פונקציית הסתברות של מ"מ: $0 \le p\left(X=x\right) \le 1$

ניסויים ב"ת והתפלגויות בדידות:

ניסוי ברנולי Ber(p): ניסוי בעל שתי תוצאות

האפשריות: 'הצלחה'- S ו'כישלון'- F :

$bin\ /\ n$ $\boxed{Var(X)=pq}$ $\boxed{EX=p}$.

התפלגות בינומית Bim(n,p): ההסתברות לקבל בדיוק k הצלחות בסדרה של n ניסויים שלכל אחד

הסתברות הצלחה p . $\boxed{P(k)=\binom{n}{k}p^k q^{n-k}}$

סכום בינומים: אם $X \sim B(n,p)$ וגם $Y \sim B(m,p)$

(אותו p בשניהם) אז : $X+Y \sim B(m+n,p)$

התפלגות גיאומטרית Geo(p):

ההסתברות להצלחה ראשונה בנסיון ה- k כאשר

ההסתברות ההצלחה בכל ניסוי היא p . $\boxed{p(k)=q^{k-1}p}$

חוסר זיכרון: אם ידוע שלא הייתה הצלחה עד לניסוי ה-m אז ההסתברות כי היא תקרה בניסוי m+k הינה עדיין כמו בנוסחה.

התפלגות בינומית שלילית NB(m,p):

ההסתברות לקבל את ההצלחה ה- m -ית בדיוק

בניסיון ה- k כאשר לכל ניסוי הסת' הצלחה p .

$\boxed{p(k)=\binom{k-1}{m-1}p^m q^{k-m}}$ כמו סכום של גיאומטרי

התפלגות היפר-גיאומטרית HG(N,D,n): בכד

נמצאים N כדורים שחורים ו- M כדורים לבנים.

מוציאים n כדורים באקראי ללא החזרה. ההסתברות שבין הכדורים שהוצאו נמצאים k כדורים שחורים.

$$\boxed{P(k)=\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \quad n<N,\ D<N}$$

מרובה סוגים: כאשר $n_1+..+n_k=n$ ובוחרים n מ N.

$$P(n_j \text{ of type } 1\le j\le k)=\frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdot\ldots\cdot\binom{N_k}{n_k}}{\binom{N}{n}}$$

קירוב בינומי להיפר גיאומטרי: כאשר N גדול מאוד ביחס ל- n , אזי תוצאת החישוב עם ההחזרה (הבינומית) קרובה לתוצאת החישוב בלי החזרה (ההיפר גיאומטרית). $p=D/N,\ n=n$.

$$\boxed{X \sim HG(n,N,M) \Rightarrow X \sim Bin\left(n,\frac{M}{N+M}\right)}$$

התפלגות פואסון בדידה Pois(λ): מספר האירועים ביחידת זמן נתונה, אם ידוע כי הם מתרחשים בקצב ממוצע קבוע $\lambda>0$ ובאופן ב"ת בפרק הזמן מאז

האירוע האחרון. $\boxed{p(k)=\frac{e^{-\lambda}\lambda^k}{k!}}$

מירוץ פואסון: $\boxed{X_1|X_1+X_2=n \sim Bin\left(n,\lambda_1/\lambda_1+\lambda_2\right)}$

קירוב פואסוני לבינומי: כאשר n גדול ו-p קטן אז ניתן לקרב את הבינומי לפואסוני עם $\lambda=np$.

הערה! ניתן לעשות קירוב נורמלי לבינומי-מג"מ!

סכום פואסונים ב"ת: $X \sim Pois(\lambda)$, $Y \sim Pois(\mu)$

$(X+Y) \sim Pois(\lambda+\mu)$.

פיצול פואסון: אם $Z \sim Pois(\mu)$ ואז מתרחש אחד

משני מאורעות בהסתברויות $p_a+p_b=1$ אז נקבל

מ"מים ב"ת שמתפלגים $Pois(p_a\cdot\mu), Pois(p_b\cdot\mu)$.

ב"ת!: מתקיים פיצול לכל שתי אפשרויות של "האנשים" שמגיעים. אלו שני פואסון ב"ת!

התפלגות אחידה Uni(k,n): כל ערך מתוך טווח ערכים מסוים, מתקבל באותה הסתברות.

סטטיסטי מספיק- דוגמא:

$$f_X=\prod\frac{1}{\mu}e^{-\frac{1}{\mu}(x_i-\theta)}=\left(\frac{1}{\mu}\right)^n e^{-\frac{1}{\mu}\sum x_i}e^{-\frac{n\theta}{\mu}}$$

ס"מ ל- μ : $S=\sum X_i$. ס"מ ל- θ :

מוסיפים $I\{X_{(1)}>\theta\}$ $S=X_{(1)}$. ס"מ ל- (θ,μ)

$S=\left(X_{(1)},\sum X_i\right)$ ($h(\bar{x})=1$).

