סיכומים למבחן בקורס סטטיסטיקה למדעי המחשב

סמסטר א' 2008-9 (ד"ר סהרון רוסט)

תוכן עניינים:

הגדרות בסיסיות בהסתברות בדידה:

תוחלת:

בהינתן X מ״מ מהתפלגות F, ומרחב המדגם מקיים $\Omega \subset \mathbb{R}$, התוחלת: $E(X) = \sum_{x\in\Omega} xF(x)$. תכונות:

א. לכל $a \in \mathbb{R}$: $E(aX) = aE(X)$

ב. אדיטיביות: $E(X_1) + E(X_2) = E(X_1 + X_2)$

ג. עבור $f(X) = aX + b$ פונ׳ לינארית, מתקיים: $E\big(f(X)\big) = aE(X) + b$

שונות:

מידת ה״פיזור״ של X: $var(X) = E\left[\big(X - E(X)\big)^2\right] = E(X^2) - E(X)^2$. תכונות:

א. לכל $a \in \mathbb{R}$: $var(aX) = a^2var(X)$

ב. $var(X_1 + X_2) = var(X_1) + var(X_2) - 2cov(X_1, X_2)$. עבור מ״מ בלתי מתואמים: $cov(X_1, X_2) = 0$, כלומר, מתקיימת אדיטיביות תחת אי תלות. נוסחה ל-cov: $cov(X_1, X_2) = E\big[\big(X_1 - E(X_1)\big)\cdot\big(X_2 - E(X_2)\big)\big]$

מקדם המתאם:

מקדם המתאם (של פירסון): $-1 \le \rho(X_1, X_2) = \frac{cov(X_1,X_2)}{\sqrt{var(X_1)\cdot var(X_2)}} \le 1$.

אם $\rho = 0$ אזי לא קיים קשר לינארי בין שני המ״מ (בלתי לתלויים לינארית). אם $\rho = 1$ – קשורים לינארית, $\rho = -1$ - קשורים לינארית בפונ׳ לינארית הפוכה.

דוגמאות למ״מ בדידים חשובים:

התפלגות	פונ׳ התפלגות	תוחלת $E(X)$	שונות $var(X)$
ברנולי עם פרמטר p (לרוב מרחב מדגם פשוט: $\Omega = \{0,1\}$) $X\sim Bernoulli(p)$ סיפור: הסיכוי להצלחה בהטלת מטבע בודדת	$F(1) = p,$ $F(0) = 1 - p$	p	$p(1-p)$
בינומי עם פרמטרים n,p $X\sim Bin(n,p)$ סיפור: הסיכוי ל-k עצים ב-n הטלות מטבע (p – סיכוי לעץ)	$F(X=k) = \binom{n}{k}p^k(1-p)^{n-k},$ $0 \le k \le n$	np	$np(1-p)$
נשים לב כי אם $X_1, \dots, X_n \sim Bern(p)$ וב״ת, אזי: $\sum_{i=1}^n X_i \sim Bin(n,p)$. מכאן גם ניתן לחשב את התוחלת והשונות לבינומי			
פואסוני עם פרמטר λ $X\sim Pois(\lambda)$ סיפור: הסיכוי להתרחשות k אירועים בפרק זמן קבוע בקצב λ, כאשר מס׳ אירועים בכל פרק זמן אינו תלוי בזמנים אחרים	$F(X=k) = \frac{\lambda^k}{k!}\cdot e^{-\lambda},$ $k \ge 0$	λ	λ
גיאומטרי עם פרמטר p $X\sim G(p)$ סיפור: הסיכוי שהעץ הראשון שנקבל יהיה בהטלה ה-k	$F(X=k) = (1-p)^{k-1}\cdot p,$ $k \ge 1$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
היפרגיאומטרית עם פרמטרים N,D,n $X\sim HG(N,D,n)$ סיפור: מתוך אוכלוסיה בגודל N המכילה D פריטים מיוחדים ממנה שולפים n פריטים, הסיכוי שמתוכם k פריטים יהיו מיוחדים	$F(X=k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$	$\frac{nD}{N}$	$\frac{n\left(\frac{D}{N}\right)\left(\frac{1-D}{N}\right)(N-n)}{N-1}$

נוסחה חשובה להתפלגות מותנית להחלפת סדר ההתניה: $P(X=F \mid Y=A) = P(Y=A \mid X=F)\cdot\frac{P(X=F)}{P(Y=A)}$

פרדוקס סימפסון:

פרדוקס סטטיסטי בו הצלחת קבוצת מחקר מסויימת נראית שונה כאשר מאחדים בין קבוצות.
דוגמא שניתנה בשיעור: אחוז הקבלה של בנות לאוני' נמוך מאחוז הקבלה של בנים, אך אם מסתכלים על כל פקולטה בנפרד מקבלים שבכל פקולטה אחוז הקבלה של הבנות גדול מאחוז הקבלה של הבנים. לכאורה מקבלים כאן פרדוקס, אך ניתן לראות שהנתונים מאפשרים זאת באמצעות משחק עם הנוסחאות (בייס, ההסתברות השלמה).

ניתוחים ויזואליים של משתני מדידה מספריים:

היסטוגרמה: מהתבוננות בהיסטוגרמה ניתן לדלות מידע מתוך:

- מספר השיאים
- סימטריה / זנב ימינה או שמאלה
- ערכים קיצוניים המעידים על תצפיות מיוחדות או טעויות מדידה.

היסטוגרמה נעה: שמים את רוחב כל תא על כל נקודה ממשית בציר ה-x ולאותה נקודה נותנים כערך את מספר התצפיות ברוחב שמדדנו.

אם נקח רוחב תא קטן מדי, יהיו לנו הרבה קפיצות, ואם נקח רוחב תא גדול מדי, נאבד אינפורמציה.

טרנספורמציות של x: שינוי סקאלה כדי שתצוגת הנתונים תהיה אינפורמטיבית, כגון טרנס' log.

תמצית מספרית של נתונים מספריים:

הגדרות:

עבור אוסף תצפיות $x=(x_1,\dots,x_n)$, נגדיר:

תמצית אמפירית (*summary statistic*): של אוסף נתונים x הוא פונקציה $\mathbb{R}^n\to\mathbb{R}$ שאינה תלויה בסדר הנתונים.

סטטיסטי הסדר (*order statistic*): של אוסף נתונים x הוא אותו אוסף הערכים, אך מסודר בסדר עולה (וקטור ערכים ממויין).

שברונים ואחוזונים של התפלגות אמפירית (*Percentiles, Quantiles*):

השברון ה-q כאשר $0\le q\le 1$ עבור ההתפלגות האמפירית הוא: $X_q=X_{qn+0.5}$, והוא האחוזון ה-$100q$. למשל: $q=0.5$ הוא השברון ה-0.5, שהוא האחוזון ה-50%, כלומר החציון במקרה זה. אם $qn+0.5$ אינו שלם, נקח את הערך הנופל בין $\lfloor qn+0.5\rfloor$ ל-$\lceil qn+0.5\rceil$ באופן פרופורציונאלי למרחק ביניהם (אינטרפולציה לינארית):

$$X_q=[\lceil qn+0.5\rceil-(qn+0.5)]\cdot X_{\lfloor qn+0.5\rfloor}+[(qn+0.5)-\lfloor qn+0.5\rfloor]\cdot X_{\lceil qn+0.5\rceil}$$

נשים לב כי כל מרחק מכפילים בשברון הנגדי לו, כיוון שהמרחק הקצר = השברון שקרוב אליו משפיע יותר על X_q.

הערה: $q<\frac{1}{2n}$ או $\frac{2n-1}{2n}<q \Leftarrow X_q$ אינו מוגדר.

קריטריונים לבחירת תמציות:

(1) נקודת השבירה:

אחוז התצפיות שצריך "לקלקל" בכדי לגרום לשינוי לא חסום בתמצית. דוגמאות:

- ממוצע: נק' שבירה $\frac{1}{n}$, כי מספיק לשלוח תצפית אחת לאינסוף כדי לקלקל את הממוצע – שיהיה גם הוא אינסוף.
- חציון: בערך $\frac{1}{2}$, למשל במדגם אי זוגי נוכל לקלקל את עד $\lfloor\frac{n}{2}\rfloor$ תצפיות, והחציון עדיין בטווח של האוסף המקורי (שהוא $X_{(1)}$ עד $X_{(n)}$).

