}[\sum_{i=1}^m X_i]$.

נוכיח כי $\mathbb{E}[X_i] \leq \frac{\alpha}{m}$, וכתוצאה מזה רואים מיידית ש-$FCR \leq \alpha$ כנדרש.

טענה 24.10.1 - נותר להוכיח $\mathbb{E}[X_i] \leq \frac{\alpha}{m}$.

הוכחת 24.10.1 –

$$X_i = \sum_{k=1}^{m} \frac{\mathbb{I}\left\{i \in S, \theta_i \notin CI_i\left(\frac{\alpha}{m} R^{(i)}\right)\right\} \mathbb{I}\left\{R^{(i)} = k\right\}}{R_{CI}^S} \overset{R^{(i)}=k \ by \ 2^{nd} \ indicator}{=}$$
$$\sum_{k=1}^{m} \frac{\mathbb{I}\left\{i \in S, \theta_i \notin CI_i\left(\frac{\alpha}{m} k\right)\right\} \mathbb{I}\left\{R^{(i)} = k\right\}}{R_{CI}^S} \overset{R^{(i)} \leq R_{CI}^S}{\leq}$$
$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{I}\left\{i \in S, \theta_i \notin CI_i\left(\frac{\alpha}{m} k\right)\right\} \mathbb{I}\left\{R^{(i)} = k\right\} \overset{omitting \ condition \ on \ indicator \ can \ only \ make \ sum \ larger}{\leq}$$
$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{I}\left\{\theta_i \notin CI_i\left(\frac{\alpha}{m} k\right)\right\} \mathbb{I}\left\{R^{(i)} = k\right\}$$

נבחן את התוחלת המותנית של X_i בהינתן כל האינדקסים האחרים T_j:

$$\mathbb{E}\left[X_i \middle| T_j, j \neq i\right] \overset{by\ what\ we\ proved\ above\ on\ X_i}{\leq}$$
$$\mathbb{E}\left[\sum_{k=1}^{m} \frac{1}{k} \mathbb{I}\left\{\theta_i \notin CI_i\left(\frac{\alpha}{m}k\right)\right\} \mathbb{I}\left\{R^{(i)} = k\right\} \middle| T_j, j \neq i\right] \overset{sum\ and \frac{1}{k}\ out\ of\ expectation}{=}$$
$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{E}\left[\mathbb{I}\left\{\theta_i \notin CI_i\left(\frac{\alpha}{m}k\right)\right\} \mathbb{I}\left\{R^{(i)} = k\right\} \middle| T_j, j \neq i\right] \overset{R^{(i)} \perp T_i\ given\ all\ other\ T_1,\dots,T_{i-1},T_{i+1},\dots,T_m}{=}$$
$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{I}\left\{R^{(i)} = k\right\} \mathbb{E}\left[\mathbb{I}\left\{\theta_i \notin CI_i\left(\frac{\alpha}{m}k\right)\right\} \middle| T_j, j \neq i\right] \overset{T_1,\dots,T_m\ independent \to can\ omit\ condition}{=}$$
$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{I}\left\{R^{(i)} = k\right\} \mathbb{E}\left[\mathbb{I}\left\{\theta_i \notin CI_i\left(\frac{\alpha}{m}k\right)\right\}\right] \overset{expectation\ of\ indicator\ is\ probability}{=}$$
$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{I}\left\{R^{(i)} = k\right\} \mathbb{P}\left(\theta_i \notin CI_i\left(\frac{\alpha}{m}k\right)\right) \overset{CI(\gamma)\ is\ defined\ by\ \mathbb{P}(\theta_i \in CI_i(\gamma)) = 1-\gamma\ when\ looking\ at\ CI_i\ only}{=}$$
$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{I}\left\{R^{(i)} = k\right\} \frac{\alpha}{m} k =$$
$$\frac{\alpha}{m} \sum_{k=1}^{m} \mathbb{I}\left\{R^{(i)} = k\right\} \overset{maximum\ of\ 1\ of\ the\ indicators\ is\ positive}{\leq}$$
$$\frac{\alpha}{m}$$

הוכחנו על התוחלת המותנית $\mathbb{E}\left[X_i | T_j, j \neq i\right] \leq \frac{\alpha}{m}$. עתה:

$$\mathbb{E}[X_i] \overset{double\ expectation\ law}{=} \mathbb{E}\left[\mathbb{E}\left[X_i \middle| T_j, j \neq i\right]\right] \leq \mathbb{E}\left[\frac{\alpha}{m}\right] = \frac{\alpha}{m}$$

שיעור 25

נושא 5 - מודלים עם Splines

הגדרה 25.1 – **פונקציית Spline לינארית** נניח קיימות נקודות $x_0, \dots, x_K$ ופונקציה מוגדרת על הקטע $[x_0, \dots, x_K]$ כך שמתקיים:

1. בכל קטע $[x_{k-1}, x_k]$ הפונקציה לינארית
2. הפונקציה רציפה בכל אחת מהנקודות $x_1, \dots, x_{k-1}$.

הגדרה 25.2 – פונקציית Spline ריבועית – בדומה לפונקציה הלינארית, פונקציית Spline ריבועית היא פונקציה שבה:

1. בכל קטע $[x_{k-1}, x_k]$ הפונקציה ריבועית.
2. הפונקציה רציפה בכל אחת מהנקודות $x_1, \dots, x_K$.
3. הנגזרת של הפונקציה רציפה בכל אחת בכל אחת מהנקודות $x_1, \dots, x_K$.

הגדרה 25.3 - באופן כללי Spline מדרגה m הוא פונקציה המקיימת:

1. הפונקציה היא פולינום בדרגה m בתוך כל קטע $[x_{k-1}, x_k]$
2. הפונקציה עצמה וכל הנגזרות עד סדר $m-1$ רציפות.

הערה 25.4 – המקרה הכי נפוץ לשימוש הוא של דרגה $m = 3$, נקרא Cubic Spline.