דוגמא-2:

$$f_X=\prod\beta x_i^{\beta-1}=\beta^n\left(\prod x_i\right)^{\beta-1}$$

$$S=\prod x_i \text{ or } \sum \ln x_i$$

$$\boxed{\frac{1}{n}\sum X_i^2-\bar{X}^2=\frac{1}{n}\sum\left(X_i-\bar{X}\right)^2}$$

$$\boxed{\sum\left(X_i-\theta\right)^2=\sum\left(X_i-\bar{X}\right)^2+\sum\left(\bar{X}-\theta\right)^2}$$

התפלגות משותפת ושולית:

התפלגות משותפת: חיתוך המאורעות

$\boxed{P(X=x,Y=y)=P_{XY}(x,y)}$

התפלגות שולית: $\boxed{P_X(x)=\sum_y P_{XY}(x,y)}$

אי-תלות מ"מ בדידים:

מ"מ הם ב"ת אם"ם

$\boxed{\forall x,y:\ P_{XY}(x,y)=P_X(x)\cdot P_Y(y)}$

אם הסתברות משותפת היא מכפלת שתי פונקציות, האחת של x והשנייה של y, $f(x)\cdot g(y)$ אזי ניתן להסיק שקיימת אי-תלות בין המשתנים.

הערה: פונקציות ותת-קבוצות זרות של מ"מ בלתי תלויים הן בלתי תלויות.

תוחלת: $\boxed{E(X)=\sum_{x\in R_X} xP(X=x)}$

($\sum_x |x|\cdot P(X=x)<\infty$

$E\left(X+Y\right)=EX+EY \quad E\left(aX+b\right)=aEX+b$

$E(g(X,Y))=\sum_x\sum_y g(x,y)\cdot P(X=x,Y=y)$

$E(g(X))=\sum_x g(x)\cdot P(X=x)$

כלל הכפל: אם X,Y ב"ת אז $\boxed{E(XY)=E(X)\cdot E(Y)}$

זוג מ"מ שמקיים את הנוסחה: **בלתי מתואמים.**

לכל שני מ"מ: $\boxed{E(XY)=\sum_x\sum_y xyP(X=x,Y=y)}$

חוק המספרים הגדולים: עבור X_i ב"ת ו-

$EX_i=\mu$ אז: $\boxed{\frac{1}{n}\sum_{i=1}^{n}X_i\to\mu}$ **החלש:** לכל

$\varepsilon>0 \quad P\left(|\bar{X}_n-\mu|\ge\varepsilon\right)\to 0$ **החזק:**

$P\left(\lim_{x\to\infty}|X_n-\mu|=0\right)=1$

נוסחת הזנב: $\boxed{E(X)=\sum_{k=1}^{\infty}P(X\ge k)}$ ($X\in\mathbb{N}$)

תוחלת מותנית:

$\boxed{E(X\mid Y=y)=\sum_x x\cdot P(X=x\mid Y=y)}$,

$\boxed{E(Y)=\sum_x E(Y\mid X=x)\cdot P(X=x)}$

אם X,Y ב"ת אז $E(X\mid Y)=E(X)$

במקרה הרציף: $E(X\mid Y=y)=\int_{-\infty}^{\infty}xf_{X|Y}(x|y)dx$

נוסחת ההחלקה:

$\boxed{E(X)=\sum_y E(X\mid Y=y)\cdot P(Y=y)}$

אם $E(X\mid Y)=g(Y)$

אז $\boxed{E(X)=E(E(X|Y))=E(g(Y))}$

במקרה הרציף

$$E\left(Y\right)=\int_{-\infty}^{\infty}f_x(x)\left[\int_{-\infty}^{\infty}y\cdot f_{Y|X}(y|x)dy\right]dx=\int_{-\infty}^{\infty}f_x(x)\cdot E(Y|X=x)dx$$

נוסחאות נוספות:

$E(X+Y|Y)=E(X|Y)+Y$

$E(X+Y\mid Y=y)=E(X\mid Y=y)+y$

$E(XY\mid Y=y)=y\cdot E(X\mid Y=y) \quad E(XY\mid Y)=Y\cdot E(X\mid Y)$

$E(X^2)=E(E(X^2\mid Y=y)) \qquad E(XY)=E(E(XY\mid Y=y))$

$E(X^2|Y=y)=Var(X\mid Y=y)+(E(X|Y=y))^2$

משפט wald: X_i ב"ת $EX_i=\mu$. N מ"מ ב"ת

בכל ה- X_i , $S_N=X_1+\cdots+X_N$

$\boxed{E(S_N)=EX\cdot EN=\mu\cdot EN}$

QQ-PLOT:

i	1	…	n
i/n+1			
$\xi_{i/n+1}$			
תצפיות	$X_{(1)}$		$X_{(n)}$

ציר x: שברונית תיאורטיים- $\xi_{i/n+1}$

ציר y: תצפיות ממוינות לפי הסדר.

דוגמא:

מוצאים כיצד לבטא ξ_p ע"י $p=i/n+1$

$F_X\left(x\right)=x^\beta \quad F_X\left(\xi_p\right)=P\left(X\le\xi_p\right)=p$

$\xi_p^{\ \beta}=p \xrightarrow{\beta=2} \xi_p=\sqrt{p}$

קו ישר בגרף- הנחת ההתפלגות נכונה.

שונות וסטיית תקן:

$\boxed{Var(X)=E(X^2)-(E(X))^2=E((X-E(X))^2)}$

$Var(aX+b)=a^2\cdot Var(X)$

$\boxed{Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)}$

אם X,Y **ב"ת**: $Var(X+Y)=Var(X)+Var(Y)$ וגם

$Var(X|Y)=Var(X)$. c קבוע אמ"מ $Var(c)=0$.

