Föreläsning 4
2019-04-03
Johan Swärd

Föreläsning 4

Repetition från förra föreläsningen.

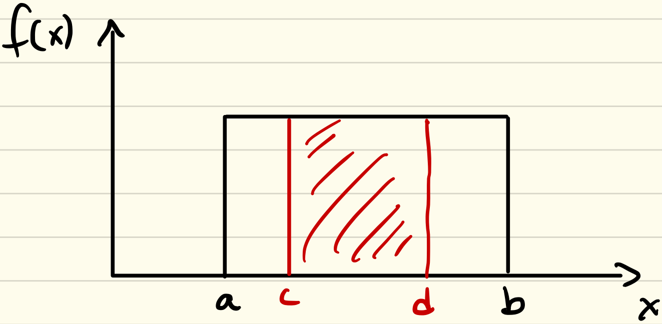
* Kontinuerliga slumpvariabler.
    - Antar oändligt många värden i något intervall.
    - $P(X = x) = 0$

	Diskreta	Kontinuerliga
$f(x)$	Sannolikhetsfunktion	Täthetsfunktion
$F(x) = P(X \leq x)$	Fördelningsfunktion	Fördelningsfunktion
$F(x) =$	$\sum_{x_i \leq x} f(x_i)$	$\int_{-\infty}^{x} f(z)\,dz$
$E[X]$	$\sum_{\text{alla } x} x f(x)$	$\int_{-\infty}^{\infty} x f(x)\,dx$
$V[X]$	$\sum_{\text{alla } x} (x - E[X])^2 f(x)$	$\int_{-\infty}^{\infty} (x - E[X])^2 f(x)\,dx$

Repetition

Repetion från förra föreläsningen.

* Täthetsfunktioner, $f(x)$
  - Alltid $f(x) \geq 0 \quad \forall x$.
  - $\int_{-\infty}^{\infty} f(x) = 1 \iff F(\infty) = 1$
* Sannolikheter ges av areor



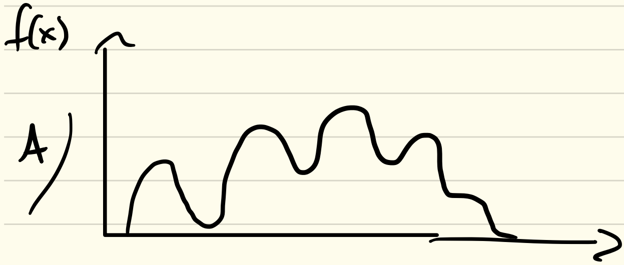
$$P(c \leq X \leq d) = \int_c^d f(x)\,dx =$$

$$= \int_{-\infty}^{d} f(x)\,dx - \int_{-\infty}^{c} f(x)\,dx = F(d) - F(c)$$

Repetition

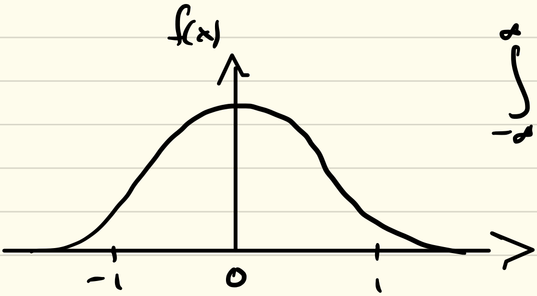
Vilka av följand funktioner är täthetsfunktioner?

A)



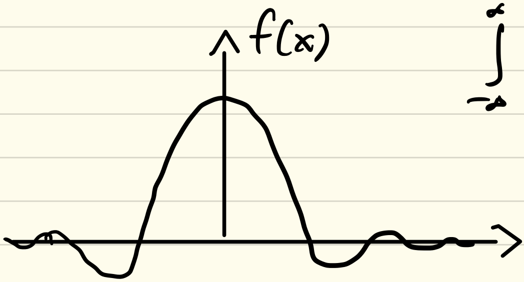
$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

B)



$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

C)



$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Repetition

* Normalfördelningen - en mycket viktig och användbar fördelning.

$$X \sim N(\mu, \sigma^2)$$

väntevärde "center"

varians "tjocklek"