נקודת השבירה האסימפטוטית: גבול נקודת השבירה כאשר $n\to\infty$. למשל: עבור ממוצע: 0; חציון: $\frac{1}{2}$; ממוצע קצוץ α: α.

(2) אינווריאנטיות (אי תלות בסקאלה):

אינווי' לטרנס' מונוטונית היא שמתקיים: $f\big(T(X)\big)=T\big(f(X)\big)$, עבור $f:\mathbb{R}\to\mathbb{R}$ פונ' מונוטונית. למשל: ממוצע: לא; חציון: כן; קצוץ α: לא.

תמציות למיקום:

באופן אבסטרקטי – "אמצע" הנתונים שלנו (כמו ממוצע או חציון). פורמלית, פונקציה $T:\mathbb{R}^n\to\mathbb{R}$ המקיימת:

- $T(X+a)=T(X)+a$
- $T(cX)=xT(X)$

דוגמאות לתמציות למיקום (מקיימות את שתי התכונות):

- ממוצע: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- ממוצע משוקלל: $\bar{X}_w = \sum_{i=1}^{n} w_i X_i$, כאשר: $\sum_{i=1}^{n} w_i = 1$.
- חציון: $X_{0.5}$ - הסדרון היחיד המקיים את התכונה השניה **לכל** c. כל שאר הסדרונים מקיימים רק עבור $c > 0$.
- ממוצע קצוץ α: $T_\alpha(X) = \frac{1}{n-2\lfloor n\alpha \rfloor} \sum_{i=\lfloor n\alpha \rfloor+1}^{n-\lfloor n\alpha \rfloor} X_{(i)}$ (הורדת הקצוות; $X_{(i)}$ הוא השברון ה-i עבור סטטיסטי הסדר, כלומר X ממויין).

תמציות לפיזור:

פונקציה $T: \mathbb{R}^n \to \mathbb{R}$ המקיימת:

- $T(X+a) = T(X)$
- $T(cX) = |c| \cdot T(X)$, $c \geq 0$, מכאן ש: $T(X) = T(-X)$
- $T(X) \geq 0$

דוגמאות לתמציות לפיזור:

- טווח: $X_{(n)} - X_{(1)}$, נקי שבירה $\frac{1}{n}$ אסימפי 0 (מספיק לשנות תצפית אחת ל-∞).
- סטיית התקן האמפירית: $\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$, נקי שבירה $\frac{1}{n}$ אסימפטוטית 0. השונות האמפירית לא מקיימת את התכונה השניה, כיוון שהכפלה ב-c גוררת הכפלה ב-c^2.
- טווח בין רביעונים: $X_{0.75} - X_{0.25}$, נקי שבירה בערך $\frac{1}{4}$, אסימפי $\frac{1}{4}$.
- MAD (המדד המשוגע): $MAD(X) = median\{|X_i - X_{0.5}|\}$, כלומר: חציון כל הערכים המוחלטים של ההפרשים מהחציון. נקי שבירה: $\frac{1}{2}$ אסימפי $\frac{1}{2}$. טענה: כל תמצית מהצורה: $T_1\{|X_i - T_2(X)|\}$, T_1, T_2 תמציות למיקום, היא תמצית לפיזור (לפי התכונות).

תמציות לנטייה:

נגדיר תחילה פוני נטייה Sk: $Sk(i) = \frac{X_{(i)} + X_{(n-i+1)}}{2} - X_{0.5}$ - לוקחים ממוצע שני סדרונים הנמצאים במרחק שווה מהאמצע, ומחסירים החציון.

נטייה ימינה (זנב ימינה): Sk פוני יורדת של i.

פונקציה $T: \mathbb{R}^n \to \mathbb{R}$ המקיימת:

- $T(X+a) = T(X)$
- $T(cX) = T(X)$, $c \geq o$
- $T(-X) = -T(X)$, ומצירוף תכונות 1 ו-2 נובע שאם המדגם סימטרי אז $T(x) = 0$ (אין נטייה).
- $Sk(i)$ יורד ב-i (נטייה ימינה) $\Leftarrow$ $T(X) > 0$.

דוגמאות לתמציות לנטייה:

- מדד $YULE$ לנטייה: $YULE(X) = \frac{(X_{0.75} + X_{0.25})/2 - X_{0.5}}{(X_{0.75} + X_{0.25})/2}$
- מדד $Pearson$: $Pear(X) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^3}{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{\frac{3}{2}}}$

מטרות ההצגה הגרפית:

1. זיהוי אשכולות (*clustering*) – קבוצות המתנהגות באופן דומה לפי מדדים שונים.
2. זיהוי וחקר תצפיות חריגות.
3. תיאור קשר בין משתנים.
4. הבנת ההתפלגות המותנה: $P(Y|X)$.

מטרה חשובה של ניתוח נתונים היא מציאת סיבתיות, אך גם בניתוח גרפי ומתמטי יתכן ונמצא קורילציה, אך לא בטוח שהיא סיבתית.

<u>מידול מתמטי של קשרים בין משתנים – רגרסיה לינארית</u>:

עבור משתנה מסביר X ומשתנה מוסבר (תלוי) Y, אנו רוצים לתאר במספר את הקשר הלינארי ביניהם (2 מספרים).

שיטת הקו החסין ($Resistant\ line$):

נתון אוסף זוגות מספרים: $\{(x_1, y_1), \dots, (x_n, y_n)\}$.

- נחלק את ציר ה-x ל-3 חלקים שווים לפי $X_{\frac{1}{3}}, X_{\frac{2}{3}}$. נסמן את הקבוצות: low, mid, hi
- ניקח חציוני y ל-low, hi: $(x_l, y_l), (x_h, y_h)$.
- נגדיר שיפוע: $b_{RL} = \frac{y_h - y_l}{x_h - x_l}$
- אנו רוצים מחצית מהנקודות מעל הקו ומחצית מתחת, נסמן: $r_i = y_i - b_{RL}x_i$. נגדיר את החותך: $a_{RL} = median\{r_i\}$

עבור הקו החסין שהתקבל $y = a_{RL} + b_{RL}x$ מתקיים ש-$y_i - (a_{RL} + b_{RL}x_i) \geq 0$ עבור בדיוק מחצית מהתצפיות, כנדרש.

נקודת השבירה היא $\frac{1}{6}$ - קלקול עד חצי (בערך) מתוך $\frac{1}{3}$ מהנתונים.

קו הריבועים הפחותים ($Least\text{-}squares\ line$):

הרעיון של בניית הקו: מגדירים קו כללי כלשהו ע"י a,b (חיתוך, שיפוע), ולפיו מגדירים r_i – הסטיה של הנקודה (x_i, y_i) מהקו. נגדיר את RSS ($Residual\ sum\ of\ squares$) להיות סכום ריבועי הסטיות הללו. קו הריבועים הפחותים יהיה a,b שיביאו את RSS למינימום. נקבל את הקו $y = a_{LS} + b_{LS}x$ כאשר $a_{LS} = \bar{y} - b_{LS}\bar{x}$, $b_{LS} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{cov_{amp}(x,y)}{var_{amp}(x)}$.

נקבל כי נקודת הממוצעים $(\bar{x}, \bar{y})$ תמיד נמצאת על קו זה.

מדידת טיב התאמת קו הריבועים הפחותים לנתונים:

מקדם המתאם של פירסון (r^2):

$r_{xy} = \rho = \frac{cov_{amp}(x,y)}{\sqrt{var(x) \cdot var(y)}}$. מכאן נקבל שעבור b_{LS} מתקיים: $b_{LS} = r_{xy} \cdot \frac{sd(y)}{sd(x)}$. תכונות r_{xy}:

- $r_{xy} = r_{yx}$
- אם $f, b > 0$ טרנס' לינארית, אזי: $r(x, f(y)) = r(f(x), y) = r(x, y)$, ומכאן שאין קשר לסקאלה של הצירים. אם $b < 0$ אז נקבל שזה שווה ל-$-r(x, y)$.
- $r(x, a + bx) = \begin{cases} 1, & b > 0 \\ -1, & b < 0 \\ 0, & b = 0 \end{cases}$. אם הקו עובר בכל הנקודות, נקבל $|1|$.
- $-1 \leq r_{xy} \leq 1$

משמעות r^2:

r^2 הוא: אחוז השונות המוסברת, כלומר, איזה אחוז מהשונות שהתחלנו ממנה ($var(y)$) מוסבר ע"י קו הריבועים הפחותים. כלומר, אם נקח $b = 0$ (ואז $a = \bar{y}$) נקבל שהשונות של y הוא RSS עבור ערכים אלו של a,b.