$\boxed{SD(X)=\sqrt{Var(X)}}$ $\quad SD(aX+b)=|a|\cdot SD(X)$

שונות מותנית: $Var(Y|X)=E\left((Y-E(Y|X))^2|X\right)$

$\boxed{Var\left(Y\right)=E\left(Var(Y|X)\right)+Var\left(E\left(Y|X\right)\right)}$

קו-ואריאנס: $Cov(X,X)=Var(X)$

$\boxed{Cov(X,Y)=E\left(\left(X-\mu_x\right)\left(Y-\mu_Y\right)\right)=E\left(XY\right)-E\left(X\right)E\left(Y\right)}$

$\boxed{Cov(X,Y)=\frac{1}{2}(Var(X+Y)-Var(X)-Var(Y))}$

1) אם X ו- Y **ב"ת** אז $Cov(X,Y)=0$.

בד"כ ההיפך לא נכון $Cov(X,Y)=0 \not\Rightarrow X,Y$ ב"ת.

2) אם $Cov(X,Y)=0$ אז X,Y בלתי מתואמים.

ב"ת $\Leftarrow$ בלתי מתואמים. (ההפך לא בהכרח נכון).

3) $Cov(X,Y)=Cov(Y,X)$ - סימטריות.

4) $Cov(aX+b,cY+d)=acCov(X,Y)$.

5) $Cov(X,Y+Z)=Cov(Y,X)+Cov(X,Z)$.

6) לכל a קבוע $Cov(X,a)=0$.

תקנון: $X^*=\frac{X-\mu_X}{\sigma_X}$, כאשר: $Corr(X,Y)=E\left(X^*Y^*\right)$

אינטגרלים: $\int_0^\infty\lambda\cdot e^{-\lambda x}dx=1$, $\int_0^\infty e^{-z^2}dz=\sqrt{\pi}$

$\int_{-\infty}^{\infty}x\cdot e^{-\frac{x^2}{2}}dx=0 \qquad \int_0^\infty S^{n-1}\cdot e^{-S}dS=(n-1)!$

התפלגויות רציפות:

פונקציית הצפיפות:

$\boxed{P\left(X\in\left[a,b\right]\right)=P\left(a\le X\le b\right)=\int_a^b f_x(x)dx}$

$\boxed{F_X(x)=P(X\le a)=\int_{-\infty}^{a}f_x(x)dx}$

$P\left(a\le X\le b\right)=P\left(X\le b\right)-P\left(X\le a\right)$

אם $f_x(x)$ פונקציית צפיפות של X מ"מ רציף אז

$\int_{-\infty}^{\infty}f_X(x)dx=1$ ותמיד $f_X(x)\ge 0$.

הסת' שלמה: $\boxed{f_X(x)=\sum_k f_{X|Y}(x|Y=k)P(Y=k)}$

תוחלת של רציף: $\boxed{E\left(X\right)=\int_{-\infty}^{\infty}x\cdot f_x(x)dx}$.

$\boxed{E\left(g(X)\right)=\int_{-\infty}^{\infty}g(x)\cdot f_x(x)dx}$

התוחלת קיימת רק כאשר $\int_{-\infty}^{\infty}|x\cdot f_x(x)|dx<\infty$.

נוסחת הזנב $\boxed{E\left(X\right)=\int_0^\infty(1-F_x(x))dx=\int_0^\infty P(X>x)dx}$

לשים-לב: באזורים בהם $F_X(x)=0$, אז $\int 1$.

שונות של רציף:

$\boxed{Var(X)=\int_{-\infty}^{\infty}x^2\cdot f_X(x)dx-\left(\int_{-\infty}^{\infty}x\cdot f_X(x)dx\right)^2}$

התפלגות	ש.מומנטים	אנ"מ
		$\hat{\mu}=\bar{X}$
$N\left(\mu,\sigma^2\right)$		$\hat{\sigma}^2=\frac{1}{n}\sum\left(X_i-\bar{X}\right)^2$
$Pois(\lambda)$		$\hat{\lambda}=\bar{X}$
$U\left(0,\theta\right)$	$\tilde{\theta}=2\bar{X}$	$\hat{\theta}=X_{(n)}$
$\exp(1/\theta)$		$\hat{\theta}=\bar{X}$
$\exp(\lambda)$		$\hat{\lambda}=1/\bar{X}$
$Bin(n,p)$		$\hat{p}=\bar{X}$
$Ber(p)$		$p=\bar{X}$

סטטיסטי מספיק:

$\theta=\left(\mu,\sigma^2\right)$ ס"מ ל- θ $\left(\sum X_i^2,\sum X_i\right)$

ס"מ לשונות כשהתוחלת ידועה: $\sum\left(X_i-\mu\right)^2$

ס"מ לתוחלת כשהשונות ידועה: $\bar{X}$

$Ber(p)$, $\exp(\lambda)$, $Pois(\lambda)$ - $\sum X_i$

פונקצית התפלגות מצטברת:

הגדרה: $\boxed{F_x(x)=P(X\le x)}$.

בדיד: $\boxed{F_x(x)=\sum_{t\le x}P(X=t)}$ **רציף:** $\boxed{F_x(x)=\int_{-\infty}^{x}f_x(x)dx}$

1) $0\le F_x(x)\le 1$. 2) $F_x(x)$ לא יורדת.