הגדרה 25.5 – נסמן $S(m; x_0, \dots, x_K)$ להיות אוסף הפונקציות שהן $Splines$ בדרגה m עם הנקודות $x_0, \dots, x_K$. כלומר אלו צריכות להיות פונקציות עם לכל היותר דרגה m בנקודות הנתונות.

טענה 25.6 – נשים לב כי אם קיימות שתי פונקציות $f_1 \in S$ וכן $f_2 \in S$ וקבועים כלשהם $c_1, c_2 \in \mathbb{R}$. אזי:

$$f = c_1 f_1 + c_2 f_2 \in S$$

מסקנה 25.7 – מרחב פונקציות ה-$Spline$ הוא מרחב לינארי וסגור לחיבור.

טענה 25.8 – נוכל לבטא Spline מדרגה m באמצעות ההגדרה הבאה: כל $f \in S(m; x_0, \dots, X_K)$ אפשר לכתוב בצורה הבאה, עבור כל $x \in [x_0, x_K]$:

$$f(x) = \sum_{r=0}^{m} C_r\, x^r + \sum_{s=1}^{k-1} d_s (x - x_s)^m{}_+$$

כאשר

$$u_+ = \begin{cases} u, & u \geq 0 \\ 0, & u < 0 \end{cases}$$

ו-$c_0, \dots, c_m \in \mathbb{R}$ קבועים כלשהם, וכן $d_1, \dots, d_{k-1} \in \mathbb{R}$ סקלרים כלשהם.

דוגמה 25.9 – Spline לינארי עם נקודות $[x_0, x_1, x_2]$, ודרגה $m = 1$. אזי:

$$f(x) = c_0 + c_1 x + d_1 (x - x_1)_+ = \begin{cases} c_0 + c_1 x, & x \leq x_1 \\ c_0 + c_1 x + d_1 (x - x_1)_+ & x \geq x_1 \end{cases}$$
$$= \begin{cases} c_0 + c_1 x, & x \leq x_1 \\ (c_0 - d_1 x_1) + (c_1 + d_1) x & x \geq x_1 \end{cases}$$

כלומר

1. השיפוע של הפונקציה מתחיל בשיפוע c_1.
2. כאשר מגיעים לנקודה x_1 השיפוע גדל לשיפוע $c_1 + d_1$.

הגדרה 25.10 – B-Spline – Spline המוגדר על קטע קטן ובכל נקודה אחרת מוגדר להיות 0.

$$B_{m,p}(x) = \sum_{j=p}^{m+p+1} \left[\prod_{\ell \in C(p,j)} \left(x_\ell - x_j \right)^{-1} \right] \left(x - x_j \right)_+^m$$

כאשר $C(p, j)$ הוא אוסף המספרים השלמים ב-$[p, m + p + 1]$ לא כולל j.

טענה 25.11 – הפונקציות האלה מקיימות את הרקורסיה הבאה:

$$B_{m,p}(x) = \frac{1}{x_{p+m+1} - x_p} \left[\left(x - x_p \right) B_{m-1,p}(x) + \left(x_{p+m+1} - x \right) \right]$$

כאשר נקודת העצירה היא:

$$B_{1,p}(x) = \frac{\left(x - x_p \right)_+}{\left(x_{p+1} - x_p \right)\left(x_{p+2} - x_p \right)} + \frac{\left(x - x_{p+1} \right)_+}{\left(x_p - x_{p+1} \right)\left(x_{p+2} - x_{p+1} \right)} + \frac{\left(x - x_{p+1} \right)_+}{\left(x_p - x_{p+2} \right)\left(x_{p+1} - x_{p+2} \right)}$$

דוגמה 25.12 – עבור נקודות שבהן הרווחים כולם שווים לאותו ערך h מתקיים:

$$B_{3,p}(x) = \frac{1}{25h^4} \left[\left[\left(x - x_p \right)_+ - 4\left(x - x_{p+1} \right) \right]_+^3 + 6\left[x - x_{p+2} \right]_+^3 - 4\left[x - x_{p+3} \right]_+^3 + \left[x - x_{p+4} \right]^3 \right]$$

משפט 25.13 – נניח שיש נקודות כלשהן $x_0, \dots, x_K$. נגדיר $x_{-1}, \dots, x_{-m}$ ו-$x_{k+1}, \dots, x_{k+1+m}$ כך ש-

$$x_{-m} < x_{-(m-1)} < \cdots < x_{-1} < x_0 < x_1 < \cdots < x_K < x_{K+1} < \cdots < x_{K+m}$$

אנחנו בעצם מעוניינים בקטע $[x_0, x_K]$. אזי כל פונקציה ב-$S(m; x_0, \dots, x_K)$ אפשר לכתוב באופן הבא:

$$f(x) = \sum_{r=-m}^{K+m} c_r B_{m,p}(x)$$

טענה 25.14 – **שימוש ב-Splines בהתאמת עקומות** – נרצה להתאים מודל

$$(X_i, Y_i)$$
$$\mu(x) = \mathbb{E}[Y_i | X_i = x]$$

במונחים של הבסיס הראשון מגדירים:

$$X_{i0} = 1$$
$$X_{i1} = X_i$$
$$X_{i2} = X_i^2$$
$$X_{i3} = X_i^3$$
$$X_{i4} = (X_i - x_1)_+^3$$
$$X_{i5} = (X_i - x_2)_+^3$$
$$X_{i6} = (X_i - x_3)_+^3$$
$$X_{i7} = (X_i - x_4)_+^3$$

אם בהתאם לכך נגדיר:

$$\mu(x) = \sum_{r=0}^{7} \beta_i X_{ir}$$

הפונקציה הזאת הינה פונקצית Spline בדרגה 3 עם נקודות פנימיות x_1, x_2, x_3, x_4.

טענה 25.15 – בחירת הנקודות – כמה שיטות:

1. לנסות לבחור את הנקודות בצורה "חכמה" באמצעות שערוך הצפיפות.
2. מרווחים קבועים.
3. אחוזונים קבועים לפי הצפיפות של x.

