PROBABILITY AND STATISTICS

הסתברות וסטטיסטיקה

מאת יוג'ין קנציפר

■ נוסחאון למבחן סוף סמסטר

הרצאה 1: מושגי יסוד

הגדרות בסיסיות

- **מרחב המידגם** Ω – אוסף כל התוצאות האפשריות של הניסוי.
- **מאורע** A – קבוצה חלקית כלשהי של תוצאות ניסוי.
- **הסתברות של המאורע** A היא $P(A)=|A|/|\Omega|$. הנוסחא מניחה **סבירות שווה** של כל אחת מן התוצאות האפשריות של ניסוי אקראי.

אלגברת מאורעות

- **איחוד** $A\bigcup B$ – אוסף כל המאורעות הכלולים ב- A או ב- B או בשניהם.
- **חיתוך** $A\bigcap B$ – אוסף כל המאורעות הכלולים ב- A וגם ב- B.
- **מינוס** $A\setminus B$ – אוסף כל המאורעות הכלולים ב- A ולא כלולים ב- B
- **השלמה** $\overline{A}$ – אוסף כל המאורעות הכלולים במרחב המדגם ולא כלולים ב- A.
- **הכלה** $A\subset B$ – כל תוצאה השייכת למאורע A שייכת גם למאורע B.

סוגי מאורעות

- **מאורע וודאי** Ω – מאורע שכולל את כל המאורעות של מרחב המדגם.
- **מאורע ריק** $\varnothing$ – תוצאה בלתי אפשרית, מאורע שאינו כולל אף מאורע של מרחב המדגם.
- **מאורעות זרים** – A ו- B הם מאורעות זרים אם הם לא כוללים מאורעות משותפים, $A\bigcap B=\varnothing$.
- **מאורעות זרים בזוגות** – A_i ו- A_j הם מאורעות זרים בזוגות אם $A_i\bigcap A_j=\varnothing$ לכל $i\neq j$.

תכונות של פונקצית הסתברות וחוקים חשובים

- $0\le P(A)\le 1$ לכל A.
- $P(\Omega)=1$ כאשר Ω הוא מרחב המדגם.
- אם A_i ו- A_j הם מאורעות זרים בזוגות, אזי $P(A_1\bigcup A_2\bigcup\ldots)=P(A_1)+P(A_2)+\ldots$.
- אם $A\subset B$ אזי $P(A)\le P(B)$.
- **חוק המשלים:** $P(\overline{A})=1-P(A)$.
- **חוק האיחוד:** $P(A\bigcup B)=P(A)+P(B)-P(A\bigcap B)$.
- כאשר $A\bigcap B=\varnothing$, מתקיים: $P(A\bigcup B)=P(A)+P(B)$.
- **חוק קומוטטיבי:** $A\bigcup B=B\bigcup A$, $A\bigcap B=B\bigcap A$.

- **חוק אסוציאטיבי:** $(A\cup B)\cup C=A\cup(B\cup C)=A\cup B\cup C$,
$(A\cap B)\cap C=A\cap(B\cap C)=A\cap B\cap C$.
- **חוק דיסטריבוטיבי:** $(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$,
$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$.
- **חוק דה מורגן:** $\overline{(A\cup B)}=\overline{A}\cap\overline{B}$, $\overline{(A\cap B)}=\overline{A}\cup\overline{B}$.

הרצאה 2: הסתברות מותנית

הסתברות מותנית ואי תלות

- **נוסחא להסתברות מותנית:** $P(A/B)=P(A\cap B)/P(B)$.
- **אי תלות** – המאורעות A ו- B הם מאורעות בלתי תלויים אם $P(A\cap B)=P(A)\cdot P(B)$.
במילים אחרות, עבור מאורעות בלתי תלויות מתקיים: $P(A/B)=P(A)$.

הסתברות שלמה

- **הסתברות שלמה** – אם $A_1,A_2,\dots,A_n$ הם מאורעות זרים בזוגות, כך ש- $A_i\cap A_j=\varnothing$ עבור כל זוג $i\neq j$, ואיחודם הוא כל מרחב המדגם $\bigcup_{i=1}^{n}A_i=\Omega$ (זאת אומרת $A_1\cup A_2\cup\ldots=\Omega$), אזי לכל מאורע B מתקיים:
$P(B)=\sum_{i=1}^{n}P(B/A_i)\cdot P(A_i)=P(B/A_1)\cdot P(A_1)+P(B/A_2)\cdot P(A_2)+\ldots+P(B/A_n)\cdot P(A_n)$.