הסבר: אם ניקח שיפוע 0, ברור ש-$r^2 = 0$. כמו כן נקבל קו אופקי, שלא מסביר כלום מהשונות של y, כלומר לא "מפחית" מהשונות שהיתה קודם. לכן ה-RSS שנקבל הוא $var(y)$. כלומר, 0 אחוז מהשונות הוסבר ע"י הקו, ולכן סכום ריבועי המרחקים מהקו הוא השונות שהיתה ונשארה.

בקיצור: פיזור גדול סביב הקו = אחוז מוסבר קטן = r^2 קטן.

תכונות וחולשות r^2:

- זהו מדד לאיכות המודל הלינארי. רצוי להשוות לפי מדד זה רק בין מודלים שהותאמו על אותם נתונים.
- מדד זה רגיש לגודל הרעש – תכונה רצויה, אך גם רגיש לכמות הנתונים – תכונה לא רצויה (עם 2 נק' בלבד נקבל $r^2 = 1$ תמיד).

ניתוח שאריות וטרנס' ללינאריות:

נסמן את השאריות $\delta_i = y_i - \hat{y_i}$, כאשר $\hat{y_i} = a_{LS} + b_{LS}x_i$.

אם ישנו סדר או מגמה בשאריות, אזי הקשר בין x ל-y כנראה **לא** לינארי. נניח f היא פונ' לא לינארית המתארת את הקשר בין x ל-y, אזי עבור $z_i := f(x_i)$ נוכל להתאים קו ישר $\hat{y_i} = a_{LS} + b_{LS}z_i$, עם $a_{LS} = 0, b_{LS} = 1$, ונקבל התאמה מושלמת.

כללי אצבע בהסתכלות על שאריות:

- אם לשארית צורה קמורה עם בטן כלפי מטה, אז:
  - אם $b_{LS} > 0$: יש לנסות את הטרנס' x^p עבור $p > 1$ או y^q עבור $q < 1$.
  - אם $b_{LS} < 0$: יש לנסות את הטרנס' x^p עבור $p < -1$ או y^q עבור $-1 < q < 0$.
- אם לשארית צורה קעורה (בטן מעלה): x^p עבור $-1 < p < 1$, או y^q עבור $q < -1, 1 < q$.

רגרסיה רב משתנית (אותה גברת בשינוי לאדרת ממש מכוערת ומגעילה):

כעת מתקיים כי $x_i \in \mathbb{R}^p$ עבור $p \geq 1$ טבעי כלשהו – p משתנים מסבירים ל-y. נכתוב וקטור עמודה $x_i = (x_{i1}, \dots, x_{ip})$ ונרצה להתאים את המודל $\hat{y_i} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ ($\beta = (\beta_0, \dots, \beta_p)$ הם הפרמטרים שנרצה להתאים).

קריטריון RSS: $RSS(\beta) = \sum_{i=1}^{n}(y - \hat{y})^2$ (כמו קודם, רק עם ה-$\hat{y}$ החדש). לפי כתיב מטרציוני נקבל שזו נורמה ריבועית אוקלידית של וקטור: $RSS = ||y - x\beta||^2$, כאשר כאן: $y = (y_1, \dots, y_n)^t$ – וקטור שורה של כל ה-y_i; $x \in Mat_{n\times(p+1)}$ - מטריצה שכל שורותיה הן x_i כמוגדר לעיל.

נרצה גם כאן למצוא β שיביא את RSS למינימום, ונקבל: $\hat{\beta} = (x^t x)^{-1} \cdot x^t \cdot y$.

משתנים מקריים רציפים:

פונקצית הסתברות מצטברת ופונקצית צפיפות:

תהי F פונ' הסתברות מצטברת עם התכונות:

- $F(a) = 0, F(b) = 1$ עבור a, b כלשהם על ציר ה-x.
- F לא יורדת.

נאמר שמ"מ X מתפלג ע"פ F: $X \sim F$ אם לכל $x \in \Omega$ מתקיים: $P(X \leq x) = F(x)$.

פונ' הצפיפות $f(x)$ תהיה: $f(x) = \frac{\partial F(x)}{\partial(x)}$. תכונות הצפיפות:

- $f(x) \geq 0$ לכל $x \in \Omega$
- $\int_a^b f(x)dx = 1$

דוגמאות:

- מ"מ אחיד רציף סטנדרטי: $X \sim U(0,1)$, $\Omega = [0,1]$, מתקיים: $f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & o/w \end{cases}$, $F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & 1 < x \end{cases}$.
- מ"מ נורמלי סטנדרטי: $X \sim N(0,1)$, $\Omega = \mathbb{R}$, $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $F_X(c) = \Phi(c) = \int_{-\infty}^{c} f_X(x)dx$.
- מ"מ נורמלי כללי: $X \sim N(\mu, \sigma^2)$: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, $F_X(c) = \Phi\left(\frac{c-\mu}{\sigma}\right)$.

התפלגות משותפת מותנה של מספר מ"מ רציפים:

יהיו X, Y מ"מ רציפים. נגדיר:

- התפלגות משותפת: $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$
- התפלגות מותנה: $F_{Y|X}(y|x) = P(Y \leq y | X \leq x) = P(Y \leq y, X \leq x)/F_X(x)$
- צפיפות משותפת: $f_{X,Y}(x,y) = \partial^2 F_{X,Y}(x,y)/\partial x \partial y$

אם X, Y ב"ת אזי: $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, $F_{Y|X}(y|x) = F_Y(y)$, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

תוחלת ושונות מ"מ רציפים:

יהי $X\sim F_X$ מ"מ רציף עם צפיפות f_X, אזי:

- תוחלת: $E(X)=\int_\Omega xf_X(x)dx$
- שונות: $var(X)=\int_\Omega (X-EX)^2 f_X(x)dx$ (חלק חשוב בחישוב: $E\big(g(X)\big)=\int_\Omega g(x)f_X(x)dx$)

דוגמאות:

- $X\sim U(0,1)$: $E(X)=\frac{1}{2}, var(X)=\frac{1}{12}$.
- $X\sim U(a,b)$: $E(X)=\frac{a+b}{2}, var(X)=\frac{(b-a)^2}{12}$
- $X\sim N(0,1)$: $E(X)=0, var(X)=1$ (מן הסתם). $X\sim N(\mu,\sigma^2)$: ברור.

אחוזונים ושברונים של התפלגויות רציפות:

נתונה פונ' התפלגות מצטברת F_X. השברון ה-q הוא: $X_q\ s.t.\ F_X(X_q)=q$, כלומר $X_q=F_X^{-1}(q)$. דוגמאות:

- $X\sim U(0,1)$: $X_q=q$.
- $X\sim U(a,b)$: $X_q=a+q(b-a)$.
- $X\sim N(0,1)$: $Z_q\coloneqq \Phi^{-1}(q)$, כלומר $\Phi(Z_q)=q$, ערכים לפי טבלת ההתפלגות Z.
- $X\sim N(\mu,\sigma^2)$: $P(X\le c)=\Phi\left(\frac{X-\mu}{\sigma}\right)\Rightarrow q=\Phi\left(\frac{X-\mu}{\sigma}\right)\Rightarrow X_q=\sigma Z_q+\mu$.

<u>משפט הגבול המרכזי לקירוב נורמלי</u>:

משפט הגבול המרכזי (*Central limit theorem*):

יהיו $X_1,\dots,X_n$ מ"מ ממשיים (לאו דווקא רציפים), ב"ת ושוו התפלגות בעלי תוחלת μ ושונות σ^2, אזי: $\frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}\xrightarrow[n\to\infty]{p}N(0,1)$, כלומר:

עבור n גדול מספיק ניתן להגיד כי $\frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}\dot\sim N(0,1)$ או בקיצור: $\bar X\dot\sim N\left(\mu,\frac{\sigma^2}{n}\right)$. דוגמאות:

- $X_i\sim Bernoulli(p)$, $\sum_{i=1}^n X_i\sim Bin(n,p)$. לפי משפט הגבול המרכזי נקבל את הקירובים:
  $\bar X\dot\sim N\left(p,\frac{p(1-p)}{n}\right)$ ומכאן שגם $\frac{1}{n}Bin(n,p)\dot\sim N\left(p,\frac{p(1-p)}{n}\right)$ (כי $\bar X=\frac{\sum X}{n}$).
- $X_i\sim U(0,1)$: $\bar X\dot\sim N\left(\frac{1}{2},\frac{1}{12n}\right)$.