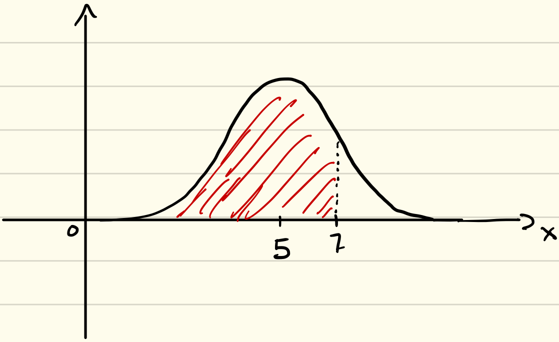
* För att kunna använda en tabell för alla normalfördelningar så måste vi kunna standardisera:

$$X \sim N(\mu, \sigma^2)$$

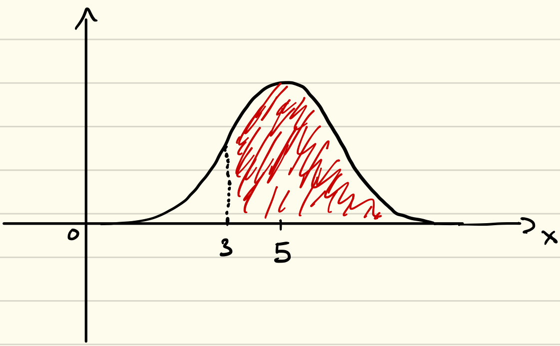
$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Exempel:

Låt $X \in N(5,2)$. Vad är $P(X \le 7)$?

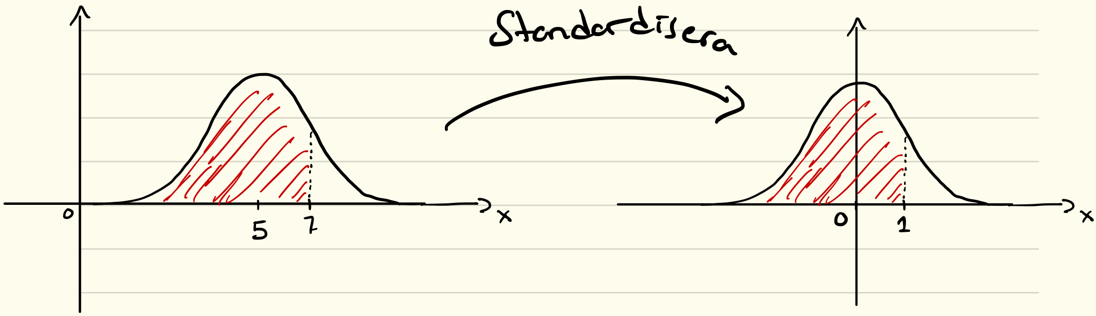


Vad är $P(X \ge 3)$?



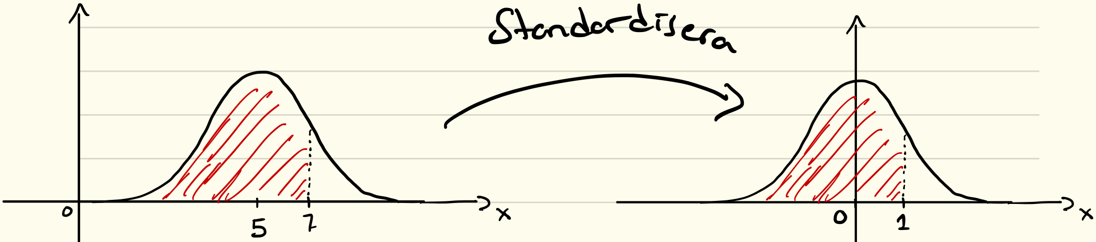
Exempel:

Låt $X \in N(5,4)$. Vad är $P(X \le 7)$?