3) $\lim_{x\to-\infty}F_x(x)=0$ $\lim_{x\to\infty}F_x(x)=1$.

4) $F_x(x)$ רציפה מימין. 5) $f_x(x)=\frac{d}{dx}F_x(x)$.

6) $F_X\left(x\right)\sim Uni[0,1]$ לכל פונקציה.

אם $E\left(X\right)<\infty$ אז $\lim_{x\to\infty}x\cdot\left(1-F_x(x)\right)=0$.

התפלגות נורמלית: שימוש בטבלה עבור מ"מ

נורמלי $Z\sim N(0,1)$ $\boxed{f_Z\left(z\right)=\frac{1}{\sqrt{2\Pi}}e^{-z^2/2}}$

$F_X\to P\left(Z\le a\right)=\Phi\left(a\right) \qquad P\left(a\le Z\le b\right)=\Phi\left(b\right)-\Phi\left(a\right)$

סכום נורמלים: $X\sim N(\mu_1,\sigma_1^2)$ ו- $Y\sim N(\mu_2,\sigma_2^2)$

ב"ת אז $\boxed{X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)}$.

מנת נורמלים: מתפלגת קושי. **נורמלי ב-^{2}:** גאמא

משפט הגבול המרכזי: $X_1,X_2,\ldots,X_n$ ב"ת, ש"ה,

$E\left(X_i\right)=\mu$, $Var\left(X_i\right)=\sigma^2$ נגדיר: $\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$

אז: $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\underset{n\to\infty}{\Rightarrow}N\left(0,1\right)$ כלומר $\boxed{\bar{X}_n\sim N\left(\mu,\frac{\sigma^2}{n}\right)}$

כלומר $\boxed{\lim_{n\to\infty}P\left(\frac{1}{\sigma/\sqrt{n}}\sum_{i=1}^{n}\left(X_i-\mu\right)\le a\right)=\Phi(a)}$

$S_n=\sum_{i=1}^{n}X_i$ $\boxed{N(n\cdot\mu,n\sigma^2)}$ $\boxed{\frac{S_n-n\mu}{\sqrt{n}\sigma}\to N\left(0,1\right)}$

התפלגות אחידה (יוניפורמית): $U[0,1]\to f_X(x)=1$

$E(X)=\frac{a+b}{2} \qquad F_X\to P(c\le X\le d)=\frac{d-c}{b-a}$

$\boxed{f_x(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & otherwise\end{cases}} \qquad Var(X)=\frac{(b-a)^2}{12}$

סטטיסטי הסדר: אם $X_1,X_2,\ldots,X_n\sim U\left[0,1\right]$ ב"ת:

$\boxed{(\min|\max=x)\sim Uni[0,x]} \quad f_{\min}\left(x\right)=n\left(1-x\right)^{n-1} \quad f_{\max}\left(x\right)=nx^{n-1}$

$E(X_{\min})=\frac{b}{n+1} \qquad E(X_{\max})=\frac{n}{n+1}$

סטנדרטיזציה: $X\sim U[a,b]\Leftrightarrow\frac{X-a}{b-a}\sim U[0,1]$

מעבר לאקספוננט: $\left(-\frac{\ln(1-U)}{\lambda}\right)\sim\exp(\lambda)$, $U[0,1]$

התפלגות אקספוננציאלית: זמן בין מאורעות פואסון.

$Var\left(X\right)=1/\lambda^2 \quad E\left(X\right)=1/\lambda \quad P(X>a)=e^{-\lambda a}$

$\boxed{f_X(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0 & otherwise\end{cases}} \qquad P\left(a\le X\le b\right)=e^{-\lambda a}-e^{-\lambda b}$

מינימום של אקספוננט הוא אקספוננט:

$X\sim Exp(\lambda), Y\sim Exp(\mu)\Rightarrow Min(X,Y)\sim Exp(\lambda+\mu)$

תחרות בין אקספוננציאלים: X,Y ב"ת

$X\sim Exp(\lambda), Y\sim Exp(\mu)\Rightarrow\boxed{P(X<Y)=\frac{\lambda}{\lambda+\mu}}$

סכום עם גיאומטרי: $X_i\sim\exp(\lambda)$, $N\sim Geo(p)$

אז: $S_N=\sum_{i-1}^{N}X_i\sim\exp\left(\lambda p\right)$

תכונת חוסר זיכרון: אם נתון כי עד רגע t לא היה מופיע, אז הזמן עד למופע הבא מתחיל מהתחלה

$\boxed{P(T>t+s|T>t)=P(T>s)} \quad E\left(X-Y|X>Y\right)=EX$.

משפט: נתון מ"מ T , נגדיר $G(t)=P(T>t)$. אם

$G(t+s)=G(t)\cdot G(s)$, קיים λ כך ש $T\sim Exp(\lambda)$.

התפלגות Gamma: $X_1,X_2,\ldots,X_n$ מ"מ ב"ת,

$X_i\sim Exp\left(\lambda\right)$, נגדיר $Y=\sum_{i=1}^{n}X_i$, $Y\sim Gamma\left(n,\lambda\right)$.

$$\boxed{f_X(x)=\begin{cases}\frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{\Gamma(n)}=\frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} & x\ge 0\\ 0 & otherwise\end{cases}}$$

$\boxed{Gamma\left(1,\lambda\right)=\exp\left(\lambda\right)}$ $\boxed{\Gamma(n)=(n-1)\Gamma(n-1)}$ $\Gamma\left(\frac{1}{2}\right)=\sqrt{\Pi}$

כאשר n שלם : $\Gamma\left(n\right)=\left(n-1\right)!$

חיבור Gamma: אם $X\sim Gamma\left(r,\lambda\right)$,

$Y\sim Gamma\left(s,\lambda\right)$ ב"ת אז $\boxed{X+Y\sim Gamma(r+s,\lambda)}$.