מתי הקירוב הנורמלי מתאים:

- הערך של μ לא משנה לאיכות הקירוב ואילו של σ^2 חשוב.
- עבור בינומי מקובל להשתמש אם $\mu=np\ge 10$ וגם תוחלת הכשלונות $n(1-p)\ge 10$.
- מקובל $n\ge 30$ חסם תחתון לכלל ההתפלגויות הרציפות, אך אין כללי אצבע ברורים.

הערות:

- חוק המספרים הגדולים: $(\overline{X_n}-\mu)\xrightarrow[n\to\infty]{p}0$ (נובע מאי שוויון צ'בישב).
- ממוצע של מ"מ המתפלגים $N(\mu,\sigma^2)$ לא מתפלג בקירוב $N(\mu,\frac{\sigma^2}{n})$, אלא בדיוק כך.

<u>מבוא להסקה סטטיסטית</u>:

מושגי יסוד באמידה:

- **אוכלוסיה**: אוסף סופי / אינסופי של פריטים בעלי תכונה / תכונות המעניינות אותנו.
- **התפלגות האוכלוסיה**: תיאור הסתברותי של הערך X של תכונה המעניינת אותנו בפרט מקרי באוכלוסייה. למשל:
  - X קטגוריאלי בעל שני ערכים (כמו מגדר): $X\sim Bernoulli(p)$, יהיה 1 בהסתברות p ו-0 בהסתברות $1-p$.
  - $X\sim N(\mu,\sigma^2)$: תיאור יכול להיות מדוייק רק עבור אוכ' אינסופית לא בת מניה. לרוב יהווה קירוב מספק. במקרה של אוכלוסיה סופית, ההתפלגות האמיתית של X היא פשוט ההתפלגות האמפירית של האוכלוסיה.
- **פרמטר**: תמצית של התכונה המעניינת אותנו באוכלוסיה, למשל תוחלת. נסמן פרמטר המעניין אותנו באות המתאימה, או באופן גנרי: θ.
- **מדגם מקרי** (*random sample*): בגודל n הוא אוסף n פרטים הנדגמים בסיכוי שווה <u>עם החזרה</u> מהאוכלוסיה. המדגם הוא מ"מ מקרי n מימדי: $X=(X_1,\dots,X_n)$ וערכיו הם גם וקטור n מימדי: $x=(x_1,\dots,x_n)$.
- **סטטיסטי** (*statistic*): תמצית מספרית של המדגם. נסמנו ב-$S(X)$, והוא גם מ"מ. את ערכו במדגם נסמן $S(x)$.
- **אומד לפרמטר** (*estimator*): סטטיסטי שמכוון להעריך את אורכו של הפרמטר באוכלוסיה. למשל: ממוצע המדגם המקרי כאומד ל-μ.

תכונות אומדים ואיכותם:

הטיה (*Bias*): עד כמה ה"מיקום" של האומד הוא נכון.

אומד S חסר הטיה לפרמטר θ אם: $\forall\theta: E_\theta\big(S(X)\big)=\theta$, כלומר: תוחלת האומד לפרמטר היא הפרמטר עצמו. אם האומד אינו חסר הטיה, נגדיר את ההטיה שלו (*Bias*), שהיא: $B(\theta)=\theta-E_\theta\big(S(X)\big)$. דוגמא: $\hat{p}=\frac{\sum X_i}{n}$, האם $\hat{p}$ חסר הטיה: $E(\hat{p})=\frac{\sum E(X_i)}{n}=\frac{np}{n}=p \Leftarrow$ אכן חסר הטיה.

הערה: ניתן להשתמש גם במדדים אחרים למיקום לחישוב הטיה.

מדד פיזור:

עד כמה ניתן "לסמוך" על הערך של $S(X)$ במדגם. נהוג להשתמש ב-$var_\theta\big(S(X)\big)$ או ב-$\sigma_\theta\big(S(X)\big)$. במקרה ברנולי: $var(\hat{p})=\frac{p(1-p)}{n}$.

שילובים של הטיה ופיזור בקביעת איכות האומד:

המדד המקובל: $MSE_\theta\big(S(X)\big)=E_\theta\big(S(X)-\theta\big)^2$

- אם נשתמש באומד חסר הטיה נקבל כי $MSE_\theta\big(S(X)\big)=var\big(S(X)\big)$.
- אם האומד מוטה נקבל כי $MSE_\theta\big(S(X)\big)=var\big(S(X)\big)+B^2(\theta)$.

הערות:

- שימוש להשוואת אומדים: יהיו S,T אומדים; נאמר כי <u>S שולט על T</u> אם $\forall\theta: MSE_\theta\big(S(X)\big)\le MSE_\theta\big(T(X)\big)$. לדוגמא, עבור מקרה ברנולי נהוג לאמוד את p ב-$\hat{p}=\bar{X}$, אך אומד לפלס: $L(X)=\frac{\sum X_i+1}{n+2}$ הוא אומד לרוב מוטה אך עם MLE נמוך יותר (חסר הטיה ב-$p=\frac{1}{2}$).
- רווחי סמך ברמת סמך (ביטחון) α: ערכים $\varepsilon_1,\varepsilon_2$ כך שברמת ביטחון $1-\alpha$ מתקיים: $\theta\in[S(X)-\varepsilon_1, S(X)+\varepsilon_2]$.

תכונה כללית של ממוצע המדגם $\bar{X}$ למקרים נורמלי וברנולי:

- $E(\bar{X})=E(X)$ - חסר הטייה.
- $var(\bar{X})=\frac{var(X)}{n}$.

אומד חסר הטיה לשונות במקרה הנורמלי: $\widehat{\sigma^2}=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$.

עקיבות:

יהי $T_n(X_n)$ אומד מוצע ל-θ. עקיבות: $T_n(X_n)\xrightarrow[n\to\infty]{p}\theta$, כלומר: עם מספיק נתונים, האומד שלנו יהיה נכון.

הגדרת התכנסות בהסתברות:

נאמר כי $T_n(X_n)\xrightarrow[n\to\infty]{p}\theta$ אמ"מ: $P(|T_n(X_n)-\theta|>\varepsilon)\xrightarrow[n\to\infty]{}0, \forall\varepsilon$. למשל, $\bar{X}$ כאומד ל-μ הוא תמיד עקיב (בנוסף להיותו תמיד חסר הטיה). כמו כן $\hat{p}$ במקרה הבינומי. דוגמא פחות טרוויאלית: $T_n(X_n)=max\,\{X_i\}$ הוא עומד עקיב במקרה האחיד של $X_i\sim U(0,\theta)$.

- עקיבות **אינה** גוררת חוסר הטיה.
- כמו כן חוסר הטיה **אינה** גוררת עקיבות. למשל בהתפלגות נורמלית נקח את $T_n(X_n)=X_1$. $E(X_1)=\mu$ אך הוא לא עקיב.

גישות ליצירת אומדים:

פונקצית הנראות ($Likelihood$):

נניח $X_i \sim f_\theta$ (או P_θ במקרה הבדיד; נתייחס באופן כללי ל-f_θ פונקצית צפיפות / הסתברות), ו-$x = (x_1, \dots, x_n)$ מדגם מקרי. אזי:

$L(\theta; x) = f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n)$ - היא פונקצית הנראות. לרוב נשתמש בטרנס' log של הנראות: $\ell(\theta; x) = log\big(L(\theta; x)\big)$. דוגמאות:

- ברנולי: $P_p(X) = \begin{cases} p, & X = 1 \\ 1-p, & X = 0 \end{cases}$, ולכן: $L_p = p^{\Sigma x} \cdot (1-p)^{n-\Sigma x_i}$, $\ell_p = \sum x_i \cdot log(p) + (n - \sum x_i)\, log(1-p)$.
- נורמלי עם שונות ידועה: $L_\mu = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$, $\ell_\mu = -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$

אומד נראות מקסימלי ($\hat{\theta}_{MLE}$) ותכונותיו:

הערך של θ שיביא למקסימום את (לוג) פונקצית הנראות: $\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{argmax} L(\theta)$. כדי למצוא אנ"מ נגזור את $\ell(\theta; x)$ (או את $L(\theta; x)$) כפונקציה של θ, ונשווה לאפס למציאת קיצון, וע"י נגזרת שניה נוודא שזהו אכן מקסימום (נגזרת שניה קטנה מאפס). דוגמאות:

- בינומי: מתברר שאכן $\hat{p} = \frac{\sum X_i}{n}$ הוא $\hat{p}_{MLE}$.
- נורמלי (שונות ידועה): $\bar{X} = \frac{\sum X_i}{n} = \hat{\mu}_{MLE}$ גם כן.
- המקרה האחיד: $L_\theta = \begin{cases} \left(\frac{1}{\theta}\right)^n, & 0 \le x_1, \dots, x_n \le \theta \\ 0, & o/w \end{cases}$, ולכן נרצה את הערך החוקי המינימלי של θ שיקיים $x_1, \dots, x_n \le \theta$. זה יתקיים בוודאות עבור $max\,\{x_i\}$, ולכן $\hat{\theta}_{MLE} = max\,\{x_i\}$.