הערה 25.16 – כמה נקודות לשים? אפשר להשתמש במדד R^2 מרגרסיה (כאשר b הוא מספר האיברים במודל מלבד החותך בפרמטריזציה הראשונה):

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} = 1 - \frac{SSE}{SST}$$
$$adjusted\ R^2 = 1 - \frac{SSE/(n-b)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}^2(current\ model)}{\hat{\sigma}^2(intercept\ only\ model)}$$

ניתן להשתמש גם ב-AIC וב-BIC.

דף נוסחאות

מטריצות

פירוק ספקטרלי $A \in \mathbb{R}^{m\times m}$ סימטרית. קיים $A = U\Lambda U^T$

כאשר: $\Lambda, U \in \mathbb{R}^{m\times m}$. Λ אלכסונית על אלכסון ע"ע של A. U עמודה j הינה ו"ע של A התואם לע"ע λ_j. U אורתוגונלית.

$$A^k = U\Lambda^k U^T, \qquad A^{1/2} = U\Lambda^{1/2}U^T$$

פירוק חולסקי $G = LL^T$ כאשר L משולשית תחתונה

פירוק לוג חולסקי (שימושי כאשר יש דרישה של א"ש על איברים מסוימים). באיבר ts שצריך להיות חיובי נדרוש$\log(L_{ts})$

נגזרות

עבור מטריצה A שכל קואורדינטה בה ניתנת לייצוג כפונקציה של וקטור פרמטרים $\vec{\theta}$ ומסומנת $A(\vec{\theta})$, ניתן להגדיר את הנגזרת החלקית על קואורדינטה ה-r של $\vec{\theta}$ להיות:

הגדרה	$\frac{\partial A}{\partial\theta_r} = \left[\frac{\partial A_{ij}}{\partial\theta_r}\right]_{1\le i,j\le n}$
סכום	$\frac{\partial}{\partial\theta_r}[A(\theta)+B(\theta)] = \frac{\partial A}{\partial\theta_r} + \frac{\partial B}{\partial\theta_r}$
מכפלה	$\frac{\partial}{\partial\theta_r}A(\theta)B(\theta) = \frac{\partial A}{\partial\theta_r}B(\theta) + A(\theta)\frac{\partial B}{\partial\theta_r}$
הופכית	$\frac{\partial}{\partial\theta_r}A(\theta)^{-1} = -A(\theta)^{-1}\frac{\partial A}{\partial\theta_r}A(\theta)^{-1}$
פרמטר	עבור מטריצה $E^{(rs)}$ מטריצה שבה אפסים למעט $E_{rs}=1$ המוגדרת על ידי $E_{ij}^{(rs)} = \delta_{ir}\delta_{js}$. לכל מטריצה ניתן להגדיר $\frac{\partial A}{\partial A_{rs}} = E^{(rs)}$.
דטרמיננטה	$cof(A)_{rs} = \frac{\partial}{\partial A_{rs}}\lvert A\rvert = (-1)^{r+s}\lvert A^{(rs)}\rvert$ $\frac{\partial}{\partial\theta_r}\lvert A(\theta)\rvert = tr\left([cof(A)]^T\cdot\frac{\partial A}{\partial\theta_r}\right)$
לוג דטרמיננטה	$\frac{\partial}{\partial\theta_r}\log(\det(A(\theta))) = tr\left(A^{-1}\frac{\partial A}{\partial\theta_r}\right)$

$\boldsymbol{cof(A)}$– מטריצה שבמיקום הij שלה יש את הדטרמיננטה של A^{ij}.

 # תכונות של trace: $tr(AB) = tr(BA)$

חוק קרמר: עבור מערכת משוואות $Ax = b$

מתקיים: $x_i = \frac{\det(\tilde{A}_i)}{\det(A)}$ כאשר $\tilde{A}_{ib}$ = החלפת עמודה i של b. נובע $A^{-1} = [cof(A)]^T/\det(A)$.

 # הופכי של מטריצת בלוקים: מטריצת בלוקים ריבועית $\begin{bmatrix} A & B\\ C & D\end{bmatrix}$ וההופכי של המטריצה מוגדר על ידי: $E = D - CA^{-1}B$

$$\begin{bmatrix} A & B\\ C & D\end{bmatrix}^{-1} = \begin{bmatrix} A^{-1}+A^{-1}BE^{-1}CA^{-1} & -A^{-1}BE^{-1}\\ -E^{-1}CA^{-1} & E^{-1}\end{bmatrix} = \begin{bmatrix} G & H\\ K & L\end{bmatrix}$$

 # שוויון Woodbury:

$$(P+ZQZ^T)^{-1} = P^{-1} - P^{-1}Z(Q^{-1} + Z^TPZ)^{-1}Z^TP^{-1}$$

מספר האיברים השונים במטריצה סימטרית $n(n+1)/2$

כלל השרשרת הרב משתני: עבור $f:\mathbb{R}^p \to \mathbb{R}^q, g:\mathbb{R}^q \to \mathbb{R}^s$

מתקיים: $\frac{\partial h_u}{\partial x_j} = \sum_{t=1}^q \dot{g}_{ut}(f(x))\dot{f}_{tj}(x)$

מטריצת שונויות

$[Cov(Z)]_{rs} = Cov(Z_r, Z_s)$	$Cov(Z,Z) = Cov(Z)$
$[Cov(Z,W)]_{rs} = Cov(Z_r, W_s)$	$Cov(X+a) = CovX$
$Cov(AZ) = A\cdot Cov(Z)\cdot A^T$	
$Cov(AZ,BW) = A\cdot Cov(Z,W)\cdot B^T$	
$Cov(Z) = \mathbb{E}[(Z-\mathbb{E}[Z])(Z-\mathbb{E}[Z])^T]$	
$Cov(Z,W) = \mathbb{E}[(Z-\mathbb{E}[Z])(W-\mathbb{E}[W])^T]$	
$Cov(X+Y) = Var(X)+Var(Y)+2Cov(X,Y)$	
$Cov(aX+bY,cW+dV) = acCov(X,W)$	
$+adCov(X,V)+bcCov(Y,W)+bdCov(Y,V)$	

התפלגויות

<u>התפלגות $\boldsymbol{t}$:</u> $Z\sim N(0,1)$ ו-$Q\sim\chi_d^2$ ו-$Z\perp Q$. אז $\frac{Z}{\sqrt{Q/d}}\sim t_d$

<u>כי-בריבוע $\boldsymbol{\chi^2}$</u> אם $Z_1,\dots,Z_d\sim N(0,1)$ אז $\sum_{i=1}^d Z_i^2\sim\chi_d^2$

<u>כי בריבוע לא ממורכז $\boldsymbol{\chi^2(\lambda)}$:</u> אם $Z_1,\dots,Z_d\sim N(\mu_i,1)$ אז $\sum_{i=1}^d Z_i^2\sim\chi_d^2(\sum_{i=1}^d\mu_i^2)$.

