Övning 8 ~ Normalfördelningen

I kapitel 6 ska vi lära oss om

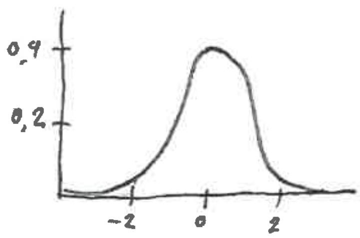
* Den standardiserade normalfördelningen, $N(0,1)$
* Den allmänna normalfördelningen, $N(\mu,\sigma)$
* Linjärkombinationer av oberoende normalfördelade stokastiska variabler
* Centrala gränsvärdessatsen

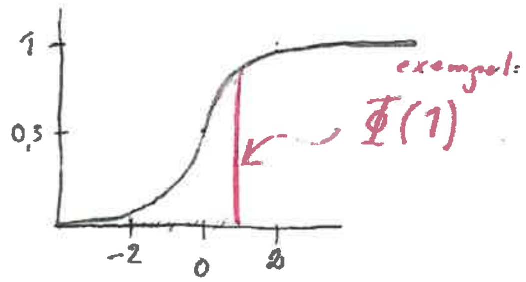
Normalfördelningen är en väldigt viktig fördelning inom sannolikhetsteori och statistik, som ofta används för att beskriva variationen hos olika företeelser. Stora delar av den statistikteori som finns bygger på den här fördelningen.

Några viktiga formler:

$\boxed{6.2}$ Täthetsfunktion och Fördelningsfunktion för $X \in N(0,1)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

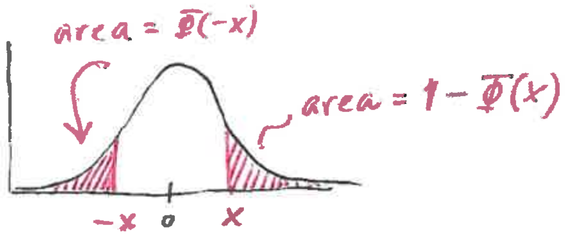




Jämn funktion ger $\varphi(x) = \varphi(-x)$

$\boxed{6.3}$

$$\Phi(-x) = 1 - \Phi(x)$$

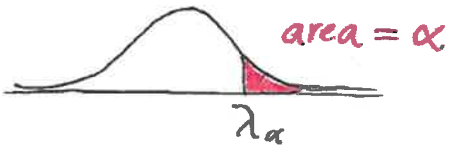


$\boxed{6.4}$ Sannolikheten att X ligger mellan två värden a och b blir

$$P(a < X \leq b) = \Phi(b) - \Phi(a)$$

$\boxed{6.5}$ λ_α – α-kvantilen för en standardiserad normalfördelning

$$P(X > \lambda_\alpha) = \alpha$$



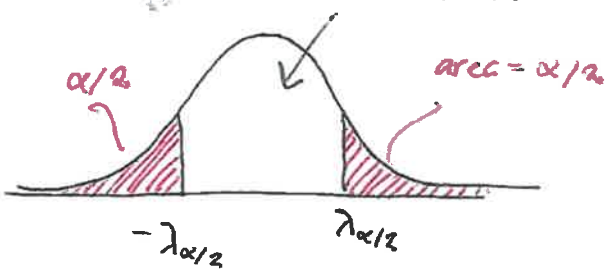
"Arean under täthetsfunktionen $\varphi(x)$ till höger om λ_α är alltså lika med α."

α	0.0005	0.001	0.005	0.01	0.025	0.05	0.10
λ_α	3.29	3.09	2.58	2.33	1.96	1.64	1.28

Pga symmetri har vi också

$\boxed{6.6}$

$$P(-\lambda_{\alpha/2} < X < \lambda_{\alpha/2}) = 1 - \alpha$$



Uppgifter

6.1 X är $N(0,1)$. Bestäm

a) $P(X \leq 1.82)$

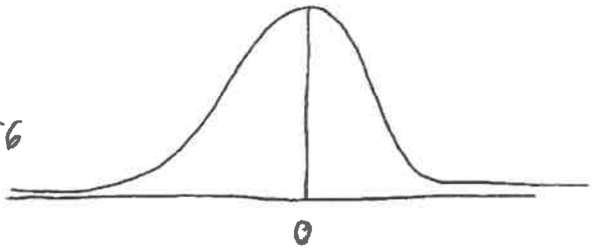
b) $P(X \leq -0{,}35)$

c) $P(-1{,}2 < X < 0{,}5)$

d) a så att $P(X > a) = 5\%$

e) a så att $P(|X| < a) = 95\%$

Rita bild!!