תכונות אנ"מ:

- <u>עקיבות</u>: אם $X \sim F_\theta$ אז $\hat{\theta}_{MLE} \xrightarrow[n\to\infty]{p} 0$.
- כל מיני תכונות אופטימליות אסימפטוטית.
- <u>אינווריאנטיות פונקציונלית</u>: עבור g פונקציה אחד לאחד מתקיים: $\widehat{g(\theta)}_{MLE} = g(\hat{\theta}_{MLE})$. למשל בנורמלי: $\widehat{\mu^2}_{MLE} = \bar{X}^2$.

רווחי סמך:

רווחי סמך ומשמעותם:

נתון $X \sim F_\theta$, פרמטר θ ואומד $T(X)$. רווח סמך ברמת סמך $1 - \alpha$ מוגדר ע"י הערכים $\varepsilon_1, \varepsilon_2 > 0$ כך ש:

$$P_\theta(\theta \in [T(X) - \varepsilon_1, T(X) + \varepsilon_2]) \ge 1 - \alpha \Leftrightarrow P_\theta(T(X) \in [\theta - \varepsilon_2, \theta + \varepsilon_1]) \ge 1 - \alpha$$

בד"כ ילקח $\alpha = 0.05$, אז נאמר שיש לנו רמת סמך של 95% ($1 - \alpha$).

דוגמא: במקרה הנורמלי, $\bar{X}$ אומד ל-μ, נדרוש $\varepsilon_1 = \varepsilon_2$ ונקבל: $C = \left[\bar{X} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$.

משמעות רווח הסמך:

- האם $\mu \in C$ בוודאות? **ודאי שלא בוודאות**.
- האם $\mu \in C$ ברמת ביטחון של $1 - \alpha$? **ודאי שכן**, מהגדרה.
- האם זה אומר שלכל מדגם מקרי עתידי בגודל n נקבל סיכוי של $1 - \alpha$ ש-$\mu \in C$? **לא**, אך נוכל לבדוק האם הסיכוי גבוה מ-$1 - \alpha$ או נמוך ממנו: יהיו $X_1, X_2 \sim N(\mu, \sigma^2)$, אזי $\bar{X}_1, \bar{X}_2 \sim N(\mu, \frac{\sigma^2}{n})$ הם מ"מ ב"ת, מכאן ש: $\bar{X}_1 - \bar{X}_2 \sim N(0, \frac{2\sigma^2}{n})$. כעת נוכל לחשב: $P(\bar{X}_2 \in C) = P\left(\frac{(\bar{X}_1 - \bar{X}_2)\sqrt{n}}{\sqrt{2}\sigma} \in \left[-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{2}}, \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{2}}\right]\right) = 2\Phi\left(\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{2}}\right) - 1$. את התשובה לזה נשווה לרמת הסמך $1 - \alpha$ של $\bar{X}_1$, ונדע מה הסיכוי.

רווח סמך לתוחלת נורמלית עם שונות לא ידועה:

נחליף את השימוש בסטיית התקן σ עם טעות התקן $\hat{\sigma}$, כאשר $\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. במקרה זה נשתמש ב**התפלגות t של סטודנט**, ונקבל:

$\frac{(\bar{X} - \mu)}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$ - כלומר עם $n - 1$ דרגות חופש. בחישוב רווח סמך עם שונות לא ידועה נחליף את $Z_{1-\frac{\alpha}{2}}$ ב- $t_{n-1,1-\frac{\alpha}{2}}$. כאשר $n \to \infty$ אז $t_{n-1} \to Z$.

רווחי סמך לפרמטר בינומי:

$X_i \sim Bernoulli(p)$, רוצים למצוא רווח סמך ל-p כאשר נשתמש באומד $\hat{p}$. החישוב המדויק ארוך ומסורבל, לכן יש שתי גישות:

- קירוב נורמלי: $C = \left[\hat{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + \cdots\right] \Leftarrow$ משמעות: ברמת ביטחון של **בערך** (בגלל הקירוב ושימוש באומד לשונות) $1-\alpha$ נצפה ש-p יפול בר״ס.
- שימוש באומד השמרני: פתרון לאי הדיוק הנובע משימוש באומד לשונות: נחליף את השימוש ב-$\hat{p}$ תחת השורש במקסימום האפשרי שלו: $\hat{p}(1-\hat{p}) \le 0.25$. נקבל ר״ס דומה, רק שנציב 0.25 במקום $\hat{p}(1-\hat{p}) \Leftarrow$ משמעות: ברמת ביטחון של **לפחות** $1-\alpha$...

אינווריאנטיות רווחי סמך תחת טרנס׳ מונוטונית:

אם C הוא ר״ס עבור θ ברמת ביטחון של $1-\alpha$ ו-f היא פונ׳ מונוטונית, אזי $f(C)$ הוא ר״ס ברמת ביטחון $1-\alpha$ עבור $f(\theta)$.
נקח את הדוגמא הבודקת את אחוז הסטודנטים באוכ׳ שגובהם מתחת ל-170 ס״מ. בדרך הרגילה נשתמש בר״ס בינומי על סמך השיעור האמפירי במדגם, אך אז אנו מתעלמים מהנתון שההתפלגות האמיתית של גובהי הסטודנטים היא נורמלית, ולכן נקבל ר״ס גדול יותר מאשר טרנס׳ מונוטונית של ר״ס של גובהי הסטודנטים.

<u>בדיקת השערות</u>:

מושגי יסוד:

- **השערת האפס** (H_0): ברירת המחדל, העובדה שאנו רוצים להפריך באמצעות הנתונים, וזאת רק אם נסיק בצורה ברורה שאינה נכונה.
- **אלטרנטיבה** (H_A / H_1): ברור.
- **מבחן סטטיסטי ברמה** α: נניח $H_0: \theta = \theta_0$. מבחן סטטיסטי ברמה α יהיה אזור הדחיה C_α כך ש-$P_{H_0}(S(X) \in C_\alpha) \le \alpha$. המבחן: אם $S(X) \in C_\alpha$, נדחה את השערת האפס, אחרת נקבלה. קיימים הרבה איזורי דחיה (המייצגים אלטרנטיבות), נבחר את האחד המייצג את האלט׳ המעניינת אותנו. הערה: במקרה הרציף יהיה זה שוויון ממש.

כיצד נבחר בין אזורי דחיה אפשריים:

<u>דוגמא במקרה הנורמלי</u>: ונניח אנו דוחים עבור ערכים גבוהים, אז: $C_\alpha = [d_\alpha, \infty)$, נמצא את d_α:

$$P_{H_0}(\bar{X} \ge d_\alpha) = \alpha \Leftrightarrow P_{H_0}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \ge \frac{d_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha \Leftrightarrow \frac{d_\alpha - \mu_0}{\sigma/\sqrt{n}} = Z_{1-\alpha} \Leftrightarrow d_\alpha = \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$$

כיוון שלא משתמשים כלל בערך של האלט׳ $\theta_1 = \mu_1$, אזור דחיה זה יהיה נכון לכל $\mu_1 > \mu_0$ שנבחר כאלט׳. באופן דומה עבור אלט׳ של ערכים נמוכים מ-μ_0 או אלט׳ דו כיוונית – פשוט איחוד שני איזורי דחיה זרים, ושימוש ב-$Z_{1-\frac{\alpha}{2}}$. במקרה שונות לא ידועה: שימוש ב-$\widehat{\sigma^2}$ והתפלגות t_{n-1}.