<u>תוחלת:</u> $\mathbb{E}[\chi_d^2(\lambda)] = d+\lambda$

- אם $Y\sim\chi_d^2(\lambda)$ אז $Y = Y_1+Y_2$ כאשר $Y_1\sim\chi_1^2(\lambda)$ ו-$Y_2\sim\chi_{d-1}^2$ והמשתנים בלתי-תלויים $Y_1\perp Y_2$.
- <u>צפיפות</u> $f_Y(y) = \int_0^\infty f_{Y_1}(y_1)f_{Y_2}(y-y_1)dy_1 = \int_0^\infty f_{\chi_1^2(\lambda)}(y_1)f_{\chi_{d-1}^2}(y-y_1)dy_1$
- צפיפות מצטברת:

$$F_Y(y) = \int_0^\infty f_{Y_1}(y_1)\mathbb{P}(Y=y|Y_1=y_1)dy_1 = \int_0^\infty f_{Y_1}(y_1)\mathbb{P}(Y_2\le y-y_1)dy_1$$

- עבור $W\sim N_d(0,V)$ עם V הפיכה. $Q = W^TV^{-1}W$. אזי $Q\sim\chi_d^2$. כאשר $d = \dim(W)$
- עבור $W\sim N(0,I)$ ו-P סימטרית אידמפוטנטית ו-$rank(P) = k$ אזי $\|PW\|^2\sim\chi_k^2$

התפלגות רב נורמלית

$W = \vec{\mu}+AZ$ $\mu\in\mathbb{R}^k$ קבוע $A\in\mathbb{R}^{k\times m}$ קבועים $Z\in\mathbb{R}^m$ מקרי $Z_1,\dots,Z_m\overset{iid}{\sim}N(0,1)$.

1. **תוחלת ושונות** $Cov[W] = AA^T$ $\mathbb{E}[W]=\mu$
2. **יוצרת מומנטים** של W הינה: $M_W(t) = \exp\left\{\mu^Tt+\frac{1}{2}t^TAA^Tt\right\} \overset{V=AA^T}{=} \exp\left\{\mu^Tt+\frac{1}{2}t^TVt\right\}$
3. אם $W_1 = \mu+A_1Z$, $W_2 = \mu+A_2Z$ וגם $A_1A_1^T = A_2A_2^T = V$ אז $W_1\sim W_2\sim N(\mu,V)$
4. **הצפיפות** אם V הפיכה: $f_W(x) = (2\pi)^{-\frac{d}{2}}\cdot|V|^{-\frac{1}{2}}\cdot\exp\left\{-\frac{1}{2}(x-\mu)^TV^{-1}(x-\mu)\right\}$
5. **מכפלה בקבוע** $c^TW\sim N_d(c^T\mu, c^TVc)$
6. **כל איבר בוקטור** $W_r\sim N(\mu_r, V_{rr})$
7. **מכפלה במטריצה** $CW\sim N(C\mu, CVC^T)$
8. **חלוקה לתתי וקטורים** $W_1 = (W_1\dots W_r)^T, W_2 = (W_{r+1}\dots W_d)^T$ וחלוקה של μ בהתאמה. V מטריצת בלוקים הפיכה $V = \begin{pmatrix} V_{11}\in\mathbb{R}^{r\times r} & V_{12}\in\mathbb{R}^{r\times(d-r)}\\ V_{21}\in\mathbb{R}^{(d-r)\times r} & V_{22}\in\mathbb{R}^{(d-r)\times(d-r)}\end{pmatrix}$ אז $W_2|W_1\sim N(\mu_2+V_{21}V_{11}^{-1}(W_1-\mu_1), V_{22}-V_{21}V_{11}^{-1}V_{12})$
אם $V_{12}=0$ אזי: $Cov(W_1,W_2)=0$ ולכן: $W_2|W_1\sim N(\mu_2,V_{22})$ לא תלוי ב-W_1 ולכן $W_1\perp W_2$.

התפלגות *Wishart*

עבור $Y_{d\times n}$, שכל עמודה מתפלגת $Y_i\sim N_d(v_i,C)$, כאשר v_i הן העמודות של $M_{d\times n}$ ונסמן $H = MM^T$ הפובה למחצה כך ש$(H_{rs} = \sum_{i=1}^n v_{ri}v_{si})$ אזי: $S_{d\times d} = YY^T = \sum_{i=1}^n Y_iY_i^T\sim W_d(n,C,H)$

1. **מכפלה בוקטור** $l\in\mathbb{R}^d$ אז $l^TSl\sim\chi_k^2\left(\frac{l^THl}{l^TCl}\right)$
2. **הביטוי $\boldsymbol{U^TAU}$:** עבור $U_i\sim N_d(\xi,G)$ המהוות עמודות של $\mathcal{U}\in\mathbb{R}^{k\times d}$, ו- $A\in\mathbb{R}^{k\times k}$ קבועה, אזי שני התנאים הבאים שקולים: <u>תנאי א'</u> $U^TAU\sim W_d(k,G,\cdot)$, <u>תנאי ב':</u> $l^TU^TAUl\sim\sigma_l^2\chi_k^2(\cdot)$ עבור כל $l\in\mathbb{R}^d$ עם $\sigma_l^2 = l^TGl$.
3. עבור $U_i\sim N_d(\xi,G)$ המהוות עמודות של $\mathcal{U}\in\mathbb{R}^{k\times d}$:

(א) מטריצות $A_1,A_2\in\mathbb{R}^{k\times k}$ קבועות $Q_i = \mathcal{U}^TA_i\mathcal{U}$, אזי $Q_1,Q_2\sim W_d$ מתפלגים וב"ת אא"ם לכל $\ell\in\mathbb{R}^d$ המשתנים המקריים $T_i = \ell^TU^TA_iU\ell$, מתפלגים $\chi_d^2(\cdot)$ ו-T_1,T_2 בלתי-תלויים.

