הסקה סטטיסטית- יורם לוזון תשפ"ד

רגב אימרה יחזקאל

August 9, 2024

לשאלות/ הצעות שיפורים מוזמנים לשלוח מייל ל-$regevel2006@gmail.com$.

תוכן העניינים

1 הרצאה 1.

1.1 מבוא.

מה זה סטטיסטיקה? סטטיסטיקה זה המדע של לקשר בין תצפיות לבין מודל.

נניח עכשיו אני רוצה לענות על השאלה "האם ירד גשם מחר?", מה אני אעשה?

נניח יש לי תצפיות, אני פשוט אכניס אותם למודל וממנו אקבל מסקנות.

תצפיות $\leftarrow$ מודל $\leftarrow$ מסקנות

נניח אני קובע שגשם זה אירוע שקורא כאשר בורא עולם מטיל מטבע, זה מאורע ברנולי ולכן הכי חשוב לי זה להסיק את p על ידי התצפיות.

איך אני קובע מה זה p? ישנן כמה שאלות שאני יכול לשאול:

1. מה הסיכוי שהפרמטר $p_0 \le p \le p_1$?
2. תן לי טווח עבורו הסיכוי שp מחוץ לטווח קטן מε.
3. משה טוען $p < 0.2$. מה הסיכויים שמשה טועה (נסמן את זה H_0)? מה הסיכוי שH_0 לא נכון?

הגדרה 1.1. אומד זה דבר שמתווך בין תצפיות ובין מודל שאני בונה.

שתי שאלות:

א) מהו אומד טוב?

ב) האם המודל שלי בכלל נכון?

דוגמה לאומד לא טוב:

נניח ואנחנו רוצים לתכנן חופשה מרגיעה לבה סינג סה, ואנחנו במשך שבוע שואלים את סוכנת המכירות שלנו, אזולה, האם יש מלחמה בבה סינג סה. כל יום אזולה אומרת לנו "אין מלחמה בבה סינג סה", ולכן אנחנו תיירים משוכנעים, טסים לבה סינג סה, מדינה שידועה במלחמתה נגד אומת האש.

מה השתבש לנו באומד? בעצם בגלל שאלנו את אזולה במשך שבוע לגבי המלחמה ומזה הסקנו על כל הזמן.

לכן המציאו את הגישה הבייסיאנית שאומרת שאם טעיתי והמודל שלי לא עבד, זה בגלל שהערכתי לא נכון את הסיכויים להצלחה ושאני צריך לשנות את הגדרות המודל שלי.

מה נעשה בקורס הזה:

- חזרה על הסתברות (היום):
  - התפלגויות למיניהן.
  - פונקציות יוצרות מומנטים.
  - משפט הגבול המרכזי.
- סטטיסטיקה:
  - אומד אופטימלי (4 עד 5 שיעורים).
  - רווח סמך (1 עד 2 שיעורים).
  - בדיקת השערות (2 עד 3 שיעורים).

נתונים $\leftarrow$ אומד $\leftarrow$ מודל $\leftarrow$ מסקנות

1.2 מודל הסתברות.

נניח יש לי מרחב מדגם- A קבוצה, וקבוצה $B \subseteq A$ או אוסף של תתי קבוצות של A כגון $\{B_i\}_{i\in I}$.

על אוסף תתי קבוצות של A שנסמן $\sigma = \{B_i\}_{i\in I}$ נדרוש 3 דברים:

1. $\emptyset \in \sigma$
2. $B \in \sigma \Rightarrow B^c \in \sigma$
3. $B_j \in \sigma \Rightarrow \bigcup_j B_j \in \sigma$

הגדרה 1.2. σ העונה על שלושת הדרישות הנ"ל נקראת **סיגמה אלגברה**.

על הσ אלגברה שלי נגדיר פונקציה $\mathbb{P} : B \to [0,1]$ עם הדרישות הצנועות הבאות:

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(B) = 1$
3. $\mathbb{P}(B^c) = 1 - \mathbb{P}(B)$
4. $B_i \bigcap B_j = \emptyset \Rightarrow \mathbb{P}(B_i \cup B_j) = \mathbb{P}(B_i) + \mathbb{P}(B_j)$

הגדרה 1.3. כעת נגדיר פונקציה מאוד מיוחדת, שנקראת **משתנה מקרי** באופן הבא:

- $f : A \to \mathbb{R}$
- $x \in A \to f(x)$
- $P(f(x) = y) = \mathbb{P}(B | x \in B \iff f(x) = y)$

1.3 התפלגויות בדידות.

- ברנולי: $A = [H, T]$, נסמן $\mathbb{P}(H) = p, \mathbb{P}(T) = q$ ונקבל $q = 1 - p$.
- בינום: מתוך T הטלות, מה הסיכוי שבדיוק k פעמים יצא לי עץ? $\mathbb{P}(k) = \left(\begin{array}{c} T \\ k \end{array} \right) p^k (1-p)^{T-k}$.
- פואסון: $\mathbb{P}(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

בהתפלגויות רציפות: $\mathbb{P}(x \in [y, y+dy]) = \int f(y)dy$ כאשר נדרוש $0 \leq f(y)$ אינטגרבילית וגם $\int\limits_{-\infty}^{\infty} f(y)dy = 1$.

1.4 התפלגויות רציפות.

- אחידה: $\mathbb{P}(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & else \end{cases}$.
- אקפוננציאלית: $\mathbb{P}(x) = \begin{cases} \tau e^{-\tau x} & x \geq 0 \\ 0 & else \end{cases}$.
- נורמלית: $\mathbb{P}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$.

1.5 מומנטים.

הגדרה 1.4. נגדיר **תוחלת**, שזה מעין "הממוצע" של התפלגות.

- בדיד: $\mathbb{E}[f(x)] = \sum \mathbb{P}(x) f(x)$.
- רציף: $\mathbb{E}[f(x)] = \int \mathbb{P}(x) f(x) dx$.

אותי מעניין ספציפית פונקציות מהצורה $f(x)=x^k$, לכן נגדיר את המושג הבא:

הגדרה 1.5. המומנט ה-k של x להיות

$$M_k(x)=\mathbb{E}\left[x^k\right]$$

(נשים ♡ כי $M_0=1$).

הגדרה 1.6. דרך המומנטים נגדיר את **השונות** של x להיות

$$\mathbb{V}\left[x\right]=\mathbb{E}[(x-\mathbb{E}[x])^2]=\mathbb{E}\left[x^2\right]-\mathbb{E}\left[x\right]^2=M_2(x)-M_1^2(x)$$

משפט 1.7. אם כל המומנטים של שתי התפלגויות שווים אז ההתפלגויות שוות (למעט אולי בקבוצה שמידתה 0).

הוכחה: ראינו כבר בהסתברות.

נראה כמו הרבה טרחה לחשב מומנטים, הרבה אינטגרלים ומהר מאוד זה הופך להיות מעצבן.

הגדרה 1.8. בשביל להימנע מחישוב של הרבה אינטגרלים, נגדיר **פונקציה יוצרת מומנטים** על ידי

$$M(t)=\mathbb{E}[e^{tx}]$$

חשוב לנו לראות ש$M(t)$ הזאת עומדת בשמה ושהיא אכן יוצרת מומנטים. נפתח את $M(t)$ לטור הטיילור שלה: $e^{tx}=\sum\limits_{k=0}^{\infty}\frac{x^k}{k!}t^k$. לכן נכתוב

$$M(t)=\mathbb{E}\left[\sum_{k=0}^{\infty}\frac{x^k}{k!}t^k\right]=\sum_{k=0}^{\infty}\mathbb{E}\left[\frac{x^k}{k!}t^k\right]=\sum_{k=0}^{\infty}\frac{t^k}{k!}\mathbb{E}\left[x^k\right]=\sum_{k=0}^{\infty}\frac{t^k}{k!}M_k(x)$$

כעת, הקסם קורה:

$$\begin{aligned}M^{(0)}(0)=&M_0(x)=1\\M^{(1)}(0)=&M_1(x)\\&\vdots\\M^{(k)}(0)=&M_k(x)\end{aligned}$$

עכשיו כשיש לי דרך קלה לחשב מומנטים, הגיע הזמן להגדיר כמה דברים על פונקציות יוצרות מומנטים:

נניח ויש לי שתי התפלגויות X,Y ונגדיר $Z=X+Y$. נתבונן בפונקציה יוצרת מומנטים של Z, כאשר מניחים שX ו-Y בלתי תלויים

$$M_Z(t)=\mathbb{E}\left[e^{tZ}\right]=\mathbb{E}\left[e^{t(X+Y)}\right]=\mathbb{E}\left[e^{tX}e^{tY}\right]=\mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right]=M_X(t)M_Y(t)$$

זה שימושי! עכשיו אם שואלים אותי לחשב $\mathbb{V}\left[X+Y\right]$, אל חשש, לא צריך קונבולוציה ושום דבר אלא לדעת מה המומנטים של X ו-Y ואם הם בלתי תלויים אז לפתח את השונות עם ההגדרה לפי מומנטים ומשם קלי קלות!

הערה 1.9. $M(cX)=(M(X))^c$.

דוגמה 1.10. נחשב פונקציה יוצרת מומנטים לברנולי.

פתרון. לפי ההגדרה, אם $X\sim\mathrm{Ber}(p)$ אזי

$$M_X(t)=\mathbb{E}\left[e^{tX}\right]=e^{t\cdot 0}\cdot q+e^{t\cdot 1}\cdot p=(1-p)+pe^t$$

דוגמה 1.11. נחשב פונקציה יוצרת מומנטים לבינום:

פתרון. מצד אחד אפשר לחשב לפי ההגדרה

$$\mathbb{E}\left[e^{tX}\right]=\sum P(x)e^{tX}$$

אבל לא תסיימו לחשב את זה עד המבחן. מצד שני, אפשר לזכור שבינום הוא סכום של T ברנולים בלתי תלויים, ולכן

$$M_X(t)=((1-p)+pe^t)^T$$

דוגמה 1.12 (לאמיצים בלבד!): חשבו את הפונקציה יוצרת מומנטים להתפלגות נורמלית עבור $\sigma,\mu\in\mathbb{R}$:

פתרון. $e^{\mu t}e^{\frac{1}{2}\sigma^2t^2}$ ואת ההוכחה אשאיר כתרגיל בית קצרצר לקורא ☺.

שאלה 1.13 אם X מתפלג נורמלית עם σ_1,μ_1, וגם Y מתפלג נורמלית עם σ_2,μ_2. איך נראית ההתפלגות של $X+Y$?

פתרון: לחשב ישר $P(X+Y)$ בלי טריקים זה מאוד מסובך לי עד לנקודה שאני אפילו לא אגיע למועד ב ☹. מצד שני, ניעזר בפונקציה יוצרת מומנטים ונקבל

$$M_{X+Y}(t)=M_X(t)\cdot M_Y(t)=\left(e^{\mu_1t+\frac{1}{2}\sigma_1^2t^2}\right)\left(e^{\mu_2t+\frac{1}{2}\sigma_2^2t^2}\right)=e^{\mu_1t+\frac{1}{2}\sigma_1^2t^2+\mu_2t+\frac{1}{2}\sigma_2^2t^2}=e^{\frac{t^2}{2}\left(\sqrt{\sigma_1^2+\sigma_2^2}\right)^2}e^{t(\sigma_1\mu_1+\sigma_2\mu_2)}$$

שזה פונקציה יוצרת מומנטים של התפלגות נורמלית עם $\sigma'=\sqrt{\sigma_1^2+\sigma_2^2}$ ועם $\mu'=\sigma_1\mu_1+\sigma_2\mu_2$.

1.6 סיכום.

1. התפלגות נורמלית היא נקודת שבת של הפעולה $\mathbb{P}(Z) = \mathbb{P}(X+Y)$.
2. טענה: יש רק נקודת שבת אחת לפעולה הנ"ל ולכן כשאני אסכום מספיק משתנים מקריים אקבל בהכרח התפלגות נורמלית.

הערה 1.14. - טענה 2 היא ניסוח אחר של משפט הגבול המרכזי.

☻

2 הרצאה 2.

תזכורת: פונקציה יוצרת מומנטים היא
$$M(t)=\mathbb{E}\left[e^{tX}\right]$$

הגדרה 2.1. **פונקציה יוצרת קומולנטים** תוגדר להיות
$$K(t)=\ln(M(t))$$

דוגמה 2.2. פונקציה יוצרת מומנטים להתפלגויות נפוצות:

- ברנולי: $(1-p)+pe^t$.
- גאומטרי (כשמתחילים לבדוק החל מהניסיון ה0): $\frac{p}{1-(1-p)e^t}$.
- פואסון: $e^{\lambda\left(e^t-1\right)}$.
- נורמלית: $e^{\left[\mu t+\frac{1}{2}\sigma^2t^2\right]}$.

משפט 2.3. אם לשתי התפלגויות יש את אותם מומנטים אז ההתפלגויות זהות (פרט אולי לקבוצה בעלת מידה 0).

משפט 2.4. לפונקציה יוצרת מומנטים מתקיים:

א אם $Y=aX+b$ אזי
$$M_Y(t)=\mathbb{E}\left[e^{(aX+b)t}\right]=e^{bt}\mathbb{E}\left[e^{xat}\right]=e^{bt}M_X(at)=e^{bt}\left(M_X(t)\right)^a$$

ב אם X,Y מ"מ ב"ת אזי
$$M_{X+Y}(t)=\mathbb{E}\left[e^{t(X+Y)}\right]=\mathbb{E}\left[e^{tX}e^{tY}\right]=\mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right]=M_X(t)M_Y(t)$$

ג המומנט הk של $P(X)$ הוא $M_t^{(k)}(0)$.

ניזכר: $\ln(1+z)=z-\frac{1}{2}z^2+O(z^3)$, לכן נשאף לפתח למשהו דומה את הפונקציה יוצרת קומולנטים:

$$K(t)=\ln(M(t))=\ln\left(\underbrace{M_0(X)}_{=1}+\underbrace{M_1(X)\cdot t+\frac{M_2(X)}{2}\cdot t^2+O(t^3)}_{=z}\right)=$$
$$=\left(M_1(X)\cdot t+\frac{M_2(X)}{2}\cdot t^2+O(t^3)\right)-\frac{1}{2}\left(M_1(X)\cdot t+\frac{M_2(X)}{2}\cdot t^2+O(t^3)\right)^2+\underbrace{O(z^3)}_{=O(t^3)}$$
$$=M_1(x)t+\frac{\left[M_2(x)-M_1^2(x)\right]t^2}{2}+O(t^3)=\mathbb{E}\left[X\right]t+\mathbb{V}\left[X\right]t+O(t^3)$$

כלומר בחינם קיבלתי את התוחלת והשונות!

מסקנה 2.5. לפונקציות יוצרות מומנטים/ קומולנטים:

א סכום של שתי התפלגויות נורמליות הוא נורמלי.

ב בפונקציה יוצרת קומולנטים של התפלגות נורמלית, כל המקדמים מעבר לt^2 הם 0.

ג אם לשתי פונקציות התפלגות יש אותה פונקציה יוצרת מומנטים אז הן זהות (עד לקבוצה שמידתה 0).

ד אם יש לי התפלגות עם תוחלת μ ושונות σ^2 וכל שאר המקדמים של t^k בפונקציה יוצרת קומולנטים הם 0 אז בהכרח זו התפלגות נורמלית עם תוחלת μ ושונות σ^2.

2.1 משפט הגבול המרכזי (Yoram's version).

משפט 2.6. אם x_i מ"מ ב"ת עם תוחלת 0 וס"ת σ לכל אחד, ושאר המומנטים חסומים, אזי

$$z=\sum_{i=1}^{n}\frac{x_i}{\sqrt{n}}-(\text{התוחלת שלהם})$$
$$\lim_{n\to\infty}k(z)=\frac{1}{2}\sigma^2t^2$$

הוכחה. תהי $k_z(t)$ פונקציה יוצרת קומולנטים של z. אזי:

$$k_\ell(x)=K^{(\ell)}(t)=\begin{cases}k_0(x)= & 0\\ k_1(x)= & \text{תוחלת}\\ k_2(x)= & \text{שונות}\\ k_\ell(x)= & t^\ell\text{ המקדם של}\end{cases}$$

כעת אם $z=x_1+x_2$ עבור x_1,x_2 ב"ת:

$$\begin{aligned}M_z(t)=&M_{x_1}(t)\cdot M_{x_2}(t)\\ K_z(t)=&K_{x_1}(t)+K_{x_2}(t)\end{aligned}$$

ואם $z=ax$:

$$\begin{aligned}M_z(t)=&M_x(at)\\ K_z(t)=&K_x(at)\end{aligned}$$

ובפרט אם $z=\sum_{i=1}^{n}\frac{x_i}{\sqrt{n}}$ נקבל $K_z(t)=\sum_{i=1}^{n}K_{x_i}\left(\frac{t}{\sqrt{n}}\right)$. כעת:

$$\begin{aligned}K_0(Z)=&\sum_{i=1}^{n}K_{x_i}(0)=0\\ K_1(Z)=&\frac{d}{dt}\left[\sum_{i=1}^{n}K_{x_i}\left(\frac{t}{\sqrt{n}}\right)\right]_{t=0}=\sum_{i=1}^{n}\frac{d}{dt}K_{x_i}\left(\frac{t}{\sqrt{n}}\right)_{t=0}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}K_1(x_i)\underset{\star}{=}0\\ K_2(Z)=&\frac{1}{n}\sum_{i=1}^{n}K_2(x_i)\underset{\star\star}{=}\frac{1}{n}\sum_{i=1}^{n}\sigma^2=\frac{1}{n}\cdot n\cdot\sigma^2=\sigma^2\\ K_3(Z)=&\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}K_3(x_i)\underset{\star\star\star}{\leq}\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}M_x=\frac{n}{n^{\frac{3}{2}}}M_x=\frac{1}{n^{\frac{1}{2}}}M_x\underset{n\to\infty}{\longrightarrow}0\\ &\vdots\\ K_{m\geq4}(Z)=&\frac{1}{n^{\frac{m}{2}}}\sum_{i=1}^{n}K_m(x_i)\underset{\star\star\star}{\leq}\frac{1}{n^{\frac{m}{2}}}\sum_{i=1}^{n}M_x=\frac{n}{n^{\frac{m}{2}}}M_x=\frac{1}{n^{\frac{m}{2}-1}}M_x\underset{n\to\infty}{\longrightarrow}0\end{aligned}$$

מסקנה: אם $x_1,\ldots,x_n$ ב"ת עם ממוצע 0 וס"ת σ אז עבור $z=\sum_{i=1}^{n}\frac{x_i}{\sqrt{n}}$ אז $K_0(z)=0,K_1(z)=0,K_2(z)=\sigma^2,K_{n\geq3}(z)\underset{n\to\infty}{\longrightarrow}0$, כלומר z שואף להתפלגות נורמלית עם ממוצע 0 וס"ת σ, כנדרש. ■

$\star$: הממוצע של כל ה-x_i הוא 0.

