מבוא להסתברות וסטטיסטיקה 2, 88-162

פרופ' ע. וישנה

סמסטר ב', מועד א', תשס"ט – פתרון מלא

1. שטחה של מושבת חיידקים עגולה הגדלה בלי הפרעה מתפלג מערכית, עם תוחלת $\theta = 2\pi$. נסמן ב־ R את רדיוס המושבה.

(א) מצא את פונקציית הצפיפות של R. **פתרון**. נסמן ב־ S את שטח המושבה, אז $S \sim \mathrm{Exp}(2\pi)$ עם פונקציית צפיפות $f_S(s) = \frac{1}{2\pi}e^{-s/2\pi}$ כאשר $s \geq 0$. מכיוון ש־ $S = \pi R^2$ (ולא $5\log(R^7)$, כפי שסברו כמה מכם), הטרנספורמציה היא $R = h(S)$ כאשר $r = h(s) = \sqrt{s/\pi}$, עם נגזרת $h'(s) = \frac{1}{2\sqrt{\pi}}s^{-1/2}$. לכן $f_R(r) = f_S(s)\left|h'(s)\right|^{-1} = re^{-r^2/2}$.

(ב) מה התוחלת של R? **פתרון**. אם Z משתנה נורמלי סטנדרטי, אז $E(R) = \int_0^\infty r^2e^{-r^2/2}dr = \frac{\sqrt{2\pi}}{2}\int_{-\infty}^\infty r^2\frac{1}{\sqrt{2\pi}}e^{-r^2/2}dr = \sqrt{\frac{\pi}{2}}E(Z^2) = \sqrt{\frac{\pi}{2}}.$

(ג) מה השונות של R? **פתרון**. אין צורך לחשב ישירות, כי R^2 הוא כפולה של S: $E(R^2) = E(S/\pi) = 2$, ולכן $V(R) = E(R^2) - E(R)^2 = 2 - \frac{\pi}{2} = \frac{4-\pi}{2}$.

2. משקלה של ככר לחם תקנית מתפלג נורמלית, עם תוחלת 750 גרם וסטיית תקן 20 גרם.

(א) מה הסיכוי לכך שעשרה ככרות לחם תקניים ישקלו יחד פחות מ־ 7450 גרם? **פתרון**. הסכום S של משקלי עשרה ככרות לחם מתפלג נורמלית, עם סכום התוחלות והשונויות: $S \sim N(750, 10\cdot 20^2)$. לפי הטבלה, $P(S < 7450) = P(\frac{S-7500}{20\sqrt{10}} < \frac{7450-7500}{20\sqrt{10}} = -\frac{\sqrt{10}}{4}) \approx 0.2146$.

(ב) מאפייה מייצרת ככרות שמשקלם מתפלג $N(\mu, 20^2)$. כתוב את ההשערות (H_0 ו־H_1) הנחוצות כדי להוכיח שהמאפיה עוברת על החוק, ומוכרת ככרות מתחת למשקל התקני. **פתרון**. $H_0 : \mu = 750$, $H_1 : \mu < 750$. תוצאה חריגה תאפשר לדחות את H_0 וכך להרשיע את המאפיה.

(ג) בדוק, ברמת מובהקות 95%, את ההשערה שהמאפייה עוברת על החוק, אם בחנות נקנו עשרה ככרות שמשקלם הכולל 7450 גרם. **פתרון**. ראינו לעיל שהסיכוי לתופעה כזו הוא כ־20%. רק תופעה שהסיכוי לה הוא 5% או פחות יכולה להביא לדחיית השערה ברמת מובהקות של 0.95.

3. במשך השנה, אורכה של פרסומת בטלוויזיה מתפלג נורמלית עם תוחלת 60 שניות וסטיית תקן 12 שניות.

(א) מה אחוז הפרסומות שאורכן מעל 90 שניות? **פתרון**. נסמן ב־ X אז האורך של פרסומת בודדת. ההתפלגות היא $X \sim N(60, 12^2)$, והסיכוי ל־ $X > 90$ הוא $P(X > 90) = P(\frac{X-60}{12} > \frac{90-60}{12} = 2.5) \approx 0.0062$.