(ב) מטריצה $A\in\mathbb{R}^{k\times k}$ קבועה ו-$b\in\mathbb{R}^k$ וקטור קבוע. אזי: $\mathcal{U}^Tb\sim N(\cdot,\cdot),\ \mathcal{U}^TA\mathcal{U}\sim W_d(k,G,\cdot)$ ב"ת אא"ם לכל $\ell\in\mathbb{R}^d$ מתקיים: $\ell^T\mathcal{U}^Tb\sim N(\cdot,\cdot)$, $\ell^T\mathcal{U}^TA\mathcal{U}\ell\sim\chi_k^2(\cdot)$ אשר ב"ת.

(4) **סכום:** עבור $S_i\sim W_d(k_i,C)$ ב"ת. אזי $S_1+S_2\sim W_d(k_1+k_2,C)$.

(5) **מכפלה במטריצה:** עבור $S\sim W_d(k,C,\cdot)$ ו-$B\in\mathbb{R}^{q\times d}$ מטריצה קבועה. אזי $BSB^T\sim W_q(k,BCB^T,\cdot)$.

(6) עבור $S\sim W_d(k,\Gamma)$, אזי לכל $1\le r\le d$ מתקיים: $\frac{(\Gamma^{-1})_{rr}}{(S^{-1})_{rr}}\sim\chi_{k-d+1}^2$ הביטוי ב"ת בS_{ij} (לא כולל r)

(7) עבור $S\sim W_d(k,C)$ ו-$l\in\mathbb{R}^d$ לא מקרי. אזי $\frac{l^TC^{-1}l}{l^TS^{-1}l}\sim\chi_{k-d+1}^2$

(8) $S\sim W_d(k,C)$ ו-C לא סינגולרית. אזי $\frac{\det(S)}{\det(C)} = \prod_{r=1}^d V_r$ כאשר $V_r\sim\chi_{k-d+r}^2$ ו-$V_1,\dots,V_d$ בלתי-תלויים.

(9)$S\sim W_d(k,C)$ עם C הפיכה. אזי $\frac{\det(S)}{\det(C)} = \prod_{r=1}^d V_r$ לכל $\forall r: V_r\overset{indep}{\sim}\chi_{k-d+r}^2$

(10) $S_1\sim W_d(k_1,C)$ ו-$S_2\sim W_d(k_2,C)$ בלתי-תלויים. כאשר $k_1\ge d$ אזי

$$\frac{\det(S_1)}{\det(S_1+S_2)} = \prod_{r=1}^d H_r, \forall r: H_r\overset{indep}{\sim}Beta\left(\frac{k_1-r+1}{2},\frac{k_2}{2}\right)$$

עקומת גדילה

המודל: $Y_{ij} = \beta_{1i}+\beta_{2i}t_j+\epsilon_{ij}, i = 1,\dots,n, j = 1,\dots,J$

$Y_{i_{J\times1}}$ – מודדים J ערכים.

רישום מטריציוני: $Y_i = \vec{\beta}_iX_i+\vec{\epsilon}_i$ כאשר $X_{J\times2} = [\vec{1}\quad\vec{t}]$

$\begin{bmatrix}\beta_{1i}\\ \beta_{2i}\end{bmatrix}\sim N_2\left(\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}, A\right), \epsilon_{ij}\sim N(0,\sigma^2)$ β וϵ_{ij} ב"ת. ϵ_{ij} ב"ת.

$$Y_i\sim N(X\beta, XAX^T+\sigma^2I)$$

אומדים:

<u>אומד לβ:</u> $\hat{\beta} = \frac{1}{n}\sum_{i=1}^n\hat{\beta}_i, \hat{\beta}_i = (X^TX)^{-1}X^TY_i$

<u>אומד לY:</u> $\hat{Y}_i = X\hat{\beta}_i$ <u>אומד לϵ:</u> $e_i = Y_i-\hat{Y}_i$ מתקיים $X^Te_i = 0$

<u>אומד לσ^2:</u> $\widehat{\sigma^2} = \frac{1}{n(J-2)}\sum_{i=1}^n e_i^Te_i$

<u>אומד לC:</u> אנ"מ $\hat{C} = \frac{1}{n}\sum_{i=1}^n(\hat{\beta}_i-\hat{\beta})(\hat{\beta}_i-\hat{\beta})^T$ (ח"ה $\frac{1}{n-1}$)

התפלגויות:

<u>β:</u> $\boldsymbol{Cov(\hat{\beta}_i)} = C = A+\sigma^2(X^TX)^{-1}$

$\boldsymbol{E(\hat{\beta}_i)} = \beta_i+(X^TX)^{-1}X^T\epsilon_i$

$\boldsymbol{\hat{\beta}_i}\sim N\left(\beta, \overbrace{A+\sigma^2(X^TX)^{-1}}^{C}\right)$

$$\frac{n-1}{2(n-2)}\left(\hat{\beta}-\beta^{(0)}\right)^T\hat{C}^{-1}\left(\hat{\beta}-\beta^{(0)}\right)\sim F_{2,n-2}$$

מודל הליניארי המעורב

<u>המודל:</u> $\boldsymbol{Y_{ij} = X_{ij}^T\beta_i+Z_{ij}^Tb_i+\varepsilon_{ij}}$, עבור $i = 1,\dots,n, j = 1,\dots,J_i$