$\star\star$: השונות של כל ה-x_i היא σ^2.

$\star\star\star$: ניקח $M_x=\max_i\{K_m(x_i)\}$.

2.2 סטטיסטיקה.

מודל- הנחה על ההתפלגות. המודל מכיל פרמטרים θ. לי יש תצפיות $x_1, \ldots, x_k$. באופן כללי אני אניח שהתצפיות שוות התפלגות ובלתי תלויות.

הרעיון: בהינתן $x_1, \ldots, x_k$ לפענח מה זה θ.

דוגמה 2.7. התפלגות ההכנסות:

$$\mathbb{P}(x) = \begin{cases} x^{-\theta} & x \geq \alpha_0 \\ 0 & x < \alpha_0 \end{cases}$$

אבל אני לא יודע מה זה θ, α_0. הבעיה: $x_1, \ldots, x_k$ זה הרבה נתונים, לכן אני אחליף אותם באומד $T(x_1, \ldots, x_k)$.

נשים ♡ כי אומד הוא משתנה מקרי. לכן אני יכול לחשב את $\mathbb{P}(T|\theta)$.

נניח שאני יודע את x_i. קל לראות כי $\mathbb{P}(T|x_i) \neq \mathbb{P}(T|\theta)$.

דוגמה 2.8. נניח יש לי 10 משתני ברנולי $x = [0, 0, 1, 1, 0, 0, 0, 1, 0, 1]$ כאשר $p(0) = q, p(1) = p$. לכן $\mathbb{P}(x) = q^6 p^4$.

נגדיר אומד $T = \sum_{i=1}^{10} x_i$ ונקבל כי $\mathbb{P}(T = 4|x) = 1 \neq \binom{10}{4} p^4 q^6 = \mathbb{P}(T = 4|\theta = p)$. בנוסף,

$$\mathbb{P}(x|T = 4) = \frac{\mathbb{P}(x \cap T = 4)}{\mathbb{P}(T = 4)} = \frac{p^4 q^6}{\binom{10}{4} p^4 q^6} = \frac{1}{\binom{10}{4}}$$

וזה היה תלוי לגמרי ב-T, כלומר לא צריך לדעת את p, q וזה ממש טוב.

מנגד, אם נגדיר אומד $T = x_1$ אזי

$$\mathbb{P}(x|T = 0) = \frac{\mathbb{P}(x \cap T = 0)}{\mathbb{P}(T = 0)} = \frac{\mathbb{P}(x)}{q} = \frac{p^4 q^6}{q} = p^4 q^5$$

אומד לא טוב כי גם אחרי שעשיתי את כל החישובים אני עדיין תלוי בp, q, כלומר לא חסכתי לעצמי את הטרחה בלמצוא את הפרמטרים של המודל שלי.

הגדרה 2.9. בהינתן מודל ודגימות $i.i.d$ (בלתי תלויים ושווי התפלגות) מהמודל $\{x_i\}$ ואומד $T(\{x_i\})$. אומרים ש-T הוא **אומד מספיק** אם מתקיים

$$\mathbb{P}(\{x_i\}|\theta) = \mathbb{P}(\{x_i\}|T)\mathbb{P}(T|\theta)$$

טענה 2.10. T אומד מספיק $\iff$ מתקיים

$$\mathbb{P}(\{x_i\}|\theta, T) = \mathbb{P}(\{x_i\}|T)$$

משפט 2.11. (ניימר פישר): T הוא אומד מספיק $\iff$ מתקיים $\mathbb{P}(\{x\}|\theta) = h(x)g(\theta, T)$.

הוכחה. $\Leftarrow$ נניח T אומד מספיק, לכן לפי ההגדרה

$$\mathbb{P}(\{x\}|\theta) = \underbrace{\underbrace{\mathbb{P}(\{x\}|T)}_{\mathbb{P}(\{x\}|T(\{x\})}}_{h(x)} \underbrace{\mathbb{P}(T|\theta)}_{g(\theta,T)}$$

$\Rightarrow$ נניח שמתקיים $\mathbb{P}(\{x\}|\theta) = h(x)g(\theta, T)$, אזי

$$\mathbb{P}(\{x\}|T = t, \theta) = \frac{\mathbb{P}(\{x\} \cap T = t|\theta)}{\mathbb{P}(T|\theta)}$$

נחלק ל2 אפשרויות:

אפשרות א: $t \neq T(\{x\})$, אזי $\mathbb{P}(\{x\}|T = t, \theta) = 0$ וזה לא תלוי ב-θ.

אפשרות ב: $t = T(\{x\})$, אזי מתקיים

$$\mathbb{P}(x|T, \theta) = \mathbb{P}(x|\theta) = h(x)g(\theta, T)$$

$$\mathbb{P}(T = t|\theta) = \sum_{\{x\}, T(x)=t} \mathbb{P}(x|\theta) = \sum_{\{x\}, T(x)=t} h(x)g(\theta, T)$$

$$\frac{\mathbb{P}(x, T|\theta)}{\mathbb{P}(T|\theta)} = \frac{h(x)g(\theta, T)}{\sum_{\{x\}, T(x)=t} h(x)g(\theta, T)} = \frac{h(x)}{\sum_{\{x\}, T(x)=t} h(x)} \Rightarrow \theta = \mathbb{P}(x|T) \text{ לא תלוי ב}$$

כנדרש. ■

2.3 משפחת ההתפלגויות האקספוננציאליות.

הגדרה 2.12. בהינתן אוסף של דגימות $x = \{x_1, x_2, ..., x_n\}$ ואומד $T = T(x) = T(\{x_1, x_2, ..., x_n\})$ ופרמטרים θ לא ידועים נגדיר את **משפחת הפונקציות האקספוננציאליות** להיות $\mathbb{P}(x|\theta) = h(x) \cdot e^{\mu(\theta)^T T(x) - A(\mu)}$ כאשר $A(\mu)$ פונקציה כלשהי, $\mu(\theta)^T$ מסמל שחלוף במידה ו-μ הוא ווקטור (הוא יכול להיות גם סקלר בודד).

דוגמה 2.13. ברנולי:

$$\mathbb{P}(x_i) = x_i p + (1 - x_i)(1 - p)$$
$$T(x) = \sum_{i=1}^{n} x_i$$

האם זה מחלק ממשפחה אקספוננציאלית? נבדוק:

$$\begin{aligned}\mathbb{P}(x) &= \binom{n}{x} \cdot \left(\frac{p}{1-p}\right)^{T(x)} \cdot (1-p)^n \\ &= \binom{n}{x} e^{T(x)\ln\left(\frac{p}{1-p}\right)} \cdot e^{n\ln((1-p))} \\ &= \binom{n}{x} \cdot e^{T(x)\ln\left(\frac{p}{1-p}\right) - n\ln\left(\frac{1}{1-p}\right)}\end{aligned}$$

ולכן זה ממשפחת הפונקציות האקספוננציאליות

$$h(x) = \binom{n}{x}, T(x) = \sum_{i=1}^{n} x_i, \mu(p) = \ln\left(\frac{p}{1-p}\right), A(\mu) = n\ln\left(\frac{1}{1-p}\right)$$

דוגמה 2.14. $x_i \sim \mathcal{N}(\mu, \sigma^2)$:

$$\mathbb{P}(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

וניקח $x = \{x_1, \ldots, x_n\}$ ונקבל

$$\begin{aligned}\mathbb{P}(\{x_1, \ldots, x_n\}) &= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right] = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i^2 - 2x_i\mu + \mu^2)}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_i \frac{x_i^2}{2\sigma^2} + \sum_{i=1}^{n}\frac{x_i\mu}{\sigma^2} - \frac{\mu^2 n}{2\sigma^2}} \underbrace{=}_{\star} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-T_1 \cdot \frac{1}{2\sigma^2} + T_2 \cdot \frac{\mu}{\sigma^2} - \frac{\mu^2 n}{2\sigma^2}}\end{aligned}$$

ואחרי שנגדיר

$$A(\mu, \sigma) = \frac{\mu^2 n^2}{2\sigma^2} \ln\left(\sqrt{2\pi}\sigma\right)$$

נקבל

$$\mathbb{P}(\{x\}) = 1 \cdot e^{\begin{bmatrix} T_1 & T_2 \end{bmatrix} \begin{bmatrix} -\frac{1}{2\sigma^2} \\ \frac{\mu}{\sigma^2} \end{bmatrix} - A(\mu,\sigma)}$$

ולכן זה ממשפחת ההתפלגויות האקספוננציאליות.

$\star$: נסמן $T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} \sum x_i^2 \\ \sum x_i \end{bmatrix}$.

2.4 סיכום.

כאשר נרצה להגיע לאומד מספיק, נרצה להגיע לצורה $\mathbb{P}(\{x_i\}) = h(x) \cdot e^{\mu^T T - A(\mu,\sigma)}$ ואז T יהיה אומד מספיק.

☻

3 הרצאה 3.

תזכורת: יש לי התפלגות עם פרמטרים θ לא ידועים, ויש לי פרמטרים $x_1, \ldots, x_n$ והמטרה שלי בחיים זה מתוך התצפיות לאמוד את הפרמטרים.

סימון: מכאן והלאה $\hat{\theta}$ הוא הערכה של θ בהתבסס על התצפיות.

דברים שאני רוצה מ-$\hat{\theta}$:

1. אומד מספיק.
2. אומד לא מוטה.
3. שהשונות שלו תהיה מינימלית.

דוגמה 3.1. ב3 זריקות של עץ או פלי יש לי פרמטר θ שאומר שבהסתברות p יצא לי H.

נגדיר 3 אומדים:

1. $\hat{\theta}_1 = \frac{\#H}{3}$ (כלומר מספר הפעמים שיצא לי H חלקי 3).
2. $\hat{\theta}_2 = \frac{I(H_1)+I(H_2)}{2}$.
3. $\hat{\theta}_3 = \frac{I(H_1)+2I(H_2)+2I(H_3)}{6}$.

אם יצא לי זריקות $[H, H, T]$ נקבל $\hat{\theta}_1 = \frac{2}{3}, \hat{\theta}_2 = 1, \hat{\theta}_3 = 0.5$ בעוד שאנחנו יודעים ש-$\theta = 0.3$.

נחשב את התוחלות של האומדים שלנו:

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_1] &= p \\
\mathbb{E}[\hat{\theta}_2] &= p \\
\mathbb{E}[\hat{\theta}_3] &= \frac{5}{6}p
\end{aligned}$$

כלומר האומד השלישי שלנו מוטה, והראשון והשני לא מוטים.

שאלה: מהו $\mathbb{P}(x_i|\theta, \hat{\theta}_1)$?

תשובה:

$$\mathbb{P}(x_i|\theta, \hat{\theta}_1) = \frac{\mathbb{P}(x_i, \hat{\theta}_1|\theta)}{\mathbb{P}(\hat{\theta}_1|\theta)} = \frac{\mathbb{P}(x_i, |\theta)}{\mathbb{P}(\hat{\theta}_1|\theta)} = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}$$

כאשר $k = \#H = 3\hat{\theta}_1$.

מצד שני, אם אני רוצה להשתמש ב$\hat{\theta}_2$ כדי לאמוד הסיכוי לרצף מסוים, נקבל

$$\mathbb{P}([\underbrace{\quad}_{1}, \underbrace{\quad}_{2}, \underbrace{\quad}_{3}]) = \mathbb{P}([\underbrace{\quad}_{1}, \underbrace{\quad}_{2}])\mathbb{P}([\underbrace{\quad}_{3}])$$

ואילו $\mathbb{P}([\underbrace{\quad}_{3}])$ תלוי ב-p ולכן הוא לא אומד מספיק טוב בשבילנו.

3.1 איך מייצרים אומד-DIY.

3.1.1 אומד נראות מירבית.

נניח זרקתי n פעמים מטבע בסיכוי p כאשר p לא ידוע, המטרה למצוא אומד ל-p.

התצפיות שלי הם $x_1, \ldots, x_n$. יש לי סדרה $[1, 1, 0, 0, 0, 1, \ldots]$.

אני אקבל $P(x_1) = p^{x_1}(1-p)^{1-x_1}$. לכן:

$$\begin{aligned}
\mathbb{P}(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \left[p^{x_i}(1-p)^{(1-x_i)}\right] \\
&= L(p|x_1, \ldots, x_n)
\end{aligned}$$

כאשר נגדיר את L להיות הנראות שלי, כלומר כמה סביר ש-p יהיה ערך כלשהו בהינתן הדגימות $x_1, ..., x_n$.

אני רוצה לקבל נראות מירבית לכן אני ארצה לגזור את L לקבל מקסימום, אבל לגזור מכפלה זה קשה, לכן במקום לחפש

$$\max(L(p|x_1, \ldots, x_n))$$

על ידי גזירה והשוואה אפס אני אחפש את

$$\log(L(p|x_1, \ldots, x_n))$$

על ידי גזירה והשוואה לאפס:

$$\log(L) = \sum_{i=1}^{n} \left(x_i \log(p) + (1-x_i)\log(1-p)\right) = k\log(p) + (n-k)\log(1-p)$$
$$\frac{d}{dp}\log(L) = 0 \Rightarrow \frac{k}{p} - \frac{n-k}{1-p} = 0 \Rightarrow p = \frac{k}{n}$$

כאשר $k = \sum_{i=1}^{n} x_i$. זה אומד נראות מירבי.

אלגוריתם 3.2. איך אני יכול בבית למצוא אומד נראות מירבית?

1. חשב נראות.
2. חשב $\log$ נראות.
3. גזור.
4. השווה ל-0.
5. וודא בעזרת נגזרת שנייה שאכן קיבלת מקסימום.

דוגמה 3.3. בהתפלגות נורמלית $\mathbb{P}(x_i) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ ולכן

$$\mathbb{P}(x_1,\ldots,x_n|\mu,\sigma) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right] = L(\mu,\sigma)$$
$$\log(L(\mu,\sigma)) = n\log\left(\frac{1}{\sqrt{2\pi\sigma}}\right) - \sum_{i=1}^{n}\left[\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

ואני רוצה לחשב את μ:

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{n}\left[\frac{(x_i-\mu)}{\sigma^2}\right] = 0 \Rightarrow \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu = 0 \Rightarrow \hat{\mu} = \sum_{i=1}^{n}\frac{x_i}{n}$$

ועכשיו אני רוצה לחשב את σ:

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n}\left[\frac{(x_i-\hat{\mu})^2}{\sigma^3}\right] = 0 \Rightarrow \sigma^2 = \frac{\sum_{i=1}^{n}\left[(x_i-\hat{\mu})^2\right]}{n} \Rightarrow \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}\left[(x_i-\hat{\mu})^2\right]}{n}}$$

3.1.2 שיטת המומנטים.

אלגוריתם. למציאת אומד בשיטת המומנטים:

1. חשב מומנטים.
2. חשב מומנטים מתוך ההתפלגות.
3. השווה ביניהם.

אם יש לי תצפיות $x_1,\ldots,x_n$, אזי

$$\widehat{M_1} = \sum_{i=1}^{n}\frac{x_i}{n}$$
$$\widehat{M_2} = \sum_{i=1}^{n}\frac{x_i^2}{n}$$
$$\vdots$$
$$\widehat{M_k} = \sum_{i=1}^{n}\frac{x_i^k}{n}$$

דוגמה 3.4. זרקתי n מטבעות עם סיכוי p כאשר $H=1, T=0$.

המומנט הראשון של ההתפלגות

$$M_1 = p$$

ולכן נשווה ביניהם ונקבל

$$p = \sum_{i=1}^{n}\frac{x_i}{n}$$

דוגמה 3.5. הגרלתי מספרים מהתפלגות נורמלית עם ממוצע μ וסתיית תקן σ. אני אקבל

$$M_1 = \mu$$
$$M_2 = \sigma^2 + \mu^2$$

ומההשוואה נקבל

$$\sum_{i=1}^{n} \frac{x_i}{n} = \mu$$
$$\sum_{i=1}^{n} \frac{x_i^2}{n} = \sigma^2 + \mu^2$$

3.1.3 אומד בייסיאני.