(ב) איך מתפלג אורכו של מקבץ בן חמש פרסומות, בהנחה שהפרסומות בלתי תלויות זו בזו? **פתרון**. סכום של משתנים נורמליים: $N(300, 5\cdot 12^2)$. בכמה מקבצים יש לפחות פרסומת אחת שאורכה מעל 90 שניות? **פתרון**. המספר A של הפרסומות באורך 90 שניות ומעלה מתפלג בינומית $\mathrm{Bin}(5, p)$ כאשר $p = 0.0062$. לכן הסיכוי ל־ $A \geq 1$ הוא $1 - (0.9932)^5 \approx 0.03354$

(ג) ביום מסויים שודרו 40 פרסומות, שאורכן הממוצע היה $60+\delta$ שניות. מה צריך להיות הערך של δ כדי שאפשר יהיה להוכיח, ברמת מובהקות 95%, שאורך הפרסומות באותו יום היה חריג במכוון? **פתרון**. תחת השערת האפס (שלפיה תוחלת האורך לא השתנתה) הממוצע של 40 פרסומות מתפלג $N(60,\frac{12^2}{40})$, ולכן $P(60+\delta=\bar{X}>60+\delta)=0.05$ כאשר $\delta=1.645\cdot\frac{12}{\sqrt{40}}=3.12$

4. חוקרי מס הכנסה מעוניינים לבדוק האם יש קשר בין העלמת מס להפקעת מחירים. הם דוגמים 60 דוכנים בשוק, ומגלים שהיו 15 מקרים של העלמת מס, ו־ 24 מקרים של הפקעת מחירים.

(א) מהו המספר הצפוי של דוכנים שגם מעלימים מס וגם מפקיעים מחירים, אם העבירות אינן תלויות זו בזו? **פתרון**. בהנתן מספר מעלימי המס ומספר מפקיעי המחירים, מספר הדוכנים העונים לשני התנאים מתפלג היפרגאומטרית $H(60;15,24)$. אבל חישוב התוחלת פשוט יותר: הסיכוי של כל דוכן להיות בשתי הקבוצות גם יחד הוא $\frac{15}{60}\cdot\frac{24}{60}=\frac{1}{10}$. לכן צפויים $60\cdot\frac{1}{10}=6$ דוכנים כאלה.

(ב) כתוב את השערת האפס ואת ההשערה האלטרנטיבית. באיזה מבחן יש לבדוק את ההשערה? (כמה דרגות חופש?) **פתרון**. השערת האפס היא שאין תלות בין העלמת מס לבין הפקעת מחירים; ההשערה האלטרנטיבית היא שיש קשר. המבחן המתאים הוא χ^2 עם דרגת חופש אחת: $(2-1)(2-1)=1$.

(ג) בפועל התברר שהיו 11 דוכנים שעברו את שתי העבירות. בדוק את ההשערה ברמת מובהקות של 90% ושל 95%. **פתרון**. הנתונים הנצפים הם $\begin{matrix}11&4\\13&32\end{matrix}$, והנתונים המצופים הם $\begin{matrix}6&9\\18&27\end{matrix}$. לכן הסטטיסטי המתאים מקבל את הערך $\frac{(11-6)^2}{6}+\frac{(4-9)^2}{9}+\frac{(13-18)^2}{18}+\frac{(32-27)^2}{27}=\frac{250}{27}\approx 9.25$. הערך הזה גדול מן הערכים המתאימים בטבלה, ולכן דוחים את ההשערה: העבירות תלויות.

(ד) מה תהיה ההשפעה על התוצאות אם יתברר שלצד 60 הדוכנים שנבדקו בשוק, היו עוד 60 שלא נבדקו כלל? **פתרון**. פשיטא: דוכנים שלא נבדקו, 60 או 6000, אינם יכולים לשנות את התוצאה המחושבת תמיד מתוך המדגם.

5. ארנבת משוטטת במבוך שיש בו שני חדרים. אם היא בחדר הגדול, הסיכוי שתהיה שם לאחר שעה הוא 0.9; אחרת היא עוברת לחדר הקטן. אם היא בחדר הקטן, הסיכוי שתשאר שם לאחר שעה הוא 0.6.