משפט בייס

- **משפט בייס** – עבור מאורעות A ו- B בעלי הסתברות חיובית מתקיים:
$$P(A/B)=\frac{P(B/A)\cdot P(A)}{P(B)}.$$
- **משפט בייס בשילוב נוסחא להסתברות שלמה:**
$$P(A_i/B)=\frac{P(B/A_i)\cdot P(A_i)}{\sum_{j=1}^{n}P(B/A_j)\cdot P(A_j)}.$$

הרצאה 3: קומבינטוריקה

עצרת של מספר, פונקציית גאמא ונוסחת סטירלינג

- **עצרת של מספר:** מוגדר עבור מספרים שלמים חיוביים $n=1,2,3,\ldots$ ושווה ל- $n!=1\cdot2\cdot3\cdot\ldots\cdot n$.
- **פונקציית גאמא:** $\Gamma(\alpha+1)=\int_0^{+\infty}dt\ t^{\alpha}e^{-t}$ (מוגדרת עבור $\alpha>-1$). עבור $\alpha=n$ מתקיים:
$n!=\Gamma(n+1)$.
- **נוסחת סטירלינג:** עבור $n>>1$ מתקיים בקירוב $n!\approx\sqrt{2\pi}\cdot n^{n+1/2}e^{-n}$.

עקרון הכפל

- **ניסוח ראשון:** אם ניסוי מתבצע ב- k שלבים בזה אחר זה כאשר בשלב הראשון יש n_1 תוצאות אפשריות, בשלב השני יש n_2 תוצאות אפשריות, ..., בשלב ה- k יש n_k תוצאות אפשריות, אזי מספר התוצאות האפשריות בניסוי כולו שווה ל- $n_1 \cdot n_2 \cdot \ldots \cdot n_k$.
- **ניסוח שני – כלל שרשרת:** לכל $A_1, A_2, \ldots, A_n$ מתקיים:
$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2 / A_1) \cdot P(A_3 / A_1 \cap A_2) \cdot \ldots \cdot P(A_n / A_1 \cap \ldots \cap A_{n-1})$.

מספר סידורים ונוסחא מולטינומית

- של n איברים <u>שונים</u> בשורה שווה ל- $P_n = n!$.
- של n איברים <u>שונים</u> במעגל שווה ל- $(n-1)!$.
- של n איברים בשורה שמתוכם n_1 איברים זהים מסוג ראשון, n_2 איברים זהים מסוג שני, .., n_k איברים זהים מסוג ה- k (כך ש- $n_1 + n_2 + \ldots + n_k = n$) שווה ל-
$P_n(n_1, \ldots, n_k) = \dfrac{n!}{n_1! n_2! \ldots n_k!}$.
- נוסחא מולטינומית

$$(x_1 + x_2 + \ldots + x_k)^n = \sum_{\substack{(n_1 \geq 0, \ldots, n_k \geq 0): \\ n_1 + \ldots + n_k = n}} P_n(n_1, \ldots, n_k)\ x_1^{n_1} \ldots x_k^{n_k} = \sum_{\substack{(n_1 \geq 0, \ldots, n_k \geq 0): \\ n_1 + \ldots + n_k = n}} \frac{n!}{n_1! \ldots n_k!}\ x_1^{n_1} \ldots x_k^{n_k}$$

מספר בחירות

בלי החזרה	עם החזרה	
$P_n^k = \dfrac{n!}{(n-k)!}$	n^k	**מסודרת**
$C_n^k = \dfrac{n!}{k!(n-k)!}$	$C_{n+k-1}^k = \dfrac{(n+k-1)!}{k!(n-1)!}$	**לא מסודרת**

הרצאה 4: משתנה מקרי חד ממדי

פונקציית הסתברות $P_X(x)$ ופונקציית התפלגות מצטברת $F_X(t)$

- **תכונות של פונקציית הסתברות:**
  א. $0 \leq P(x) \leq 1$ לכל ערך x.
  ב. $\sum_x P(x) = 1$.
- **פונקציית התפלגות מצטברת:** $F_X(t) = P(X \leq t) = \sum_{x \leq t} P_X(x)$.