Exempel:

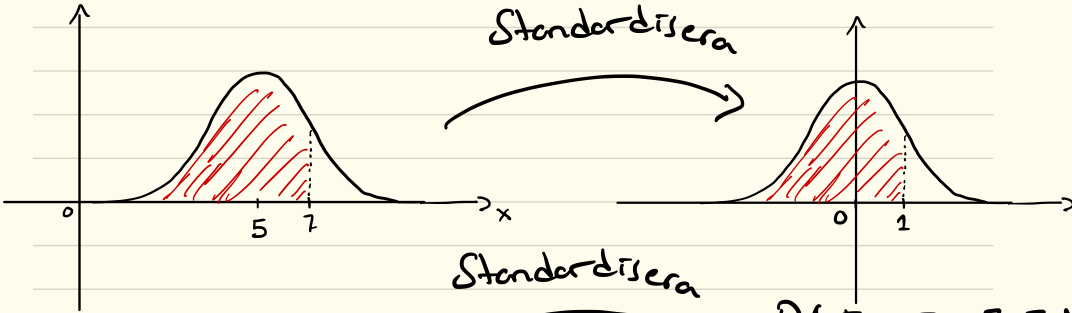
Låt $X \in N(5,4)$. Vad är $P(X \leq 7)$?



$$P(X \leq 7) \xrightarrow{\text{Standardisera}} P\left(\underbrace{\frac{X-5}{2}}_{\in N(0,1)} \leq \frac{7-5}{2}\right)$$

Exempel:

Låt $X \in N(5,4)$. Vad är $P(X \le 7)$?

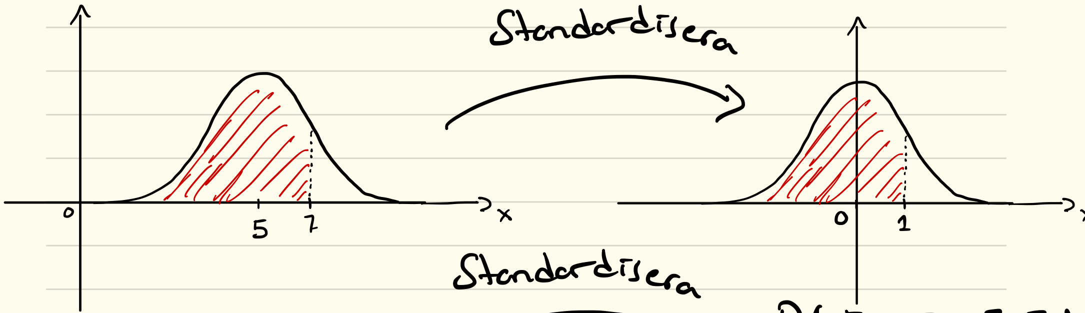


$$P(X \le 7) \xrightarrow{\text{Standardisera}} P\Big(\underbrace{\frac{X-5}{2}}_{\in N(0,1)} \le \frac{7-5}{2}\Big)$$

$$P\left(\frac{X-5}{2} \le 1\right) = \Phi(1) \overset{\text{Tabell}}{=} 0.8413$$

Exempel:

Låt $X \in N(5,4)$. Vad är $P(X \le 7)$?



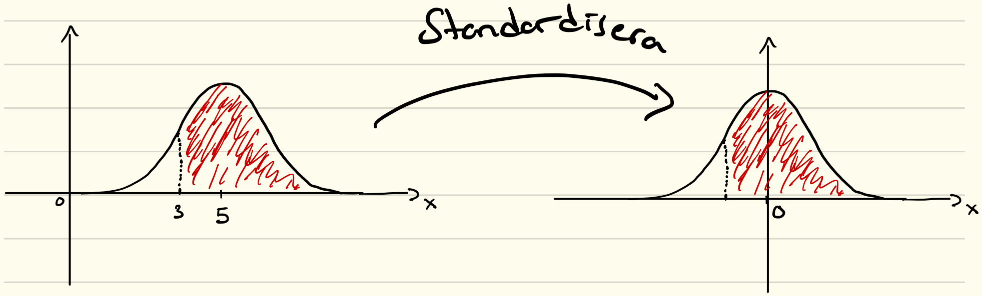
$$P(X \le 7) \xrightarrow{\text{Standardisera}} P\left(\underbrace{\frac{X-5}{2}}_{\in N(0,1)} \le \frac{7-5}{2}\right)$$

$$P\left(\frac{X-5}{2} \le 1\right) = \Phi(1) \overset{\text{Tabell}}{=} 0.8413$$

De färgade areorna under båda graferna är lika stora och har värdet 0.8413.