(א) מה שעור הזמן שהארנבת מבלה בחדר הגדול? **פתרון**. מטריצת מרקוב המתארת את התהליך היא $\begin{pmatrix}0.9&0.4\\0.1&0.6\end{pmatrix}$, וכדי למצוא את ההתפלגות הסטציונרית יש לפתור את המערכת $x+y=1$, $0.9x+0.4y=x$; לכן $x=0.8$ הוא שעור הזמן המבוקש.

(ב) לאחר שהוסיפו גלגל מרוץ בחדר הקטן, הארנבת משנה את התנהגותה. מעכשיו, כשהיא בחדר הגדול, הסיכוי שלה להשאר בו הוא 0.9 אם היתה בו גם בשעה הקודמת, אבל רק 0.8 אם היתה לפני שעה בחדר הקטן. הסיכוי להשאר בחדר הקטן נותר ללא שינוי. תאר (באופן גרפי) תהליך מרקוב מתאים למצב החדש. **פתרון**. כדי לתאר את הזכרון שנוסף לארנבת, יש להפריד בין המצבים 'שהות בחדר הגדול שעה שניה ברציפות', לבין 'שהות בחדר הגדול לאחר מעבר מן החדר הקטן', בתוספת למצב הקודם 'שהות בחדר הקטן'. מן המצב הראשון או השני אפשר לעבור רק למצב

הראשון או האחרון, ומן המצב האחרון אפשר לעבור רק לשני ולאחרון. המטריצה היא

$$\begin{pmatrix} 0.9 & 0.8 & 0 \\ 0 & 0 & 0.4 \\ 0.1 & 0.2 & 0.6 \end{pmatrix}.$$

(ג) מה שעור הזמן שהארנבת מבלה בחדר הגדול לאחר השינוי? **פתרון**. מערכת המשוואות היא $0.9x+0.8y=x, 0.4z=y, 0.1x+0.2y+0.6z=z, x+y+z=1$, והפתרון הוא $x=\frac{16}{23}, y=\frac{2}{23}, z=\frac{5}{23}$. לכן שעור הזמן בחדר הגדול הוא $\frac{16+2}{23}\approx 78\%$.

מבוא להסתברות וסטטיסטיקה 2, 88-162

פרופ' ע. וישנה

סמסטר ב', מועד ב', תשס"ט – פתרון מלא

יש לענות על 4 מתוך 5 השאלות.

משך המבחן: שעתיים וחצי (לאחר הארכה).

חומר מותר בשימוש: דפי עזר (מצורפים) וטבלאות התפלגות.

1. בשתי שקיות יש שני קילוגרם סוכר, קילוגרם בכל אחת. מעבירים כמות אקראית (בעלת התפלגות אחידה, $X \sim U[0,1]$) משקית אחת לשניה. נסמן ב־Y את יחס המשקלים בין השקית הקלה לכבדה (כך שתמיד $Y \leq 1$).

(א) מצא את פונקציית הצפיפות $f_Y(y)$ של Y (באיזה טווח היא מוגדרת?). **פתרון**. נסמן $h(x) = \frac{1-x}{1+x}$, אז $Y = h(X)$, ומתקיים $0 \leq Y \leq 1$. עבור פונקציה מיוחדת זו מתקיים גם $X = h(Y)$. לפי נוסחת המעבר, $f_Y(y) = f_X(x) \cdot |h'(x)|^{-1} = 1 \cdot \frac{(1+x)^2}{2} = \frac{2}{(1+y)^2}$.

(ב) מה התוחלת של Y? **פתרון**. $E(Y) = \int_0^1 \frac{2y}{(1+y)^2}dy = 2[\frac{1}{1+y} + \log(1+y)]_0^1 = 2\log(2) - 1 \approx 0.386$

2. ידוע שמשקלם של גורי חתולים בני שבועיים מתפלג נורמלית עם סטיית תקן 10.