חילוק Gamma: עבור X,Y כנ"ל יתקיים

$\frac{X}{X+Y}\sim Beta\left(r,s\right)$. המנה והסכום בשני הנ"ל ב"ת.

רווח סמך:

ר"ס לתוחלת במדגם מהתפלגות נורמאלית:

כאשר σ ידוע:

$$Q=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$
$$\mu=X\pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

כאשר σ לא ידוע:

$$Q=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$$
$$\mu\in\bar{X}\pm t_{(n-1),1-\alpha/2}\frac{S}{\sqrt{n}}$$

תכונות:

1. זהו רווח סמך הקצר ביותר ברמת סמך נתונה.
2. אורכו קבוע מראש רק כאשר השונות ידועה.
3. אורך הרווח עולה עם הגדלת רמת הסמך והשונות.
4. אורך הרווח קטן עם עליה בגודל המדגם

גודל המדגם הדרוש כדי שבביטחון של $1-\alpha$ הסטייה בין ממוצע המדגם ל μ לא תעלה על d:

$$2z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\le 2d\rightarrow \boxed{n\ge\frac{z_{1-\alpha/2}\sigma^2}{d^2}}$$

ר"ס ל- θ במדגם מהתפלגות אחידה:

$$Q\left(X_{(n)},\theta\right)=\frac{X_{(n)}}{\theta}\quad \theta\in\left[X_{(n)},\frac{X_{(n)}}{\alpha^{1/n}}\right]$$

ר"ס לפרופורציה במדגם מהתפלגות נורמאלית:

$$X_i\sim Ber(p),\ \hat{p}=\bar{X}=\frac{1}{n}\sum X_i=\frac{X}{n}\quad X\sim Bin(n,p)$$
$$Q=\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\sim N(0,1)\quad p\in\hat{p}\pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

גודל המדגם כדי שבביטחון של $1-\alpha$ הסטייה בין הפרופורציה במדגם ל p לא תעלה על d:

$$n\ge\frac{z^2_{1-\alpha/2}\hat{p}(1-\hat{p})}{d^2},\quad p(1-p)\le 0.25,\quad \boxed{n\ge\frac{z^2_{1-\alpha/2}}{4d^2}}$$

ר"ס לשונות במדגם מהתפלגות נורמאלית:

כאשר μ ידוע:

$$Q=\frac{1}{\sigma^2}\sum(X_i-\mu)^2\sim\chi^2_{(n)}$$
$$\sigma^2\in\left[\frac{\sum(X_i-\mu)^2}{\chi^2_{(n),1-\alpha/2}},\frac{\sum(X_i-\mu)^2}{\chi^2_{(n),\alpha/2}}\right]$$

כאשר μ לא ידוע:

$$Q=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{(n-1)}\quad S^2=\frac{1}{n-1}\sum(X_i-\bar{X})^2$$
$$\sigma^2\in\left[\frac{(n-1)S^2}{\chi^2_{(n-1),1-\alpha/2}},\frac{(n-1)S^2}{\chi^2_{(n-1),\alpha/2}}\right]$$

תכונות:

1. אינו רווח סמך אופטימאלי עבור רמת סמך נתונה.
2. רווח סמך עבור סטיית תקן: שורש על שני הקצוות.

ר"ס ל- α במדגם מהתפלגות $F_X(x)=x^\alpha$ (Gamma)

$$\hat{\alpha}=\frac{n}{-\sum\ln X_i}\quad Y=-\ln(X),\ Y\sim\exp(\alpha),$$
$$Q=2\alpha\sum Y_i\sim Gamma\left(n,\frac{1}{2}\right)=\chi^2_{(2n)},\ f_X=\alpha x^{\alpha-1}$$
$$\alpha\in\left[\frac{\chi^2_{(2n),\alpha/2}}{2\sum Y_i},\frac{\chi^2_{(2n),1-\alpha/2}}{2\sum Y_i}\right]\quad\sum Y_i\sim Gamma(n,\alpha)$$

ר"ס להפרש תוחלות בשני מדגמים ב"ת:

כאשר σ ידוע:

$$Q=\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}}\sim N(0,1)$$
$$\mu_X-\mu_Y\in(\bar{X}-\bar{Y})\pm z_{1-\alpha/2}\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}$$

כאשר σ לא ידוע והשונויות זהות:

$$S_P^2=\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$$
$$Q=\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{S_P^2\left(\frac{1}{n}+\frac{1}{m}\right)}}\sim t_{(n+m-2)}$$
$$\mu_X-\mu_Y\in(\bar{X}-\bar{Y})\pm t_{(n+m-2),1-\alpha/2}\sqrt{S_P^2\left(\frac{1}{n}+\frac{1}{m}\right)}$$