52801 מודלים סטטיסטיים למתקדמים 2023א
מרצה: פרופ' דוד צוקר
סיכום: אריאל וישנה

ממדים: $Y_i \in \mathbb{R}^{J_i}, \vec{\beta} \in \mathbb{R}^p, b_i \in \mathbb{R}^q, \epsilon_i \in \mathbb{R}^{J_i}$ $Z_i \in \mathbb{R}^{J_i \times q}$, $Z_{ij} \in \mathbb{R}^q, X_i = [X_{i1}^T](i \in 1,\dots,J_i) \in \mathbb{R}^{J_i \times p}, X_{ij} \in \mathbb{R}^p$

$\boldsymbol{J_i}$ דגימות שלקחנו לבן אדם $\boldsymbol{i}$, $\boldsymbol{p}$ הגורמים הקבועים, $\boldsymbol{q}$ הגורמים המקריים

<u>תצוגה וקטורית</u>: $\boldsymbol{Y_i = X_i\beta + Z_i b_i + \varepsilon_i}$

ממדים: $Y_i \in \mathbb{R}^{J_i}, X \in \mathbb{R}^{J_i \times p}, Z \in \mathbb{R}^{J_i \times q}, b_i \in \mathbb{R}^q, \varepsilon_i \in \mathbb{R}^{J_i}$

<u>תצוגה מטרציונית</u>: $\boldsymbol{Y = X\beta + Zb + \epsilon}$

ממדים: $\vec{Y} \in \mathbb{R}^N, X \in \mathbb{R}^{N \times p}, \beta \in \mathbb{R}^{p \times 1}, Z \in \mathbb{R}^{N \times nq}, b \in \mathbb{R}^{q \times 1}$, $\varepsilon \in \mathbb{R}^{N \times 1}, X\beta \in \mathbb{R}^N, Z\vec{b} \in \mathbb{R}^N$

Z – מטריצת בלוקים עם Z_i על האלכסון. V בלוקים עם V_i על האלכסון.

<u>התפלגויות:</u>

	תצוגה מטריציות	תצוגה וקטורית
$\boldsymbol{Y}$	$Y \sim N(X\beta, V)$	$Y_i \sim N(X_i\beta, V_i)$
	$V = ZGZ^T + R = diag(V_i)$	$V_i = Z_i G Z_i^T + R_i$
$\boldsymbol{b}$	$b \sim N(0, \mathcal{G})$	$b_i \sim N\left(0, G\left(\theta^{(1)}\right)\right)$
	$\mathcal{G} = diag(G) \in \mathbb{R}^{nq \times nq}$	$G \in \mathbb{R}^{q \times q}$
$\boldsymbol{\varepsilon}$	$\varepsilon \sim N(0, R)$	$\varepsilon_i \sim N\left(0, R_i\left(\theta^{(2)}\right)\right)$
	$R = diag(R_i) \in \mathbb{R}^{N \times N}$	$R_i \in \mathbb{R}^{J_i \times J_i}$
$\boldsymbol{\hat{\beta}}$	$\hat{\beta} \dot{\sim} N(\beta, (X^T V^{-1}(\theta) X)^{-1})$	

$$\begin{bmatrix} \hat{\beta} \\ \hat{\theta} \end{bmatrix} \dot{\sim} N\left(\begin{bmatrix} \beta \\ \theta \end{bmatrix}, \begin{bmatrix} (X^T V^{-1} X)^{-1} & 0 \\ 0 & \left(\mathfrak{I}^{\theta\theta}\right)^{-1} \end{bmatrix}\right)$$

כאשר $\mathfrak{I}^{\theta\theta} = \mathbb{E}\left[\ell^{\theta\theta}\right]_{rs} = \left[-\frac{1}{2} tr\left(V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right)\right]_{rs}$

* ההתפלגויות המטרציונית נובעות מהנחת אי-תלות בין היחידות

<u>ר"ס ובדיקת השערות</u>

<u>β_r</u>-ר"ס להשפעה של משתנה מסביר ה r :

$$\hat{\beta}_r \pm Z_{1-\alpha/2}\sqrt{([X^T V^{-1} X]^{-1})_{rr}}$$

* עבור מדגם קטן-בינוני משתמשים בהת' t:

$$\hat{\beta}_r \pm t_{d;(1-\alpha)}\sqrt{([X^T V^{-1} X]^{-1})_{rr}}\,, d = 2\left[\mathbb{V}(\hat{\beta}_r)\right]^2 / \mathbb{V}\left[\widehat{\mathbb{V}}(\hat{\beta}_r)\right]$$

דרגות חופש נאמדות לפי *Satterwhite*

בדיקת השערות: $H_0: \beta_r = 0, H_1: \beta_r \neq 0$

נשתמש ב:

$$\zeta_r = \hat{\beta}_r / \sqrt{([X^T V^{-1} X]^{-1})_{rr}} \overset{H_0}{\sim} N(0,1)$$

<u>β – ר"ס לוקטור β:</u> $(\beta - \hat{\beta})^T \left[\widehat{Cov}(\hat{\beta})\right]^{-1} (\beta - \hat{\beta}) \dot{\sim} \chi_p^2$

$$\left\{\beta \in \mathbb{R}^p : (\beta - \hat{\beta})^T \left[\widehat{Cov}(\hat{\beta})\right]^{-1} (\beta - \hat{\beta}) \leq \chi^2_{p;(1-\alpha)}\right\} = \left\{\beta \in \mathbb{R}^p : \left(X\beta - X\hat{\beta}\right)^T V^{-1} \left(X\beta - X\hat{\beta}\right) \leq \chi^2_{p;(1-\alpha)}\right\}$$

<u>$X_{ij}^T\hat{\beta}$ – האומד לתוחלת של Y:</u>

$$X_{ij}^T\hat{\beta} \dot{\sim} N\left(X_{ij}^T\beta, X_{ij}^T \Omega^{\beta\beta} X_{ij}\right)$$

