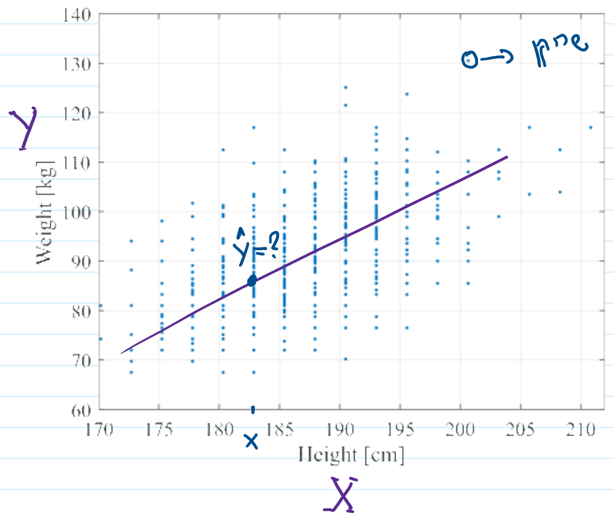
דוגמה: קשר בין גובה למשקל של שחקני בייסבול בליגת על

ממוצע

$E[X]$ ←
$E[Y]$
$Cov[X,Y]$
$Var[X]$

```
Ex = mean(X);
Ey = mean(Y);
Cxy = mean(X.*Y) - Ex*Ey;
Vx = var(X);
Yh = Ey+Cxy/Vx*(X-Ex);

e = mean(Y - Yh); % error
mse = mean(e.^2);
```



1034 שחקנים

תזכורת

$$\hat{Y} = a_{opt}x + b_{opt} = E[Y] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}(x - E[X])$$

בדוגמה: $\rho \cong 0.53$

שגיאת חיזוי - שגיאה ריבועית ממוצעת (תכונה 2.6): שגיאת החיזוי נתונה ע"י

$$mse_{min} = E\left[\left(Y - (a_{opt}X + b_{opt})\right)^2\right] = \mathrm{Var}[Y](1 - \underbrace{\rho_{XY}^2}_{\text{מקדם קורלציה}}) \qquad (2.23)$$

תזכורת

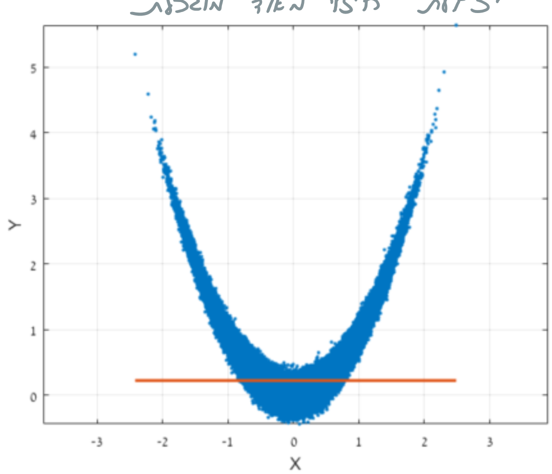
$$\rho_{XY} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}$$

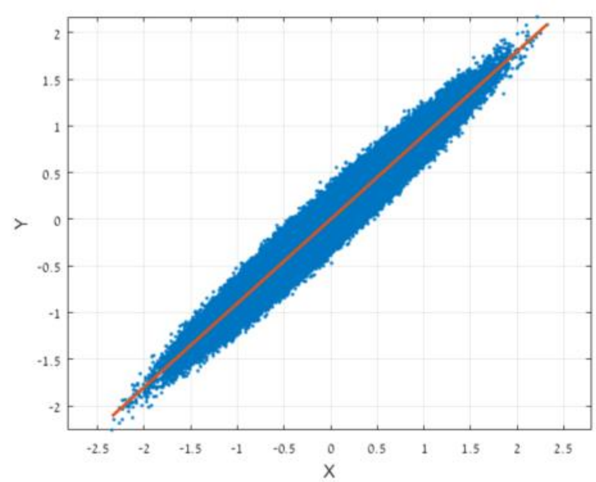
* $\rho_{XY} = \pm 1$ ← קשר לינארי מושלם