- **תכונותיה:**
  א. $F_X(t=-\infty)=0$.
  ב. $F_X(t=+\infty)=1$.
  ג. אם $t_1<t_2$, $F_X(t_1)\le F_X(t_2)$.
  ד. אם $F_X(t_0)=0$ עבור t_0 כלשהו, אזי $F_X(t)=0$ עבור כל $t\le t_0$.
  ה. $P(x_1<X\le x_2)=F_X(x_2)-F_X(x_1)$ עבור $x_1<x_2$.

מדדים של משתנה מקרי

- **תוחלת ותכונותיה:**
  א. $\mu=E[X]=\sum_x x\cdot P_X(x)$.
  ב. $E[g(X)]=\sum_x g(x)\cdot P_X(x)$.
  ג. $E[a]=a$, a קבוע.
  ד. $E[X+a]=E[X]+a$, a קבוע.
  ה. $E[b\cdot X]=b\cdot E[X]$, b קבוע.
  ו. $E[X_1+X_2]=E[X_1]+E[X_2]$.

- **שונות, סטיית תקן ותכונותיה:**
  א. $\sigma^2=Var[X]=E[(X-\mu)^2]$.
  ב. $\sigma^2=Var[X]=E[X^2]-(E[X])^2$.
  ג. $\sigma=\sqrt{Var[X]}$.
  ד. $Var[X]\ge 0$ לכל משתנה מקרי X.
  ה. $Var[X]=0$ אך ורק כאשר X הוא קבוע.
  ו. $Var[X+a]=Var[X]$, a קבוע.
  ז. $Var[b\cdot X]=b^2\cdot Var[X]$, b קבוע.
  ח. אם X ו-Y הם משתנים מקריים בלתי תלויים (מתייחסים לשני ניסויים בלתי תלויים), אזי מתקיים: $Var[X+Y]=Var[X]+Var[Y]$.

- **חציון:**
  א. הוא הערך M המקיים את שני התנאים הבאים: $P(X<M)\le\frac{1}{2}$ ו-$P(X\le M)\ge\frac{1}{2}$.
  ב. הוא הערך הראשון, שעבורו פונקצית התפלגות מצטברת היא לפחות $\frac{1}{2}$.

- **ערך החלוקה ה-p :**
  א. הוא ערך M_p המקיים את שני התנאים הבאים: $P(X<M_p)\le p$ ו-$P(X\le M_p)\ge p$.
  ב. הוא הערך הראשון, שעבורו פונקצית התפלגות מצטברת היא לפחות p, $0<p<1$.

- **שכיח:** הערך של משתנה מקרי X בעל שכיחות גבוהה ביותר

הרצאה 5: התפלגויות בדידות מיוחדות

התפלגות אחידה

- **סימון:** $X \sim U(a,b)$, a ו- b מספרים שלמים.
- **פונקצית הסתברות:** $P(X=k)=\dfrac{1}{b-a+1}$ עבור $k=a,a+1,\dots,b$.
- **תוחלת:** $E[X]=\dfrac{a+b}{2}$.
- **שונות:** $Var[X]=\dfrac{(b-a+1)^2-1}{12}$.
- **פונקצית התפלגות מצטברת:** $F_X(k)=P(X\le k)=\dfrac{k-a+1}{b-a+1}$ עבור $k=a,a+1,\dots,b$.

התפלגות ברנולי – "הצלחה" או "כשלון"

- **סימון:** $X \sim Ber(p)$, $0\le p\le 1$.
- **פונקצית הסתברות:** $P_X(x)=\begin{cases}1-p, & x=0\\ p, & x=1\end{cases}$.
- **תוחלת:** $E[X]=p$.
- **שונות:** $Var[X]=p(1-p)$.