הקדמה: נניח זרקתי את המטבע שלי המון המון פעמים ויצא לי $[T,T,T,T,T,T,T,T,T]$ אז אני אקבל $\mathbb{P}(H) = \sum_{i=1}^{N} \frac{x_i}{N} = \frac{0}{9} = 0$.

הטלתי את המטבע עוד 5 פעמים וההטלות שלי כולן הן

$$[T,T,T,T,T,T,T,T,T,T,T,T,T,T,T]$$

נניח ואני לא רוצה לאמוד מה ה-p הסביר ביותר, אלא אני רוצה לקבל את ההתפלגות של p.

$$L(p) = c \cdot p^k (1-p)^{N-k}$$

כאשר c קבוע. לכן נכתוב

$$\mathbb{P}(\theta|\{x_i\}) = \frac{\mathbb{P}_0(\theta)\mathbb{P}(\{x_i\}|\theta)}{\mathbb{P}(\{x_i\})} = \frac{\mathbb{P}_0(\theta)L(\theta)}{\mathbb{P}(\{x_i\})}$$

כאשר נסמן $\mathbb{P}_0(\theta)$ להיות ה$prior$ שלנו. נסמן את ה-$\mathbb{P}_{AP}(\theta|\{x_i\}) = \underbrace{\mathbb{P}_0(\theta)}_{\text{תלוי באמונה}} \cdot \underbrace{p^k(1-p)^{N-k}}_{\text{תלוי בתצפיות}}$ להיות ההתפלגות $apostrior$.

הבעיה היא איך מלכתחילה אני יודע איך לבחור את $\mathbb{P}_0(p)$?

נבחר $\mathbb{P}_0(p) = \underbrace{p^a(1-p)^b}_{\text{ההתפלגות הצמודה של בינום}} \cdot c$. לכן נקבל

$$\mathbb{P}_{AP}(p) = c \cdot p^a(1-p)^b p^k (1-p)^{N-k} = \beta(a,b) p^k (1-p)^{N-k} = c \cdot p^{k+a-1}(1-p)^{N-k+b-1}$$

☻

4 הרצאה 4.

היום אנחנו רוצים למצוא את האומד הכי טוב, כלומר אומד מספיק וחסר הטיה, עם שונות מינימלית.

המטרה היא לקבל מינימום לביטוי

$$\mathbb{E}\left[\left(\hat{\theta}-\theta\right)^{2}\right]=\underbrace{\mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}\left[\hat{\theta}\right]\right)^{2}\right]}_{\mathbb{V}[\hat{\theta}]}+\mathbb{E}\Bigg[\underbrace{\left(\mathbb{E}\left[\hat{\theta}\right]-\theta\right)^{2}}_{:=B\left(\hat{\theta}\right)^{2}}\Bigg]+\underbrace{2\mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}\left[\hat{\theta}\right]\right)\left(\mathbb{E}\left[\hat{\theta}\right]-\theta\right)\right]}_{=0}$$

כלומר

$$\mathbb{E}\left[\left(\hat{\theta}-\theta\right)^{2}\right]=B\left(\hat{\theta}\right)^{2}+\mathbb{V}\left[\hat{\theta}\right]$$

האומד הכי טוב הוא כאשר $\mathbb{E}\left[\left(\hat{\theta}-\theta\right)^{2}\right]$ מינימלי, כלומר $\mathrm{ERR}(\hat{\theta})=B\left(\hat{\theta}\right):=\hat{\mu}-\theta,\mathbb{V}\left[\hat{\theta}\right]$ מינימליים.

שאלה 4.1. האם $\sum_{i=1}^{n}\frac{x_i}{n}$ הוא אומד חסר הטעיה?

$$\mathbb{E}\left[\sum_{i=1}^{n}\frac{x_i}{n}\right]=\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}x_i\right]=\frac{1}{n}\cdot n\cdot\mu=\mu$$

לכן

$$\mathbb{E}\left[\sum_{i=1}^{n}\frac{x_i}{n}-\mu\right]=0$$

כלומר הממוצע הוא אומד חסר הטעיה של התוחלת.

שאלה 4.2. מה לגבי השונות?

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\sum_{i=1}^{n}\frac{x_i}{n}\right)^{2}\right]&=\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\left(x_i-\sum_{i=1}^{n}\frac{x_i}{n}\right)^{2}\right]=\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}\left(\sum_{i=1}^{n}\frac{x_i}{n}\right)^{2}-2\sum_{i=1}^{n}\mathbb{E}\left[x_i\cdot\sum_{i=1}^{n}\frac{x_i}{n}\right]\right]\\&=\mathbb{E}\left[x^2\right]-\mathbb{E}\left[\left(\sum_{i=1}^{n}\frac{x_i}{n}\right)^{2}\right]=\mathbb{E}\left[x^2\right]-\mathbb{E}\left[\left(\sum_{i=1}^{n}\frac{x_i}{n}\sum_{j=1}^{n}\frac{x_j}{n}\right)\right]=\\&=\mathbb{E}\left[x^2\right]-\frac{1}{n^2}\left(\mathbb{E}\left[\left(\sum_{i=j=1}^{n}\mathbb{E}\left[x_ix_j\right]\right)\right]+\sum_{i\neq j=1}^{n}\mathbb{E}\left[x_ix_j\right]\right)\\&=\mathbb{E}\left[x^2\right]-\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[x^2\right]-\frac{1}{n^2}\sum_{i\neq j=1}^{n}\mathbb{E}\left[x_ix_j\right]\\&=\mathbb{E}\left[x^2\right]-\frac{1}{n}\mathbb{E}\left[x^2\right]-\frac{n(n-1)}{n^2}\mathbb{E}\left[x_ix_j\right]\\&=\frac{n-1}{n}\mathbb{V}\left[x\right]\end{aligned}$$

כלומר בהינתן אוסף תצפיות $x_1,\ldots,x_n$ אומד לא מוטה של השונות יהיה

$$\hat{V}=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-T\right)^2$$

דוגמה 4.3. יש לי דגימות $x_1,\ldots,x_4$ מניסוי ברנולי, ושני אומדים

$$\begin{aligned}\theta_1&=\frac{x_1+x_2}{2}\\\theta_2&=\frac{x_1+x_2+x_3}{3}\end{aligned}$$

אני רצה לאמוד תוחלת.

שניהם חסרי הטעיה.

שניהם לא אומדים מספיקים.

שונות:

$$\mathbb{V}\left[\hat{\theta}_1\right]=\mathbb{V}\left[\frac{x_1+x_2}{2}\right]=\frac{1}{2^2}\left(\mathbb{V}\left[x_1+x_2\right]\right)=\frac{1}{4}\left(\mathbb{V}\left[x_1\right]+\mathbb{V}\left[x_2\right]\right)=\frac{1}{4}\left(p+p\right)=\frac{p}{2}$$
$$\mathbb{V}\left[\hat{\theta}_2\right]=\mathbb{V}\left[\frac{x_1+x_2+x_3}{3}\right]=\frac{1}{3^2}\mathbb{V}\left[x_1+x_2+x_3\right]=\frac{1}{9}\left(\mathbb{V}\left[x_1\right]+\mathbb{V}\left[x_2\right]+\mathbb{V}\left[x_3\right]\right)=\frac{1}{9}\left(p+p+p\right)=\frac{p}{3}$$

כלומר $\mathbb{V}\left[\hat{\theta}_1\right]<\mathbb{V}\left[\hat{\theta}_2\right]$ אבל האם $\hat{\theta}_2$ הכי יעיל?

4.1 יעילות של אומד.

בהינתן θ ותצפיות $x_1,\dots,x_n$ ואומד $\hat{\theta}(x_1,\dots,x_n)$ נגדיר

$$\mathrm{EFF}(\hat{\theta})=\min_{\tilde{\theta}}\frac{\mathrm{ERR}(\tilde{\theta})}{\mathrm{ERR}(\hat{\theta})}\leq 1$$

המטרה: למצוא אומד עבורו $\mathrm{EFF}(\hat{\theta})=1$.

הגדרה 4.4. אם האומד לא מוטה

$$\mathrm{EFF}(\hat{\theta})=\min_{\tilde{\theta}}\frac{\mathbb{V}\left[\tilde{\theta}\right]}{\mathbb{V}\left[\hat{\theta}\right]}$$

אומד שהוא מספיק ולא מוטה עם יעילות של 1 יקרא $UMVEA$.

טענה 4.5. לכל פרמטר θ קיים חסם תחתון ל-$\mathbb{V}\left[\theta\right]$.

הגדרה 4.6. בהינתן תצפיות X ופרמטר θ וסיכוי $f(X|\theta)$ נגדיר

$$W(X)=\frac{\partial}{\partial\theta}\log(f(X|\theta))=\frac{\frac{\partial}{\partial\theta}f(X|\theta)}{f(X|\theta)}$$

טענה 4.7. $\mathbb{E}\left[W\right]=0$.

הוכחה. נלך לפי ההגדרה:

$$\begin{aligned}\mathbb{E}\left[W\right]&=\int f(X|\theta)W(X)dX\\&=\int f(X|\theta)\frac{\frac{\partial}{\partial\theta}f(X|\theta)}{f(X|\theta)}dX\\&=\frac{\partial}{\partial\theta}\int f(X|\theta)dX\\&=\frac{\partial}{\partial\theta}(1)\\&=0\end{aligned}$$

כנדרש. ■

טענה 4.7. $\mathbb{V}\left[W\right]=\mathbb{E}\left[-\frac{\partial}{\partial\theta}W\right]$.

הוכחה. שוב פעם נלך לפי ההגדרה:

$$\mathbb{V}\left[W\right]=\mathbb{E}\left[W^2\right]-\mathbb{E}\left[W\right]^2=\mathbb{E}\left[W^2\right]$$
$$\begin{aligned}\mathbb{E}\left[-\frac{\partial}{\partial\theta}W\right]&=\mathbb{E}\left[-\frac{\partial}{\partial\theta}\frac{\frac{\partial}{\partial\theta}f(X|\theta)}{f(X|\theta)}\right]\\&=\mathbb{E}\left[-\frac{\frac{\partial^2}{\partial\theta^2}f(X|\theta)\cdot f(X|\theta)-\left[\frac{\partial}{\partial\theta}f(X|\theta)\right]^2}{f(X|\theta)^2}\right]\\&=\mathbb{E}\left[-\frac{\frac{\partial^2}{\partial\theta^2}f(X|\theta)}{f(X|\theta)}\right]+\mathbb{E}\left[W^2\right]\\&\underset{\star}{=}\mathbb{E}\left[W^2\right]\end{aligned}$$

כנדרש.

$\star$: $\mathbb{E}\left[-\frac{\frac{\partial^2}{\partial\theta^2}f(X|\theta)}{f(X|\theta)}\right] = -\int \frac{\frac{\partial^2}{\partial\theta^2}f(X|\theta)}{f(X|\theta)}\cdot f(X|\theta)dX = \frac{\partial^2}{\partial\theta^2}\int f(X|\theta)dX = \frac{\partial^2}{\partial\theta^2}(1) = 0$ ■

הגדרה 4.8. אינפורמצית פישר של התפלגות θ תהיה

$$I(\theta) = \mathbb{V}\left[W\right] = \mathbb{E}\left[-\frac{\partial}{\partial\theta}W\right]$$

משפט 4.9. לכל אומד חסר הטיה T של $\tau(\theta)$ מתקיים

$$\mathbb{V}\left[\tau\right] \geq \frac{1}{I(\tau(\theta))}$$

הערה 4.10. לפני ההוכחה, נשים ♡ לכמה דברים:

1. מתקיים

$$\begin{aligned}\frac{\partial}{\partial\theta}\log(f(x_1,\ldots,x_n)) =& \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log(f(x_i|\theta))\\ W(x_1,\ldots,x_n,\theta) =& \sum_{i=1}^{n}W(x_i|\theta)\\ \mathbb{V}\left[W(x_1,\ldots,x_n|\theta)\right] =& \mathbb{V}\left[\sum_{i=1}^{n}W(x_i|\theta)\right]\\ I(x_1,\ldots,x_n) =& n\cdot I(x_1)\end{aligned}$$

2. נניח שניתן לכתוב את $W = \frac{\partial}{\partial\theta}\log(f(x|\theta)) = k(\theta,n)\left(T(x)-\tau(\theta)\right)$.
ידוע כי $\mathbb{E}\left[W\right]=0$, לכן

$$\begin{aligned}&\mathbb{E}\left[k(\theta,n)\left(T(x)-\tau(\theta)\right)\right] = 0\\ &k(\theta,n)\left(\mathbb{E}\left[T(x)\right]-\tau(\theta)\right) = 0\\ &\Leftarrow T(x)\text{ אומד לא מוטה}\\ &\Leftarrow T(x)\text{ אומד מספיק}\end{aligned}$$

3. הערה מהסתברות:

$$\mathrm{cov}(x,y) = \mathbb{E}\left[xy\right]-\mathbb{E}\left[x\right]\mathbb{E}\left[y\right] \leq \sigma(x)\sigma(y)$$

ואם מתקיים $\mathbb{E}\left[y\right]=0$

$$\mathrm{cov}(xy) = \mathbb{E}\left[xy\right]$$

ואז

$$\mathbb{E}\left[xy\right]^2 \leq \mathbb{V}\left[x\right]\mathbb{V}\left[y\right]$$

הוכחה. יהי T אומד חסר הטייה של $\tau(\theta)$.
(א) $I(\tau)\mathbb{V}\left[\tau\right]\geq 1 \Rightarrow \mathbb{V}\left[\tau\right]\geq\frac{1}{I(\tau)}$ וקיים שוויון $\iff$ **(ב)** $W(x,\tau) = k(\theta,n)\left(T(x)-\tau(\theta)\right)$.

$$I(\tau)\mathbb{V}\left[\tau\right] = \mathbb{V}\left[W\right]\mathbb{V}\left[T\right] \geq \mathbb{E}\left[WT\right]^2$$

כלומר אם נוכיח ש-$\mathbb{E}\left[WT\right]^2 = 1$ נקבל את **(א)**.

$$\begin{aligned}E(WT) &= \int f(X|\theta)T(x)\frac{\frac{\partial}{\partial\tau}f(X|\theta)}{f(X|\theta)}dX\\ &= \frac{\partial}{\partial\tau}\int T(X)f(X|\theta)dX\\ &= \frac{\partial}{\partial\tau}\mathbb{E}\left[T(X)\right]\\ &= \frac{\partial\tau}{\partial\tau}\\ &= 1\end{aligned}$$

כלומר

$$I(\tau)\mathbb{V}\left[\tau\right] \geq 1$$

כנדרש. ■

תזכורת:

$$\text{cov}(xy)^2 = \mathbb{V}[x]\,\mathbb{V}[y] \iff y = ax+b$$

כלומר

$$1 = \text{cov}(T,W) \underset{W=k(T-c)}{=} \mathbb{V}[T]\,\mathbb{V}[W]$$

$$\frac{\partial}{\partial \tau}\log(f(x|\theta)) = k(T-c)$$

$$\mathbb{E}[W] = 0 = k(T-c)$$

אם T אומד חסר הטייה של τ אזי $c=\tau$. ואז

$$W = K(\theta,n)\,(T-\tau)$$

כלומר W מהצורה הנ"ל הוא בהכרח הUMVUE היחיד.

$$\frac{\partial}{\partial \tau}\log(f(x|\theta)) = k(\theta,n)(T(x)-\tau) \setminus \int()$$

$$\log(f(x|\theta)) = \int k(\theta,n)(T(x)-\tau)d\tau + c(x) \setminus e^{()}$$

$$f(x|\theta) = \underbrace{e^{c(x)}}_{:=h(x)} e^{\int k(\theta,n)(T(x)-\tau)d\tau}$$

$$f(x|\theta) = h(x)e^{\int k(\theta,n)(T(x)-\tau)d\tau}$$

מה עושים אם אין לי צורה אקספוננציאלית? נניח $p(x) = cx^{-\beta}e^{-nx}$ ולזה אין צורה אקספוננציאלית.

משפט 4.11. אם $\hat{\theta}$ אומד של θ ו-T אומד מספיק של θ אז $\theta^* = \mathbb{E}\left[\hat{\theta}|T\right]$, $\text{ERR}(\theta^*) \le \text{ERR}(\hat{\theta})$.

הוכחה. ממש לפי ההגדרה:

$$\mathbb{V}[x] \ge 0 = \mathbb{E}\left[x^2\right] - \mathbb{E}[x]^2$$

$$\Downarrow$$

$$\mathbb{E}\left[x^2\right] \ge \mathbb{E}[x]^2$$

$$\Downarrow$$

$$\mathbb{E}\left[\left(\hat{\theta}-\theta\right)^2|T\right] \ge \mathbb{E}\left[\hat{\theta}-\theta|T\right]^2 = \left[\mathbb{E}\left[\hat{\theta}|T\right] - \mathbb{E}[\theta|T]\right]^2 = (\theta^*-\theta)^2$$

$$\Downarrow$$

$$\mathbb{E}\left[\mathbb{E}\left[\left(\hat{\theta}-\theta\right)^2|T\right]\right]_T \ge \mathbb{E}\left[(\theta^*-\theta)^2\right]$$

$$\Downarrow$$

$$\text{ERR}(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta}-\theta\right)^2\right] \ge (\theta^*-\theta) = \text{ERR}(\theta^*)$$

כנדרש. ■

5 הרצאה 5.