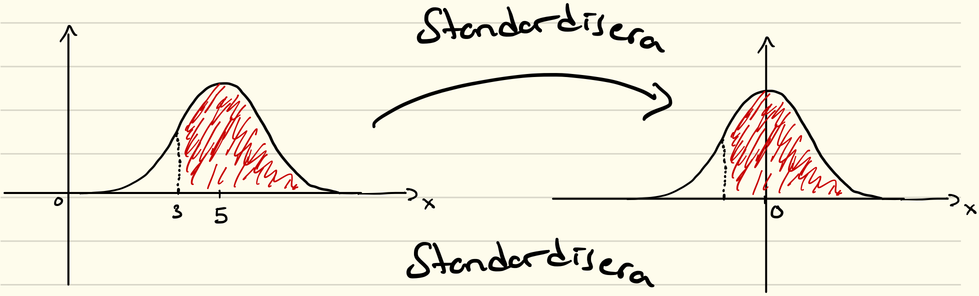
Exempel:

Låt $X \in N(5,4)$. Vad är $P(X \geq 3)$?



Exempel:

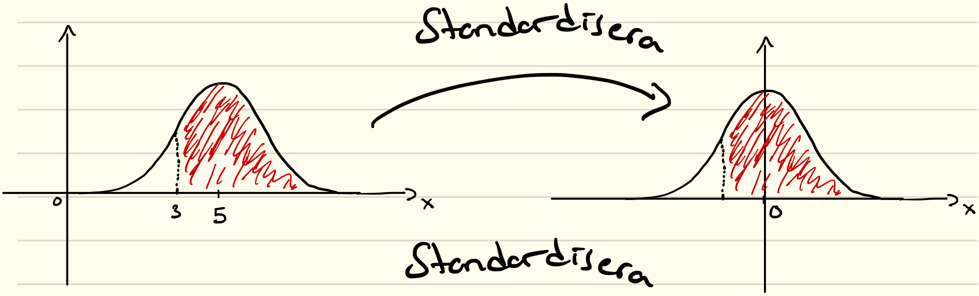
Låt $X \in N(5,4)$. Vad är $P(X \geq 3)$?



$$P(X \geq 3) = 1 - P(X \leq 3) \xrightarrow{\text{Standardisera}} 1 - P\Big(\underbrace{\frac{X-5}{2}}_{\in N(0,1)} \leq \frac{3-5}{2}\Big)$$

Exempel:

Låt $X \in N(5,4)$. Vad är $P(X \geq 3)$?

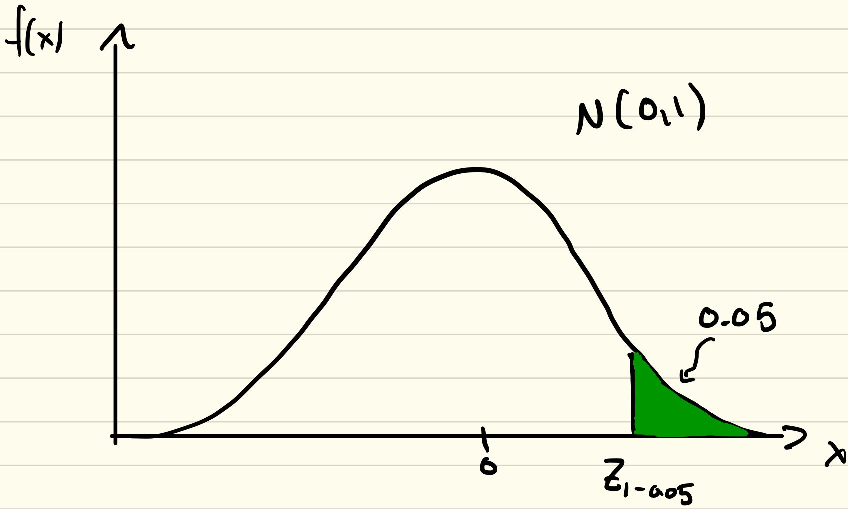


$$P(X \geq 3) = 1 - P(X \leq 3) \xrightarrow{\text{Standardisera}} 1 - P\left(\underbrace{\frac{X-5}{2}}_{\in N(0,1)} \leq \frac{3-5}{2}\right)$$

