

פתרון אפשרי למבחן סיום הקורס "הסתברות וסטטיסטיקה":
תכנית רוטשילד ויצמן

1. נתון כי משתנה מקרי X קטן ממש בכל נקודת מדגם ממשנתה אחר Y , וגם נתון כי הערך המכסימלי שהמשתנה X מקבל גדול מהערך המינימלי שהמשתנה Y מקבל. האם יתכן כי X ו- Y בלתי תלויים?

לא יתכן כי המשתנים בלתי תלויים. נימוק בלתי פורמלי הוא שכאשר X מקבל ערך קרוב לערכו המקסימלי אבל גדול מהערך המינימלי של Y , מתקבלת אינפורמציה חדשה על הערך שהמשתנה Y מקבל, לכן יש תלות.
הוכחה מדויקת אפשר לנסח כדלקמן. יהיה a ערך שבין הערך המקסימלי של X והמינימלי של Y , למשל הממוצע של $\max X$ ושל $\min Y$. יהיה A המאורע $\{X \geq a\}$ ויהיה B המאורע $\{Y > a\}$. אז A מאורע חלקי ל- B , בפרט, ההסתברות של החיתוך של A ו- B שווה להסתברות של A , אבל זו שונה, ממש גדולה, ממכפלת ההסתברויות של A ו- B , כאשר הסיבה לאי השוויון היא כי ההסתברות של B קטנה ממש מ-1. כאשר מכפלת ההסתברויות של שני מאורעות שונה מההסתברות של החיתוך, המאורעות בלתי תלויים. משתנים מקריים בלתי תלויים אם כל המאורעות המוגדרים על ידי אי שוויונות על ערכי המשתנים הם בלתי תלויים. הראנו שני מאורעות כאלה שאינם בלתי תלויים לכן המשתנים אינם בלתי תלויים. (הערה: בהוכחה לעיל אי אפשר לבחור את a להיות $\max X$ ולבחור את A להיות $\{X = a\}$ מכיוון שאז A עלולה להיות בעלת הסתברות אפס.)

2. נתונים $X_1, X_2, \dots, X_n$ משתנים מקריים בלתי תלויים ושווי התפלגות. חוק הגבול המרכזי אומר כי ההתפלגות של המשתנה המקרי הנתון על ידי הסכום $X_1 + X_2 + \dots + X_n$ שואפת, כאשר n שואף לאינסוף, להתפלגות של משתנה מקרי נורמלי. הסבר את משמעות המונח "שואפת" בהקשר זה.

משמעות השאיפה היא כי אם מנרמלים את המשתנה המקרי שהוא הסכום $X_1 + X_2 + \dots + X_n$ על ידי החסרת התוחלת של הסכום (כלומר מחסירים n כפול התוחלת של X_1) ואז מחלקים בסטיית התקן של הסכום (כלומר מחלקים בשורש n כפול סטיית התקן של X_1) מקבלים משתנה מקרי שהתפלגותו מקיימת את התכונה הבאה. לכל מספר z על הישר, האינטגרל של ההתפלגות ממינוס אינסוף ועד z מתכנס, כאשר n שואף לאינסוף, לאינטגרל ממינוס אינסוף ועד z של ההתפלגות של המשתנה המקרי הנורמלי עם תוחלת 0 וסטיית תקן 1.

3. במאמר שפרסמו מרדכי הלפרט ובוועז סנג'רו (מחקרי משפט כו, 2010, עמודים 788-734) מצביעים המחברים על טעויות משפטיות חמורות ששופטים עושים כאשר הם מרשיעים נאשם, למשל במעשה רצח, בהסתמך רק על הודאה שלו באשמה. הנימוק שהשופטים משתמשים בו הוא כי לא סביר שאדם יודה במעשה חמור שכזה אם לא ביצע אותו. לדברי כותבי המאמר יש כאן כשל בהבנת תורת ההסתברות. לדבריהם "האמת המפתיעה היא שגם ספק קלוש באמיתות ההודאה עלול לגרום במקרים מסוימים לספק עצום בוודאות ההרשעה". האם אתה מסכים עם טענה זו? האם תוכל לתאר מהם המקרים המסוימים אליהם מתכוונים, כנראה, הכותבים? הסבר.

אני מסכים. הנה דוגמא. נניח כי רק אדם אחד ממיליון עלול, בגלל בעיות נפשיות, להודות ברצח שלא ביצע, ונניח כי רוצח שנתפש מודה ברצח בהסתברות של 95%. ננתח מקרה בו בוצע רצח, החוקרים עוצרים אדם באופן אקראי והוא מודה ברצח. השופט עלול לחשוב שמכיוון שהסיכוי שאדם יודה ברצח שלא עשה הוא אחד למיליון, הרי הספק באמיתות ההודאה נראה קטן ויש סיכוי גדול מאוד שהמודה הוא הרוצח. אבל במדינה לה 7 מיליון תושבים יש (בהסתברות) שבעה שמוכנים להודות ברצח. האדם שהודה הוא או אחד מהשבעה או הרוצח. מכיוון שהאדם נעצר באופן אקראי הסיכויים שהוא הרוצח הם קצת פחות מאשר אחד משמונה (הסיכויים יהיו בדיוק אחד משמונה אם רוצח שנתפש מודה בהסתברות 1), וזה ספק עצום לגבי הרשעה. אכן, זו טעות המופיעה לעתים קרובות כאשר בוחנים מקרים של הסתברות מותנית באופן אינטואיטיבי. מילת המפתח כאן היא "אקראי". אם המודה לא נבחר אקראית, כלומר היה לגביו חשד, או מניע, התמונה משתנה. ככל שיש פחות נסיבות המקשרות את המודה לרצח, כלומר נסיבות

המגדילות, בלי תלות בהודאתו, את הסיכוי כי האדם שנתפש הוא הרוצח, כן יגדל האפקט של הטעות. (הערה: המאמר נמצא ברשת. הוא ארוך, עם דוגמאות, ועם הסברים מפורטים על אותו חלק של תורת ההסתברות הרלבנטי לניתוח, כלומר הסתברות מותנית.)