תכונות של אומדים.

1) אומד מספיק:

$$\mathbb{P}(\{x_i\}|\hat{\theta},\theta)=\mathbb{P}(\{x_i\}|\hat{\theta})$$

ומשפט הפירוק של ניימן פישר.

2) אומד לא מוטה $\Leftarrow B(\hat{\theta})=0 \Leftarrow \hat{\mu}=\theta$.

3) $\mathbb{V}\left[\hat{\theta}\right]$ יהיה מינימלי.

$$\mathrm{EFF}(\hat{\theta})=\min_{\tilde{\theta}}\frac{\left(\mathbb{V}\left[\tilde{\theta}\right]\right)}{\mathbb{V}\left[\hat{\theta}\right]}\leq 1$$

כאשר $\tilde{\theta}$ אומד מספיק

3) חסם קרמר ראו:

$$\mathbb{V}\left[\hat{\theta}\right]\geq\frac{1}{I(\theta)}$$

כאשר

$$I(\theta)=\mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log\left(P(x|\theta)\right)\right]$$

דוגמה 5.1. אינפורמציית פישר למשתנה מקרי פואסון:

$$X\sim\mathrm{Pois}(\lambda)$$

לפיכך

$$\mathbb{P}(X=k|\lambda)=\frac{e^{-\lambda}\lambda^k}{k!}$$

כלומר

$$\begin{aligned}
\mathbb{P}(k|\lambda)&=\frac{e^{-\lambda}\lambda^k}{k!}\\
\log\left(P(k|\lambda)\right)&=-\lambda+k\log\lambda-\log(k!)\\
\frac{\partial}{\partial\lambda}\log\left(P(k|\lambda)\right)&=-1+\frac{k}{\lambda}\\
\frac{\partial^2}{\partial\theta^2}\log\left(P(k|\lambda)\right)&=-\frac{k}{\lambda^2}\\
\mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log\left(P(x|\theta)\right)\right]&=\mathbb{E}\left[\frac{k}{\lambda^2}\right]=\frac{1}{\lambda^2}\mathbb{E}\left[k\right]=\frac{\lambda}{\lambda^2}=\frac{1}{\lambda}
\end{aligned}$$

כלומר

$$I(\lambda)=\frac{1}{\lambda}$$

ולכן לכל אומד יתקיים

$$\mathbb{V}\left[\hat{\theta}\right]\geq\lambda$$

כלומר אם סתם הלכתי ברחוב, בקמפוס ופגשתי באומד עם שונות λ אזי זהו האומד הכי טוב שאני יכול לקבל.

דוגמה 5.2. למשפחה אקספוננציאלית:

$$\begin{aligned}
\mathbb{P}(k)&=\binom{N}{k}\rho^k(1-\rho)^{N-k}\\
&=\binom{N}{k}\left(\frac{\rho}{1-\rho}\right)^k(1-\rho)^N\\
&=\binom{N}{k}e^{-k\ln\left(\frac{1}{\rho}-1\right)+N\ln(1-\rho)}
\end{aligned}$$

משפט 5.3. (ראו בלקלייל) אם T אומד מספיק ו-$\hat{\theta}$ אומד לא מוטה, אזי

$$\theta^*=\mathbb{E}\left[\hat{\theta}|T\right]$$

אומד עם שונות נמוכה יותר משל $\hat{\theta}$, כלומר $\mathbb{V}\left[\hat{\theta}\right]\geq\mathbb{V}\left[\theta^*\right]$ וגם $\mathbb{E}\left[\theta^*\right]=\theta$.

הגדרה 5.4. משפחה של התפלגויות היא **שלמה** אם מתקיים

$$\forall\theta : \mathbb{E}\left[U(x)\right]_{\theta} = 0 \Rightarrow \mathbb{V}\left[x\right] = 0$$

הגדרה 5.5. אומד יקרא **נלוה** אם הוא אומד שלא תלוי בθ.

משפט 5.6. (בסו) במשפחה שלמה אם T_1מספיק ו-T_2 נלוה אז T_1, T_2 בלתי תלויים.

הוכחה. יהי T_1 אומד נלווה ו-T_2 אומד מספיק, כך שT_1, T_2 בלתי תלויים. לכל θ מתקיים

$$\begin{aligned}
\mathbb{E}\left[\varphi(T_1)|T_2\right]_{\theta} &= \mathbb{E}\left[\varphi(T_1)\right]_{\theta} \\
\mathbb{E}\left[\varphi(T_1)\right]_{\theta} &= c \ \ (\text{לא תלוי ב}\theta) \\
\mathbb{E}\left[\varphi(T_1)|T_2\right]_{\theta} &= c \ \ (\text{לא תלוי ב}\theta) \\
\mathbb{E}\left[\mathbb{E}\left[\varphi(T_1)|T_2\right]_{\theta}\right]_{T_2} &= \mathbb{E}\left[\varphi(T_1)\right]_{\theta} \\
\forall\theta : \mathbb{E}\left[\eta(T_2)\right] &= c \\
\forall\theta : \mathbb{E}\left[\eta(T_2) - c\right] &= 0
\end{aligned}$$

כלומר

$$\eta(T_2) = c$$

ואז נקבל

$$\mathbb{E}\left[\varphi(T_1)|T_2\right]_{\theta} = \mathbb{E}\left[\varphi(T_1)\right]_{\theta}$$

כלומר T_1, T_2 ב"ת, כנדרש. ■

משפט 5.7. **(להמן שפה)** במשפחה שלמה, אם $\hat{\theta}$ הוא אומד לא מוטה של θ ו-T אומד מספיק של θ אז

$$\theta^* = \mathbb{E}\left[\hat{\theta}|T\right]$$

הוא אומד לא מוטה בעל שונות מינימלית.

הוכחה. יהי $\tilde{\theta}$ אומד לא מוטה כלשהו של θ ו-T אומד מספיק. יהי $\theta^{\#} = \mathbb{E}\left[\tilde{\theta}|T\right]$. לפי ראו בלקלייל $\mathbb{E}\left[\theta^{\#}\right] = \theta$ וגם $\mathbb{V}\left[\theta^{\#}\right] \leq \mathbb{V}\left[\tilde{\theta}\right]$.

$$\mathbb{E}\left[\theta^*\right] = \theta \Rightarrow \mathbb{E}\left[\theta^*\right] - \mathbb{E}\left[\theta^{\#}\right] = 0 \Rightarrow \forall\theta : \mathbb{E}\left[\theta^* - \theta^{\#}\right] = 0 \Rightarrow \theta^* - \theta^{\#} = 0 \Rightarrow \theta^* = \theta^{\#}$$

לכן לכל אומד לא מוטה $\tilde{\theta}$ מתקיים

$$\mathbb{V}\left[\theta^*\right] \leq \mathbb{V}\left[\tilde{\theta}\right]$$

כנדרש. ■

5.1 רווח סמך.

יש לי θ רציף.

יש לי $\hat{\theta}$: אומד מספיק, לא מוטה ובעל שונות מינימלית.

נניח $\hat{\theta} = 5.3$, מה הסיכוי ש-$\theta = 5.3$?

תשובה: 0! כיוון ש-θ רציף אז ההסתברות שהוא יהיה שווה בדיוק ל-5.3 היא 0.

נשנה טיפה את השאלה שלי: נשאל מה הסיכוי $\mathbb{P}(|\theta - \hat{\theta}| \le \Delta)$?

מהגדרת הערך המוחלט נקבל

$$\begin{aligned}\mathbb{P}(|\theta - \hat{\theta}| \le \Delta) &= \mathbb{P}(\theta \ge \hat{\theta} - \Delta) \cdot \mathbb{P}(\theta \le \hat{\theta} + \Delta)\\ &= \mathbb{P}(\hat{\theta} \ge \theta - \Delta) \cdot \mathbb{P}(\hat{\theta} \le \theta + \Delta)\end{aligned}$$

הבעיה היא שאני לא יודע לחשב את $\mathbb{P}(\theta|\hat{\theta})$, אבל הפתרון הוא שאני יודע לחשב את $\mathbb{P}(\hat{\theta}|\theta)$.

דוגמה 5.8. יהיו $x_1, \ldots, x_n \sim \mathcal{N}(\mu, \sigma)$ וניקח את $T = \sum_{i=1}^{n} \frac{x_i}{n}$ לפיכך $T \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$. אני רוצה את $\mathbb{P}(\mu|T)$, אבל לא למדתי איך לחשב את זה!

אני יודע רק $\mathbb{P}(T|\mu)$.

$$\mathbb{P}(T) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{(T-\mu)^2}{2\sigma^2/n}}$$

ננסה לחשב את $\mathbb{P}(|T - \mu| < \Delta)$:

$$\begin{aligned}\mathbb{P}(|T - \mu| < \Delta) &= \int_{\mu-\Delta}^{\mu+\Delta} \mathbb{P}(T)dT\\ &= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{\mu-\Delta}^{\mu+\Delta} e^{-\frac{(T-\mu)^2}{2\sigma^2/n}} dT\\ &= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{-\Delta}^{\Delta} e^{-\frac{T'^2}{2\sigma^2/n}} dT'\end{aligned}$$

לכן נרצה

$$\mathbb{P}(\hat{\theta} \in [\theta - \Delta, \theta + \Delta]) \ge 1 - \alpha$$

כאשר α נקראה **רמת האמינות**. בד"כ $\alpha = 0.05$.

למה בעצם אני רוצה את התחום $[\theta - \Delta, \theta + \Delta]$ ולא אינטרוול אחר שהוא לא סימטרי?

משפט 5.9. האינטרוול הסימטרי $[\theta - \Delta, \theta + \Delta]$ הוא האינטרוול הכי קטן עבור רמת וודאות α.

5.2 התפלגות Γ.

הגדרה 5.10. נאמר כי $X \sim \Gamma(\alpha, \beta)$ **מתפלג גאמה** אם מתקיים $\mathbb{P}(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$ (וזוהי פונקציית ההסתברות Γ).

$$\begin{aligned}\mathbb{E}\left[y^k\right] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^k y^{\alpha-1} e^{-\beta y} dy\\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha+k-1} e^{-\beta y} dy\\ &= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{\beta^\alpha}{\Gamma(\alpha+k)} \int_0^\infty y^{\alpha+k-1} e^{-\beta y} dy\\ &= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\end{aligned}$$

6 הרצאה 6.

6.1 מלא מלא התפלגויות חדשות!

תזכורת: יש לי פרמטר θ שאני לא יודע לחשב, לכן אני מחשב פרמטר $\hat{\theta}$ שבא מהתצפיות שלי $x_1, \ldots, x_k$.

אומד לא מוטה הוא $\hat{\theta}$ כך ש-$\mathbb{E}\left[\hat{\theta}\right] = \theta$.

אני רוצה ש-$\hat{\theta}$ יהיה הכי קרוב ל-θ, כלומר לבדוק מה זה $\mathbb{P}(|\hat{\theta} - \theta| < \varepsilon)$ שזה $\mathbb{P}(\hat{\theta} \geq \theta - \varepsilon, \hat{\theta} \leq \theta + \varepsilon)$.

הגדרה 6.1. לקטע $[\theta - \varepsilon, \theta + \varepsilon]$ קוראים **רווח סמך**.

6.1.1 התפלגות Γ.

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \Rightarrow 1 = \frac{1}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x} dx \underset{\substack{x = \beta y \\ dx = \beta dy}}{\Longrightarrow} 1 = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-\beta y} dy$$

פונקציית ה-CDF:

$$\mathbb{P}(X \leq t) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^t y^{\alpha-1} e^{-\beta y} dy$$

פונקציית ה-PDF:

$$\mathbb{P}(X = t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$$

דוגמה 6.2. מומנטים של התפלגות $\Gamma(\alpha, \beta)$:

$$\begin{aligned}
\mathbb{E}\left[y^k\right] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^k y^{\alpha-1} e^{-\beta y} dy \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{k+\alpha-1} e^{-\beta y} dy \\
&= \frac{\Gamma(\alpha+k)}{\beta^{\alpha+k}} \frac{\beta^\alpha}{\Gamma(\alpha)} \underbrace{\int_0^\infty y^{k+\alpha-1} e^{-\beta y} \cdot \frac{\beta^{\alpha+k}}{\Gamma(\alpha+k)} dy}_{=1} \\
&= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\beta^k}
\end{aligned}$$

כלומר

$$\boxed{M_k = \beta^{-k} \prod_{i=0}^{k-1} (\alpha + i)}$$

פונקציה יוצרת מומנטים :

$$\begin{aligned}\mathbb{E}\left[e^{ty}\right] &= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\int_{0}^{\infty} e^{ty}y^{\alpha-1}e^{-\beta y}dy \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\int_{0}^{\infty} y^{\alpha-1}e^{(t-\beta)y}dy \\ &= \frac{\beta^{\alpha}}{(\beta-t)^{\alpha}}\underbrace{\frac{1}{\Gamma(\alpha)}\int_{0}^{\infty} y^{\alpha-1}e^{(t-\beta)y}\cdot(\beta-t)^{\alpha}dy}_{=1}\end{aligned}$$

כלומר

$$\boxed{M_k(t) = \left(\frac{\beta}{\beta-t}\right)^{\alpha}}$$

דוגמה 6.3. $x_i \sim \Gamma(\alpha_i,\beta)$ מתפלגים Γ עם α_i ואותו β ב"ת, יהי $z=\sum_{i=1}^{n} x_i$, איך z מתפלג?

פתרון. ניעזר בפונקציה יוצרת מומנטים :

$$\mathbb{E}\left[e^{tz}\right] = \prod_{i=1}^{n}\left(\frac{\beta}{\beta-t}\right)^{\alpha_i} = \left(\frac{\beta}{\beta-t}\right)^{\Sigma_i\alpha_i}$$

כלומר

$$z \sim \Gamma\left(\sum_{i=1}^{n}\alpha_i,\beta\right)$$

כנדרש.

6.1.2 התפלגות χ^2.

נניח $x_i \sim \mathcal{N}(0,1)$, בואו נוכיח כי $x^2 \sim \Gamma\left(\frac{1}{2},\frac{1}{2}\right)$. זה אומר כי $\hat{\theta} = \sum_{i=1}^{n} x_i^2$ מתפלג $\hat{\theta} \sim \Gamma\left(\frac{n}{2},\frac{1}{2}\right)$.

הוכחה :

$$\mathbb{P}_{\text{PDF}}(x^2<t) = P(-\sqrt{t}\le x\le\sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$$\begin{aligned}\mathbb{P}_{\text{CDF}}(x^2=t) &= \frac{d}{dt}P_{cdf}(x^2\le t) \\ &= \frac{1}{\sqrt{2\pi}}\left[e^{-\frac{x^2}{2}}\Big|_{\sqrt{t}}\cdot\frac{d\sqrt{t}}{dt} - e^{-\frac{x^2}{2}}\Big|_{-\sqrt{t}}\cdot\frac{d\left(-\sqrt{t}\right)}{dt}\right] \\ &= \frac{1}{\sqrt{2\pi}}\left[e^{-\frac{t}{2}}\frac{1}{2\sqrt{t}} + e^{-\frac{t}{2}}\frac{1}{2\sqrt{t}}\right] \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{t}{2}}t^{\frac{1}{2}-1} \\ &= \Gamma\left(\frac{1}{2},\frac{1}{2}\right)\end{aligned}$$

לכן אם $x_i \sim \mathcal{N}(0,1)$ אזי $x_i^2 \sim \chi_1^2 = \Gamma\left(\frac{1}{2},\frac{1}{2}\right)$ ואז $\sum_{i=1}^{n} x_i^2 = \chi_n = \Gamma\left(\frac{n}{2},\frac{1}{2}\right)$.

הערה 6.4. התפלגות χ_k^2 (ה-k אומר k דרגות חופש) היא ההתפלגות

$$\sum_{i=1}^{k} x_i^2, x_i \sim \mathcal{N}(0,1) = \Gamma\left(\frac{k}{2},\frac{1}{2}\right)$$

6.1.3 התפלגות F.

הקדמה: יש לי כמה אומדים ואני עכשיו אקבל את

אומד א (של משה): $\hat{\theta}_1 = \sum\limits_{i=1}^{k} \frac{x_i^2}{k}$

אומד ב (של יורם): $\hat{\theta}_2 = \sum\limits_{i=1}^{n} \frac{y_i^2}{n}$

כאשר $x_1, \ldots x_k, y_1, \ldots, y_n \sim \mathcal{N}(0,1)$. אני עכשיו רוצה לבדוק של מי האומד היותר טוב, כלומר לחשב את $\mathbb{P}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right)$.

דוגמה ריאליסטית לסיטואציה כזאת:

מודל א': הסיכוי שנבחרת x תנצח נבחרת y הוא מספר הפעמים שאמבפה משתתף בנבחרת x פחות מספר הפעמים בנבחרת y.

מודד ב': הסיכוי שנבחרת x תנצח נבחרת y הוא מספר השערים הממוצע של נבחרת x חלקי מספר השערים הממוצע של נבחרת y.