כאשר σ לא ידוע והשונויות שונות:

$$\mu_X-\mu_Y\in(\bar{X}-\bar{Y})\pm t_{(df),1-\alpha/2}\sqrt{\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)}$$
$$df=\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)^2\Big/\frac{S_X^4}{n^2(n-1)}+\frac{S_Y^4}{m^2(m-1)}$$

ר"ס להפרש תוחלות בשני מדגמים מזווגים:

בונים מדגם של הפרשים

$$\mu_X-\mu_Y=\mu_D\quad D_i\sim N(\mu_D,\sigma_D^2)\quad\bar{D}=\frac{1}{n}\sum D_i$$
$$S_D^2=\frac{1}{n-1}\sum(D_i-\bar{D})^2\quad\mu_D\in D\pm t_{(n-1),1-\alpha/2}\frac{S_D}{\sqrt{n}}$$

ר"ס ליחס שונויות בשני מדגמים ב"ת:

$$Q=\frac{S_X^2}{S_Y^2}\frac{\sigma_Y^2}{\sigma_X^2}=\frac{S_X^2}{S_Y^2}\frac{1}{\lambda}\sim F_{(n-1,m-1)}$$
$$\lambda=\frac{\sigma_X^2}{\sigma_Y^2}\in\left[\frac{S_X^2}{S_Y^2}F^{\alpha/2}_{(m-1,n-1)},\frac{S_X^2}{S_Y^2}F^{1-\alpha/2}_{(m-1,n-1)}\right]$$

ר"ס להפרש פרופורציות בשני מדגמים ב"ת:

$$Q=\frac{(\hat{p}_X-\hat{p}_Y)-(p_X-p_Y)}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}+\frac{\hat{p}_Y(1-\hat{p}_Y)}{m}}}\quad p_x-p_y\in$$
$$(\hat{p}_X-\hat{p}_Y)\pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}+\frac{\hat{p}_Y(1-\hat{p}_Y)}{m}}$$

התפלגויות דגימה:

התפלגות נורמאלית:

$$X_1,\cdots,X_n\sim N(\mu,\sigma^2)$$

$$\boxed{X_{0.95}=Z_{0.95}\cdot\sigma+\mu}\quad\boxed{Z_{0.15}=-Z_{0.85}}$$

התפלגות הממוצע:

$$\bar{X}\sim N(\mu,\sigma^2/n)\rightarrow\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

הסתברות שהסטייה המוחלטת לכל היותר d:

$$P(|\bar{X}-\mu|\le d)=2\Phi\left(\frac{d}{\sigma/\sqrt{n}}\right)-1$$

גודל המדגם כדי ש: $P(|\bar{X}-\mu|\le d)\ge 1-\alpha$ -

$$n\ge\left(Z_{1-\alpha/2}\frac{\sigma}{d}\right)^2$$

חוק התקנון: $\mu-\sigma,\underbrace{\sigma-2\sigma}_{},\underbrace{2\sigma-3\sigma}_{}$

34%   14%   2%

התפלגות חי-בריבוע:

תכונות:

1. צפיפות לא סימטרית עם זנב ימני.
2. התפלגות חיובית.
3. עבור n>50, קרוב לנורמאלי.
4. אדטיביות: $\chi^2_{(n)}+\chi^2_{(m)}=\chi^2_{(n+m)}$

קשר ל-Gamma ולנורמאלי:

$$W_n=\sum Z_i^2\sim\chi^2_{(n)}\overset{D}{=}Gamma\left(\frac{n}{2},\frac{1}{2}\right)$$
$$W_n=\frac{1}{\sigma^2}\sum(X_i-\mu)^2\sim\chi^2_{(n)}$$
$$\frac{W_n}{n}\xrightarrow{P}1\quad\frac{W_n-n}{\sqrt{2n}}\xrightarrow{D}N(0,1)$$
$$E(W_n)=n\quad Var(W_n)=2n$$

התפלגות שונות המדגם:

תוחלת ידועה: $\hat{\sigma}^2=\frac{1}{n}\sum(X_i-\mu)^2$

לא ידועה: $\hat{\sigma}^2=\frac{1}{n}\sum(X_i-\bar{X})^2$

$$\frac{\sum(X_i-\mu)^2}{\sigma^2}\sim\chi^2_{(n)}\quad\frac{\sum(X_i-\bar{X})^2}{\sigma^2}\sim\chi^2_{(n-1)}$$

$$E\left(\chi^2_{(n-1)}\right)=n-1\quad Var\left(\chi^2_{(n-1)}\right)=\sigma^2$$

ההסתברות שאנ"מ לשונות במדגם נורמאלי יסטה ממנה בסטייה יחסית שינה עולה על ε:

$$\hat{\sigma}^2=\frac{1}{n}\sum(X_i-\mu)^2\quad P\left(\left|\frac{\hat{\sigma}^2-\sigma^2}{\sigma^2}\right|\le\varepsilon\right)$$
$$=\boxed{P(W_n\le n(1+\varepsilon))-P(W_n\le n(1-\varepsilon))}$$

התפלגות t:

$$\boxed{T=\frac{Z}{\sqrt{W_k/k}}\sim t_{(k)}}$$
$$E(T)=0,\ Var(T)=k(k-2)$$

תכונות:

1. סימטרית סביב 0.
2. זנבות עבים משל הנורמאלית.