ר"ס $X_{ij}^T\hat{\beta} \pm t_{d;(1-\alpha)}\sqrt{X_{ij}^T \widehat{\Omega}^{\beta\beta} X_{ij}}$ כאשר $\Omega^{\beta\beta} = (X^T V^{-1}(\theta) X)^{-1}$

<u>$b_i|Y_i$ – אמידת גודל האפקט:</u>

$$b_i | Y_i \sim N\left(G Z_i^T V_i^{-1}(Y_i - X\beta), G - G^T Z_i^T V_i^{-1} Z_i G\right)$$

הפיתוח:

$$\begin{bmatrix} Y_i \\ b_i \end{bmatrix} \sim N\left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, \begin{bmatrix} V_i & Z_i G \\ (Z_i G)^T & G \end{bmatrix}\right), Cov(Y_i, b_i) = Z_i G$$

* אמידה בנקודה t^* : $\hat{b}_i = \widehat{\mathbb{E}}[b_i|Y_i] = \hat{G} Z_i \hat{V}_i^{-1}(Y_i - X\hat{\beta})$

<u>אומדים במקרה הכללי</u>

<u>אומד לβ:</u> $\hat{\beta} = \beta + KZb + K\varepsilon,\ K = (X^T V^{-1} X)^{-1} X^T V^{-1}$

<u>אומד לb:</u> $\hat{b} = \mathcal{G} Z^T V^{-1}(Y - X\beta) = (DKZb + F)b + (H + K)\varepsilon$

$$D = -\mathcal{G} Z^T V^{-1} X, F = \mathcal{G} Z^T V^{-1} Z, H = \mathcal{G} Z^T V^{-1}$$

<u>הוקטור $[(\hat{\beta} - \beta) \quad (\hat{b} - b)]^T \sim N(0, C)$:</u>

$$C = Cov\left(\begin{bmatrix} \hat{\beta} - \beta \\ \hat{b} - b \end{bmatrix}\right) = Cov\left(\begin{bmatrix} KZb + K\varepsilon \\ (DKZ + F - I)b + (H + K)\varepsilon \end{bmatrix}\right)$$

<u>אומדים כאשר $\boldsymbol{X_i = Z_i = \vec{1}_{J_i}}$:</u> המודל $Y_i = \vec{1}_{J_i}\mu + \vec{1}_{J_i} b_i + \varepsilon_i$

התפלגויות: $b_i \sim N(0, \sigma_b^2), \varepsilon_i \sim N_{J_i}(0, \sigma_\varepsilon^2 I)$

<u>אומד לV_i^{-1}:</u> $V_i^{-1} = \frac{1}{\sigma_\varepsilon^2}\left(I_{J_i} - \xi 1_{J_i} 1_{J_i}^T\right), \xi = \frac{w}{1 + wJ_i}, w = \frac{\sigma_b^2}{\sigma_\varepsilon^2}$

<u>אומד לb_i:</u> $\hat{b}_i = J_i\sigma_b^2 / (\sigma_\varepsilon^2 + J_i\sigma_b^2)\,(\bar{Y}_{i\cdot} - \hat{\mu})$

<u>אומד לμ_i:</u> $\hat{\mu}_i = \hat{\mu} + \hat{b}_i = \frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + J_i\sigma_b^2}\bar{Y}_{i\cdot} + \left(1 - \frac{J_i\sigma_b^2}{\sigma_\varepsilon^2 + J_i\sigma_b^2}\right)\hat{\mu}$

$\bar{Y}_{i\cdot}$ - ממוצע ביחידה i , $\hat{\mu} = \mathbb{E}\left[Y_{ij}\right]$ - האומד הכללי של Y_{ij} בכל המדגם

(1) <u>J_i גדול</u> ← $\hat{\mu}_i \approx \bar{Y}_{i\cdot}$ (2) <u>σ_b^2 גדול</u> ← $\hat{\mu}_i \approx \bar{Y}_{i\cdot}$.
(3) <u>σ_ε^2 גדול</u> ← $\widehat{\mu_i} \approx \hat{\mu}$

<u>בדיקת השערות:</u> $\hat{\mu}_i - \mu_i \sim N\left(\vec{1}_2^T C \vec{1}_2\right)$, ר"ס: $\widehat{\mu_i} \pm z_{1-\alpha/2}\sqrt{\vec{1}_2^T C \vec{1}_2}$

<u>**בדיקת השערות עבור המודל $\boldsymbol{Y_{ij} = \beta_1 + \beta_2 t_{ij} + b_{i1} + b_{i2} t_{ij} + \varepsilon_{ij}}$**</u>

Y_{ij}^* - הערך של Y עבור פרט i בזמן t^*.נגדיר

$$\Psi_{ij} = \mathbb{E}\left[Y_{ij}^* \middle| b_i\right] = [1 \quad t^*][\beta_1 \quad \beta_2]^T + [1 \quad t^*][b_{i1} \quad b_{i2}]^T$$

ואת a_i^T כך ששני האיברם הראשוניםוגם האיבר ה$2i + 1, 2i + 2$ יהיו $[1 \quad t^*]$. למשל: $a_1^T = [1 \quad t^* \quad 1 \quad t^* \quad 0 \quad \dots \quad 0]$

אז $\widehat{\Psi}_{ij} - \Psi_{ij} = a_i^T[(\hat{\beta} - \beta) \quad (\hat{b} - b)]^T$

<u>התפלגות $\widehat{\Psi}_{ij} - \Psi_{ij}^*$:</u> $\left(\widehat{\Psi}_{ij} - \Psi_{ij}^*\right) \sim N\left(0, a_i^T C a_i\right)$

<u>רווח סמך ל-Ψ_{ij}^* בזמן t^*:</u> $\widehat{\Psi}_{ij}^* \pm z_{1-\alpha/2}\sqrt{a_i^T \hat{C} a_i^T}$

<u>התפלגות:</u> $Y_{ij} - \widehat{\Psi}_{ij}^* \dot{\sim} N\left(0, a_i^T C a_i + \varepsilon_{ij}\right)$

$$\widehat{\Psi}_{ij}^* \pm z_{1-\alpha/2}\sqrt{a_i^T \hat{C} a_i^T + \sigma_\varepsilon^2}$$

