נושאים באנליזה סטטיסטית מרובת משתנים: תרגיל מס' 3

תאריך הגשה: 18.7.2022

1. גילוי אותות ברעש גאוסי באמצעות Mardia's skewness and kurtosis tests

נתונה בעיית ההחלטה הבאה עבור גילוי אות אקראי לא-גאוסי ברעש גאוסי:

$$H_0: \quad \mathbf{X}_n = \mathbf{W}_n, \quad n = 1, \ldots, N \quad \text{(signal does not exist)}$$
$$H_1: \quad \mathbf{X}_n = \mathbf{S}_n + \mathbf{W}_n, \quad n = 1, \ldots, N \quad \text{(signal exists)}$$

כאשר $\left\{\mathbf{X}_n \in \mathbb{R}^p\right\}_{n=1}^N$ היא סדרת אובזרבציות, $\left\{\mathbf{S}_n \in \mathbb{R}^p\right\}_{n=1}^N$ הוא תהליך סיגנל אקראי i.i.d. בלתי-נצפה לא-גאוסי עם פילוג סימטרי (מסביב לראשית) לא ידוע ו- $\left\{\mathbf{W}_n \in \mathbb{R}^p\right\}_{n=1}^N$ היא סדרת רעש i.i.d. עם פילוג גאוסי. מניחים כי:

- התהליכים $\left\{\mathbf{S}_n \in \mathbb{R}^p\right\}_{n=1}^N$ ו- $\left\{\mathbf{W}_n \in \mathbb{R}^p\right\}_{n=1}^N$ הם בלתי תלויים סטטיסטית (בת"ס).
- איברי וקטור הסיגנל $\mathbf{S}_n \triangleq \left[S_{1,n}, \ldots, S_{p,n}\right]^T$ הם בת"ס עם וריאנס זהה $\sigma_{\mathbf{S}}^2$.
- איברי וקטור הרעש $\mathbf{W}_n \triangleq \left[W_{1,n}, \ldots, W_{p,n}\right]^T$ הם חסרי קורלציה עם תוחלת 0 ווריאנס זהה $\sigma_{\mathbf{W}}^2$.

מעוניינים לגלות את האות באמצעות מבחני הנורמליות של Mardia.

א. חשבו את סף ההחלטה שעבורו תתקבל הסתברות גילוי שווא (PFA) של 0.05 כאשר משתמשים ב- Mardia's skewness test.

ב. האם תחת המקרה המדובר Mardia's skewness test הוא קונסיסטנטי? הוכיחו את תשובתכם.

ג. חשבו את סף ההחלטה שעבורו תתקבל הסתברות גילוי שווא (PFA) של 0.05 כאשר משתמשים ב- Mardia's kurtosis test.

ד. מצאו תנאי מספיק והכרחי על היחס:

$$\frac{1}{p\sigma_{\mathbf{S}}^4} \sum_{k=1}^{p} \mathrm{E}\left[S_k^4\right]$$

שעבורו Mardia's kurtosis test הוא קונסיסטנטי.

ה. נגדיר יחס-אות-לרעש (SNR) באופן הבא:

$$\mathrm{SNR} = 10 \log_{10}\left(\frac{\sigma_{\mathbf{S}}^2}{\sigma_{\mathbf{W}}^2}\right) \quad [\mathrm{dB}].$$

- באמצעות ספף ההחלטה שמצאתם בסעיף ג', שערכו את הסתברות הגילוי (PD) של Mardia’s kurtosis test עבור ערכי ה-SNR הבאים: $\text{SNR} \in \{-10,-5,0,5,10\}$ באמצעות 1000 סימולציות מונטה-קרלו. הניחו כי הפילוג של איברי וקטור האות נתון לפי:

$$S_k = \begin{Bmatrix} 1 & w.p.\ \frac{1}{2} \\ -1 & w.p.\ \frac{1}{2} \end{Bmatrix}, \quad k = 1,\dots,p$$

- ציירו גרף של הסתברויות הגילוי ששיערכתם כפונקציה של ה-SNR. הסבירו את השינוי בהסתברות הגילוי הנגרם כתוצאה משינוי ביחס-האות-לרעש.

2. אינווריאנטיות תחת טרנספורמציות אפיניות של מבחן ה-GLRT לחוסר תלות סטטיסטית

יהיו $\mathbf{X} \in \mathbb{R}^p$ ו- $\mathbf{Y} \in \mathbb{R}^q$ וקטורים אקראיים.

בהינתן N דגימות מהפילוג המשותף $P_{\mathbf{XY}}$: $(\mathbf{X}_1,\mathbf{Y}_1),\dots,(\mathbf{X}_N,\mathbf{Y}_N)$, נגדיר את המבחן הבא:

$$T(\mathbf{X}_1,\dots,\mathbf{X}_N;\mathbf{Y}_1,\dots,\mathbf{Y}_N) \triangleq \left(N - \frac{p+q+3}{2}\right)\left(\log\left|\hat{\mathbf{\Sigma}}_{\mathbf{X}}\right| + \log\left|\hat{\mathbf{\Sigma}}_{\mathbf{Y}}\right| - \log\left|\begin{bmatrix} \hat{\mathbf{\Sigma}}_{\mathbf{X}} & \hat{\mathbf{\Sigma}}_{\mathbf{XY}} \\ \hat{\mathbf{\Sigma}}_{\mathbf{YX}} & \hat{\mathbf{\Sigma}}_{\mathbf{Y}} \end{bmatrix}\right|\right)$$

כאשר:

- $\hat{\mathbf{\Sigma}}_{\mathbf{X}} \triangleq \frac{1}{N}\sum_{n=1}^{N}(\mathbf{X}_n - \hat{\boldsymbol{\mu}}_{\mathbf{X}})(\mathbf{X}_n - \hat{\boldsymbol{\mu}}_{\mathbf{X}})^T$, $\hat{\mathbf{\Sigma}}_{\mathbf{Y}} \triangleq \frac{1}{N}\sum_{n=1}^{N}(\mathbf{Y}_n - \hat{\boldsymbol{\mu}}_{\mathbf{Y}})(\mathbf{Y}_n - \hat{\boldsymbol{\mu}}_{\mathbf{Y}})^T$
- $\hat{\mathbf{\Sigma}}_{\mathbf{XY}} \triangleq \frac{1}{N}\sum_{n=1}^{N}(\mathbf{X}_n - \hat{\boldsymbol{\mu}}_{\mathbf{X}})(\mathbf{Y}_n - \hat{\boldsymbol{\mu}}_{\mathbf{Y}})^T$
- $\hat{\boldsymbol{\mu}}_{\mathbf{X}} \triangleq \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}_n$, $\hat{\boldsymbol{\mu}}_{\mathbf{Y}} \triangleq \frac{1}{N}\sum_{n=1}^{N}\mathbf{Y}_n$

בנוסף, נגדיר $\mathbf{X}'_k \triangleq \mathbf{A}\mathbf{X}_k + \mathbf{b}$ ו- $\mathbf{Y}'_k \triangleq \mathbf{C}\mathbf{Y}_k + \mathbf{d}$, כאשר $\mathbf{A} \in \mathbb{R}^{p\times p}$ ו- $\mathbf{B} \in \mathbb{R}^{q\times q}$ הן מטריצות דטרמיניסטיות הפיכות כלשהן ו- $\mathbf{b} \in \mathbb{R}^p$ ו- $\mathbf{d} \in \mathbb{R}^q$ הם וקטורים דטרמיניסטיים כלשהם.

הוכיחו כי המבחן מקיים את תכונת האינווריאנטיות:

$$T(\mathbf{X}_1,\dots,\mathbf{X}_N;\mathbf{Y}_1,\dots,\mathbf{Y}_N) = T(\mathbf{X}'_1,\dots,\mathbf{X}'_N;\mathbf{Y}'_1,\dots,\mathbf{Y}'_N)$$

3. אנווריאנטיות תחת טרנספורמציות אפיניות של מבחן ה-Mutual Information לחוסר תלות סטטיסטית:

יהיו $\mathbf{X} \in \mathbb{R}^p$ ו- $\mathbf{Y} \in \mathbb{R}^q$ וקטורים אקראיים.

בהינתן N דגימות מהפילוג המשותף $P_{\mathbf{XY}}$: $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_N, \mathbf{Y}_N)$, נגדיר את המבחן הבא:

$$T(\mathbf{X}_1, \ldots, \mathbf{X}_N; \mathbf{Y}_1, \ldots, \mathbf{Y}_N) = \sqrt{1 - \exp\left(-2\hat{I}(\mathbf{X}; \mathbf{Y})\right)}$$

כאשר, $\hat{I}(\mathbf{X}; \mathbf{Y})$ הוא משערך האינפורמציה ההדדית בין $\mathbf{X}$ ו- $\mathbf{Y}$ הנתון לפי:

$$\hat{I}(\mathbf{X}; \mathbf{Y}) \triangleq \frac{1}{N} \sum_{n=1}^{N} \log \frac{\hat{f}_{\mathbf{XY}}(\mathbf{X}_n, \mathbf{Y}_n)}{\hat{f}_{\mathbf{X}}(\mathbf{X}_n) \hat{f}_{\mathbf{Y}}(\mathbf{Y}_n)}$$

ו- $\hat{f}_{\mathbf{XY}}(\cdot,\cdot)$, $\hat{f}_{\mathbf{X}}(\cdot)$, $\hat{f}_{\mathbf{Y}}(\cdot)$ הן צפיפויות הפילוג האמפיריות המשותפת והשוליות, בהתאמה, אשר מתקבלות באמצעות Kernel density estimation.

נגדיר $\mathbf{X}'_k \triangleq \mathbf{A}\mathbf{X}_k + \mathbf{b}$ ו- $\mathbf{Y}'_k \triangleq \mathbf{C}\mathbf{Y}_k + \mathbf{d}$, כאשר $\mathbf{A} \in \mathbb{R}^{p \times p}$ ו- $\mathbf{B} \in \mathbb{R}^{q \times q}$ הן מטריצות דטרמיניסטיות הפיכות כלשהן ו- $\mathbf{b} \in \mathbb{R}^p$ ו- $\mathbf{d} \in \mathbb{R}^q$ הם וקטורים דטרמיניסטיים כלשהם.

הוכיחו כי המבחן מקיים את תכונת האינווריאנטיות:

$$T(\mathbf{X}_1, \ldots, \mathbf{X}_N; \mathbf{Y}_1, \ldots, \mathbf{Y}_N) = T(\mathbf{X}'_1, \ldots, \mathbf{X}'_N; \mathbf{Y}'_1, \ldots, \mathbf{Y}'_N)$$

4. השוואה בין מבחן ה-GLRT ומבחן ה-Mutual Information לחוסר תלות סטטיסטית:

יהיו X ו- Y משתנים אקראיים גאוסיים סטנדרטיים המקיימים את הקשר הלא-לינארי הבא:

$$Y = \cos(X) + W$$

כאשר W רעש גאוסי סטנדרטי. בהינתן $N = 1000$ מדידות מהפילוג המשותף של X ו- Y, שערכו את ה-p-value באמצעות 1000 מבחני פרמוטציה עבור המבחנים שהוגדרו בשאלות 3 ו-4. הסבירו את ההבדל בין ערכי ה- p-value ששיערכתם.