a) $P(X \leq 1.82) = \Phi(1{,}82) = 0{,}9656$

b) $P(X \leq -0{,}35) =$

$= \{\text{se formel } 6.3\} = 1 - \Phi(0{,}35) = 0{,}3632 \qquad (1 - 0{,}6368)$

Formel 6.4 c) $P(-1{,}2 < X < 0{,}5) = \Phi(0{,}5) - \Phi(-1{,}2) \qquad \{\text{formel } 6.3 \text{ igen}\}$

$= \Phi(0{,}5) + \Phi(1{,}2) - 1 = 0{,}5764$

d) Formel 6.3 säger oss att

$P(X > \lambda_\alpha) = \alpha$

Vi har fått α, så det är bara att kolla i tabellen.

$\alpha = 0{,}05$ ger $\lambda_\alpha = 1.64$.

e) Här kan vi antingen räkna som i c, eller se direkt att vi har fått det uppdukat för formel 6.6.

$$P(|X| < a) = P(-a < X < a) = \underbrace{1 - \alpha}_{95\%}$$

$\Rightarrow \alpha = 0{,}05$

Kolla i formel $\lambda_{\alpha/2}$. Vad är $\alpha/2$? 0,025.

Vad är $\lambda_{0{,}025}$? Ja, 1.96.

Innan vi kan göra uppgifter med den allmänna normalfördelningen behöver vi lite mer formler.

$\boxed{X \in N(\mu, \sigma)}$

Täthetsfunktion och fördelningsfunktion:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \, dt$$

Sats $\boxed{6.1}$:

$X \in N(\mu, \sigma)$ om och endast om $Y = (X-\mu)/\sigma \in N(0,1)$.

Dessutom gäller att

$$f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) \quad \text{och} \quad F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Sats $\boxed{6.2}$

Om $X \in N(\mu, \sigma)$ så gäller att

$E[X] = \mu$, $V(X) = \sigma^2$ $D(X) = \sigma$

(expectation) (variance) (deviation)

Formel 6.9

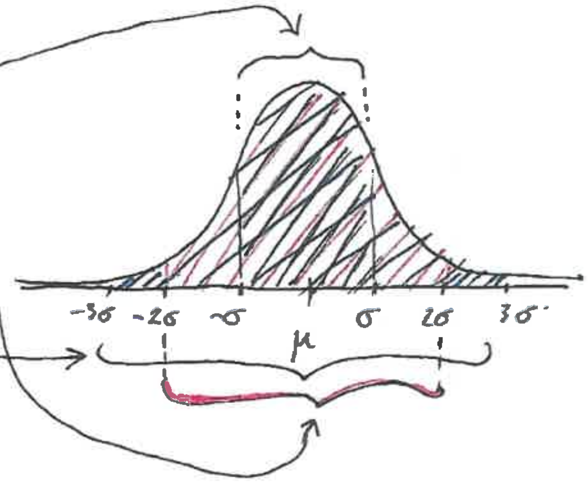
$$P(X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) \qquad P(X > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Formel 6.10 + 6.11 ger

$$P(\mu - \sigma < X < \mu + \sigma) = 0,68$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0,954$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0,997$$

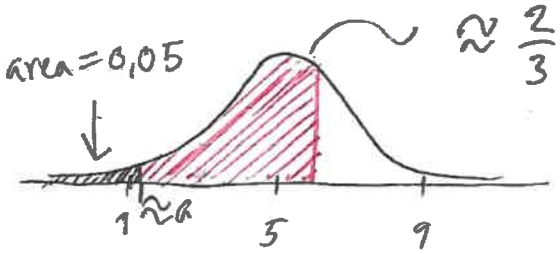


6.4. X är N(5,2). Bestäm

a) $P(X\le 6)$, ~~b) $P(4.8<X<7.2)$~~, c) a så att $P(X\le a)=5\%$

Börja med att rita!!

N(5,2) betyder att medelvärdet är 5, och standardavv. 2.



a) Här använder vi sats 6.1 och säger att vår nya stokastiska variabel $Y = \frac{X-\mu}{\sigma} \in N(0,1)$.

Vår nya fiffiga variabel Y gör så att vi kan använda fördelningsfunktionen för den standardiserade normal f.