<u>אומדים ל$\boldsymbol{\theta}$ בשיטות נומריות</u>

$$L(\vec{\beta}, \vec{\theta}) = (2\pi)^{-N/2} \cdot \det(V)^{-1/2} \cdot \exp\left\{-\frac{1}{2}(Y - X\beta)^T V^{-1}(Y - X\beta)\right\}$$

$$\ell(\vec{\beta}, \vec{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(det(V)) - \frac{1}{2}(Y - X\beta)^T V^{-1}(Y - X\beta)$$

<u>**שיטות נומריות**</u> תהליכים איטרטיביים, $\vec{\phi}^{(0)}$ כרצוני

שיטת ניוטון-רפסון: $\vec{\phi}^{(m+1)} = \phi^{(m)} - \nabla^2\ell\left(\vec{\phi}^{(m)}\right)^{-1}\nabla\ell\left(\phi^{(m)}\right)$

* לרוב מתכנס, אם לא אז כנראה יחס תצפיות לפרמטרים קטן.

שיטת פישר: $\vec{\phi}^{(m+1)} = \phi^{(m)} - \left(\mathbb{E}\left[\nabla^2\ell\left(\vec{\phi}^{(m)}\right)^{-1}\right]\right)\nabla\ell\left(\phi^{(m)}\right)$

$\boldsymbol{\nabla^2\ell}$ הוא ההסיאן: $[\nabla^2\ell]_{st} = \frac{\partial\ell}{\partial\phi_s\partial\phi_t}$	$\boldsymbol{\nabla\ell}$ הוא הגרדיאנט: $\nabla\ell = \left[\frac{\partial\ell}{\partial\phi_1} \quad \cdots \quad \frac{\partial\ell}{\partial\phi_k}\right]^T$

* <u>במקרה שלנו:</u> $\vec{\phi} = \left[\vec{\beta} \quad \vec{\theta}^{(1)} \quad \vec{\theta}^{(2)}\right]^T \in \mathbb{R}^{k = u_1 + u_2 + p}$

נגזרות חלקיות: $\dot{V}_s = \partial V / \partial\theta_s, \qquad \frac{\partial}{\partial\theta_s}V^{-1} = -V^{-1}\dot{V}_s V^{-1}$

$$\boldsymbol{\ell^\beta} = [X^T V^{-1}(Y - X\beta)]_r, \qquad \boldsymbol{\ell^{\beta\beta}} = -(X^T V^{-1} X)_{rs}$$

$$\boldsymbol{\ell^\theta} = -\frac{N}{2}tr\left(V^{-1}\dot{V}_s\right) + \frac{1}{2}(Y_i - X\beta)^T V^{-1}\dot{V}_s V^{-1}(Y_i - X\beta)$$

$$\boldsymbol{\ell^{\beta\theta}} = -(X^T V^{-1}\dot{V}_S V^{-1}(Y - X\beta), \qquad \boldsymbol{\ell^{\theta\beta}} = \ell^{\beta\theta^T}$$

$$\boldsymbol{\ell^{\theta\theta}} = -\frac{N}{2}tr\left(-V^{-1}\dot{V}_s V^{-1}\dot{V}_r\right) - \frac{N}{2}tr\left(V^{-1}\ddot{V}_{rs}\right) + \frac{1}{2}(Y - X\beta)^T Q(Y - X\beta)$$

$$Q = (\partial/\partial\theta_s)V^{-1}\dot{V}_r V^{-1} = \left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r V^{-1} + V^{-1}\ddot{V}_{rs}V^{-1} - V^{-1}\dot{V}_r V^{-1}\dot{V}_s V^{-1}$$

תוחלות לנגזרות:

$$\mathbb{E}\left[\ell^{\beta\beta}\right] = -(X^T V^{-1} X), \qquad \mathbb{E}\left[\ell^{\theta\beta}\right] = \mathbb{E}\left[\ell^{\beta\theta}\right] = 0$$

$$\mathbb{E}\left[\ell^{\theta\theta}\right] = tr\left(\left(-V^{-1}\dot{V}_s V^{-1}\right)\dot{V}_r + V^{-1}\ddot{V}_{rs} - V^{-1}\dot{V}_r V^{-1}\dot{V}_s\right)$$

<u>משפט:</u> עבור $W_1, \dots, W_n$ וקטורי תצפיות ב"ת מתפלגים בהתפלגות עם פרמטר $\vec{\phi} \in \mathbb{R}^k$. יהי $\hat{\phi}$ אנ"מ ל-ϕ. ומטריצת האינפורמציה של פישר: $\mathfrak{I}(\phi) = -\mathbb{E}[\nabla^2\ell]$ אזי $\hat{\phi} \dot{\sim} N(\phi, \mathfrak{I}(\phi)^{-1})$.

<u>מקרים ספציפיים</u>

המקרה הפשוט: $X_i = Z_i = [1 \quad t_j]$

Generlaized Least Squares: אם $V = I$ אז היינו מקבלים אומד ריבועים פחותים. במקרה שלנו:

$$\hat{\beta} = (X^T V X)^{-1} X^T V^{-1} Y$$

נספח – תאריכי

שיעורים

מספר	תאריך	הערות
1	23.10.2022	
2	26.10.2022	
3	30.10.2022	
4	02.11.2022	
5	06.11.2022	
6	09.11.2022	
7	13.11.2022	
8	16.11.2022	
9	20.11.2022	
10	23.11.2022	
11	27.11.2022	
12	30.11.2022	
13	04.12.2022	
14	07.12.2022	
15	11.12.2022	
16	14.12.2022	
17	18.12.2022	חנוכה
	21.12.2022	לא התקיים שיעור
	25.12.2022	לא התקיים שיעור
18	28.12.2022	
19	01.01.2023	
20	04.01.2023	
21	08.01.2023	
22	11.01.2023	
23	15.01.2023	
24	18.01.2023	
25	22.01.2023	