$$1 - P\left(\frac{X-5}{2} \leq -1\right) = 1 - \Phi(-1) \overset{\text{Symmetri}}{=} 1 - (1 - \Phi(1))$$

$$= 1 - 1 + \Phi(1) = \Phi(1) = 0.8413$$

Kvantiler:



$P(X \leq z_{1-0.05}) = 1 - 0.05$

$P(X \geq z_{1-0.05}) = 0.05$

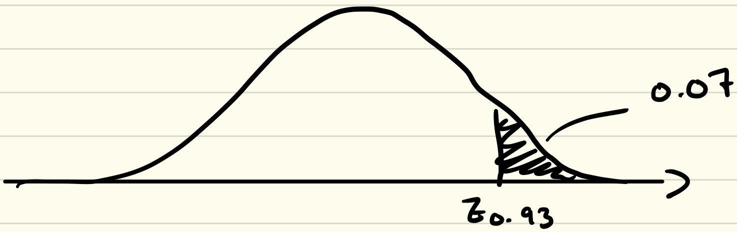
$z_{1-\alpha}$ kallas α-kvantilen

Det värde där $1-\alpha$ av arean ligger till vänster och α av arean ligger till höger.

Exempel: Vilket x-värde överstigs enbart av 7% av alla utfallen i en normalfördelning med $\mu = 0$ och $\sigma^2 = 1$?

Lösning:

$X \in N(0,1)$



$$P(X \geq z_{0.93}) = 0.07$$

$$\Leftrightarrow \quad P(X \leq z_{0.93}) = 0.93$$

Tabell ger $z_{0.93} \approx 1.476$.

Repetition

* Om $\mathbb{X}$ är diskret:

$$\mu = E[\mathbb{X}] = \sum_{\text{alla } x} x f(x)$$

$$\sigma^2 = Var[\mathbb{X}] = \sum_{\text{alla } x} (x-\mu)^2 f(x)$$

* Om $\mathbb{X}$ är kontinuerlig:

$$\mu = E[\mathbb{X}] = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$\sigma^2 = Var[\mathbb{X}] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx.$$

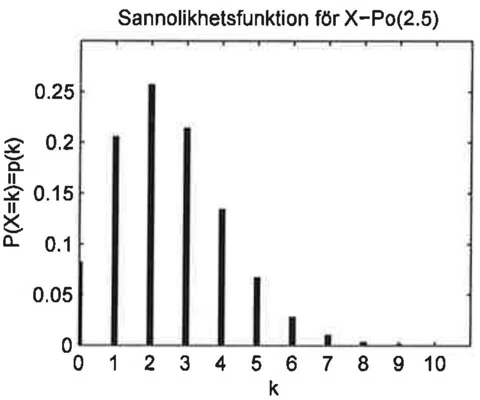
Lunds universitet | Matematikcentrum | Matematisk statistik