<u>דוגמא למקרה הבינומי</u>: $H_0: p = p_0$, $H_1: p > p_0$. במקרה זה נשתמש בקירוב הנורמלי, אך כלל ההחלטה (גדול שווה ל-10) ייבדק על p_0 (כאן מתבטאת העובדה שאנו בודקים את ההסתברות ליפול באזור הדחיה <u>תחת השערת האפס</u>). למקרה זה נקבל: $C_\alpha = \left[p_0 + Z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}, \infty\right)$.

סוגי השערות:

- **השערה פשוטה**: השערה הכוללת רק ערך אחד לפרמטר, למשל: $H_0: \theta = \theta_0$.
- **השערה מורכבת**: השערה הכוללת אוסף ערכים Θ של θ, למשל: $H_1: \theta < \theta_0$. לא נעסוק בהשערות אפס מורכבות.

אלטרנטיבה מהצורה $H_1: \theta > \theta_0$ היא חד צדדית, ואילו מהצורה $H_1: \theta \ne \theta_0$ היא דו-צדדית.

רמת מובהקות, סוגי טעויות ועוצמה:

רמת מובהקות של תוצאה (*p-value*):

נניח $H_0: \theta = \theta_0$, $H_1: \theta > \theta_0$, $S(X)$ אומד. בהינתן מדגם מקרי x ותוצאה $S(x)$, נרצה למצוא את הרמה המינימלית $p - value = \alpha^*$ כך ש-$S(x) \in C_{\alpha^*}$ - כלומר: מה הסיכוי לקבל את התוצאה שקיבלנו או קיצונית ממנה.

<u>במקרה הנורמלי, שונות ידועה</u>: $\alpha^*(\bar{x}) = 1 - \Phi\left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right)$, ונדחה את H_0 בכל רמה שגדולה או שווה ל-$\alpha^*(\bar{x})$. נשים לב כי נוסחה זו נובעת מכך ש-$\bar{x} = \mu_0 + Z_{1-\alpha^*} \cdot \frac{\sigma}{\sqrt{n}}$, כאשר $\bar{x}$ הוא ה-d_{α^*} שלנו.

משמעות *p-value*:

ערכים נמוכים $\Leftarrow$ אזור דחיה גדול יותר $\Leftarrow$ הסיכוי לדחות את השערת האפס גדל. אם נשנה משהו בנתונים, למשל נגדיל את השונות, נקבל בהכרח *p-value* גדול יותר, שמשמעו – פחות סיכוי לדחות את השערת האפס. לא להתבלבל: דחיית השערת האפס היא עבור רמת ביטחון α כלשהי.

בדוגמת גבהי סטודנטים, $H_0: \mu = 170, H_1: \mu > 170, \bar{x} = 175$, $\sigma^2 = 16$, אם נשנה את $\sigma^2 = 144$ נקבל $p - value$ גדול יותר. אינטואיטיבית: כשהשונות גדלה, פחות מפתיע אותנו לקבל $\bar{x}$ במרחק 5 ס"מ מהתוחלת מאשר קודם, לכן פחות סביר לדחות את השערת האפס.

סוגי טעויות ועוצמת מבחן:

$H_0: \theta = \theta_0, H_1: \theta = \theta_1$, סטטיסטי $S(X)$, רמה α ואזור דחיה C_α.

- **טעות מסוג ראשון**: הסיכוי לדחות את H_0 כשהיא נכונה. זהו הסיכוי ש-$S(X) \in C_\alpha$ תחת P_{H_0} וזה שווה ל-α.
- **טעות מסוג שני**: הסיכוי שלא נדחה את H_0 כשהאלטרנטיבה נכונה: $P_{H_A}(S(X) \notin C_\alpha) = \beta$.
- **עוצמת מבחן**: הסיכוי שכן נדחה את H_0 כשהאלטרנטיבה נכונה: $\pi = 1 - \beta = P_{H_A}(S(X) \in C_\alpha)$.

בחירת אזורי דחיה והלמה של ניימן-פירסון:

הלמה של ניימן-פירסון (NP):

מבחן ברמת α בעל עוצמה מקסימלית:

להשערות פשוטות (θ_0, θ_A) הוא איזור דחיה C_α^* כך שלכל איזור דחיה אחר $\widetilde{C_\alpha}$ ברמה α מתקיים: $\pi(C_\alpha^*) \geq \pi(\widetilde{C_\alpha})$. נגדיר יחס נראות:

$$\Lambda(x) := \frac{f_{\theta_0}(x)}{f_{\theta_A}(x)} = \frac{L(\theta_0; x)}{L(\theta_A; x)}$$

הלמה של ניימן-פירסון:

מבחן בעל עוצמה מקסימלית להשערות הפשוטות לעיל מוגדר ע"י: $C_\alpha^* = \{x \mid \Lambda(x) \leq d_\alpha^*\}$, כאשר d_α^* הוא הערך הקריטי של C_α^*, כלומר: $P(x \in C_\alpha^*) = d_\alpha^*$. תמיד נדחה לערכים נמוכים של יחס הנראות (בתרגול היחס הוגדר הפוך, כלומר θ_A במונה, לכן אמרנו שנדחה תמיד עבור ערכים גבוהים של $\Lambda(x)$).

במקרה הנורמלי: $H_0: \mu = \mu_0$, $H_A: \mu = \mu_A > \mu_0$, נקבל ש-$\Lambda(x)$ הוא פונקציה יורדת של $\bar{x}$. נקבל: $C_\alpha^* = \{x \mid \bar{x} \geq d_\alpha^*\}$. לכן נדחה את H_0 אמ"מ $\bar{x} \geq \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$.

הוכחת הלמה: לא יכלל בסיכום.

מבחנים בעלי עוצמה מקסימלית לפי NP:

נורמלי עם שונות ידועה: כמתואר לעיל; עבור $H_A: \mu = \mu_A < \mu_0$: $C_\alpha^* = \{x \mid \bar{x} \leq \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\}$.

נורמלי עם שונות לא ידועה: אותו דבר רק עם התפלגות t_{n-1} ו-$\widehat{\sigma^2}$.

ברנולי/בינומי: $H_0: p = p_0, H_A: p = p_A > p_0$, נקבל כי מבחן NP הוא מהצורה: $C_\alpha^* = \{x \mid \hat{p} \geq d_\alpha^*\}$ (גם כאן $\hat{p} = \frac{\sum X_i}{n}$ הוא הסטטיסטי למבחן יחס נראות מקסימלי). נשתמש בקירוב הנורמלי כדי לקבל מבחן עוצמה מקסימלית ברמה בקירוב α: $C_\alpha^* = \{x \mid \hat{p} \geq p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\}$.

פואסון: $H_0: \lambda = \lambda_0, H_A: \lambda = \lambda_A > \lambda_0$, נקבל כי מבחן NP יהיה x עצמו: $C_\alpha^* = \{x \mid x \geq d_\alpha^*\}$.

הכללת NP – מבחנים בעלי עוצמה מקסימלית במידה שווה (UMP):

עבור $H_0: \theta = \theta_0$ ואלטרנטיבה מורכבת: $H_A: \theta \in \Theta$, נאמר כי אזור דחיה C_α^* מגדיר מבחן ברמה α בעל עוצמה מקסימלית במידה שווה אם:

- $P_{H_0}(x \in C_\alpha^*) = \alpha$
- C_α^* מגדיר מבחן עוצמה מקסימלית לכל אלטרנטיבה פשוטה: $\forall \theta_A \in \Theta: H_A: \theta = \theta_A$.

במקרה הנורמלי, שונות ידועה: אזור הדחיה שקיבלנו לא היה תלוי כלל ב-μ_A, כלומר התקיים לכל $\mu_A \in \{\mu_A \mid \mu_A > \mu_0\}$, ולכן נותן UMP.

כנ"ל שאר הדוגמאות (נורמלי עם שונות לא ידועה, בינומי/ברנולי, פואסון) נותנות UMP.

במקרה הדו-צדדי: אלט' מורכבת דו צדדית **אינה** UMP מכיוון שמבחן NP תלוי בכיוון האלט'.

הסקה סטטיסטית בבעיות רגרסיה:

ריבועים פחותים: נתונות תצפיות $(x_1, y_1), \ldots, (x_n, y_n)$, מעריכים מודל: $\hat{y} = a + bx$, מגדירים פונקצית f (RSS) שהיא סכום ריבועי ההפרשים: $f(a, b) = \sum_{i=1}^{n}(y - a - bx_i)^2$, ורוצים את הערכים של a,b שיביאו את f למינימום. נסתכל על a,b כפרמטרים שרוצים למדוד, והאומדים שלהם הם $\hat{a}_{LS}, \hat{b}_{LS}$. נעשה הסקה סטטיסטית על התפלגות מותנה:

נניח $Y|X = y \sim N(a + bx, \sigma^2)$, כאשר σ^2 לא ידוע. נתאר זאת כך: $Y|X = a + bx + \varepsilon$, כאשר $\varepsilon \sim N(0, \sigma^2)$ (הרעש).