התפלגות בינומית – מספר כולל של "הצלחות" בסדרת ניסוי ברנולי

- **סימון:** $X \sim B(n,p)$, $0\le p\le 1$, n מספר שלם חיובי
- **פונקצית הסתברות:** $P(X=k)=C_n^k p^k(1-p)^{n-k}$ עבור $k=0,1,\dots,n$.
- **תוחלת:** $E[X]=np$.
- **שונות:** $Var[X]=np(1-p)$.
- **משפט הפירוק:** סכום X של n משתני ברנולי $X_1,X_2,\dots,X_n$ בלתי תלויים בעלי אותו פרמטר p מתפלג בינומית עם פרמטרים n ו- p:

$$X=\sum_{k=1}^{n}X_k=X_1+X_2+\dots+X_n \sim B(n,p).$$

- **קירוב פואסוני להתפלגות בינומית:** אם פרמטר n גדול מאוד ($n>>1$) ו- p קטן מאוד ($p<<1$) כך שהכפל $\lambda=n\cdot p$ מקבל ערך "בינוני", בקירוב מתקיים:

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

התפלגות גיאומטרית – מספר ניסוי ברנולי עד ל"הצלחה הראשונה"

- **סימון:** $X \sim G(p)$, $0 \le p \le 1$.
- **פונקצית הסתברות:** $P(X=k)=p(1-p)^{k-1}$ עבור $k=1,\dots,\infty$.
- **פונקצית התפלגות מצטברת:** $F_X(k)=P(X\le k)=1-(1-p)^k$ עבור $k=1,\dots,\infty$.
- **הסתברות:** $P(X>k)=(1-p)^k$ עבור $k=1,\dots,\infty$.
- **תוחלת:** $E[X]=\frac{1}{p}$.
- **שונות:** $Var[X]=\frac{1-p}{p^2}$.
- **תכונת "חוסר זיכרון":** $P(X>k+n \;/\; X>k)=P(X>n)$

התפלגות בינומית שלילית – מספר ניסוי ברנולי עד ל"הצלחה ה- m"

- **סימון:** $X \sim NB(p,m)$, $0\le p\le 1$ ו- $m=1,2,\dots,\infty$
- **פונקצית הסתברות:** $P(X=k)=C_{k-1}^{m-1}p^m(1-p)^{k-m}$ עבור $k=m,m+1,\dots,\infty$.
- **תוחלת:** $E[X]=\frac{m}{p}$.
- **שונות:** $Var[X]=\frac{m(1-p)}{p^2}$.

התפלגות היפרגיאומטרית – מספר פריטים מיוחדים בבחירה ללא החזרה מאוסף מעורב

- **סימון:** $X \sim Hyp(N,D,n)$, $n,D\le N$.
- **פונקצית הסתברות:** $P(X=k)=\frac{C_D^k\cdot C_{N-D}^{n-k}}{C_N^n}$ עבור $0\le k\le \min(n,D)$.
- **תוחלת:** $E[X]=n\frac{D}{N}$.
- **שונות:** $Var[X]=n\frac{D}{N}\left(1-\frac{D}{N}\right)\left(1-\frac{n-1}{N-1}\right)$.
- **קירוב בינומי להתפלגות היפרגיאומטרית:** אם $D,N>>1$ אך $\frac{D}{N}=p$ שומר על ערך סופי וגם $k\le n<<D<N$, בקירוב מתקיים: $P(X=k)\approx C_n^k p^k(1-p)^{n-k}$.
- **קירוב בינומי להתפלגות היפרגיאומטרית:** אם $D,N>>1$ אך $\frac{D}{N}=p$ שומר על ערך סופי וגם $k\le n<<D<N$, בקירוב מתקיים: $P(X=k)\approx C_n^k p^k(1-p)^{n-k}$.

התפלגות היפרגיאומטרית שלילית – מספר הוצאות ללא החזרה עד הפריט המיוחד ה- m

- **סימון:** $X_m \sim NegHyp(m;N,D)$, $1\le m\le D$, $1\le D<N$.
- **פונקצית הסתברות:** $P(X_m=k)=C_{k-1}^{m-1}\dfrac{C_{N-k}^{D-m}}{C_N^D}$ עבור $k=m,\dots,N-D+m$.
- **תוחלת:** $E[X_m]=m\dfrac{N+1}{D+1}$.
- **שונות:** $\operatorname{var}[X_m]=m\dfrac{N+1}{D+1}\dfrac{N-D}{D+2}\left(1-\dfrac{m}{D+1}\right)$.