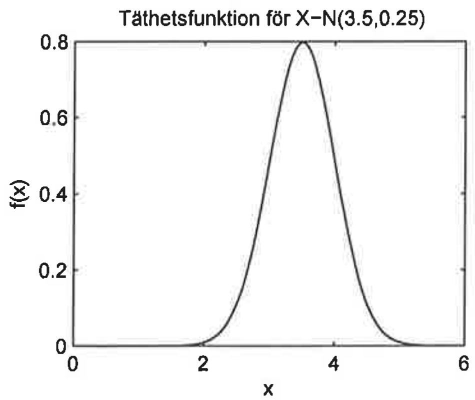
BIOSTATISTISK GRUNDKURS, MASB11
OH-BILDER 2007-11-12

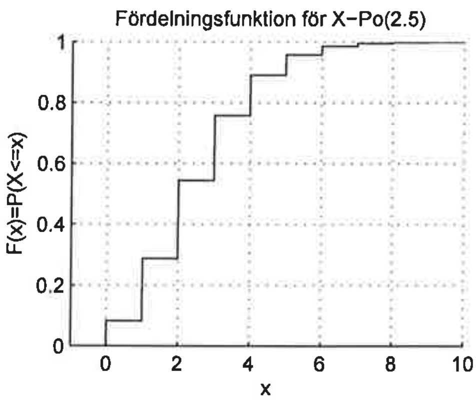
	X diskret	**X kontinuerlig**
Sannolikhetsfördelning	$f(x) = P(X = x)$	$f(x)$ **(täthetsfunktion)**
Fördelningsfunktion $F(x) = P(X \leq x)$	$F(x) = \sum_{x_i \leq x} f(x_i)$	$F(x) = \int_{-\infty}^{x} f(t)dt$
Väntevärde $\mathbf{E(X)}=\mu$	$E(X) = \sum_{\textbf{alla x}} x \cdot f(x)$	$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$

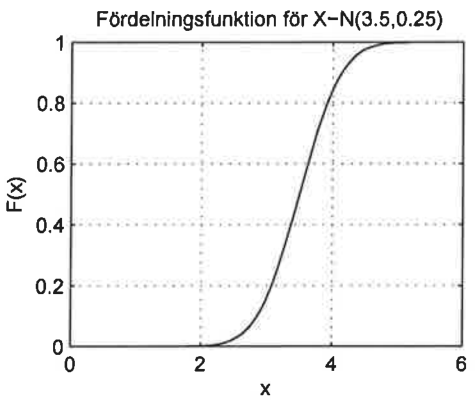
Varians $Var(X) = \sigma^2 = E[(X - \mu)^2]$
"förväntad kvadratisk avvikelse från väntevärdet"

Standardavvikelse $D(X) = \sigma = \sqrt{Var(X)}$









Extrauppgifter (Tavlan)

1. Låt $X \in N(0,1)$. Bestäm

a) $P(X \leq 1.82)$

b) $P(X \leq -0.35)$

c) $P(-1.2 < X < 0.5)$

d) a så att $P(X > a) = 5\%$.

2. Låt $X \in N(5,2)$. Bestäm

a) $P(X \leq 6)$

b) $P(1.8 < X < 7.2)$

c) a så att $P(X \leq a) = 0.05$.

Dagens föreläsning

* Räkneregler för väntevärde och varians.

---

* Först lite om att beräkna $E[X]$.

Exempel:

$X \sim U(a,b)$ dvs. $f_X(x) = \begin{cases} \frac{1}{b-a} & , a \le x \le b \\ 0 & , \text{annars} \end{cases}$

$E[X]$?

Dagens föreläsning

* Räkneregler för väntevärde och varians.

---

* Först lite om att beräkna $E[X]$.

Exempel:

$X \sim U(a,b)$ dvs. $f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{annars} \end{cases}$

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

$$= \int_a^b \frac{x}{b-a}\,dx = \left[\frac{x^2}{2(b-a)}\right]_{x=a}^{b} = \frac{b^2-a^2}{2(b-a)}$$

$$= \frac{1}{2} \cdot \frac{(b+a)(b-a)}{(b-a)} = \frac{a+b}{2}.$$

Linjära transformationer

Om $g(X)$ är en funktion av X
så är $E[g(X)]$:

$$E[g(X)] = \sum_{\text{alla } x} g(x) f(x), \quad \text{om } X \text{ är diskret}$$

och

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx, \quad \text{om } X \text{ kontin.}$$

Exempel:

$$g(x) = x^2.$$

$$X = \begin{cases} \frac{1}{10}, & 0 \leq x \leq 10 \\ 0, & \text{annars.} \end{cases}$$

$E[g(X)]$?