נאמוד את a,b לפי נראות מקסימלית, ונקבל: $\ell(a,b;(x_1,y_1),\dots,(x_n,y_n),\sigma^2)=-\frac{n}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}f(a,b)$. במקרה זה חישבנו יחס נראות לפי H_A במונה, ולכן אנו רוצים ערכים גבוהים של היחס – כיוון ש-$f(a,b)$ עם סימן שלילי, זה שקול לערכים נמוכים של $f(a,b)$ - כמו שקיבלנו בעבר.

מכאן ש: $\widehat{a,b}=\underset{a,b}{argmin} f(a,b)$ הם אנ"מ ל-a,b. וכפי שקיבלנו בעבר:

$$\hat{b}=\frac{\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^n (X_i-\bar{X})^2},\ \hat{a}=\bar{Y}-\hat{b}\bar{X}$$

מודל: $y_i=a+bx_i+\varepsilon_i$, $\varepsilon_i\sim N(0,\sigma^2)$. ההשערות שנבדוק: $H_0: b=0$ (כלומר הקו הוא קבוע ללא שיפוע, ונזכר שזה אומר ש-r^2, אחוז השונות המוסבר, הוא 0), והאלטרנטיבה: $H_A: b\neq 0$.

כיוון שע"פ המודל $\forall i: E(\varepsilon_i)=0$, מקבלים כי $E(\hat{b})=b$, כלומר $\hat{b}$ הוא אומד חסר הטיה ל-b. לבסוף נקבל: $\hat{b}\sim N\left(b,\frac{\sigma^2}{\sum_{i=1}^n (X_i-\bar{X})^2}\right)$.

מכאן ש: $H_0: \hat{b}\sim N\left(0,\frac{\sigma^2}{\sum_{i=1}^n (X_i-\bar{X})^2}\right)$, ואומד חסר הטיה ל-$\sigma^2$: $\widehat{\sigma^2}=\frac{1}{n-2}\sum_{i=1}^n (y_i-\bar{y})^2$ ($n-2$ כיוון שיש לנו 2 פרמטרים שאומדים במקרה זה), ולכן נקבל את אזור הדחיה: $C_\alpha^*=\left(-\infty,-t_{n-2,1-\frac{\alpha}{2}}\cdot\frac{\hat{\sigma}/\sqrt{\sum(X_i-\bar{X})^2}}{\sqrt{n}}\right]\cup\left[t_{n-2,1-\frac{\alpha}{2}}\cdot\frac{\hat{\sigma}/\sqrt{\sum(X_i-\bar{X})^2}}{\sqrt{n}},\infty\right)$.

מבחנים לקשר בין 2 התפלגויות / מ"מ:

נניח $X\sim F, Y\sim G$, ל-F פרמטר(ים) θ, ל-G פרמטר(ים) ψ, בד"כ $H_0: \theta=\psi$. לדוגמא במקרה הנורמלי:

$X\sim N(\mu_X,\sigma_X^2), Y\sim N(\mu_Y,\sigma_Y^2)$, ניקח אומדים לתוחלת, $H_0: \mu_X=\mu_Y$, $H_A: \mu_X>\mu_Y$ (יכולה להיות כל אלט' אחרת). נתונים מדגמים $x=(x_1,\dots,x_n)$, $y=(y_1,\dots,y_m)$.

- אם X,Y בלתי תלויים, נערוך מבחן לא מזווג של שני המדגמים.
- אם ישנה תלות, נערוך מבחן מזווג, כאשר המקרה המעניין הוא $n=m$, x_i,y_i הן תצפיות על אותו אובייקט (כלומר ישנה תלות אך ורק בין x_i ל-y_i, וכל שאר היחסים הם בלתי תלויים.

מבחנים לא מזווגים לשני מדגמים:

$\bar{X}\sim N\left(\mu_X,\frac{\sigma_X^2}{n}\right), \bar{Y}\sim N\left(\mu_Y,\frac{\sigma_Y^2}{n}\right)$, נקח את ההפרש ביניהם שגם הוא מ"מ: $\bar{Z}:=\bar{X}-\bar{Y}\sim N\left(\mu_X-\mu_Y,\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)$.

נקבל כי $H_0: E(Z)=0$, והאלטרנטיבה היא $H_A: E(Z)>0$. קיים מבחן UMP כמו במקרה הנורמלי הרגיל.

(א) שונויות ידועות: $C_\alpha^*=\{x,y\mid \bar{x}-\bar{y}\geq Z_{1-\alpha}\cdot\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\}$

(ב) שונויות לא ידועות:

- השונויות לא ידועות אך ידוע שהן שוות, נסמן $\sigma_X^2=\sigma_Y^2=:\sigma^2$: $var(\bar{X}-\bar{Y})=\frac{n+m}{nm}\sigma^2$. נעריך את σ^2 משני המדגמים יחד (בגלל שוויון השונויות) לקבלת אומד: $\widehat{\sigma_{XY}^2}=\frac{\sum_{i=1}^n (X_i-\bar{X})^2+\sum_{j=1}^m (Y_j-\bar{Y})^2}{n+m-2}=\frac{(n-1)\widehat{\sigma_X^2}+(m-1)\widehat{\sigma_Y^2}}{n+m-2}$, ונערוך מבחן t עם $n+m-2$ ד"ח:

$$C_\alpha^*=\{x,y\mid \bar{x}-\bar{y}\geq t_{n+m-2,1-\alpha}\cdot\sqrt{\frac{n+m}{nm}\widehat{\sigma_{XY}^2}}\}$$

- לא מניחים שוויון שונויות: $\sigma_{\bar{X}-\bar{Y}}^2=\frac{\widehat{\sigma_X^2}}{n}-\frac{\widehat{\sigma_Y^2}}{m}$. דרגות החופש יחושבו לפי תיקון $Welch\text{-}SatterWiate$: $\nu=\frac{\left(\widehat{\sigma_{\bar{X}-\bar{Y}}^2}\right)^2}{\frac{\widehat{\sigma_X}^4}{n^2(n-1)}+\frac{\widehat{\sigma_Y}^4}{m^2(m-1)}}$, ונקבל:

$$C_\alpha^*=\{x,y\mid \bar{x}-\bar{y}\geq t_{\nu,1-\alpha}\cdot\sqrt{\widehat{\sigma_{\bar{X}-\bar{Y}}^2}}\}$$

הסקה במדגמים מזווגים:

$n=m$, (x_i,y_i) אינם בלתי תלויים (ישנה אי תלות בין כל שאר הזיווגים). נסמן $Z_i=X_i-Y_i$, ונקבל כי $Z\sim N(\mu_X-\mu_Y,\sigma_{X-Y}^2)$. נקבל את אותו ניסוח השערות כמו קודם: $H_0: \mu_Z=0, H_A: \mu_Z>0$. נשים לב שמהתלות מתקיים: $\sigma_{X-Y}^2\neq\sigma_X^2-\sigma_Y^2$ ($var(X-Y)=var(X)+var(Y)-\underbrace{2cov(X,Y)}_{\neq 0}$).

(א) שונות ידועה: UMP יהיה: $C_\alpha^*=\{x,y\mid \bar{x}-\bar{y}\geq Z_{1-\alpha}\cdot\sqrt{\frac{\sigma_{X-Y}^2}{n}}\}$

(ב) שונות לא ידועה: $\widehat{\sigma_{X-Y}^2}=\widehat{\sigma_Z^2}=\frac{\sum_{i=1}^n (Z_i-\bar{Z})^2}{n-1}$, ומכאן: $C_\alpha=\{x,y\mid \bar{z}=\bar{x}-\bar{y}\geq t_{n-1,1-\alpha}\cdot\sqrt{\frac{\widehat{\sigma_{X-Y}^2}}{n}}\}$

נסכם את המקרים השונים:

	מזווגים	לא מזווגים
X_i, Y_i ב"ת	**נכון**: עדיין מקבלים מבחן ברמה α אך מפסידים עוצמה (פחות דרגות חופש בזיווג)	**נכון**
X_i, Y_i תלויים	**נכון**	**לא נכון**: רמת המבחן אינה α, עוצמת המבחן קטנה כיוון ש: $var(X-Y) < var(X) + var(Y)$ (כי $cov(X,Y) > 0$).