* $\rho_{XY} = 0$ ← חוסר קשר לינארי . "חוסר קורלציה" ⟺ $Cov[X,Y]=0$

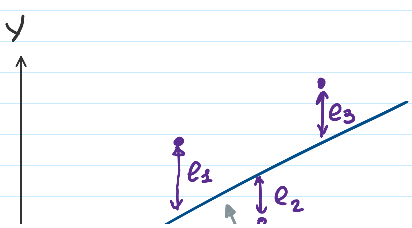
דוגמאות לקשר לינארי ולא לינארי

יכולת חיזוי לינארי מאוד מוגבלת



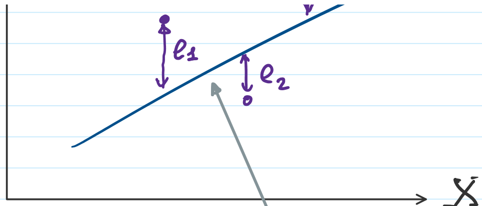


הרחבה בנושא שגיאת חיזוי



1) מפי הממוצע של e_i

$e_1 + e_2 + e_n$ מתקיים גמיד



מתקיים תמיד

$$\frac{e_1+e_2+e_3}{3} \boxed{=0}$$

$$E\left[Y-(a_{opt}X+b_{opt})\right]=0$$

2) mse - מינ׳ השגיאה עבור הקו

$$mse_{min} = \frac{e_1^2+e_2^2+e_3^2}{3} = Var[Y]\left(1-\rho_{XY}^2\right)$$

משמעות של תוחלת ושונות

מקרה פרטי מעניין הוא חיזוי ע"י קבוע,

$$mse(b) = E[(Y-b)^2].$$

מה b עבורו mse הוא מינימלי?

$$\begin{aligned} mse(b) &= E[(Y-b)^2] \\ &= E[Y^2-2bY+b^2] \\ &= E[Y^2]-2bE[Y]+b^2 \end{aligned}$$

פתיחת סוגריים

$$\frac{\partial mse(b)}{\partial b} = -2E[Y]+2b=0$$

נגזרת

סיכום: $b = E[Y]$

צפוי" expected"

שגיאת mse עבור החיזוי

$$mse_{min} = E[(Y-E[Y])^2] = Var[Y]$$

משתנה אקראי רציף

מקרה: לאפיין ניסוי עם תוצאה רציפה

CDF: הגדרה זהה למ"א בדיד $F_X(x)=P(X\le x)$

PDF:

$$f_X(x) = \frac{\partial F_X(x)}{\partial x}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(p)\,dp$$

תכונה:

* $P(a<X\le b) = \int_a^b f_X(x)dx$

$= F_X(b) - F_X(a)$

* $P(X=a)=0$

תכונות של PDF.

* $p(X=a) = 0$

ישנם אינסוף ערכי X אפשריים

תכונות של PDF:

PDF - צפיפות ההסתברות

* $\cancel{1 \geq} f_X(x) \geq 0$

* $\int_{-\infty}^{\infty} f_X(x)\,dx = 0 \iff F_X(\infty)=1$

$\left.\begin{matrix} F_X(x) \\ P_X[X] \end{matrix}\right\}$ הסתברויות ולכן ≤ 1

תוחלת: **תוחלת** (הגדרה 3.2):

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

תוחלת של פונקציה של משתנה אקראי (תכונה 3.3):

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

$g(x) = x^2$

דוגמה: $E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx$

שונות: החישוב מבוסס על תוחלת, לכן אין הבדל.

חישוב עזר של שונות

$Var[X] = E[X^2] - E^2[X]$

התפלגות אקספוננציאלית (מעריכית)

הגדרה: $X \sim Exp(\lambda)$. λ פרמטר של ההתפלגות

PDF

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & x \geqslant 0 \\ 0 & \text{אחרת} \end{cases}$$

3: חשב CDF של X, כאשר $X \sim Exp(\lambda)$.

CDF

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx = \begin{cases} 1-\exp(-\lambda x) & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

$\Rightarrow F_X(\infty) = \int_{-\infty}^{\infty} f_X(x)dx = 1$

תוחלת:

$$\int x\exp(ax)dx = \exp(ax)\left[\frac{x}{a} - \frac{1}{a^2}\right] \longrightarrow E[X] = \lambda \int_0^{\infty} x\exp(-\lambda x)dx = \frac{1}{\lambda}$$

הוכחה:

$$E[X^2] = \int_0^{\infty} \lambda x^2 \exp(-\lambda x)dx = \frac{1}{\lambda^2}\Big[-2e^{-y} - 2ye^{-y} - y^2e^{-y}\Big]_0^{\infty} = \frac{2}{\lambda^2}$$

שונות:

$$Var[X] = \frac{1}{\lambda^2} = E[X^2] - E^2[X]$$

הערה 3.5 ! בספרות (וגם ב-Matlab) מופיע הגדרה נוספת:

$= \frac{2}{\lambda^2}$

$E^2[x] = \frac{1}{\lambda^2}$

הערה 3.5 ! בספרות (וגם ב-Matlab) מופיע הגדרה נוספת:

$$f_X(x) = \begin{cases} \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) & x \geqslant 0 \\ 0 & \text{אחרת} \end{cases},$$

כאשר $\mu = \frac{1}{\lambda}$.