Linjära transformationer

Om $g(X)$ är en funktion av X

så är $E[g(X)]$:

$$E[g(X)] = \sum_{\text{alla } x} g(x) f(x), \quad \text{om } X \text{ är diskret}$$

och

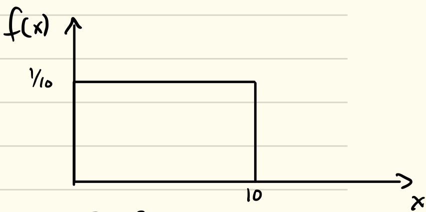
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx, \quad \text{om } X \text{ kontin.}$$

Exempel:

$$g(x) = x^2.$$

$$f(x) = \begin{cases} \frac{1}{10}, & 0 \le x \le 10 \\ 0, & \text{annars.} \end{cases}$$

$E[g(X)]$?



Lösning:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx = \int_0^{10} \frac{x^2}{10}\, dx = \left[\frac{x^3}{30}\right]_{x=0}^{10} = \frac{1000}{30}$$

$$\approx 33.333\ldots$$

Linjära transformationer

Låt $E[X] = \mu$ och låt X vara en diskret slumpvariabel.

Vad är då $E[X + b]$?

Linjära transformationer

Låt $E[X] = \mu$ och låt X vara en diskret slumpvariabel.

Vad är då $E[X+b]$?

$$E[X+b] \overset{\text{Definition}}{=} \sum_{\text{alla } x} (x+b) f(x)$$

$$= \underbrace{\sum_{\text{alla } x} x f(x)}_{\mu} + \sum_{\text{alla } x} b f(x)$$

$$= \mu + b \underbrace{\sum_{\text{alla } x} f(x)}_{=1} = \mu + b$$

Alltså: $\boxed{E[X+b] = E[X] + b}$

Linjära transformationer

Låt $E[X] = \mu$ och låt X vera en diskret slumpvariabel.

Vad är då $E[aX + b]$?

Linjära transformationer

Låt $E[X] = \mu$ och låt X vara en diskret slumpvariabel.

Vad är då $E[aX + b]$?

$$E[aX+b] = \sum_{\text{alla } x} (ax+b) f(x)$$

$$= \sum_{\text{alla } x} ax f(x) + \sum_{\text{alla } x} b f(x)$$

$$= a\left(\sum_{\text{alla } x} x f(x)\right) + b$$

$$= a\mu + b$$

Alltså:

$$\boxed{E[aX+b] = aE[X] + b}$$

Linjära transformationer

* Variansen kan skrivas på (minst) tre former:

$$\mathrm{Var}[X] = \boxed{\sum_{\text{alla } x} (x - E[X])^2 f_X(x)}$$

$$= \sum_{\text{alla } x} \Big(\underbrace{x^2 f_X(x)}_{E[X^2]} - \underbrace{2x E[X] f_X(x)}_{2E[X] \underbrace{\sum_{\text{alla } x} x f_X(x)}_{E[X]}} + \underbrace{E[X]^2 f_X(x)}_{E[X]^2} \Big)$$

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

$$= \boxed{E[X^2] - E[X]^2}$$

$$= \boxed{E\big[(X - E[X])^2\big]}$$

Linjära transformationer

Vad blir $Var(X + b)$?

Linjära transformationer

Vad blir $Var(X+b)$?

$$E[(X+b-E[X+b])^2] =$$

$$= E[(X+b-\mu-b)^2] =$$

$$= E[(X-\mu)^2] = Var[X]$$

Alltså: $\boxed{Var[X+b] = Var[X]}$

Linjära transformationer

$\text{Var}[aX + b]$?

Linjära transformationer

$Var[aX + b]$?

$$Var[aX+b] = E[(aX+b-E[aX+b])^2]$$

$$= E[(aX - a\mu)^2] = E[(a(X-\mu))^2]$$

$$= E[a^2(X-\mu)^2] = a^2 E[(X-\mu)^2]$$

$$= a^2 Var[X]$$

Alltså $\boxed{Var[aX+b] = a^2 Var(X)}$

Notera att det blir samma om X är kontinuerlig i summorna byts ut av integraler.

Summor av slumpvariabler

Låt X och Y vara två slumpvariabler. Då gäller:

$$E[X+Y] = E[X] + E[Y]$$

Om X och Y är oberoende

$$Var(X+Y) = Var(X) + Var(Y)$$

* Samma regler går att generalisera till fler termer:

Låt $X_1, \ldots, X_N$ vara slumpvariabler.