הסקה על קשרים בין התפלגויות בדידות קטגוריאליות:

המקרה הבינומי (לא מזווג): כל ה-$n+m$ בלתי תלויים

$X \sim Bernoulli(p_X), Y \sim Bernoulli(p_Y)$, $H_0: p_X = p_Y$, $H_A: p_X > p_Y$ (בה"כ), n תצפיות X, m תצפיות Y. לפי הקירוב הנורמלי:

$$\widehat{p_X} - \widehat{p_y} \dot\sim N\left(p_X - p_Y, \frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}\right) \Rightarrow H_0: p_X = p_Y =: p, \qquad \widehat{p_X} - \widehat{p_Y} \dot\sim N\left(0, \frac{p(1-p)}{n} + \frac{p(1-p)}{m}\right)$$

במקרה זה השונות תחת H_0 אינה ידועה, ולכן נאמוד אותה:

- <u>אומד הרפתקני</u>: $\hat{p} = \frac{n\widehat{p_X} + m\widehat{p_Y}}{n+m}$ ולכן $\widehat{\sigma^2_{\widehat{p_X} - \widehat{p_Y}}} = \frac{n+m}{nm}\hat{p}(1-\hat{p})$.
- <u>אומד שמרני</u>: $\hat{p} = 0.5$ ולכן $\widehat{\sigma^2_{\widehat{p_X} - \widehat{p_Y}}} = \frac{n+m}{nm} \cdot \frac{1}{4}$.

מכאן שאזור הדחיה יהיה: $C_\alpha = \left\{x, y \mid \widehat{p_x} - \widehat{p_y} \geq Z_{1-\alpha} \cdot \sqrt{\widehat{\sigma^2_{\widehat{p_X} - \widehat{p_Y}}}}\right\}$

במידה ומספר ההצלחות הכולל ידוע מראש:

נסמן את מספר ההצלחות הכולל לשני המ"מ יחד כ-$n_+ + m_+$. נקבל כי תחת H_0: $n_+ \sim HG(n+m, n_+ + m_+, n)$ - גודל המדגם הוא $n+m$, מספר ההצלחות (ה"אכלוסיה המיוחדת") הוא $n_+ + m_+$, ואנו לוקחים מדגם של n (המדגם X לצורך העניין, כי ההצלחות שלו מעניינות אותנו).

במקרה זה נוכל להסיק לגבי נכונות השערת האפס בשלושה דרכים:

- הסקה מדוייקת בעזרת HG (מבחן *Fisher exact test*).
- קירוב נורמלי: $n \cdot \widehat{p_X} = n_+ \dot\sim N\left(\frac{n(n_+ + m_+)}{n+m}, \frac{n(n_+ + m_+)(n+m-n_+ + m_+)(n+m-n)}{(n+m)^2(n+m-1)}\right)$.
- מבחן χ^2 לטבלאות.

התפלגות חי בריבוע (χ^2):

הגדרה:

יהיו $X_1, \ldots, X_k$ מ"מ בלתי תלויים מהתפלגות נורמלית סטנדרטית, $X_i \sim N(0,1)$, אזי: $\sum_{i=1}^{k} X_i^2 \sim \chi_k^2$, כאשר k הוא מס' דרגות החופש.

אם $X \sim \chi_k^2, Y \sim \chi_l^2$, אזי: $X + Y \sim \chi_{k+l}^2$.

נתון מדגם בגודל N עם שני ערכים של משתנים בדידים: U עם k ערכים אפשריים, Z עם l ערכים אפשריים. נרצה לבדוק אי תלות:

$H_0: U, Z$ – בלתי תלויים. נסכם את הנתונים בטבלת *observed* בגודל $k \times l$:

	$Z = z_1$	$Z = z_2$	…	$Z = z_l$	סה"כ
$U = u_1$	n_{11}	n_{12}		n_{1l}	n_1
$U = u_2$	n_{21}	n_{22}		n_{2l}	n_2
…	…				…
$U = u_k$	n_{k1}	n_{k2}	…	n_{kl}	n_k
סה"כ	m_1	m_2		m_l	N

נתונה על הערכים השוליים $m_1, \ldots, m_l, n_1, \ldots, n_k$ ע"י בניית טבלה שניה, טבלת ה-*expected*, מגודל זהה שערכיה יהיו $e_{ij} = \frac{n_i m_j}{N} = \frac{n_i}{N} \cdot \frac{m_j}{N} \cdot N$.

טענה:

אם כל התאים בטבלת ה-*expected* (וה-*observed*?) מכילים לפחות 5 תצפיות אזי תחת H_0 (אלטרנטיבה דו-צדדית):

$$\sum_{i=1}^{k \times l} \frac{(o_i - e_i)^2}{e_i} \dot\sim \chi^2_{(k-1)(l-1)}$$

רגרסיה מול χ^2:

גישה אחרת לבדיקת תלות בין X, Y תהיה לעשות דיסקרטיזציה של הנתונים, למשל ל-{נמוך, בינוני, גבוה}, לעשות טבלה מתאימה (3 שורות ו-3 עמודות), ולערוך מבחן χ^2 לאי-תלות עם $4 = (3-1)(3-1)$ דרגות חופש. הרווח: בודקים אי תלות כלשהי, לאו דווקא לינארית; ההפסד: הסדר בין הקטגוריות. לכן גם מרוויחים עוצמה וגם מפסידים עוצמה למבחן.

מבחן χ^2 של פירסון לאיכות התאמה:

יהי X מ״מ בדיד המקבל ערכים מ-$\{a_1,\dots,a_k\}$, תהי $F = mult(p_1,\dots,p_k)$ התפלגות מולטינומית נתונה על a_i, ו-$H_0: X\sim F$. יהי x מדגם מקרי בגודל n בו רואים: n_1 תצפיות מסוג a_1,...,n_k תצפיות מסוג a_k, כאשר $\sum n_i = n$. ניתן להסתכל על זה בתור טבלה עם שורה אחת:

$observed$:

סה״כ	a_1 a_2 ... a_k
n	n

$expected$:

סה״כ	a_1 a_2 ... a_k
n	n

אם כל תא בטבלת ה-exp יכיל 5 תצפיות ומעלה, אזי תחת H_0:

$$\sum_{i=1}^{n}\frac{(n_i-np_i)^2}{np_i}\dot{\sim}\chi^2_{k-1}$$

הערה: כיוון שאין סכומי עמודות להתנות עליהם, לא נכפיל את דרגות החופש ב-$0 = (1-1)$.

נושאים נוספים:

השוואות מרובות: לא נכלל בסיכום.

מבחנים סטטיסטיים לא פרמטריים: לא נכלל בסיכום.

אמידה בייסיאנית:

לפי הגישה הבייסיאנית, הפרמטר θ אותו אומדים הוא לא מספר אלא מ״מ. נסמן ב-$p(\theta)$ את ההתפלגות המקדימה (אפריורית) של θ - ה״אמונה״ שלנו לגבי הפרמטר לפני שראינו את הנתונים. עם קבלת הנתונים נעדכן את ההתפלגות להתפלגות האפוסטריורית. עדכון ההתפלגות בהינתן x:

$$Pr(\theta \mid x) = \frac{p(x,\theta)}{p(x)} = \frac{f(x|\theta)\cdot p(\theta)}{p(x)} \propto f(x|\theta)\cdot p(\theta)$$

אם יש לנו ידע טוב על θ, $p(\theta)$ יהיה רווחי בגישה זו. כמו כן ניתן להסיק על θ ישירות. מצד שני, לא תמיד יהיה לנו ידע על θ והחישובים מסובכים.

<u>דוגמא למקרה הנורמלי, שונות ידועה</u>:

$x|\mu\sim(\mu,\sigma^2)$, x מדגם בגודל n, ומעריכים את μ. לפי אמידה בייסיאנית: נניח $\mu\sim N(0,\sigma_\mu^2)$ (נצפה שהתוחלת תהיה קרובה ל-0), נחשב ונקבל:

$$\mu|x\sim N\left(\frac{\sigma_\mu^2}{\sigma_x^2+\sigma_\mu^2}\cdot x, \frac{\sigma_\mu^2\cdot\sigma_x^2}{\sigma_x^2+\sigma_\mu^2}\right)$$