מי מהם מודל יותר טוב? למי אני אמור להאמין יותר?

ניזכר:

$$\begin{aligned} \hat{\theta}_1 &\sim \Gamma\left(\tfrac{k}{2}, \tfrac{1}{2}\right) \\ \hat{\theta}_2 &\sim \Gamma\left(\tfrac{n}{2}, \tfrac{1}{2}\right) \end{aligned}$$

לכן

$$\begin{aligned} \mathbb{E}\left[\hat{\theta}_1\right] &= \mathbb{E}\left[\sum_{i=1}^{k} \frac{x_i^2}{k}\right] = \frac{1}{k} \sum_{i=1}^{k} \mathbb{E}\left[x_i^2\right] = 1 \\ \mathbb{E}\left[\hat{\theta}_2\right] &= \mathbb{E}\left[\sum_{i=1}^{n} \frac{y_i^2}{n}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[y_i^2\right] = 1 \end{aligned}$$

נסמן את ההתפלגויות של $\hat{\theta}_1$ ו-$\hat{\theta}_2$ כתור g ו-f בהתאמה, ונחשב את ה-CDF של $\frac{\hat{\theta}_1}{\hat{\theta}_2}$:

$$\mathbb{P}_{\text{CDF}}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2} \le t\right) = \mathbb{P}\left(\hat{\theta}_1 \le t \cdot \hat{\theta}_2\right) = \int_0^{\infty}\int_0^{t\cdot\hat{\theta}_2} f\left(\hat{\theta}_1\right) g\left(\hat{\theta}_2\right) d\hat{\theta}_1 d\hat{\theta}_2 = \int_0^{\infty} f\left(\hat{\theta}_1\right) \int_0^{t\cdot\hat{\theta}_2} g\left(\hat{\theta}_2\right) d\hat{\theta}_1 d\hat{\theta}_2$$

$$\begin{aligned} \mathbb{P}_{\text{PDF}}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2} = t\right) &= \frac{d}{dt}\mathbb{P}_{CDF}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2} \le t\right) = \int_0^{\infty} f\left(\hat{\theta}_2\right) g\left(t\cdot\hat{\theta}_2\right) \hat{\theta}_2 d\hat{\theta}_2 \\ &= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}+\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} e^{-\frac{1}{2}\hat{\theta}_2} \cdot \hat{\theta}_2^{\left(\frac{n}{2}-1\right)} \cdot e^{-\frac{1}{2}t\hat{\theta}_2} \left(t\hat{\theta}_2\right)^{\left(\frac{n}{2}-1\right)} \hat{\theta}_2 d\hat{\theta}_2 \\ &= \frac{t^{\frac{k}{2}-1}\left(\frac{1}{2}\right)^{\frac{n}{2}+\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} e^{-\frac{1}{2}(t+\hat{1})\theta_2} \cdot \hat{\theta}_2^{\left(\frac{n}{2}-1+\frac{k}{2}\right)} d\hat{\theta}_2 \\ &\underbrace{=}_{\star} \frac{\Gamma\left(\frac{n}{2}+\frac{k}{2}\right)}{(t+1)^{\frac{n}{2}+\frac{k}{2}}} \frac{t^{\frac{k}{2}-1}\left(\frac{1}{2}\right)^{\frac{n}{2}+\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)} \underbrace{\int_0^{\infty} e^{-\frac{1}{2}(t+\hat{1})\theta_2} \cdot \hat{\theta}_2^{\left(\frac{n}{2}-1+\frac{k}{2}\right)} \cdot \frac{(t+1)^{\frac{n}{2}+\frac{k}{2}}}{\Gamma\left(\frac{n}{2}+\frac{k}{2}\right)} d\hat{\theta}_2}_{=1} \\ &= \frac{t^{\frac{k}{2}-1}}{(t+1)^{\frac{n}{2}+\frac{k}{2}}} \cdot \frac{\Gamma\left(\frac{n}{2}+\frac{k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)} \end{aligned}$$

$\star$: זוהי התפלגות $\Gamma\left(\frac{n}{2}+\frac{k}{2}, \frac{t+1}{2}\right)$, לכן האינטגרל עליה שווה ל-1!

לכן נגדיר את התפלגות F להיות

$$F(t,k,n) = \frac{t^{\frac{k}{2}-1}}{(t+1)^{\frac{n}{2}+\frac{k}{2}}} \cdot \frac{\Gamma\left(\frac{n}{2}+\frac{k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

6.1.4 התפלגות T.

נניח יש לי $x_i \sim \mathcal{N}(\mu, 1)$ אז $\hat{\theta_1} = \sum_{i=1}^{n} \frac{x_i}{n} \sim \mathcal{N}(\mu, \frac{1}{\sqrt{n}})$. מה אם יש לי בנוסף $x \sim \mathcal{N}(0, \sigma)$?

עכשיו $\hat{\theta_2} = \hat{\theta_1} + x \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ כלומר $\mathbb{P}(\hat{\theta_2}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\hat{\theta_1}-\mu)^2}{\frac{2\sigma^2}{n}}}$ כאשר σ בכלל לא ידוע לי! אני רוצה לתקן את $\hat{\theta_2}$:

$$z = \frac{\hat{\theta_2} - \mu}{\hat{\sigma}}$$

כאשר

$$\hat{\sigma} = \sqrt{\left(\sum_{i=1}^{n} x_i^2\right) + x^2}$$

ואז

$$\mathbb{P}(\hat{\theta_2}) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} e^{-\frac{z^2}{2n}}$$

כאשר הכל אני יודע!

נניח עכשיו יש לי $x \sim \mathcal{N}(0, 1)$ וסדרת משתנים מקריים $y_1, \ldots, y_n \sim \mathcal{N}(0, 1)$, ונגדיר אומד

$$T = \frac{x}{\sqrt{\frac{1}{n} \sum\limits_{i=1}^{n} y_i^2}}$$

איך T מתפלג?

$$\begin{aligned}
\mathbb{P}_{\text{CDF}}(-z \le T \le z) &= \mathbb{P}_{\text{CDF}}(T^2 \le z^2) \\
&= \mathbb{P}\left(\frac{x^2}{\frac{1}{n} \sum\limits_{i=1}^{n} y_i^2} \le z^2\right) \\
&= \mathbb{P}\left(F_{1,m}(x) \le z^2\right)
\end{aligned}$$

כעת נסמן את ההתפלגות של T ב-g (ונזכור ש-g זוגית):

$$\int_{-z}^{z} g(T)dT = \int_{0}^{z^2} F_{1,n}(x)dx \quad \Big\backslash \frac{d}{dz}$$

$$\Downarrow$$

$$z \cdot g(z) + z \cdot g(-z) = 2F_{1,n}(z^2) \cdot z$$

$$\Downarrow$$

$$g(z) = z \cdot F_{1,n}(z^2) = \frac{\Gamma\left(\frac{1}{2} + \frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}$$

6.2 רווחי סמך המשך.

הגדרה 6.5. רווח סמך ברמת אמינות של α הוא מקטע $[U,L]$ המקיים $\mathbb{P}\left(\hat{\theta}\in[L,U]\right)\geq 1-\alpha$.

הלכה למעשה אנחנו מבקשים

$$U=U(\hat{\theta})$$
$$L=L(\hat{\theta})$$
$$\mathbb{P}\left(\theta\in[L,U]\right)\geq 1-\alpha,\forall\theta$$

משפט 6.6. בהינתן התפלגות אונימודלית ($f(x)$ מונוטונית עולה ב-$(-\infty,a)$ וגם $f(x)$ מונוטונית יורדת ב-(a,∞)) אז הרווח סמך הקטן ביותר הוא המרווח $I=f(x)\geq\eta$ ואז

$$\int_I f(x)=1-\alpha$$

הוכחה. ניקח רווח סמך $I=[L,U]$ באמינות α ויהי רווח סמך אחר $\tilde{I}=\left[\tilde{L},\tilde{U}\right]$ באמינות α. ברור כי:

א) $\int\limits_L^U f(x)=1-\alpha$. אורך רווח הסמך הוא $U-L$.

ב) בה"כ נניח $L<\tilde{L}$ וגם $U<\tilde{U}$

לפיכך:

$$1-\alpha=\int_{\tilde{L}}^{\tilde{U}} f(x)dx=\int_{\tilde{L}}^{L} f(x)dx=+\int_{L}^{\tilde{U}} f(x)dx$$
$$1-\alpha=\int_{L}^{U} f(x)dx=\int_{L}^{\tilde{U}} f(x)dx=+\int_{\tilde{U}}^{U} f(x)dx$$
$$\Downarrow$$
$$|L-\tilde{L}|\eta\geq\int_{\tilde{L}}^{L} f(x)dx=\int_{\tilde{U}}^{U} f(x)dx\geq|U-\tilde{U}|\eta$$
$$\Downarrow$$
$$|L-\tilde{L}|=|U-\tilde{U}|$$

כלומר קיבלתי קווח סמך יותר גדול (או שווה). ■

7 הרצאה 7.

תזכורת: יש לי פרמטר θ לא ידוע ואני רוצה לאמוד אותו בעזרת תצפיות $x_1,\ldots,x_n$. יש לי את $f_\theta(x)$ - ההתפלגות של x. אני רוצה לתת חסם תחתון L וחסם עליון U כך שבהסתברות של לפחות $1-\alpha$ אני אהיה בטוח ש-θ יהיה ביניהם, כלומר $\mathbb{P}(L\leq\theta\leq U)>1-\alpha$.

לכן אם Ω הוא האזור בו θ נמצא אז אני רוצה $\int\limits_{\Omega}f(T)\geq 1-\alpha$.

דוגמה 7.1. $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ אונימונטלית (עולה עד לאיזה נקודה כלשהי ואז יורדת).

נגדיר $T=\sum\limits_{i=1}^{n}\frac{x_i}{n}$ כאשר σ ידוע. $f(T)=f(\{x_i\}|T(x_1,\ldots,x_n)=T)$. מה רווח הסמך הקטן ביותר עבור μ?

פתרון. ידוע כי

$$f(T)=\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{\frac{2\sigma^2}{n}}}$$

אני רוצה

$$\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{(T-\mu)^2}{\frac{2\sigma^2}{n}}}\geq c$$

כלומר

$$-\frac{(T-\mu)^2}{\frac{2\sigma^2}{n}}\geq\ln\left(\frac{c\cdot\sqrt{2\pi}\sigma}{\sqrt{n}}\right)$$

$$\Downarrow$$

$$(T-\mu)^2\leq\underbrace{\frac{2\sigma^2}{n}\ln\left(\frac{c\sqrt{n}}{\sqrt{2\pi}\sigma}\right)}_{:=\eta}$$

ולכן

$$T-\sqrt{\eta}\leq\mu\leq T+\sqrt{\eta}$$

הוא אזור סימטרי סביב T. איך מחשבים את c? אני רוצה Ω כך ש-$\mathbb{P}(\Omega)=\int\limits_{\Omega}f(T)dT\geq 1-\alpha$ כלומר $\mathbb{P}(\Omega_L),\mathbb{P}(\Omega_U)\leq\frac{\alpha}{2}$.

תזכורת: יהי מ"מ x_i ונגדיר $\mathbb{V}_i=\mathbb{V}\left[x_i\right]$ (השונות האמיתית, לא על בסיס התצפיות). בדוגמה שלנו $\mathbb{V}=\sigma^2$. נגדיר

$$T_0=\sum_{i=1}^{n}\mathbb{V}_i$$

לכן

$$\mathbb{V}(T_0)=\sum_{i=1}^{n}\mathbb{V}_i=\mathbb{V}(x_1)\cdot n$$

כאשר n הוא מספר האיברים שסכמתי.

בהשראה מ-T_0 נגדיר

$$T_1=\frac{T_0}{n}=\sum_{i=1}^{n}\frac{x_i}{n}$$

ואם נסמן $\mathbb{E}(x)$ להיות התוחלת של x ונסמן $\mathbb{V}(x)$ להיות השונות של x, נקבל

$$\mathbb{E}(T_1)=\mathbb{E}(x)$$
$$\mathbb{V}(T_1)=\frac{\mathbb{V}(x)}{n}$$

לכן, לכל התפלגות בעלת מומנטים חסומים נקבל

$$\mathbb{E}(T_1)=\mathbb{E}(x)$$
$$\mathbb{V}(T_1)=\frac{\mathbb{V}(x)}{n}$$

ובדוגמה שלנו

$$\mathbb{E}(T_1)=\mu$$
$$\mathbb{V}(T_1)=\frac{\sigma^2}{n}$$

מה הבעיה? נגיד X הוא המרחק שאני הולך ביום ביחידות מידה של m (מטרים). $\mathbb{E}(x)$ יהיה ביחידות של m, אבל $\mathbb{V}(x)$ ביחידות מידה של m^2, ולמדוד דברים במטרים בריבוע זה לא נוח (זה אומר שהלכתי דונם?).

הגדרה 7.2. כדי לא להתבלבל עם יחידות המידה שלי, נגדיר את **סטיית התקן** להיות $S.T.D.(x) = \sqrt{\mathbb{V}(x)}$.

בדוגמה שלנו $S.T.D.(x) = \sigma$ ואכן התוצאה שלנו ביחידות של מטרים, כמו שאנחנו רוצים, ובנוסף

$$S.T.D.(T_1) = \frac{S.T.D.(x)}{\sqrt{n}}$$

ובנוסף

$$x \sim \mathcal{N}(\mu, \sigma)$$
$$T_1 \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$$

לכן אם אני רוצה את T_1 עד כדי 2 סטיות תקן אני אקבל

$$T_1 = \mu \mp 2\frac{S.T.D.(x)}{\sqrt{n}}$$

הגדרה 7.3. $\frac{S.T.D.(x)}{\sqrt{n}}$ יקרא **standard error**.

דוגמה 7.4. משה רוצה לאמוד את התפלגות אורכי החדק של פילים. הוא יודע שהתפלגות אורכי החדק מתפלגת $PIL(x)$. הוא אסף 13 דגימות והגיע למסקנה שאורך החדק הממוצע הוא 0.8 ∓ 0.1 בסיכוי של 0.95.

(מפה אנחנו למדים כי $\mathbb{E}(T) = 0.8, U - L = 0.2$)

משה ניסה לפרסם את התוצאות שלו ואמרו לו שהדיוק לא מספיק טוב ולכן אסף עוד 13 דגימות וקיבל ממוצע חדש של 0.81. מהם U, L?

פתרון. $U = 0.81 + \frac{0.1}{\sqrt{2}}, L = 0.81 - \frac{0.1}{\sqrt{2}}$.

מה מוסר ההשכל? הגדלתי את המדגם שלי פי-2, ולכן הstandard error קטן פי $\sqrt{2}$.

עדיין באותה דוגמה של ההתפלגות הנורמלית:

$x_1, \ldots, x_n$ מ"מ.

$T = \sum_{i=1}^{n} \frac{x_i}{n}$ אומד שמבוסס על התצפיות. לפיכך μ, σ פרמטרים.

אני רוצה להגדיר $z = \frac{T-\mu}{\sigma}$. הוא לא משתנה מקרי, הוא לא פרמטר, מה הוא?

נקרא לו $pivot$. נניח σ ידוע אבל μ לא ידוע. מהו $\mathbb{P}(z)$?

קודם כל, אני יודע כי

$$\begin{aligned}\mathbb{P}(T) &= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\left(\frac{T-\mu}{\sigma}\right)^2 \cdot \frac{n}{2}} \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{nz^2}{2}}\end{aligned}$$

האם $\mathbb{P}(z)$ תלוי ב-T? לא!

האם $\mathbb{P}(z)$ תלוי ב-μ? לא!

לכן z לא תלוי בשום גורם לא ידוע.

האם $\mathbb{P}(z)$ תלוי ב-T? לא!

הגדרה 7.5. נגדיר **pivot** להיות פונקציה של מ"מ (ליתר דיוק של אומד שמבוסס על משתנה מקרי) ושל הפרמטרים שההתפלגות שלה לא תלויה בשום גורם לא ידוע.

בדוגמה שלנו z הוא ה-pivot ו-$\mathbb{P}(z)$ הוא ההתפלגות של ה-pivot.

דוגמה 7.6. נניח שאני יודע

$$\int_{-3.08}^{3.08} \mathbb{P}(z) dz = 0.999$$

ויהי משתנה נורמלי x עם $\sigma = 1$. מדדתי אותו 31 פעמים וקיבלתי ממוצע $T = \sum_{i=1}^{n} \frac{x_i}{n} = 2.615$.

מה הרווח סמך של μ ל-$\alpha = 0.001$?

פתרון. נגדיר $z = \frac{T-\mu}{\sigma} = 2.165 - \mu$ ואני יודע כי $-\frac{3.08}{\sqrt{31}} \leq z \leq \frac{3.08}{\sqrt{31}}$ כלומר $\mu = 2.615 \mp \frac{3.08}{\sqrt{31}}$.

דוגמה 7.7. משה גילה שאורך החדקים מתפלג נורמלית עם $\sigma = 0.615$ והוא מדד $T = 0.81$ על פני 26 פילים. תן למשה רווח סמך עם סיכוי של 0.999 לאורך חדק.

פתרון. נגדיר $z = \frac{T-\mu}{\sigma}$. אני רוצה $-\frac{3.08}{\sqrt{26}} \le z \le \frac{3.08}{\sqrt{26}}$ כלומר $-\frac{3.08}{\sqrt{26}} \le \frac{0.81-\mu}{0.15} \le \frac{3.08}{\sqrt{26}}$ ומכאן אלגברה.