Då gäller:

$$E\left[\sum_{n=1}^{N} X_n\right] = \sum_{n=1}^{N} \mu_n$$

och om de är oberoende

$$V\left[\sum_{n=1}^{N} X_n\right] = \sum_{n=1}^{N} V[X_n]$$

Summor av slumpvariabler

Om $X_1, \ldots, X_N$ alla är normalfördelade, då är också summan av dem normalfördelad.

Exempel:

$X_n \sim N(\mu, \sigma^2)$. X_n oberoende.

Vad har $\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n$ för fördelning?

Summor av slumpvariabler

Om $X_1, \dots, X_N$ alla är normal-fördelade, då är också summan av dem normalfördelad.

Exempel:

$X_n \sim N(\mu, \sigma^2)$. X_n oberoende.

Vad har $\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n$ för fördelning?

$$E[\bar{X}] = E\left[\frac{1}{N} \sum_{n=1}^{N} X_n\right] = \frac{1}{N} \sum_{n=1}^{N} E[X_n]$$

$$= \frac{1}{N} \cdot N\mu = \mu$$

$$V[\bar{X}] = V\left[\frac{1}{N} \sum_{n=1}^{N} X_n\right] = \frac{1}{N^2} V\left[\sum_{n=1}^{N} X_n\right]$$

oberoende

$$= \frac{1}{N^2} N \cdot \sigma^2 = \frac{\sigma^3}{N}.$$

Viktiga räkneregler för väntevärden och varianser:

- $E(aX+b) = aE(X)+b$
- $Var(aX+b) = a^2Var(X)$

- $E(X+Y) = E(X)+E(Y)$
- $Var(X+Y) = Var(X)+Var(Y)$ om X och Y är oberoende

Dessa regler gäller för ALLA slumpvariabler oavsett fördelning!

Om X och Y är normalfördelade gäller dessutom att fördelningen för summan också är normalfördelad.

VIKTIGT SPECIALFALL: Om $X_1, \ldots, X_n$ alla är oberoende och $\mathbf{N}(\mu, \sigma^2)$ gäller

$$\sum_{i=1}^{n} X_i \sim \mathbf{N}(n \cdot \mu,\, n \cdot \sigma^2)$$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim \mathbf{N}(\mu,\, \frac{\sigma^2}{n})$$

EXEMPEL: "Starar" Efter att ha studerat ett stort antal häckande starar (Sturnus vulgaris) har man kommit fram till att variabeln ägg/kull (X) kan beskrivas av en sannolikhetsfördelning som ser ut på följande sätt:

Antal ägg/kull (x)	4	5	6	7	8
$f(x)$	0.10	0.20	0.30	0.20	0.2

1. Vi väljer slumpmässigt två kullar. Beräkna sannolikheten att det är minst 15 ägg i de båda kullarna tillsammans.
2. Vi väljer slumpmässigt 10 kullar. Vad är slh att minst en av dem har 8 ägg?
3. Beräkna $E(X)$ och $V(X)$.
4. (Nästa gång:) Vi väljer slumpmässigt ut 50 kullar. Beräkna sannolikheten att dessa i genomsnitt har mellan 6.5 och 6.9 ägg.
5. Vikten hos starhonorna anses vara normalfördelade med väntevärdet 75 g och standardavvikelse 3.2 g. Vad är slh att en slumpmässigt vald hona väger mindre än 80 g?
6. Vi väljer slumpmässigt ut fem honor. Vad är sannolikheten att medelvärdet av deras vikter överstiger 77 g?

Extrauppgifter

1. Låt $X \in N(0,1)$ och $Y = 3X + 2$.
   Beräkna $E[Y]$ och $D[Y]$.

2. Låt X och Y vara oberoende och respektive $N(1,1)$ och $N(-1,2)$.
   Vilken fördelning har $X+Y$ och $X-Y$?

3. De oberoende stokastiska variablerna X_1 och X_2 tillhör båda $N(1,2)$.
   Ange fördelningen för $\bar{X} = (X_1 + X_2)/2$.