התפלגות פואסון – מספר התרחשויות ביחידת זמן בזרם אירועים פואסוני

- **סימון:** $X\sim P(\lambda)$, $\lambda>0$.
- **פונקצית הסתברות:** $P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$ עבור $k=0,\dots,\infty$.
- **תוחלת:** $E[X]=\lambda$.
- **שונות:** $Var[X]=\lambda$.

הרצאה 6: משתנה מקרי דו ממדי

פונקצית הסתברות משותפת $P_{XY}(x,y)$ ופונקציות הסתברות שוליות $P_X(x)$ ו-$P_Y(y)$

- **תכונות של פונקצית הסתברות משותפת:**
  א. $0\le P_{XY}(x_i,y_j)\le 1$ לכל זוג ערכים אפשרי (x_i,y_j).
  ב. $\sum_{x_i}\sum_{y_j}P_{XY}(x_i,y_j)=1$.

- **פונקציות הסתברות שוליות:**
  א. $P_X(x_i)=\sum_{y_j}P_{XY}(x_i,y_j)$.
  ב. $P_Y(y_j)=\sum_{x_i}P_{XY}(x_i,y_j)$.

תלות ומתאם

- **אי תלות:** שני משתנים מקריים בעלי התפלגות משותפת P_{XY} נקראים משתנים בלתי תלויים אם <u>לכל זוג</u> (x_i,y_j) מתקיים $P_{XY}(x_i,y_j)=P_X(x_i)P_Y(y_j)$.
- **אי מתאם:** שני משתנים מקריים נקראים בלתי מתואמים אם מתקיים $E[X\cdot Y]=E[X]\cdot E[Y]$. פה $E[X\cdot Y]=\sum_{x_i,y_j}x_iy_jP(x_i,y_j)$.
- אם X,Y הם משתנים בלתי תלויים, אזי הם גם בלתי מתואמים. ההפך לא בהכרח נכון!

מדדים של משתנה דו ממדי

- **שונות משותפת:** $Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E[Y]$.
- **מקדם המתאם:** $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$.
- **תכונות של מקדם המתאם:**

א. $\rho(X,Y) = \rho(Y,X)$.

ב. לכל α, β, γ ו- δ חיוביים מתקיים $\rho(\beta X + \alpha, \gamma Y + \delta) = \text{sgn}(\beta\gamma)\rho(X,Y)$.

ג. לכל X ו- Y מתקיים: $-1 \le \rho(X,Y) \le +1$.

ד. עבור משתנים בלתי מתואמים $\rho(X,Y) = 0$.

ה. אם $Y = aX + b$ אזי: $\rho(X,Y) = +1$ עבור $a > 0$ ו- $\rho(X,Y) = -1$ עבור $a < 0$.

שונות של סכום המשתנים ורגרסיה ליניארית

- **שונות הסכום:** $Var[X+Y] = Var[X] + Var[Y] + 2Cov(X,Y)$.
- **רגרסיה ליניארית:** הניבוי <u>הליניארי</u> הטוב ביותר ל- Y בהינתן X ניתן על ידי קו הרגרסיה $Y = aX + b$ עם הפרמטרים

$$b = E[Y] - E[X]\frac{Cov(X,Y)}{Var[X]}, \quad a = \frac{Cov(X,Y)}{Var[X]} = \rho(X,Y)\sqrt{\frac{Var[Y]}{Var[X]}}.$$

עבור מדגמים גדולים, יש להשתמש בנוסחאות: $E[X] \approx \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$,

$Var[X] \approx \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$ ו- $Cov(X,Y) \approx \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)\left(Y_k - \overline{Y}\right)$.

הרצאה 7: משתנה מקרי רציף

הסתברות, פונקציית צפיפות ופונקציית התפלגות מצטברת

- **הסתברות:** $P(a < X < b) = \int_a^b f_X(x)dx$.
- **תכונותיה של פונקציית צפיפות:**

א. $f_X(x) \ge 0$. ב. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

- **פונקציית התפלגות מצטברת:** $F_X(t) = P(X \le t) = \int_{-\infty}^{t} f_X(x)dx$.
- **נוסחת הקשר:** $f_X(x) = \frac{d}{dx}F_X(x)$.
- **תוחלת:** $E[X] = \int_{-\infty}^{+\infty} x f_X(x)dx$.