(א) בהופעת קרקס הלוליין רוצה לאזן על שתי כפות מאזניים שני גורי חתולים, שנבחרו באקראי, כנגד שניים אחרים. מה התפלגות ההפרש בין משקלי הכפות? **פתרון**. מדובר במשתנה $D = (X_1 + X_2) - (X_3 + X_4)$ כאשר $X_i \sim N(\mu, 10^2)$ בלתי־תלויים, ולכן $E(D) = 2\mu - 2\mu = 0$, $V(D) = V(X_1) + V(X_2) + V(X_3) + V(X_4) = 4 \cdot 10^2$. לכן $D \sim N(0, 20^2)$.

(ב) המשקל הממוצע של חמישה זוגות שהלוליין שקל לפני הופעה הוא 220 גרם. כתוב רווח סמך ברמת מובהקות 95% לתוחלת המשקל של גור בודד. **פתרון**. נסמן ב־Y את משקלו של זוג חתולים. כמקודם, $Y \sim N(2\mu, 200)$. לפי הנתונים $\bar{Y} = 220$ עם $n = 5$ וסטיית תקן ידועה. בנוסף לזה $z_{0.975} = 1.96$. לכן רווח הסמך הוא $220 \pm 1.96 \cdot \frac{\sqrt{200}}{\sqrt{5}} \approx 220 \pm 12.4$, או $207.6 \dots 232.4$.

3. השגיאה במדידת אורך חבל מתפלגת נורמלית, $N(0, \sigma^2)$, עם סטיית תקן לא ידועה. במדגם של n מדידות, התקבלו התוצאות $X_1, \dots, X_n$ שעבורן $X_1 + \dots + X_n = \rho n$ ו־ $X_1^2 + \dots + X_n^2 = \kappa n$.

(א) בדוק את ההשערה $H_0 \colon \sigma = 3$ כנגד השערה דו־צדדית, אם $n = 20$ והערכים שהתקבלו הם $\rho = 0.1$ ו־ $\kappa = 3.6$. מהי רמת המובהקות המינימלית שעבורה דוחים את ההשערה? **פתרון**. במבחן על סטיית התקן, הסטטיסטי $T = \frac{(n-1)s^2}{\sigma_0^2}$ מתפלג χ^2_{n-1}. כאן $s^2 = \frac{1}{n-1}(\sum X_i^2 - \frac{1}{n}(\sum X_i)^2) = (\kappa - \rho^2)\frac{n}{n-1} = 3.59\frac{n}{n-1} \approx 3.79$, ולכן $T \approx 7.978$. בשורה המתאימה ל־19 דרגות חופש, המספר הזה נמצא בין הערכים המתאימים ל־0.01 ו־0.025, ולפי אינטרפולציה פשוטה, ההסתברות לקבל ערך כזה או נמוך ממנו היא בערך 0.015. מכיוון שמדובר בהשערה דו־צדדית, נדחה את ההשערה בכל רמת מובהקות עד 97%, אבל לא מעבר לכך.

(ב) מה יקרה לתשובה לסעיף א' (מבחינה איכותית) אם ההשערה האלטרנטיבית היא $H_1 : \sigma \geq 3$ במקום $H_1 : \sigma \neq 3$? **פתרון**. מכיוון שקיבלנו אומד $s < 3$, לא ניתנו לדחות את ההשערה $\sigma = 3$ כנגד $\sigma \geq 3$ בשום רמת מובהקות.

(ג) ואם יתקבלו אותם נתונים (ρ, κ) עבור $n = 40$ במקום $n = 20$, בהשערה דו־צדדית? **פתרון**. הערך החדש של T קרוב מאד לקודם, אלא שמספר דרגות החופש עלה, והערך נעשה קיצוני עוד יותר. הפעם אפשר לדחות את ההשערה גם ברמת מובהקות גבוהה יותר.

4. למחלקת התמיכה הטכנית של המשטרה מגיעות שיחות בשני נושאים (תקלות תוכנה ותקלות חומרה) מארבעה מחוזות (צפון, דרום, מרכז ושפלה). ביום חמישי האחרון, מספר השיחות בנושא הראשון היה $18, 10, 14, 20$ בהתאמה, ומספר השיחות בנושא השני היה $6, 16, 14, 14$.