Således har vi

$$P(X\le 6) = P\left(\frac{X-\mu}{\sigma} \le \frac{6-\mu}{\underset{2}{\sigma}}\right) = P\left(Y \le \frac{1}{2}\right) = \Phi(0,5)$$

$$= 0,6915.$$

b) a så att $P(X\le a) = 5\%$.

Kolla på bilden igen. Vad frågar de efter? Jo, vid vilket värde på är det så lite som 5% chans att Y antar det värdet?

$$P(X\le a) = P\Big(\underbrace{\frac{X-\mu}{\sigma}}_{Y} \le \frac{a-\mu}{\sigma}\Big) = P\left(Y \le \frac{a-\mu}{\sigma}\right) = 0,05$$

Eftersom $a-\mu<0$ så blir $\Phi\left(\frac{a-\mu}{\sigma}\right) = 1-\Phi\left(-\frac{a-\mu}{\sigma}\right)$

Formel 6.5 ger även att $P(X>\lambda_\alpha)=\alpha$, och därför blir

$$\lambda_{0,05} = -\frac{a-\mu}{\sigma}$$

$$\Rightarrow a = \mu - \sigma\cdot\lambda_{0,05} = 5 - 2\cdot 1.6449 = 1,71$$

~~6.12~~ Nästa uppgift handlar om linjärkombinationer av oberoende stokastiska variabler.

De här är så hitta praktiska eftersom normalfördelningens egenskaper bevaras under linjära transformationer.

Sats $\boxed{6.3:}$

Om $X \in N(\mu, \sigma)$ så gäller att

$$Y = aX + b \in N(a\mu + b, |a|\sigma)$$

$\boxed{6.4}$

$$X \pm Y \in N\left(\mu_X \pm \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

$\boxed{6.5}$

Om $X_1, X_2, \ldots X_n$ är oberoende och resp. $N(\mu_1, \sigma_1), \ldots N(\mu_n, \sigma_n)$ och talen $a_1, \ldots a_n$ och b är givna, så gäller

$$\sum_1^n a_i X_i + b \in N\left(\sum_1^n a_i \mu_i + b, \sqrt{\sum_1^n a_i^2 \sigma_i^2}\right)$$

6.12 En hiss i ett varuhus är markerad med "högst 10 personer eller 800 kg." Personvikten i kg hos en slumpvis uttagen vuxen varuhuskund kan antas vara $N(70,10)$. Vad är sannolikheten för att 10 vuxna överlastar hissen enligt det senare kriteriet?

Som vi ser så är det vi har fått exakt vad vi behöver för att konstruera en ny normalfördelad variabel: väntevärdet och variansen för alla tio variabler, som vi vill baka ihop till en ny.
Formel 6.5 ger (låt oss kalla vår nya variabel Y)

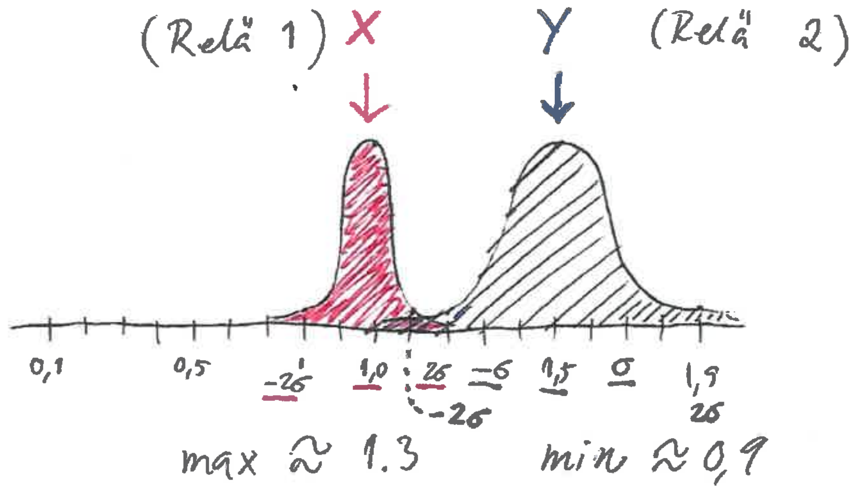
$$Y = \sum_1^{10} X_i \in N\Big(\sum_1^{10} \underbrace{\mu_i}_{70\text{ kg}}, \sqrt{\sum_1^{10} \underbrace{\sigma_i}_{10\text{ kg}}^2}\Big)$$

$$Y \in N\left(10\cdot 70, \sqrt{10\cdot 10^2}\right) = N\left(700, 10\cdot\sqrt{10}\right)$$