שונות: $Var[X] = E[X^2] - (E[X])^2 = \int_{-\infty}^{+\infty} x^2 f_X(x)dx - \left(\int_{-\infty}^{+\infty} x f_X(x)dx\right)^2$.

התפלגות אחידה

- **סימון:** $X \sim U(a,b)$, $a<b$.
- **פונקציית צפיפות:** $f_X(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b\\ 0 & \text{אחרת}\end{cases}$.
- **פונקציית התפלגות מצטברת:**
$$F_X(t)=P(X\le t)=\int_{-\infty}^{t} f_X(x)dx=\begin{cases}0, & t<a\\ \frac{t-a}{b-a}, & a\le t<b\\ 1, & t\ge b\end{cases}.$$
- **תוחלת:** $E[X]=\frac{a+b}{2}$.
- **שונות:** $Var[X]=\frac{(b-a)^2}{12}$.

התפלגות מעריכית – זמן המתנה עד התרחשותו האירוע הראשון בזרם פואסוני

- **סימון:** $Y \sim Exp(\lambda)$, $\lambda>0$.
- **פונקציית צפיפות:** $f_Y(y)=\begin{cases}\lambda\exp(-\lambda y), & y\ge 0\\ 0, & y<0\end{cases}$.
- **פונקציית התפלגות מצטברת:** $F_Y(t)=P(Y\le t)=1-\exp(-\lambda t)$
- **תוחלת:** $E[Y]=\frac{1}{\lambda}$.
- **שונות:** $Var[Y]=\frac{1}{\lambda^2}$.
- **תכונת "חוסר זיכרון":** $P(Y>s+t \;/\; Y>s)=P(Y>t)$.
- **זמן המתנה** עד תרחשותו <u>האירוע ה-</u> k בזרם אירועים פואסוני מתפלג על פי החוק: $f_T(t)=\frac{\lambda^k}{(k-1)!}t^{k-1}e^{-\lambda t}$.

התפלגות נורמלית

- **סימון:** $X \sim N(\mu,\sigma^2)$.
- **פונקציית צפיפות:** $f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ עבור $-\infty<x<+\infty$.
- **תוחלת:** $E[X]=\mu$.
- **שונות:** $Var[X]=\sigma^2$.
- **פונקציית-** Φ **:** $\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} dz e^{-z^2/2}$.

- **פונקצית התפלגות מצטברת:** $F_X(t)=P(X\le t)=\Phi\left(\frac{t-\mu}{\sigma}\right)$.
- **הסתברות:** $P(a\le X\le b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$.
- **תכונה:** סכום של משתנים נורמליים $X_1\sim N(\mu_1,\sigma_1^2)$ ו- $X_2\sim N(\mu_2,\sigma_2^2)$ בלתי תלויים, $X=X_1+X_2$, מתפלג נורמלית גם כן: $X\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

הרצאה 8: משפט הגבול המרכזי פלוס

קירוב נורמלי להתפלגות בינומית – משפט דה מואבר-לפלס

- עבור $n>>1$, **משתנה בינומי** $X_B\sim B(n,p)$ מתנהג **כמשתנה נורמלי** $X_B\mapsto X_N\sim N(np,np(1-p))$.
- **תיקון רציפות בחישוב הסתברויות:**

א. $P(X_B=k)=P(k-½<X_N<k+½)=\Phi\left(\frac{k+½-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{k-½-np}{\sqrt{np(1-p)}}\right)$.

ב. $P(X_B\le k)=P(X_N\le k+½)=\Phi\left(\frac{k+½-np}{\sqrt{np(1-p)}}\right)$.