הערה 7.8. אם σ לא ידוע אז ה-pivot של μ הוא $t = \frac{T_1-\mu}{\sqrt{T_2}}$ כאשר $T_1 = \sum_{i=1}^{n} \frac{x_i}{n}$ וגם $T_2 = \sum_{i=1}^{n} \frac{(x_i-T_1)}{n-1}$.

אלגוריתם 7.9. מתכון למציאת pivot:

1. חשב $T_1 = \sum_{i=1}^{n} \frac{x_i}{n}$.
2. חשב $T_2 = \sum_{i=1}^{n} \frac{(x_i-T_1)}{n-1}$.
3. חשב $t = \frac{T_1-\mu}{\sqrt{T_2}}$.
4. ה-pivot שלנו t מתפלג בהתפלגות t.
5. T_2 אומד ל-σ^2 והוא מתפלג χ^2.

דוגמה 7.10. דוד חקר פילים בסרי לנקה ומדד אורך חדקים של 12117 פילים.

התושבים המקומיים הסבירו לו שהתפלגות PIL הוא לא נורמלית. דוד טוען שעדיין אפשר להשתמש במתכון הנ"ל לאמוד רווח סמך לאורך חדק פיל. האם דוד צודק?

פתרון. ברור שדוד צודק! לפי משפט הגבול המרכזי למרות שכל $x_i \sim PIL$ עדיין T_1, T_2 יתפלגו נורמלית לפי משפט הגבול המרכזי.

7.1 בדיקת השערות.

משה מדד אורך חדקי פילים באפריקה.

דוד מדד אורך חדקי פילים בסרי לנקה.

משה טוען שלפיל אפריקאי יש אורך חדק שונה מפיל בסרי לנקה.

נסמן:

$$AF \sim PIL(\mu_0, \dots)$$
$$SL \sim PIL(\mu_1, \dots)$$

כלומר משה טוען כי $\mu_1 > \mu_0$.

פורמלית, משה טוען שהשערה (שתסומן H_0) $\mu_1 = \mu_0$ לא נכונה. ושהשערה (שתסומן H_1) $\mu_1 \neq \mu_0$ נכונה.

הגדרה 7.11. נגדיר את **השערת ה-0** להיות מה שאנחנו רוצים לשלול, ונסמנה ב-H_0.

אף פעם לא מוכיחים את השערת ה-0!

הגדרה 7.12. נגדיר **מדד** $M(x_1, x_2, \dots, x_n)$.

הגדרה 7.13. עבור φ מסויים, נגדיר $\Omega_0 = \{\{x_i\} | M(x_1, \dots x_n) < \varphi\}$ וגם $\Omega_1 = \{\{x_i\} | M(x_1, \dots x_n) \ge \varphi\}$. נשים ♡ כי $\Omega_0 = \Omega_1^c$.

ב-Ω_0 מקבלים את H_0, ב-Ω_1 מפריכים את H_0 (ומקבלים את H_1).

	$M \in \Omega_0$	$M \in \Omega_1$
H_0 נכונה	$1-\alpha$ ✓	✗ טעות מסוג ראשון α
H_1 נכונה	✗ טעות מסוג שני β	$\pi = 1-\beta$ ✓

ואת השמות של התוצאות (אם צדקתי כשטעיתי וכו) נקרא

	$M \in \Omega_0$	$M \in \Omega_1$
H_0 נכונה	T.P.	F.P.
H_1 נכונה	F.N.	T.N.

הגדרה 7.14. **חוזק המבחן שלי** הוא $\pi = 1 - \beta$.

על סמך האינטואיציה שלנו נשנה את Ω_1 להיות

$$\Omega_1 = \{\mathbb{P}(H_0) < 0.05\}$$

בהינתן M בוחרים φ כך ש-α קטן.

הייתי רוצה לבחור M כך ש-β מינימלי כלומר π מקסימלי.

סימון: Test Power Maximum - MP

למה 7.15. **(ניימן פירסון)** אם השערת ה0 היא

$$\begin{matrix} H_0 = \theta = \theta_0 \\ H_1 = \theta = \theta_1 \end{matrix} \quad \theta \neq \theta_0$$

אז ה-$M.P.$ מוגדר על ידי

$$M = \frac{L(T|\theta_1)}{L(T|\theta_0)}$$

כאשר T הוא UMVUE ו-L הוא הנראות.

8 הרצאה 8.

הגדרה 8.1. **השערה** H_0 היא טענה לגבי פרמטרים.

דוגמה 8.2. אני משער ש-$\mu = 3, \sigma = 2$.

הגדרה 8.3. **השערה פשוטה** היא השערה שלא משאירה משתנה לא ידוע.

הגדרה 8.4. השערה מורכבת היא השערה שמשאירה אי ודאות.

דוגמה 8.5. מיון השערות:

פשוט	מורכב
$\mu = 0, \sigma = 3$	$\mu > 0$
$\mu = 0$, σ ידוע	$\mu = 0, \sigma = ?$
$\sigma = 0$, μ ידוע	נאמוד את σ ונניח $\mu = 0$
σ ידוע, $\mu_1, \mu_2 = 63, 68$	σ ידוע, $\mu_1 - \mu_2$

בהינתן השערה אני יכול להציע מבחן $\Omega \subseteq V$ כאשר V מרחב המדגם שלי.

H_0 נכונה $\{x_i\} \in \Omega_0$.

אם H_1 לא נכונה, נטען ש-H_1 נכונה.

H_1 היא דחייה של השערת ה-0.

הגדרה 8.6. נגדיר את כל המבחנים עם דיוק $\hat{\alpha}$ להיות **אוסף כל המבחנים** עבורם $\alpha \leq \hat{\alpha}$.

הגדרה 8.7. נגדיר עבור השערה פשוטה $H_0 = \theta = \theta_0$ את **המבחן בעל העוצמה המקסימלית** $M.P.$ בתור המבחן עבורו β מינימלי.

משפט 8.8. (ניימן פירסון) בהינתן התפלגות $f_\theta(x)$ ושתי השערות

$$
\begin{aligned}
H_0 \,&: \theta = \theta_0 \\
H_1 \,&: \theta = \theta_1
\end{aligned}
$$

ודרישה לדיוק של α על מבחן אז המבחן בעל העוצמה המירבית MP היא $\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} \geq c$ ואם T הוא אומד מספיק אז המבחן הוא $\lambda(T) = \frac{f_{\theta_1}(T)}{f_{\theta_0}(T)} \geq c$ ו-c נקבע על ידי הדרישה

$$\int\limits_{\Omega_1} f_{\theta_0}(x)dx = \alpha$$

או במילילם אחרות

$$\mathbb{P}\left(\lambda(T) \geq c\right) = \alpha$$

כאשר נגדיר

$$
\begin{aligned}
\Omega_1 &= \{x | \lambda(T) \geq c\} \\
\Omega_0 &= \{x | \lambda(T) < c\}
\end{aligned}
$$

הוכחה. בהינתן המבחן של ניימן פירסון שמורכב מ-Ω_0, Ω_1 ומבחן אחר עם Ω_0', Ω_1'. נגדיר

$$
\begin{aligned}
\alpha &= \int\limits_{\Omega_1} f_{\theta_0}(T)dT & \alpha' &= \int\limits_{\Omega_1'} f_{\theta_0}(T)dT \\
\beta &= \int\limits_{\Omega_0} f_{\theta_1}(T)dT & \beta' &= \int\limits_{\Omega'_0} f_{\theta_1}(T)dT \\
\pi &= 1 - \beta & \pi' &= 1 - \beta
\end{aligned}
$$

ואנחנו רוצים להוכיח $\pi > \pi'$.

$$
\begin{aligned}
\pi - \pi' &= \int\limits_{\Omega_1} f_{\theta_1}(T)dT - \int\limits_{\Omega_1'} f_{\theta_1}(T)dT \\
&= \int\limits_{\Omega_1 \cap \Omega_1'} f_{\theta_1}(T)dT + \int\limits_{\Omega_1 \cap \Omega_0'} f_{\theta_1}(T)dT - \int\limits_{\Omega_0 \cap \Omega_1'} f_{\theta_1}(T)dT - \int\limits_{\Omega_1 \cap \Omega_1'} f_{\theta_1}(T)dT \\
&= \int\limits_{\Omega_1 \cap \Omega_0'} f_{\theta_1}(T)dT - \int\limits_{\Omega_0 \cap \Omega_1'} f_{\theta_1}(T)dT
\end{aligned}
$$

בתוך Ω_1 מתקיים $\frac{f_{\theta_1}(T)}{f_{\theta_0}(T)} \geq c$ ובתוך Ω_0 $c > \frac{f_{\theta_1}(T)}{f_{\theta_0}(T)}$, לכן

$$\begin{aligned}
\pi-\pi' &\geq c\left(\int\limits_{\Omega_1\cap\Omega_0'} f_{\theta_0}(T)dT - \int\limits_{\Omega_0\cap\Omega_1'} f_{\theta_0}(T)dT\right)\\
&= c\left(\int\limits_{\Omega_1\cap\Omega_0'} f_{\theta_0}(T)dT + \int\limits_{\Omega_1\cap\Omega_1'} f_{\theta_0}(T)dT - \int\limits_{\Omega_1\cap\Omega_1'} f_{\theta_0}(T)dT - \int\limits_{\Omega_0\cap\Omega_1'} f_{\theta_0}(T)dT\right)\\
&= c\left(\int\limits_{\Omega_1} f_{\theta_0}(T)dT - \int\limits_{\Omega_1'} f_{\theta_0}(T)dT\right)\\
&= c\left(\alpha-\alpha'\right)\\
&\geq 0
\end{aligned}$$

כלומר

$$\pi \geq \pi'$$

כנדרש. ■

דוגמה 8.9. אם יש לי $f \sim \mathcal{N}(\mu,\sigma)$, והשערת $H_0:\mu=\mu_0$ והשערת $H_1:\mu=\mu_1$. נגדיר $T=\sum\limits_{i=1}^{n}\frac{x_i}{n}$ שהוא אומד מספיק ולא מוטה.

דוגמה 8.10. לכן לפי ניימן פישר:

$$\Omega_1 = \frac{f_{\mu_1}(T)}{f_{\mu_0}(T)} \geq c$$

כלומר

$$\begin{aligned}
\Omega_1 &= \frac{\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{(T-\mu_1)^2}{\frac{2\sigma^2}{n}}}}{\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{(T-\mu_2)^2}{\frac{2\sigma^2}{n}}}}\\
&= e^{\frac{-\left(\mu_1^2-\mu_2^2-2\mu_1T+2\mu_2T\right)}{\frac{2\sigma^2}{n}}}\\
&= e^{\frac{-2(\mu_1-\mu_0)\left(\frac{\mu_1+\mu_2}{2}-2T\right)}{\frac{2\sigma^2}{n}}}\\
&\geq c
\end{aligned}$$

ו-c נקבע על ידי

$$\int\limits_{\Omega_1} f_{\theta_1}(T)dT = \alpha$$

כלומר

$$\int\limits_{\substack{\Omega_1\\ T\geq c}} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{(T-\mu_0)^2}{\frac{2\sigma^2}{n}}} = \alpha$$

ניקח $pivot$ להיות $z=\frac{T-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ ונקבל

$$\int\limits_{\substack{\Omega_1\\ T\geq c}} \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} = \alpha$$

ואם נבחר $\alpha=0.001$ נקבל $z>3.008$ ואז נקבל $\frac{T-\mu_0}{\frac{\sigma}{\sqrt{n}}}>3.008$ וזה מבחן הרבה הרבה יותר פשוט!

רשימת השערות שנצטרך לדעת להתמודד איתן:

מיון השערות פשוטות:

- σ בהינתן μ.

- μ בהינתן σ.
- p בהתפלגות בינומית.
- λ בהתפלגות פואסון.
- τ בהתפלגות מעריכית.

מיון השערות מורכבות:

- בהינתן σ או לא בהינתן σ, האם $\mu > \mu_0$.
- μ, σ לא ידועים- האם $\mu = \mu_0$.
- μ, σ לא ידועים- האם $\sigma = \sigma_0$.
- בהינתן 2 אוכלוסיות נורמליות $\mu_1 = \mu_2$ בהינתן $\sigma_1 = \sigma_2/\sigma_1 \neq \sigma_2$ ידוע/ לא ידוע.
- בהינתן $X, Y \sim \mathscr{N}(\bar{\mu}, \bar{C})$ כאשר $\bar{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$ וגם $\bar{C} = \begin{pmatrix} \mathbb{V}[X] & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & \mathbb{V}[Y] \end{pmatrix}$, האם X, Y בלתי מתואמים.

דיון: אם יש לנו התפלגות בינומית עם N ניסויים

$$\mathbb{P}(k) = \binom{N}{k} p^k (1-p)^{N-k}$$

ו-$\lambda = Np$ קבוע, $N \to \infty$, $p \to 0$ אזי ההתפלגות הופכת להתפלגות פואסון

$$\mathbb{P}(k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

ואם $\lambda, k \to \infty$ נקבל התפלגות נורמלית $P(k) \sim \mathscr{N}(\lambda, \sqrt{\lambda})$, וזה מבחן פרמטרי.

בהתפלגות בינום

$$\mathbb{E} = Np$$
$$\mathbb{V} = Np(1-p)$$

בהתפלגות פואסון

$$\mathbb{E} = \lambda$$
$$\mathbb{V} = \lambda$$

בהתפלגות נורמלית

$$\mathbb{E} = Np$$
$$\mathbb{V} = Np(1-p)$$

דוגמה 8.11. משה מבצע מספר מאוד גדול $N >> 0$ של ניסויי ברנולי ורוצה לבנות מבחן על הטענה $p = p_0$ כאשר p_0 לא זניח.

משה מציע לו להחליף את השערת H_1 במקום שיהיה $p \neq p_0$ ל-$p = p_1$.

א) האם דוד צודק.

ב) פתח מבחן ניימן פירסון.

ג) תן פתרון למבחן עם $\alpha = 0.001$.

פתרון. א) כן, כי ניימן פירסון יודע להתמודד עם השערה $p = p_0$ מול $p = p_1$.

ב) נגדיר אומד $K = \sum_{i=1}^{N} x_i$. מבחן ניימן פירסון יתן לי

$$\Omega_1 : \frac{\binom{N}{k} p_1^k (1-p_1)^{N-k}}{\binom{N}{k} p_0^k (1-p_0)^{N-k}} = \left(\frac{p_1}{p_0}\right)^k \left(\frac{1-p_1}{1-p_0}\right)^{N-k} \geq c$$

ג) רשמית: צריך

$$\sum_{k=c}^{N} \binom{N}{k} p_0^k (1-p_0)^{N-k} \leq \alpha$$

אבל $N >> 0$ לכן $K \sim \mathscr{N}(Np_0, \sqrt{Np_0(1-p_0)})$. נגדיר $pivot$ שיהיה $\frac{k - Np_0}{\sqrt{Np_0(1-p_0)}} \gtrsim 3.008$ ומכאן קל להמשיך.

9 הרצאה 9.

9.1 בדיקת השערות.

תזכורת:

למדנו השערות פשוטות ואת הלמה של ניימן פירסון: אם יש לי $\begin{matrix} H_0 & \mu=\mu_0 \\ H_1 & \mu=\mu_1 \end{matrix}$ אזי המבחן עבורו $P_{\mu_0}\left(x\in\Omega_1\right)\le\alpha,\min\left[P_{\mu_1}\left(x\in\Omega_0\right)\right]$ מוגדר על ידי

$$\frac{L\left(\mu_1|X\right)}{L\left(\mu_0|X\right)}\ge c$$

כאשר

$$L\left(\mu_0|X\right)=\alpha=\mathbb{P}_{\mu_1}\left(x\in\Omega_0\right)$$

דוגמה 9.1. σ ידוע ו-μ לא ידוע. השערה: $\mu=\mu_0$ כלומר

$$\frac{\prod_{i=1}^{n}e^{-\frac{\left(x_i-\mu_1\right)^2}{2\sigma^2}}}{\prod_{i=1}^{n}e^{-\frac{\left(x_i-\mu_0\right)^2}{2\sigma^2}}}\ge c$$

כלומר אנחנו רוצים

$$\prod_{i=1}^{n}e^{-\frac{\left(x_i-\mu_0\right)^2}{2\sigma^2}}\ge c$$

ואם נגדיר $T=\sum_{i=1}^{n}\frac{x_i}{n}$ אזי זה יהיה שקול לדרישה

$$e^{-\frac{\left(T-\mu_0\right)^2}{2\sigma^2/n}}\ge\tilde{c}$$

ואנחנו רוצים

$$P\left(e^{-\frac{\left(T-\mu_0\right)^2}{2\sigma^2/n}}\ge\tilde{c}\right)=\alpha$$

כאשר לרוב ניקח $\alpha=0.005$ לכן נקבל $\frac{\left(T-\mu_0\right)^2}{2\sigma^2/n}\ge Z_\alpha$.

דוגמה 9.2. מטילים מטבע. נשער את ההשערה $\begin{matrix} H_0: & \rho=\rho_0 \\ H_1: & \rho=\rho_1 \end{matrix}$ (השערה פשוטה) ונגדיר $k=\sum_{i=0}^{N}x_i$ הלומר מספר הפעמים שיצא עץ.