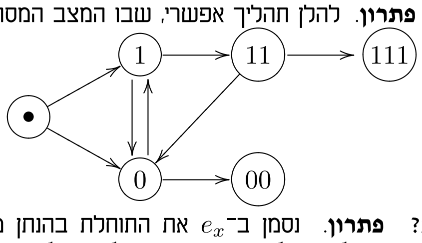
(א) כתוב את ההשערה וההשערה האלטרנטיבית לבדיקת הטענה שבכל המחוזות נתקלים בשני סוגי הבעיות באותה שכיחות. כתוב מהו המשתנה המקרי שאמצעותו אפשר לבדוק את ההשערה, ומה ההתפלגות שלו (לרבות דרגות חופש) אם היא נכונה. **פתרון**. ההשערה H_0 היא 'בכל המחוזות יש אותה הסתברות לתקלת חומרה', או 'אין קשר בין הסיכוי לתקלת חומרה לבין המחוז'. הסטטיסטי המתאים הוא $\sum_{i=1..4,j=1,2} \frac{(E_{ij}-O_{ij})^2}{E_{ij}}$ כאשר O_{ij} הוא הערך הנצפה ו־E_{ij} הערך הצפוי. התפלגותו בהנחת H_0 היא χ^2_3.

(ב) בדוק את ההשערה ברמת מובהקות 90%. **פתרון**. החישוב נותן $S = 7.24$. הערך בטבלה המתאים ל־90% הוא 6.251, ולכן דוחים את ההשערה.

(ג) לו היית מגלה שכל שיחה נרשמה במוקד פעמיים, בטעות, איך היה הדבר משפיע על המסקנות? **פתרון**. תיקון הטעות היה מחייב להחליף כל O_{ij}, ולכן גם כל E_{ij}, במחצית מן הערך הקודם, וכך גם S היה מוכפל בגורם 0.5. הערך החדש, 3.62, אינו מאפשר לדחות את ההשערה.

5. סדרת ניסויים (בלתי תלויים, עם הסתברות $\frac{1}{2}$ להצלחה) נמשכת עד להשגת שתי הצלחות רצופות, או שלושה כשלונות רצופים.

(א) רשום תהליך מרקוב מתאים. **פתרון**. להלן תהליך אפשרי, שבו המצב המסומן ב־$\bullet$ הוא מצב התחלתי, ללא הסטוריה:



(ב) כמה ניסויים ייערכו, בתוחלת? **פתרון**. נסמן ב־e_x את התוחלת בהנתן מצב x. המשוואות הן $e_0 = 1+\frac{1}{2}e_1+\frac{1}{2}e_{00}$, $e_1 = 1+\frac{1}{2}e_0+\frac{1}{2}e_{11}$, $e_{11} = 1+\frac{1}{2}e_0+\frac{1}{2}e_{000}$, כאשר $e_{11} = e_{000} = 0$ מכיוון שאלו מצבי העצירה. פתרון המשוואות הוא $e_1 = \frac{14}{5}$, $e_0 = \frac{18}{5}$, $e_{11} = \frac{12}{5}$. תוחלת מספר הטלות המטבע היא $e_\bullet = 1 + \frac{1}{2}e_0 + \frac{1}{2}e_1 = \frac{21}{5}$.

(ג) מה הסיכוי שהסדרה תופסק בשל הצלחות, ולא בשל כשלונות? **פתרון**. באופן דומה, נסמן ב־p_x את הסיכוי שהתליך יפסק בשל הצלחות (11) ולא בשל כשלונות (000), בהנחה שאנחנו במצב x. המשוואות דומות: $p_0 = \frac{1}{2}p_1 + \frac{1}{2}p_{00}$, $p_1 = \frac{1}{2}p_0 + \frac{1}{2}p_{11}$, $p_{11} = \frac{1}{2}p_0 + \frac{1}{2}p_{000}$, כאשר $p_{11} = 1$, ו־$p_{000} = 0$. הפתרון הוא $p_0 = \frac{3}{5}$, $p_1 = \frac{4}{5}$, $p_{00} = \frac{2}{5}$. לכן גם $p_\bullet = \frac{1}{2}p_0 + \frac{1}{2}p_1 = \frac{7}{10}$.