משפט הגבול המרכזי

- **משפט:** יהיו $X_1,X_2,\dots,X_n$ משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה μ ושונותה σ^2. אם n גדול מאוד, $n>>1$, המשתנה $X=\sum_{k=1}^{n}X_k$ מתפלג נורמלית בקירוב: $X=\sum_{k=1}^{n}X_k\sim N(n\mu,n\sigma^2)$.
- **הסתברויות:**

א. אם $X_1,X_2,\dots,X_n$ הם משתנים <u>רציפים</u>, בקירוב מתקיים:

$$P(k_1\le X\le k_2)=\Phi\left(\frac{k_2-n\mu}{\sigma\sqrt{n}}\right)-\Phi\left(\frac{k_1-n\mu}{\sigma\sqrt{n}}\right).$$

ב. אם $X_1,X_2,\dots,X_n$ הם משתנים <u>בודדים</u>, בקירוב מתקיים:

$$P(k_1\le X\le k_2)=\Phi\left(\frac{k_2+\frac{1}{2}-n\mu}{\sigma\sqrt{n}}\right)-\Phi\left(\frac{k_1-\frac{1}{2}-n\mu}{\sigma\sqrt{n}}\right).$$

- **משפט על ממוצע:** יהיו $X_1,X_2,\dots,X_n$ התצפיות עבור משתנה מקרי כלשהו. אם n גדול מאוד, $n>>1$, הממוצע $\overline{X}_n=\frac{1}{n}\sum_{k=1}^{n}X_k$ מתפלג נורמלית בקירוב $\overline{X}_n\sim N\left(\mu,\frac{\sigma^2}{n}\right)$.

הרצאה 9: בעיות אמידה קשורות להתפלגות נורמלית

אומדים נקודתיים חסרי הטיה

- אומד נקודתי לתוחלת μ: $\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$.
- אומד נקודתי לשונות σ^2: $\hat{S}^2=\frac{1}{n-1}\sum_{k=1}^{n}\left(X_k-\bar{X}\right)^2$.
- אומד נקודתי לשונות σ^2: $S^2=\frac{1}{n}\sum_{k=1}^{n}\left(X_k-\mu\right)^2$.

רווח סמך לתוחלת μ כאשר σ ידועה

- רווח סמך: $\bar{X}-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$ ברמת סמך $(1-\alpha)\cdot 100\%$.

תנאים – (מספר התצפיות גדול – $n\geq 30$ ולא חשוב איך מפולגות התצפיות) או (התצפיות מפולגות נורמלית ללא הגבלה על מספר)

- גודל המדגם עבור שגיאת אמידה מרבית נתונה ε: $n\geq\left(\frac{z_{\frac{\alpha}{2}}\cdot\sigma}{\varepsilon}\right)^2$ (שלם).
- פרמטר $z_{\frac{\alpha}{2}}$: $\Phi\left(z_{\frac{\alpha}{2}}\right)=1-\frac{\alpha}{2}$.

רווח סמך לתוחלת μ כאשר σ לא ידוע

- רווח סמך: $\bar{X}-z_{\frac{\alpha}{2}}\cdot\frac{\hat{S}}{\sqrt{n}}<\mu<\bar{X}+z_{\frac{\alpha}{2}}\cdot\frac{\hat{S}}{\sqrt{n}}$ ברמת סמך $(1-\alpha)\cdot 100\%$.

תנאים – (מספר התצפיות גדול – $n\geq 30$ ולא חשוב איך מפולגות התצפיות)

- פרמטר $z_{\frac{\alpha}{2}}$: $\Phi\left(z_{\frac{\alpha}{2}}\right)=1-\frac{\alpha}{2}$.

הרצאה 10: בעיות אמידה קשורות להתפלגויות t ו- χ^2

התפלגות χ^2

- סימון: $Y\sim\chi^2(n)$, פרמטר n – מספר דרגות חופש.
- פונקצית צפיפות: $f_Y(y;n)=\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}\cdot y^{(n-2)/2}\exp\left(-y/2\right)$ עבור $y\geq 0$.

- **תוחלת:** $E[Y]=n$.
- **שונות:** $Var[Y]=2n$.
- **נקודות החלוקה** $\chi^2_{\frac{\alpha}{2}}(n)$ **ו-** $\chi^2_{1-\frac{\alpha}{2}}(n)$ **:**

$$\int_0^{\chi^2_{1-\frac{\alpha}{2}}(n)} dy\, f_Y(y;n)=\frac{\alpha}{2} \quad \text{ו-} \quad \int_0^{\chi^2_{\frac{\alpha}{2}}(n)} dy\, f_Y(y;n)=1-\frac{\alpha}{2}.$$

(אפשר לחשב אותן באמצעות טבלת ההתפלגות-χ^2).