כאשר השונות לא ידועה: $\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$

$$\boxed{t_{(n),0.1}=-t_{(n),0.9}}$$

התפלגות F:

$$\boxed{F(k,m)=\frac{W_k/k}{W_m/m}}\quad E(F)=\frac{k}{m-2}$$
$$Var(F)=\frac{2m^2(k+m-2)}{k(m-2)^2(m-4)}$$

תכונות:

1. התפלגות לא סימטרית עם ד"ח במונה ובמכנה.
2. $(t_{(k)})^2\sim F(1,k)$ .3 $\frac{1}{F(k,m)}\sim F(m,k)$

$$\boxed{F^{(k,m)}_{0.05}=\frac{1}{F^{(m,k)}_{0.95}}}$$

בראש הטבלה: a

הסתברות-p. בפנים: $P(X\le a)=p$

סטטיסטיקה תיאורית ושיטות גרפיות:

שגיאת דגימה: שגיאה כתוצאה מכך שנאספו נתונים על מדגם שהינו חלק מכלל האוכלוסייה. גודלה תלוי בשיטת הדגימה ובגודל המדגם.

ממוצע קטום: ממוצע ללא הנמוך והגבוה.

שברון p באוכלוסייה: $P(X\le\xi_P)=p\ ,\xi_p$

הערך ש- p% מהאוכלוסייה מתחתיו.

חישוב השברון המדגמי:

$$r=(n+1)p-\lfloor(n+1)p\rfloor$$
$$\rightarrow\boxed{(1-r)x_{(k)}+rx_{(k+1)}}$$

רבעון ראשון-Q1:

$$Q_1=\frac{3}{4}X_{\left(\left\lfloor\frac{n+1}{4}\right\rfloor\right)}+\frac{1}{4}X_{\left(\left\lfloor\frac{n+1}{4}\right\rfloor+1\right)}$$

חציון-M:

זוגי: $M=\frac{1}{2}X_{\left(\left\lfloor\frac{n+1}{2}\right\rfloor\right)}+\frac{1}{2}X_{\left(\left\lfloor\frac{n+1}{2}\right\rfloor+1\right)}$

אי-זוגי: $M=X_{\left(\left\lceil\frac{n}{2}\right\rceil\right)}$

רבעון שלישי-Q3:

$$Q_3=\frac{1}{4}X_{\left(\left\lfloor\frac{3(n+1)}{4}\right\rfloor\right)}+\frac{3}{4}X_{\left(\left\lfloor\frac{3(n+1)}{4}\right\rfloor+1\right)}$$

תחום בין רבעונים: $IQR=Q_3-Q_1$

שונות המדגם: $S^2=\frac{1}{n-1}\sum(X_i-\bar{X})^2$

BOX-PLOT:

$$step=\frac{3}{2}(Q_3-Q_1)$$
$$\boxed{LF=Q_1-step}\quad\boxed{UF=Q_3+step}$$

LW: התצפית המינימאלית שגדולה מ-LF.

UW: התצפית המקסימאלית שקטנה מ-UF.

אמידה נקודתית:

סטטיסטי: פונקציה של התצפיות במדגם.

פרמטר: גודל שהוא מאפיין מסכם של האוכלוסייה.

אמד: סטטיסטי שאיתו אומדים את הפרמטר.

אומדן: הערך של האמד, המחושב מהמדגם.

שיטת המומנטים:

המומנט ה-k באוכלוסיה: $\mu_k=E(X^k)$

המומנט ה-k המדגמי: $M_k=\frac{1}{n}\sum X_i^k$.

השיטה: מבטאים את הפרמטר ע"י המומנטים של האוכלוסייה ואז מציבים במקום את המומנטים של המדגם

דוגמא: $f(x|\theta)=(\theta+1)x^\theta,\ 0\le x\le 1$

$$\mu_1=E(X)=\int_0^1 x(\theta+1)x^\theta dx=\frac{\theta+1}{\theta+2}$$
$$\theta=\frac{1}{1-\mu_1}-2\rightarrow\boxed{\hat{\theta}=\frac{1}{1-M_1}-2=\frac{1}{1-\bar{X}}-2}$$

אמידת התוחלת (הפרמטר הוא- μ):

$$\boxed{\hat{\theta}=M_1=\bar{X}}\quad\boxed{E(\hat{\theta})=\mu}\quad\boxed{Var(\hat{\theta})=\frac{1}{n}Var(X_1)}$$

אמידת השונות (הפרמטר הוא σ^2):

$$\boxed{\hat{\theta}=S^2=\frac{1}{n}\sum(X_i-\bar{X})^2}\quad\boxed{E\left(\frac{n}{n-1}\hat{\theta}\right)=\sigma^2}$$

שיטת הנראות המירבית:

בחירת הערך p עבורו ההסתברות של תוצאות המדגם היא הגדולה ביותר.

פונקצית נראות: פו' צפיפות משותפת-

$$L(\theta|x_1..x_n)=f_\theta(x_1..x_n)=\prod f_\theta(x_i)$$

מוצאים אמד שימקסם את $L(\theta)$ ע"י גזירה והשוואה ל-0.

תכונות:

1. אנ"מ אינו בהכרח יחיד.
2. לפעמים קל יותר למקסם את $l(\theta)=\ln L(\theta)$
3. אם $\hat{\theta}$ אנ"מ ל- θ אז לכל פו $\tau(\theta)$, אנ"מ ל- $\tau(\theta)$ יהיה: $\tau(\hat{\theta})$.