4. בסל אחד נמצאים חמישה כדורים לבנים וחמישה שחורים. בסל שני יש חמישים כדורים לבנים וחמישים שחורים. אתה רשאי לבחור אחד מהסלים ואז להוציא באקראי שני כדורים. מטרתך להגדיל ככול האפשר את ההסתברות שלפחות אחד מהכדורים שהוצאת יהיה לבן. האם משנה באיזה סל תבחר, ואם כן מהו הסל בו תבחר?

בתור מרחב מדגם המייצג את סיפור המעשה נבחר את מרחב הזוגות הסדורים של כדורים מהסל, בהסתברות שווה לכל זוג, כאשר האינטרפרטציה היא שהכדור הראשון שיצא מיוצג על ידי הראשון בזוג. לגבי הסל הראשון, את המאורע "לפחות כדור אחד לבן" אפשר להציג כאיחוד זר של המאורע "כדור ראשון יצא לבן", שלו הסתברות 0.5, והמאורע "כדור ראשון שחור ושני לבן", ומכיוון שהבחירות בלתי תלויות (כך מניחים) ההסתברות של המאורע האחרון תהיה 0.5 כפול חמש תשיעיות. מכיוון שהמאורעות זרים ההסתברות להוציא כדור לבן היא סכום שני המספרים. אותה חלוקה לגבי הסל השני נותנת הסתברויות של 0.5 ו-0.5 כפול 50 חלקי 99. כמובן שההסתברות להוציא כדור לבן כאשר הסל הוא השני קטנה יותר (5 חלקי 9 גדול מ-50 חלקי 99). אבחר לכן בסל הראשון.

5. בניסוי שנערך בארצות הברית (תוצאותיו התפרסמו ב-1986) ניתנו למחצית מקרב 20,000 אנשים גלולות אספירין מדי יום, ולמחצית השנייה פלסבו. מתוך אלה שבלעו את הגלולות קיבלו 145 התקף לב, ומתוך אלה שבלעו פלסבו קיבלו 209 התקף לב. באיזו מידה מהוות תוצאות אלה עדות ליעילות של אספירין במניעת התקפי לב?

נניח כי אין כל השפעה ללקיחת אספירין. אז בין הלוקים במחלת הלב מבין האנשים שהשתתפו בניסוי האפשרות שלקחו אספירין או פלסבו היא אקראית. כלומר, אפשר לראות את העובדה שהלוקים לוקחים אספירין או פלסבו כבחירות חוזרות בלתי תלויות, אקראיות בסיכוי שווה בין אוכלוסיית הלוקחים אספירין ואלה שלקחו פלסבו. נסמן ב- X_n את המשתנה המקרי הנותן 1 אם הלוקה בלב לקח פלסבו ומקבל ערך 0 אם הלוקה לקח אספירין. אם ההנחות לעיל מתקיימות, ההתפלגות של הסכום, בין 1 ל- n של המשתנים האלה מתכנסת (לפי משפט הגבול המרכזי) למשתנה נורמלי. 354 חזרות זהו מספר מספיק גדול כדי שההתפלגויות שמדובר בהן, כולל סטיות התקן של המשתנים, קרובות מאוד לאלה של המשתנה הנורמלי. נחשב כמה סטיות תקן רחוקה תוצאת הניסיון מהתוחלת, שהיא 177, של סכום המשתנים, ונחשב זאת לגבי המשתנה המקרי המנורמל לתוחלת 0 וסטיות תקן 1. החישוב מתבצע על ידי 177-209 חלקי שורש 354 המוכפל ב-0.5 (0.5 היא סטיית התקן בכל בחירה). התוצאה היא בערך 3.4. במשתנה מקרי נורמלי (0,1) לקבל ערך קרוב ל 3.4 הוא אירוע שסיכוייו קטנים מאלפית (תוצאה גבוהה מסטיית תקן 3 קורית בסיכוי 1.35 כפול אלפית). לכן קשה לי לקבל את ההנחה שאין השפעה לאספירין.

6. מציעים לך הגרלה כדלהלן: נטיל מטבע מאוזנת מספר רב מאוד של פעמים. אם הפרופורציה של מספר הפעמים שהמטבע תיפול על "מספר" רחוקה שתי סטיות תקן מחמישים אחוז, תקבל 10 שקלים. אם רחוקה פחות משתי סטיות תקן תשלם שקל אחד. האם תוחלת הרווח שלך חיובית?

התפלגות מנורמלת של הטלות חוזרות בלתי תלויות של מטבע מאוזנת קרובה (לפי משפט הגבול המרכזי) להתפלגות נורמלית (0,1). כאשר ההתפלגות נורמלית הסיכוי לקבל תוצאה שהיא שתי סטיות תקן מהממוצע הוא פחות מחמישה אחוזים (לפי המספרים שניתנו בכיתה הסיכוי הוא 0.046, גם ניתן בכיתה הנתון כי טעות של 5% מתקבלת בסטיית התקן 1.960 או יותר). לכן תוחלת הרווח מההגרלה המוצעת קרובה למספר שהוא פחות מאשר 10 כפול 0.05, מינוס 1 כפול 0.95. זו תוחלת שלילית בעליל ובפרט אינה חיובית. (אני מפרש כאן מספר רב של פעמים כמספר בו הערכות משפט הגבול המרכזי טובות. מספר זה להטלות מטבע אינו גדול מכמה עשרות.)