דוגמה מספרית-

דוגמה 3.3: מודל של הופעת תקלות במערכת מסויימת מתוארת ע"י התפלגות $Exp(\lambda)$, כאשר $\lambda = \frac{1}{1000}\left[\frac{1}{\text{שעה}}\right]$. שאלות:

(א) מהו סיכוי לתקלה אחרי 1000 שעות?

(ב) מהו סיכוי לתקלה אחרי 100 שעות ולפני 1000 שעות?

(ג) מהו משך הזמן עבורו הסיכוי לתקלה הוא $\frac{1}{2}$?

$X \sim exp(\frac{1}{1000})$ הוא זמן הופעת התקלה

א $P(X > 1000) = 1 - P(X \leq 1000)$
$= 1 - F_X(x) = 1 - \left(1 - e^{-\frac{1000}{1000}}\right) \cong 0.36$

ב $P(100 \leq X \leq 1000) = F_X(1000) - F_X(100) \cong 0.54$

במשתנים רציפים אין הבדל בין $>, \geq$ ולבין $<, \leq$

ג $\underbrace{F_X(x_0)}_{\text{סיכוי לתקלה לפני זמן } x_0} = \underbrace{1 - F_X(x_0)}_{\text{אחרי זמן } x_0} \Rightarrow x_0 = ?$

$$\Rightarrow x_0 = \frac{\ln\left(\frac{1}{2}\right)}{-0.001} \approx 693[\text{שעה}]$$

חציון (הגדרה 3.4): ערך x עבורו מתקיים $\Pr(X \leq x) = \Pr(X \geq x)$, נקרא ערך חציוני.

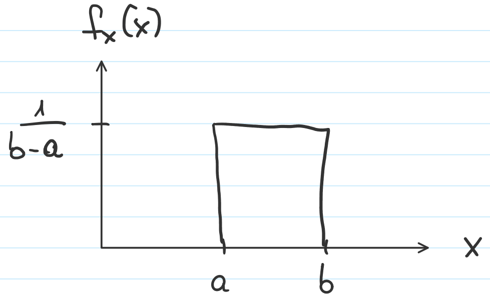
דוגמה מהחדשות: משכורות:

משכורת ממוצעת $\cong 10{,}000$ ממוצע של כל המשכורות

חציונית $\cong 8{,}500$ חצי מעל, חצי מתחת

$\boxed{\text{התפלגות אחידה}}$ - גרסה רציפה

מטרה: תופעות בעלי הסתברות זהה



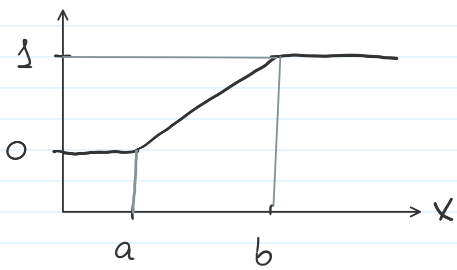
הגדרה עבור $X \sim U[a,b]$ מתקיים

PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{אחרת} \end{cases}$$

הערה: $U(a,b)$

$a < x < b$



CDF

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leqslant x \leqslant b \\ 1 & b < x \end{cases}$$

דוגמאות: חשב $E[X], E[X^2]$

תזכורת

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx}$$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$= \int_a^b x \cdot \frac{1}{b-a} dx = \frac{x^2}{2}\Big|_a^b \cdot \frac{1}{b-a}$$

$$= \frac{1}{2} \cdot \frac{b^2-a^2}{b-a} = \frac{b+a}{2}$$

$$E[X^2] = \int_a^b x^2 \frac{1}{b-a} dx$$

$$= \frac{1}{3} \frac{b^3-a^3}{b-a} = \frac{1}{3}(a^2+ab+b^2)$$

$$Var[X] = \frac{(b-a)^2}{12}$$

דוגמה מספרית:

$$X \sim U[0,2] \qquad E[X^4] = \int_0^2 \overbrace{x^4}^{g(x)=x^4} \underbrace{\frac{1}{2-0}}_{f_X(x)} dx =$$

$$= \frac{1}{2} \cdot \frac{x^5}{5}\Big|_0^2 = \frac{32}{10} = \frac{16}{5}$$

$$\boxed{\text{התפלגות גאוסית}}$$

הקדמה: משפט גבול המרכזי

בהינתן סדרה של מ"מ $X_1, X_2, \dots, X_n, \dots$ בעלי אותה התפלגות (לא משנה מהי)