דוגמא:

$$L(\theta)=\prod(\theta+1)x_i^\theta=(\theta+1)^n\left(\prod x_i\right)^\theta$$
$$l(\theta)=\ln(L(\theta))=n\ln(\theta+1)+\theta\sum\ln(x_i)$$
$$\frac{\partial L(\theta)}{\partial\theta}=\frac{n}{\theta+1}+\sum\ln(x_i)=0\rightarrow\boxed{\hat{\theta}=-\frac{n}{\sum\ln x_i}-1}$$

כאשר x מוגבל ע"י הפרמטר, מכניסים אינדיקטור

דוגמא:

$$L(\theta)=\prod\lambda e^{-\lambda(x_i-\theta)}I\{x_i\ge\theta\}$$
$$=\lambda^n e^{-\lambda\sum(x_i-\theta)}I\{X_{(1)}\ge\theta\}$$
$$=e^{n\lambda\theta}I\{X_{(1)}\ge\theta\}\lambda^n e^{-\lambda\sum x_i}\rightarrow\boxed{\hat{\theta}=X_{(1)}}$$

חוסר הטייה: ממוצע האומדנים יהיה שווה בקירוב לערך הפרמטר אותו רוצים לאמוד. $\hat{\theta}$

אמד חסר הטייה אם: $\boxed{E(\hat{\theta})=\theta}$.

השגיאה הריבועית הממוצעת-תוחלת הסטייה:

$$MSE_{\hat{\theta}}(\theta)=E(\hat{\theta}-\theta)^2$$
$$=Var(\hat{\theta})+\left(E(\hat{\theta})-\theta\right)^2$$

נעדיף את האמד שבו ה-MSE הקטן ביותר.

בהתפלגות אחידה: האנ"מ עדיף ל $n\ge3$ אבל הוא אמד מוטה. אח"ה: $\theta^*=\frac{n+1}{n}X_{(n)}$ אבל עדיין

לפי ה-MSE: $MSE_{\hat{\theta}}(\theta)\approx MSE_{\theta^*}(\theta)$

עקיבות: אומד שה-MSE שלו שואף ל-0.

$$\boxed{MSE\underset{n\to\infty}{\to}0}\quad\boxed{\hat{\theta}\underset{n\to\infty}{\to}\theta}$$ אם לכל $\varepsilon>0$ האמד

שואף לפרמטר בהסת' 1: $\boxed{\lim_{n\to\infty}P\left(|\hat{\theta}-\theta|<\varepsilon\right)=1}$

הערה: בד"כ כשה-MSE לא שואף ל-0, לא עקיב.

משפט: אם $\hat{\theta}$ עקיב ל- θ אז לפונקציה רציפה $\tau(\hat{\theta})$ עקיב ל- $\tau(\theta)$. (ליניארית שומרת אח"ה).

סטטיסטי מספיק:

פונקציה שבה כל המידע במדגם ביחס לפרמטר.

הגדרה: $S=S(X_1,,,X_n)$ סטטיסטי מספיק

אם"ם הצפיפות $f(X_1..X_n|S=s)$ או התפלגות

$P(X_1=x_1..X_n=x_n|S=s)$ לא תלויה ב θ, $\forall s$.

תכונות:

1. אם $\hat{\theta}$ אנ"מ ל θ אז $\hat{\theta}$ הוא פו' של ס"מ S.
2. ס"מ אינו יחיד, אם $S=g(S^*)$, גם: S^* ס"מ וגם $f(S)$ ס"מ כאשר f חח"ע.

משפט הפירוק: סטטיסטי S מספיק ביחס ל θ

אם $f(x_1..x_n|\theta)=h(x_1..x_n)\,g(S(x_1..x_n)|\theta)$

כלומר: מפרקים את f ל- h שתלויה רק בתצפיות ו- g שתלויה ב- θ, כאשר S היא המידע מהתצפיות בתוך g.

משפט: אנ"מ תלוי במדגם מקרי דרך ס"מ.

דוגמא:

$$\underbrace{\left(\frac{1}{\sigma}\right)^n e^{-\frac{\sum x_i}{\sigma}}}_{h(\underline{x})}\cdot\underbrace{e^{\frac{\mu n}{\sigma}}I\{X_{(1)}\ge\mu\}}_{g(X_{(1)}|\mu)}\rightarrow\boxed{S=X_{(1)}}$$

משפט Rao-Blackwell: $\hat{\theta}$ אמד ו- S ס"מ:

1. $\theta^*=E(\hat{\theta}|S)$ הוא אמד. 2. $E(\theta^*)=E(\hat{\theta})$.
3. $MSE_{\theta^*}(\theta)\le MSE_{\hat{\theta}}(\theta)$ (שוויון - $\hat{\theta}$ כבר מבוסס על הס"מ).

דוגמא: $Ber(p)$, $\exp(\lambda)$, אומד: X_1. שיפורו:

$$E(X_1|\sum x_i)=P(X_1=1|\sum x_i=s)=\frac{\sum x_i}{n}=\bar{X}$$

$Pois(\lambda)$:

$$E(X_1|S)=\frac{s}{n}\left(n-\frac{1}{n}\right)^{s-1}$$
$$=\frac{s}{n}\left(1-\frac{1}{n}\right)^{s-1}\sim Bin\left(s,\frac{1}{n}\right)$$