- **משפט:** יהיו $X_1, X_2, ..., X_n$ משתנים נורמליים סטנדרטיים בלתי תלויים. אזי המשתנה $Y=\sum_{k=1}^{n} X_k^2$ מתפלג התפלגות-χ^2 עם n דרגות חופש: $Y \sim \chi^2(n)$.

רווח סמך לשונות σ^2 – אמידה על סמך מדגמים קטנים

- **רווח סמך:** $\frac{(n-1)\cdot \hat{S}^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)\cdot \hat{S}^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$ ברמת סמך $(1-\alpha)\cdot 100\%$.

תנאים – (תצפיות מפולגות נורמלית ללא הגבלה על מספרן)

התפלגות t

- **סימון:** $T \sim t(n)$, פרמטר n – מספר דרגות חופש.
-
- **פונקצית צפיפות:** $f_T(t;n)=\frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$ עבור $-\infty < t < +\infty$.
- **תוחלת:** $E[T]=0$ אם $n>1$.
- **שונות:** $Var[T]=\frac{n}{n-2}$ אם $n>2$.
- **נקודות החלוקה** $t_{\frac{\alpha}{2}}(n)$ **:** $\int_{-\infty}^{t_{\frac{\alpha}{2}}(n)} dt\, f_T(t;n)=1-\frac{\alpha}{2}$.

(אפשר לחשב אותה באמצעות טבלת ההתפלגות-t).

- **תכונות:**

א. עבור $n\to\infty$, המשתנה $T \sim t(n)$ שואף למשתנה נורמלי סטנדרטי $T \sim N(0,1)$.

ב. $\lim_{n\to\infty} t_{\frac{\alpha}{2}}(n)=z_{\frac{\alpha}{2}}$.

- **משפט:** יהיו $X_0, X_1, X_2, \dots, X_n$ משתנים נורמליים סטנדרטיים בלתי תלויים. אזי

  המשתנה $T = \dfrac{X_0}{\sqrt{\frac{1}{n}\sum_{k=1}^{n} X_k^2}}$ מתפלג התפלגות-t עם n דרגות חופש: $T \sim t(n)$.

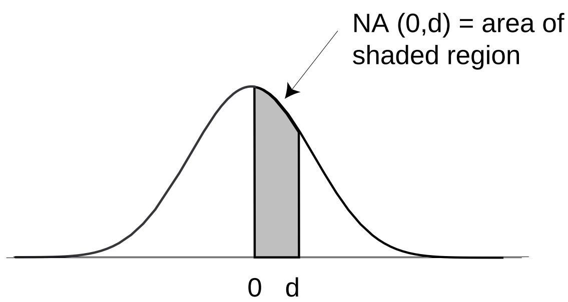
רווח סמך לתוחלת μ – אמידה על סמך מדגמים קטנים

- **רווח סמך:** $\overline{X} - t_{\frac{\alpha}{2}}(n-1) \cdot \dfrac{\hat{S}}{\sqrt{n}} < \mu < \overline{X} + t_{\frac{\alpha}{2}}(n-1) \cdot \dfrac{\hat{S}}{\sqrt{n}}$ ברמת סמך $(1-\alpha) \cdot 100\%$.

  תנאים – (תצפיות מפולגות נורמלית ללא הגבלה על מספרן**)**

Appendix to the Lecture No 7

Normal distribution table: Don't forget to add 1/2 !!



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

Table of Student's t-Probabilities: $P_\nu(T \geq t_\alpha(\nu)) = \alpha$

	α					
ν	**0.250**	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
∞	0.674	1.282	1.645	1.960	2.326	2.576

Useful properties:

i. $P_\nu(T \geq t) = P_\nu(T \leq -t)$

ii. $P_\nu(T \geq t) = 1 - P_\nu(T \geq -t)$

Table of χ^2-Probabilities

	α									
ν	**0.995**	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1			0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169