ממבחן ניימר פירסון אני אחפש

$$\frac{P\left(\rho_1|\text{הטלות שלי}\right)}{P\left(\rho_0|\text{הטלות שלי}\right)}\ge c$$

כלומר

$$\frac{\binom{N}{k}\rho_1^k\left(1-\rho_1\right)^{N-k}}{\binom{N}{k}\rho_0^k\left(1-\rho_0\right)^{N-k}}=\left(\frac{\rho_1}{1-\rho_1}\right)^k\left(\frac{1-\rho_0}{\rho_0}\right)^k\left(\frac{1-\rho_1}{1-\rho_0}\right)^{N-k}\ge c$$

לפי מבחן ניימן פירסון נחפש את ה-c עבורו

$$\sum_{i=0}^{N}\binom{N}{k}\rho_0^k\left(1-\rho_0\right)^{N-k}\ge c$$

וזה לא משהו שאנחנו אוהבים לחשב. לכן, ככל שאנחנו דוגמים יותר, ככה אנחנו יכולים יותר ויותר לשאוף להתפלגות פואסונית וממנה להתפלגות נורמלית על ידי התקנון

$$\frac{k-\rho_0N}{\sqrt{\rho_0\left(1-\rho_0\right)N}}\sim\mathcal{N}\left(0,1\right)$$

כלומר המבחן שלנו יהיה $\frac{k-\rho_0N}{\sqrt{\rho_0\left(1-\rho_0\right)N}}\ge z_a$.

הגדרה 9.3. בהינתן התפלגות ודגימות x_i ואומד מספיק $T\left(x_i\right)$ שמתפלג בגבול

$$\lim_{n\to\infty}T\sim\mathcal{N}\left(\mu_0,\sigma_0\right)$$

אזי למבחן

$$\frac{T-\mu_0}{\sigma_0}$$

ייקרא **מבחן wald** והוא יתפלג z ולכן המבחן יהיה

$$\frac{T-\mu_0}{\sigma}\sim Z_\alpha$$

9.2 השערה על σ.

יש לנו השערה $\begin{matrix} H_0 & \sigma=\sigma_0 \\ H_1 & \sigma=\sigma_1 \end{matrix}$. לפי מבחן ניימן פירסון נקבל

$$\frac{P\left(\theta_1|x_1,\ldots,x_n\right)}{P\left(\theta_0|x_1,\ldots,x_n\right)}\geq c$$

כלומר

$$\frac{\left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^n\prod\limits_{i=1}^{n}e^{-\frac{(x_i-\mu)^2}{2\sigma_1^2}}}{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n\prod\limits_{i=1}^{n}e^{-\frac{(x_i-\mu)^2}{2\sigma_0^2}}}=\left(\frac{\sigma_0}{\sigma_1}\right)^n e^{-\sum\limits_{i=1}^{n}\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)\frac{(x_i-\mu)^2}{2}}=\left(\frac{\sigma_0}{\sigma_1}\right)^n e^{-\left(\frac{\sigma_0^2-\sigma_1^2}{\sigma_1^2\sigma_0^2}-\frac{1}{\sigma_0^2}\right)\sum\limits_{i=1}^{n}\frac{(x_i-\mu)^2}{2}}\geq c$$

ונסמן את $\sum\limits_i\frac{(x-\mu)^2}{2}$ להיות ה-pivot שלנו. לכן נקבל גם

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2}\geq\underbrace{\frac{\ln\left(c\left(\frac{\sigma_0}{\sigma_1}\right)^n\right)}{\sigma_1^2-\sigma_0^2}\sigma_1^2\sigma_0^2}_{:=\tilde{c}}$$

כלומר קיבלתי

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2}\geq\tilde{c}$$

וה-c שכתוב פה הוא ה-c שמקיים

$$\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\sum\limits_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma_0^2}}\geq c$$

כלומר בסוף אני אקבל את המבחן

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma_0^2}\sim\chi_n^2\geq\tilde{c}$$

הערה 9.4. האם לומר $\frac{\mathbb{P}(\theta_1|x_1,\ldots,x_n)}{\mathbb{P}(\theta_0|x_1,\ldots,x_n)}$ זה אותו דבר כמו לומר $\frac{\mathbb{P}(x_1,\ldots,x_n|\theta_1)}{\mathbb{P}(x_1,\ldots,x_n|\theta_0)}$? לפי חוק בייס נראה כי

$$\frac{\mathbb{P}\left(\theta_1|x_1,\ldots,x_n\right)}{\mathbb{P}\left(\theta_0|x_1,\ldots,x_n\right)}\geq c\iff\frac{\frac{\mathbb{P}(x_1,\ldots,x_n|\theta_1)\mathbb{P}(\theta_1)}{\mathbb{P}(x_1,\ldots,x_n)}}{\frac{\mathbb{P}(x_1,\ldots,x_n|\theta_0)\mathbb{P}(\theta_0)}{\mathbb{P}(x_1,\ldots,x_n)}}c\iff\frac{\mathbb{P}\left(\theta_1\right)}{\mathbb{P}\left(\theta_0\right)}\cdot\frac{\mathbb{P}\left(x_1,\ldots,x_n|\theta_1\right)}{\mathbb{P}\left(x_1,\ldots,x_n|\theta_0\right)}\geq c\iff\frac{\mathbb{P}\left(x_1,\ldots,x_n|\theta_1\right)}{\mathbb{P}\left(x_1,\ldots,x_n|\theta_0\right)}\geq\underbrace{c\cdot\frac{\mathbb{P}\left(\theta_0\right)}{\mathbb{P}\left(\theta_1\right)}}_{:=\tilde{c}}$$

$$\iff\frac{\mathbb{P}\left(x_1,\ldots,x_n|\theta_1\right)}{\mathbb{P}\left(x_1,\ldots,x_n|\theta_0\right)}\geq\tilde{c}$$

כלומר לא מדובר באותו דבר!

סיכום 9.5. סיכום השערות פשוטות:

- μ לא ידוע ו-σ: נגדיר $T=\sum\limits_{i=1}^{n}\frac{x_i}{n}$ והמבחן יהיה $z\sim\frac{T-\mu}{\frac{\sigma}{\sqrt{n}}}$.
- μ ידוע, σ לא ידוע: נגדיר $T=\sum\limits_{i=1}^{n}\frac{x_i-\mu}{n}$ והמבחן יהיה $\chi^2\sim\frac{T}{\sigma^2}$.
- בינום/ פואסון: ניקח $k=\sum\limits_{i=1}^{n}x_i$ והמבחן יהיה $\mathcal{N}(0,1)\sim\frac{k-\rho_0 N}{\sqrt{\rho_0(1-\rho_0)N}}$

9.3 השערות מורכבות.

עכשיו ההשערות שלי הופכות להיות $\begin{matrix} H_0 & \theta\in\theta_0 \\ H_1 & \theta\in\theta_1 \end{matrix}$ כאשר $\theta_1\cap\theta_0=\emptyset$ וגם $\theta_1\cup\theta_0=V$ כאשר V זה העולם שלי.

אם נגדיר מבחן על ידי חלוקה של מרחב התצפיות ל-Ω_0 ו-Ω_1 ונדחה את H_0 אם $\{x_i\}\in\Omega_1$ אז העוצמה של המבחן תהיה $\pi(\theta)=\mathbb{P}_\theta\left(\Omega_1\right)$.

הבעיה היא שאפילו אם אישרתי ש-$\theta\in\theta_0$ אני עדיין לא יודע לדחות את כל ה-$\theta'\in\theta_0$ שלא שווים ל-θ. לכן נגדיר את המושג הבא שיעזור לנו לדחות את כל ה-θ':

הגדרה 9.6. רמת מובהקות α בהשערה מורכבת יהיה

$$\sup_{\theta\in\theta_0}\left(\pi(\theta)\right)$$

הגדרה 9.7. נגיד שמבחן M בהשערה מורכבת הוא **ump** אם הוא מקיים

א $\sup_{\theta\in\theta_0}\left(\pi_M(\theta)\right)=\alpha$

ב לכל מבחן אחר A שמקיים $\sup_{\theta\in\theta_0}\left(\pi_A(\theta)\right)=\alpha$ ולכל $\theta\in\theta_1$ מתקיים

$$\pi_M\left(\theta\right)\geq\pi_A\left(\theta\right)$$

הבעיה היא שכמעט תמיד אין ump ☹.

הגדרה 9.8. נגיד שמשפחה מקיימת את **תכונת m.l.r** (monotone likelihood ratio) עבור אומד מספיק T ופרמטרים $\theta_2\geq\theta_1$ אם

$$\mathbb{P}\left(\theta_2|T\right)\geq\mathbb{P}\left(\theta_1|T\right)$$

דוגמה 9.9. תהי f משפחה מעריכית חד מימדית ו-T אומד מספיק.

$$f\left(x|\theta\right)=h(x)\cdot e^{c(\theta)\cdot T+A(\theta)}$$

נניח $\theta_2\geq\theta_1$ ו-$c(\theta)$ מונוטונית, לדוגמה $c(\theta)=\ln\left(\frac{\rho}{1-\rho}\right)$ (פיתוח של בינום למשפחה מעריכית), אזי

$$\frac{\mathbb{P}\left(\theta_2|T\right)}{\mathbb{P}\left(\theta_1|T\right)}=\frac{h(x)\cdot e^{c(\theta_2)\cdot T-A(\theta_2)}}{h(x)\cdot e^{c(\theta_1)\cdot T-A(\theta_1)}}=e^{(c(\theta_2)-c(\theta_1))\cdot T+(A(\theta_1)-A(\theta_2))}$$

וזה פונקציה מונוטונית עולה ולכן $m.l.r$.

משפט 9.10. למשפחה עם $m.l.r$ יש מבחן $u.m.p$ להשערה $H_0:\theta\leq\theta_0$ והמבחן הוא מבחן ניימן פירסון עבור $\theta=\theta_0$.

ניזכר שבהשערה פשוטה $\begin{array}{l}H_0=\theta_0\\H_1=\theta_1\end{array}$ אזי מבחן ניימן פירסון יהיה $\frac{f(\theta_1|x)}{f(\theta_0|x)}$ ושאם יש לי השערה מורכבת $\begin{array}{ll}H_0: & \theta\in\theta_0\\H_1: & \theta\in\theta_1\end{array}$ אזי המבחן יהיה

$$\lambda(x)=\frac{\sup_{\theta\in\theta_1}\left(f(\theta|x)\right)}{\sup_{\theta\in\theta_0}\left(f(\theta|x)\right)}$$

ולפעמים יגדירו

$$\lambda^*(x)=\frac{\sup_{\theta\in V}\left(f(\theta|x)\right)}{\sup_{\theta\in\theta_0}\left(f(\theta|x)\right)}$$

כלומר $\lambda^*(x)=\max\left\{\lambda(x),1\right\}$. עבור איזה θ מתקבל הסופרימום. לכן נעקוב אחרי הצעדים הבאים:

1. בחר אומד נראות מירבית על V.
2. בחר אומד נראות מירבית על θ_0.
3. חשב יחס ביניהם.
4. הפעל מבחן ניימן פירסון על ה-θ של הנראות המירבית.

וזה יהיה המבחן הכי טוב שאני יכול לייצר.

דוגמה 9.11. השערה $\mu=\mu_0$ ו-σ לא ידוע בהתפלגות נורמלית.

1. נחשב נראות מירבית ל-μ ו-σ בכל התחום:

$$\hat{\mu}=\sum_{i=1}^{n}\frac{x_i}{n},\hat{\sigma^2}=\sum_{i=1}^{n}\frac{(x_i-\hat{\mu})^2}{n}$$

2. נחשב אומד נראות מירבית על θ_0:

$$\hat{\mu}=\mu_0,\hat{\sigma_0^2}=\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{n}$$

3. נחשב את היחס ביניהם:

$$\lambda=\frac{\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)e^{-\frac{(x_i-\hat{\mu})^2}{2\hat{\sigma^2}}}}{\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\hat{\sigma_0}}\right)e^{-\frac{(x_i-\hat{\mu_0})^2}{2\hat{\sigma_0^2}}}}=\left(\frac{\hat{\sigma_0}}{\sigma_0}\right)^n\frac{e^{-\frac{\sum_{i=1}^{n}\frac{(x_i-\hat{\mu})^2}{2n}\cdot n}{2\hat{\sigma^2}}}}{e^{-\frac{\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{2n}\cdot n}{2\hat{\sigma_0^2}}}}=\left(\frac{\hat{\sigma_0}}{\sigma_0}\right)^n\geq c$$

4. נעשה בשיעור הבא!

10 הרצאה 10.

תזכורת: אם יש לי השערה פשוטה $\begin{matrix} H_0 : \rho = \rho_0 \\ H_1 : \rho = \rho_1 \end{matrix}$ אז יש לי מבחן $\frac{L_x(\rho_1)}{L_x(\rho_0)} \geq c$ כאשר $L_x(\rho) = \mathbb{P}(x|\rho)$ וגם $L_x(\rho_0) = \alpha$.

דוגמה 10.1. אני רוצה לחשב את $\frac{\mathbb{P}(x|\mu_1)}{\mathbb{P}(x|\mu_1)}$ עבור $x_1, \ldots, x_n$ כאשר σ ידוע, אזי נקבל

$$\frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu_1)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu_0)^2}{2\sigma^2}}} \geq c$$

בהשערה מורכבת $\begin{matrix} H_0 : \theta \in \theta_0 \\ H_1 : \theta \in \theta_1 \end{matrix}$ נקבל

$$\frac{\max_{\theta} L_x(\rho_1)}{\max_{\theta \in \theta_0} L_x(\rho_0)} \geq c$$

דוגמה 10.2. יש לי $H_0 : \mu = \mu_0$ וגם σ לא ידוע. יש לי תצפיות $x_1, \ldots, x_n$.

א. אני מחשב נראות מירבית למונה.

ב. אני מחשב נראות מירבית למכנה.

ג. מחשב את היחס ודורש שיהיה גדול מ-c.

$$\mathbb{P}(x_1,...,x_n)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum\limits_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$L=\ln\left(\mathbb{P}(x_1,...,x_n)\right)=n\ln\left(\frac{1}{\sqrt{2\pi}}\right)-n\ln\sigma-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

א
$$\frac{\partial L}{\partial\mu}=0\Rightarrow\sum_{i=1}^{n}(x_i-\mu)=0\Rightarrow\hat{\mu}=\sum_{i=1}^{n}\frac{x_i}{n}$$

א
$$\frac{\partial L}{\partial\sigma}=0\Rightarrow-\frac{n}{\sigma}+\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3}=0\Rightarrow\hat{\sigma^2}=\sum_{i=1}^{n}\frac{(x_i-\hat{\mu})^2}{n}$$

ב
$$\frac{\partial L}{\partial\sigma}=0\Rightarrow-\frac{n}{\sigma}+\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{\sigma^3}=0\Rightarrow\hat{\sigma_0^2}=\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{n}$$

ג
$$\begin{aligned}
c &\leq\frac{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{-\frac{(x_i-\hat{\mu_0})^2}{2\hat{\sigma^2}}}}{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma_0}}\right)^n e^{-\frac{(x_i-\mu_0)^2}{2\hat{\sigma_0^2}}}}\\
&=\left(\frac{\hat{\sigma_0}}{\hat{\sigma}}\right)^n\\
&=\left(\frac{\sum\limits_{i=1}^{n}\frac{(x_i-\mu_0)^2}{n}}{\sum\limits_{i=1}^{n}\frac{(x_i-\hat{\mu})^2}{n}}\right)^{\frac{n}{2}}\\
&=\left(\frac{\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}{\sum\limits_{i=1}^{n}(x_i-\hat{\mu})^2}\right)^{\frac{n}{2}}\\
&=\left(\frac{\sum\limits_{i=1}^{n}(x_i-\hat{\mu})^2+n(\hat{\mu}-\mu_0)}{\sum\limits_{i=1}^{n}(x_i-\hat{\mu})^2}\right)^{\frac{n}{2}}\\
&=\left(1+\frac{n(\hat{\mu}-\mu_0)}{\sum\limits_{i=1}^{n}(x_i-\hat{\mu})^2}\right)^{\frac{n}{2}}\\
&=\left(1+T^2\right)^{\frac{n}{2}}\\
&\Downarrow\\
T &\geq\tilde{c}
\end{aligned}$$

$$T^2=\frac{n(\hat{\mu}-\mu_0)}{\sum\limits_{i=1}^{n}(x_i-\hat{\mu})^2}$$

ניקח $t_{n,\alpha}$ עבורו הסיכוי להיות מעל $\tilde{c}$ בהתפלגות t עם n דרגות חופש הוא α.

לדוגמה: $n=30$, $\alpha=0.01$, הולכים לטבלה ובודקים את $t_{n,\alpha}$.