Nice! Och vad kan vi göra med en stokastisk variabel som inte är $N(0,1)$? Normalisera!
Eftersom
$Y \in N(700, \sqrt{1000})$ och $\frac{Y-\mu}{\sigma} \in N(0,1)$ så blir värdet vi söker

$$P(\underbrace{Y}_{\text{vår nya variabel}} > \underbrace{800}_{\text{maxvikten}}) = P\left(\frac{Y-\mu}{\sigma} > \frac{800-\mu}{\sigma}\right) = 1-\Phi\left(\frac{800-700}{\sqrt{1000}}\right) = 1-\Phi(\sqrt{10}) = 0{,}0007827.$$

6.13. Man har två reläer som är inställda för utlösning 1 resp. 1.5 sek efter en impuls. Utlösningstiderna är ej konstanta utan $N(1, 0.1)$ resp $N(1.5, 0{,}2)$ och oberoende. Bestäm sannolikheten att det andra reläet utlöses före det första om de samtidigt utsätts för en impuls.



Tolkningen av texten innebär att relä 2 löser ut före relä 1 om $Y < X$, eller när $Y - X < 0$.
Så hur löser vi det här? Med en ny stokastisk variabel!
Vi introducerar $W = Y - X$. Fördelningen för W är

$$\mu_W = E[W] = E[Y-X] = 1{,}5 - 1 = 0{,}5 \quad \text{(pga linjäritet)}$$

$$\sigma^2 = V[W] = V[Y-X] = V[Y] + (-1)^2 \cdot V(X) = 0{,}2^2 + 0{,}1^2$$

$$\Rightarrow W \in N(0{,}5, \sqrt{0{,}05}).$$

Nu vill vi veta sannolikheten att W blir mindre än noll, dvs, sannolikheten att Y är en kortare tid än X.

$$P(W < 0) = P\left(\frac{W-\mu}{\sigma} < \frac{0-\mu}{\sigma}\right) = \Phi\left(\frac{-\mu}{\sigma}\right) = \Phi(-\sqrt{5})$$

$$= 1 - \Phi(\sqrt{5}) = 1 - 0{,}9873 = \underline{\underline{0{,}0127}}$$

DVS chansen är väldigt liten att det händer.

Och nu... Centrala Gränsvärdessatsen!! En av de mest fundamentala satserna inom Statistik och sannolikhetsteori.

sats. 6.8: CGS

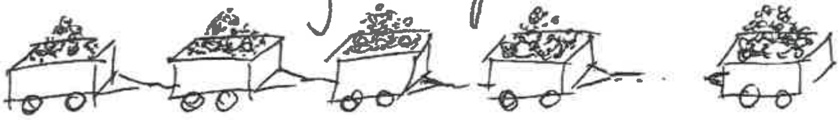
"Om $X_1, X_2 ..$ är en oändlig följd av oberoende likafördelade s.u. med väntevärdet μ och standardav. $\sigma > 0$, så gäller för $Y_n = X_1 + X_2 ... + X_n$ att

$$P\left(a < \frac{Y_n - n \cdot \mu}{\sigma\sqrt{n}} \leq b\right) \longrightarrow \Phi(b) - \Phi(a) \text{ då } n \longrightarrow \infty$$

6.20 Vid lastning av malm i en järnvägsvagn avviker den verkliga lastens vikt slumpmässigt från det nominella värdet 10 ton och kan ses som utfall av en s.u med $(10, 0{,}5^2)$

$\mu = 10, \sigma = 0{,}5$

Bestäm approx. sannolikh. för att totallasten i ett tågsätt om 25 vagnar överskrider 255 ton. Antag att vagnarnas laster är oberoende.



$Y = X_1 + X_2 + X_3 + X_4 + \ldots + X_{25}$ = vikten för de 25 vagnarna

$P(Y > 225) = ?$

$X_1, \ldots, X_{25}$ är likafördelade och ob.

$$\mu_Y = E[Y] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = n \cdot E[X] = 25 \cdot 10 = 250$$

variansen, σ^2, blir

$$\sigma^2 = V(Y) = V\left(\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \sum_{i=1}^{n} V(X_i) = n \cdot V(X) = 25 \cdot 0{,}5^2 = \frac{25}{4}$$

DVS $Y \sim N(\mu, \sigma) = N(250, 2{,}5)$ enligt CGS.

$$P(Y > 225) = P\left(\frac{Y - \mu}{\sigma} > \frac{255 - \mu}{\sigma}\right) \approx 1 - \Phi(2) = 0{,}02275.$$