בהינתן סדרה של מ"א $X_1, X_2, \dots, X_n, \dots$ בעלי אותה התפלגות (לא משנה מה)
עבור n מספיק גדול $X_1+\dots+X_n$ יתפלג גאוסית

תובנה חשובה: ברוב המקרים של הנדסה. מדובר בתופעות עם מודל גאוסי.

* PDF (הגדרה 3.6): עבור משתנה אקראי גאוסי $Y \sim N(\mu, \sigma^2)$, בעל תוחלת μ ושונות σ^2, מתקיים

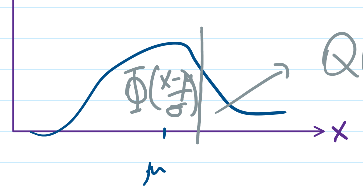
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{3.14}$$

* CDF: נתון $X \sim N(\mu, \sigma^2)$

$$\Rightarrow F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

כאשר $\Phi(x)$ זה CDF של מ"א $N(0,1)$.

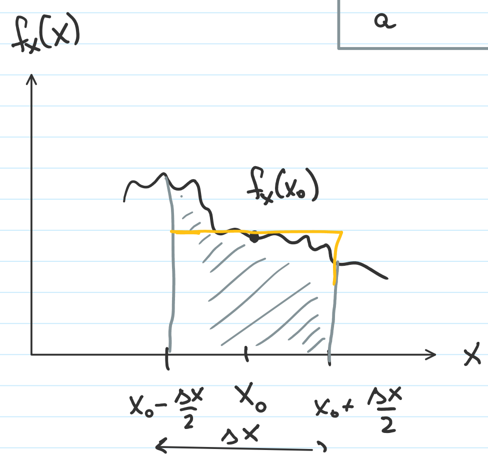
* Q-function $Q(x) = 1 - \Phi(x)$



Histogram

מטרה: להעריך PDF על סמך נתונים מספריים

בסיס לחישוב: $\boxed{\int_a^b f_X(s)\,ds = P(a \le X \le b)}$

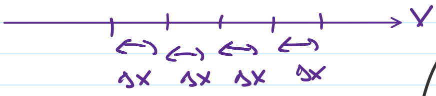


$$(1) \quad P\left(x_0 - \frac{\Delta x}{2} \le X \le x_0 + \frac{\Delta x}{2}\right) \cong f_X(x_0) \cdot \Delta x$$

$$\Rightarrow (2) \quad f_X(x_0) \cong \frac{P\left(x_0 - \frac{\Delta x}{2} \le X \le x_0 + \frac{\Delta x}{2}\right)}{\Delta x}$$

- איך השיטה עובדת?

* נתונות N תוצאות ניסוי



* נחלק תחום התוצאות למרווחים בגודל Δx

* $\left(\frac{\text{מספר תוצאות הניסוי במרווח נתון } \Delta x}{N}\right) = p\left(x_0 - \frac{\Delta x}{2} \le X \le x_0 + \frac{\Delta x}{2}\right)$

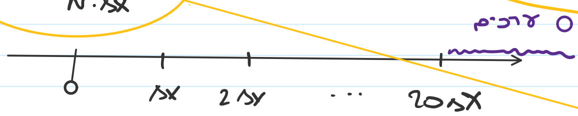
* מבצעים חישוב עבור כל מרווח Δx ומנרמלים ע"י הכפלה ב- $\frac{1}{\Delta x}$.

סיכום דוגמה חישובית.

נתונים N ערכים של משתנה X. $N = 10^6$

$\Delta x = \frac{\max(X)}{20}$

$\frac{\text{מספר ערכים}}{N \cdot \Delta x}$



$\mu = \frac{1}{\lambda}$; מספר ערכים

להגריל מספרים בעלי התפלגות exp

```
X = exprnd(1, 1, 1e6);

n = 20;
dx = max(X)/n;
for k = 1:n
h(k) = sum(dx*(k-1) < X & X < dx*k)/1e6/dx;
end

bar(dx/2+dx*(1:n),h)

histogram(X,n,'Normalization','pdf')
```

חישוב לכל מרווח

גרף

פונ' מובנת של Matlab