תזכורות:

1. אם $x_i\sim\mathcal{N}(\mu,\sigma)$ אזי $\bar{x}=\sum\limits_{i=1}^{n}\frac{x_i}{n}\sim\mathcal{N}(\mu,\frac{\sigma}{\sqrt{n}})$.
2. $\bar{x}-\mu_0\sim\mathcal{N}(\mu-\mu_0,\frac{\sigma}{\sqrt{n}})$.
   תחת השערת האפס $H_0:\mu=\mu_0$ נקבל $\bar{x}-\mu_0\sim\mathcal{N}(0,\frac{\sigma}{\sqrt{n}})$.
3. $t\sim\frac{\mathcal{N}}{\sqrt{\chi^2}}\sim\sqrt{\frac{n(\hat{\mu}-\mu_0)}{\sum\limits_{i=1}^{n}(x_i-\hat{\mu})^2}}=T$.
4. רשימת התפלגויות שאנחנו מכירים:
   - Γ: $x^{-\beta}e^{-\frac{x}{a}}$.
   - z: התפלגות נורמלית סטנדרטית.

- $x^2=\chi^2$: x מתפלג נורמלית.
- $\sum_{i=1}^{n} x_i^2=\chi_n^2$: כאשר x_i מתפלג נורמלית.
- t: $\frac{\mathcal{N}}{\sqrt{\chi^2}}$.
- F: $\frac{\chi_1^2}{\chi_2^2}$.

5. אם יש לי אומדים T,S חסרי הטיה אז גם האומד $\frac{T}{S}$ חסר הטיה.

10.1 2 אולכוסיות, מבחן טיב ההתאמה.

10.1.1 מבחן טיב ההתאמה.

שאלה 10.3. מבחן טיב ההתאמה: בהינתן התפלגות מולטינומית עם סיכוי $p_1,...,p_k$ לתוצאה $1,...,k$ ועם N הגרלות עם תוצאות $v_1,...,v_k$ כאשר $\sum_{i=1}^{N} v_i=N$.

אנחנו צריכים להציע מבחן להשערת ה-0

$$\begin{aligned}H_0 :&p_1=\rho_1\\&p_2=\rho_2\\&\vdots\\&p_k=\rho_k\end{aligned}$$

פתרון. נפתור בכמה שלבים:

א) נניח שכל אחת מהאפשרויות מתפלגת נורמלית עם ממוצע Np_i וסטיית תקן $\sqrt{Np_i(1-p_i)}$ לכן התוצאות v_i בזריקה ה-i הוא

$$v_i\sim\mathcal{N}(Np_i,\sqrt{Np_i(1-p_i)})$$

. נגדיר מ"מ חדש

$$X_i=\frac{v_i-Np_i}{\sqrt{Np_i(1-p_i)}}\sim\mathcal{N}(0,1)$$

לכן

$$X_i^2=\frac{(v_i-Np_i)^2}{Np_i(1-p_i)}\sim\chi_1^2$$

ונגדיר

$$T=\sum_{i=1}^{n}X_i^2=\sum_{i=1}^{n}\frac{(v_i-Np_i)^2}{Np_i(1-p_i)}\sim\chi_n^2$$

ומפה זה מבחנים שאנחנו יודעים לעשות.

10.1.2 2 אוכלוסיות.

בקורס מסויים, מסיבה לא ברורה, כל התלמידים ניגשו למועד ב'.

השערת ה-0: הציונים במועד א' ובמועד ב' באים מאותה התפלגות.

- אם σ ידוע ושווה במועד א ובמועד ב, האם ההשערה פשוטה?
- מה המבחן?
- מה המבחן אם σ לא ידוע?

הערה: ההניח ההתפלגות נורמלית.

תצפיות:

$$y_i=x_{i_{\text{מועד ר}}}-x_{i_{\text{מועד ב}}}$$

ההנחה שלי היא $H_0:\mu(y_i)=0$ והיא הנחה מורכבת.

$x_{i_{\text{מועד א}}}\sim\mathcal{N}(\mu_0,\sigma)$ ציוני מועד א'.

$x_{i_{\text{מועד ב}}}\sim\mathcal{N}(\mu_1,\sigma)$ ציוני מועד ב'.

לכן

$$y_i\sim\mathcal{N}(\mu_0-\mu_1,\sqrt{2}\sigma)$$

בהינתן השערת ה-0 נקבל

$$y_i \sim \mathcal{N}(0, \sqrt{2}\sigma)$$

כדי לפתח מבחן נגדיר אומד

$$T = \sum_{i=1}^{n} \frac{x_{i_{\text{א}}} - x_{i_{\text{ב}}}}{n}$$

אבל הוא לא מתפלג z לכן נתקן אותו :

$$T' = \frac{T}{\sqrt{2}\sigma} \sim \mathcal{N}(0,1)$$

ולכן המבחן יהיה

$$T' \geq z_\alpha$$

כמעט. בגלל הסימטריות של z נקבל שהמבחן יהיה

$$T' > z_{\frac{\alpha}{2}}$$
$$T' < -z_{\frac{\alpha}{2}}$$

ובמקרה ש-σ לא ידוע לנו :

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

ואז המבחן יהיה

$$\frac{T}{\sqrt{2}\hat{\sigma}} > t_{n,\frac{\alpha}{2}}$$
$$\frac{T}{\sqrt{2}\hat{\sigma}} < -t_{n,\frac{\alpha}{2}}$$

שאלה 10.4. כעת, באותו קורס היו 2 קבוצות : הקבוצה של יום שלישי והקבוצה של יום שני. כל קבוצה קיבלה מבחן באותם הגדרות כמו השאלה הקודמת.

השערת ה-0 היא שהמבחנים הם עם ממוצע זהה.

משה טוען שהפתרון הוא כמו קודם.

דוד טוען שיש צורך בתיקון.

מי צודק ואם צריך לתקן, תקן.

פתרון. בקבוצה א יש n ציונים המתפלגים $\mathcal{N}(\mu_1, \sigma_1)$ בעוד שבקבוצה א יש n ציונים המתפלגים $\mathcal{N}(\mu_2, \sigma_2)$.

השערת ה-0 שלנו היא $H_0 : \mu_1 = \mu_2$. נגדיר

$$T = \bar{X} - \bar{Y}$$

כאשר $\bar{X} \sim \mathcal{N}(\mu_1, \frac{\sigma_1}{\sqrt{n}}), \bar{Y} \sim \mathcal{N}(\mu_2, \frac{\sigma_2}{\sqrt{n}})$ ולכן $T \sim \mathcal{N}(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}})$. לכן המבחן שלי יהיה

$$-z_{\frac{\alpha}{2}} > \frac{T = \bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} > z_{\frac{\alpha}{2}}$$

שאלה 10.5. כעת השערת ה- 0 שלי היא שקבוצה 1 טובה ב-5 נקודות מקבוצה 2. תנו לי מבחן.

פתרון. כמו בשאלה הקודמת, המבחן יהיה בתוספת תקנון של T שיתפלג z :

$$-z_{\frac{\alpha}{2}} > \frac{T = \bar{X} - \bar{Y} - 5}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} > z_{\frac{\alpha}{2}}$$

שאלה 10.6. כעת השערת ה- 0 שלי היא שקבוצה 1 טובה בלפחות 5 נקודות מקבוצה 2. תנו לי מבחן.

פתרון. השערת ה-0 שלי תהפוך להיות $H_0 : \mu_0 - \mu_1 \geq 5$. כעת כיוון שאני רוצה להיות בצד שגדול מ-5, אני רוצה פשוט להיות אחרי 5 כלומר אני לא צריך במבחן שלי את החלק של $-z_{\frac{\alpha}{2}} >$ ונוכל להחליף את $z_{\frac{\alpha}{2}}$ ב-z_α כלומר המבחן יהיה

$$\frac{T = \bar{X} - \bar{Y} - 5}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} > z_\alpha$$

שאלה 10.7. כעת נזכרתי שמרוב שמיהרתי שכחתי כי σ_1, σ_2 לא ידועים. מה עושים?

פתרון. נמצא אומדי נראות מירבית $\hat{\sigma_2^2} = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}, \hat{\sigma_1^2} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$ לכן

$$\frac{T = \bar{X} - \bar{Y} - 5}{\sqrt{\frac{\hat{\sigma_1^2}}{n} + \frac{\hat{\sigma_2^2}}{n}}} \geq t_{2n,\alpha}$$

וזה יהיה המבחן.

11 שיעור חזרה:

תרגיל 11.1. יהיו $x_1,...,x_n \sim \mathcal{N}(\theta,\theta)$ בלתי תלויים.

(א) מצא אומד נראות מירבית ל-θ.

(ב) הוכיחו שהאומד שמצאתם לא אומד שלם.

פתרון.

(א) לפי ההגדרה:

$$\mathbb{P}(x) = \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{(x-\theta)^2}{2\theta^2}}$$

לכן

$$\begin{aligned}\mathbb{P}(x_1,...,x_n) = L(x_1,...,x_n) =& \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\theta} e^{-\frac{(x_i-\theta)^2}{2\theta^2}}\right)\\ =& \left(\frac{1}{\sqrt{2\pi}\theta}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\theta^2}}\end{aligned}$$

נוציא לוג ונקבל

$$\begin{aligned}\log(L) &= n\log\left(\frac{1}{\sqrt{2\pi}\theta}\right) - \sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\theta^2}\\ &= -n\log\left(\sqrt{2\pi}\theta\right) - \sum_{i=1}^{n}\frac{x_i^2}{2\theta^2} + \sum_{i=1}^{n}\frac{x_i}{\theta} - \frac{n}{2}\end{aligned}$$

נגזור ונקבל

$$\frac{\partial}{\partial\theta}\log(L) = -\frac{n}{\theta} + \frac{1}{\theta^3}\underbrace{\sum_{i=1}^{n} x_i^2}_{:=T_2} - \frac{1}{\theta^2}\underbrace{\sum_{i=1}^{n} x_i}_{:=T_1}$$

נשווה לאפס ונקבל

$$-n\theta^2 + T_2 - \theta T_1 = 0$$

כלומר

$$\theta = \frac{T_1 \pm \sqrt{T_1^2 + 4nT_2}}{-2n}$$

ובהצבה לאחור

$$\theta^* = \frac{\sum_{i=1}^{n} x_i \pm \sqrt{\left(\sum_{i=1}^{n} x_i\right)^2 + 4n\sum_{i=1}^{n} x_i^2}}{-2n}$$

(ב) נמצא פונקציה $g(T_1,T_2) \neq 0$ של T_1,T_2 שלכל θ יתקיים כי $\mathbb{E}[g(T_1,T_2)] = 0$:

נגדיר

$$\begin{aligned}\hat{\mu} &= \frac{T_1}{n}\\ \hat{\sigma} &= \sum_{i=1}^{n}\frac{(x_i-\hat{\mu})^2}{n-1}\\ g(T_1,T_2) &= \hat{\sigma} - \hat{\mu}^2\end{aligned}$$

ומתקיים כי

$$\begin{aligned}\mathbb{E}[\hat{\mu}] &= \theta\\ \mathbb{E}[\hat{\sigma}] &= \theta^2\end{aligned}$$

ולכן

$$\mathbb{E}[g(T_1,T_2)] = \mathbb{E}_\theta\left[\hat{\sigma} - \hat{\mu}^2\right] = \mathbb{E}_\theta\left[\theta^2 - \theta^2\right] = 0$$

ולכן $T = [T_1,T_2]$ לא אומד שלם!

*נשים ♡ כי האומד $\hat{\sigma} - \hat{\mu}^2$ הוא אומד נלווה (כי הוא לא תלוי ב-θ) בעוד ש-$\theta^* = \frac{\sum_{i=1}^{n} x_i \pm \sqrt{\left(\sum_{i=1}^{n} x_i\right)^2 + 4n\sum_{i=1}^{n} x_i^2}}{-2n}$ אומד מספיק ולכן הם בלתי תלויים!

תרגיל 11.2. נגדיר התפלגות על θ באופן הבא: אם $\theta = 0$ אז $f(x) = 1$ ואם $\theta = 1$ אז $f(x) = \frac{1}{2\sqrt{x}}$.

(א) מצאו אומד נראות מירבית ל-θ.

(ב) כעת משה מגריל בסיכוי θ מתוך התפלגות $f(x) = 1$ ובסיכוי $1-\theta$ מתוך $f(x) = \frac{1}{2\sqrt{x}}$ על הקטע $x \in [0,1]$. תן אומד נראות מירבית ל-$\theta \in [0,1]$.

פתרון.

(א) אם $\theta = 0$ אזי $f(x_i) = 1$ ולכן $f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i) = 1$.

מנגד לזה אם $\theta = 1$ אזי $f(x_i) = \frac{1}{2\sqrt{x_i}}$ ולכן $f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i) = \frac{1}{2^n\sqrt{x_1 \cdot x_2 \cdots x_n}}$.

אם $\frac{1}{2^{2n}} > x_1 \cdot x_2 \cdots x_n$ אזי $\theta = 1$, אחרת $\theta = 0$.

(ב) לפי נוסחאת ההסתברות השלמה

$$\mathbb{P}(A) = \sum_B \mathbb{P}(B)\mathbb{P}(A|B)$$

ולכן

$$\mathbb{P}(x) = \theta + \frac{1-\theta}{2\sqrt{x}}$$

כלומר

$$\mathbb{P}(x_1, ..., x_n) = \prod_{i=1}^{n} \left(\theta + \frac{1-\theta}{2\sqrt{x_i}}\right)$$

נוציא לוג:

$$\log\left(\mathbb{P}(x_1, ..., x_n)\right) = \sum_{i=1}^{n} \left(\theta + \frac{1-\theta}{2\sqrt{x_i}}\right)$$

ונגזור ונשווה לאפס ונקבל שצריך לפתור משהו.

תרגיל 11.3. נתונה התפלגות

$$\mathbb{P}(x) = (\alpha-1)x^{-\alpha}, x \in [1,\infty)$$

בהינתן דגימות $x_1, ..., x_n$, תוך כדי שימוש בשיטת המומנטים, מהו α?

פתרון. נגדיר $T = \sum_{i=1}^{n} \frac{x_i}{n}$. מצד אחד

$$\mathbb{E}[x] = \int_1^\infty (\alpha-1)x^{1-\alpha}dx = \frac{\alpha-1}{2-\alpha}$$

ומצד שני לפי שיטת המומנטים

$$\mathbb{E}[x] = T$$

לכו

$$\frac{\alpha-1}{2-\alpha} = T$$

ונפתור ל-α.

תרגיל 11.4. בהינתן סדרת אירועי ברנולי $x = [0,1,1,1,0,0,1,0,1,0,...]$ ופריור $\beta(a,b)$ להתפלגות תן אומד MAP לערך של ρ הסיכוי לעץ.

פתרון. נסמן ב-n את מספר הזריקות, $T = \sum_{i=1}^{n} x_i$ לכן

$$\mathbb{P}_{AP}(T) = \binom{n}{T}\rho^T(1-\rho)^{n-T} \cdot c\rho^a(1-\rho)^b = \binom{n}{T}c \cdot \rho^{T+a}(1-\rho)^{n-T+b}$$

ומכאן ממשיכים כרגיל

דוגמה 11.5. ניימן פירסון: יש לי שתי השערות $\begin{matrix} H_0 : \alpha = \alpha_0 \\ H_1 : \alpha = \alpha_1 \end{matrix}$ כאשר ההתפלגות שלי היא $\mathbb{P}(x) = \alpha e^{-\alpha x}, x \in [0,\infty)$ ודגימות $x_1, ..., x_n$.

אז כדי לדחות את השערת האפס אני צריך

$$\frac{L(\alpha_1|x_1, ..., x_n)}{L(\alpha_0|x_1, ..., x_n)} \geq c$$

ו-c הוא מקיים $L(\alpha_0|x_1, ..., x_n) = \alpha$. נגדיר $T = \sum_{i=1}^{n} \frac{x_i}{n}$ ובנוסף

$$\mathbb{P}_\alpha(x_1, ..., x_n) = \alpha^n e^{-\alpha T}$$

לכן המבחן ניימן פירסון שלי יהיה

$$\frac{\alpha_1^n e^{-\alpha_1 T}}{\alpha_0^n e^{-\alpha_0 T}} \geq c$$

כלומר המבחן הוא

$$\left(\frac{\alpha_1}{\alpha_0}\right)^n e^{(\alpha_0-\alpha_1)T} \geq c$$

ובנוסף c הוא ה-c שמקיים

$$\int\limits_c^\infty \alpha_0^n e^{-\alpha_0 T} dT = 0.05$$

תזכורת- מבחן וול: עבור $n \to \infty$ ו-$z = \frac{T-\mu}{\sigma/\sqrt{n}}$ נפעיל מבחן z על z.

תזכורת: אם יש לי $f(x|\theta)$ אזי אני יכול להגדיר $W_\theta = \frac{\partial}{\partial\theta}\log(f(x|\theta))$ ונגדיר את האינפורנציית פישר להיות

$$I_\theta = \mathbb{E}_\theta\left[-\frac{\partial}{\partial\theta}W_\theta\right]$$

דוגמה 11.6. רווחי סמך: נניח יש לי

$$\mathbb{P}(T) = \frac{1}{1+(1+T)^2}$$

כאשר T אומד של θ ו-θ לא ידוע.

נגדיר $[a,b]$ להיות הקטע עבורו

$$\forall x \in [a,b] : \frac{1}{1+(1+x)^2} \geq c$$

וגם

$$\int\limits_a^b \frac{1}{1+(1+x)^2} dx = 1-\alpha$$

כלומר רווח סמך הוא $[a,b]$ המקיימים $\int\limits_a^b \frac{1}{1+(1+x)^2}dx = 1-\alpha$ ורווח סמך מינימלי הוא אחד המקיים גם $c \leq \frac{1}{1+(1+x)^2} : \forall x \in [a,b]$.