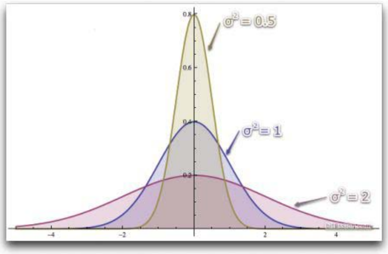
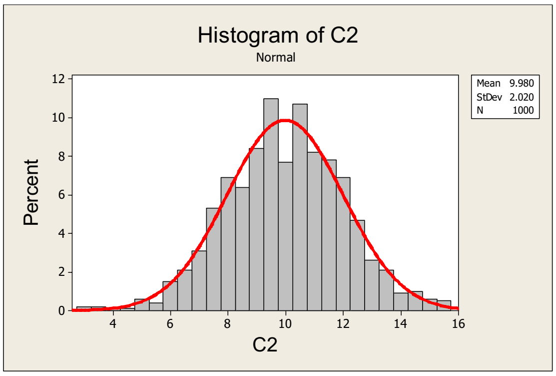
Normalfördelningen



uwe.menzel@math.uu.se

Normalfördelningen är viktig



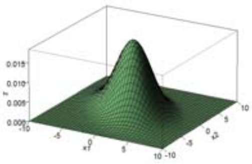
Moivre 1733 (myntkast)
Laplace 1812
Ljapunov

Approximation av en diskret fördelning

Normalfördelningen

Philosophically speaking, the Normal distribution represents one of the empirically verified elementary "truths about the general nature of reality," and its status can be compared to the one of fundamental laws of natural sciences.

http://www.statsoft.com/textbook/esc.html



Exempel

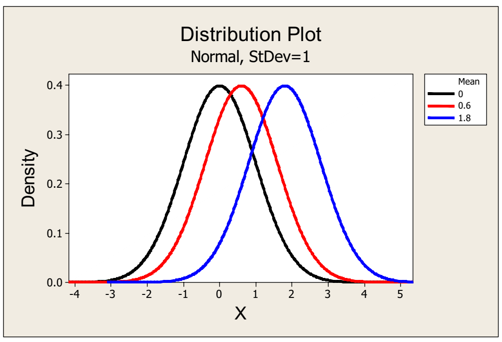
- Kroppslängd däggdjur:
  - lognormal $Y = \exp(X)$ (X normal, Y lognormal)
- mätfel
- blodtryck (kvinnor och män var för sig)
- blodsocker
- proteinkoncentration i blodet

Täthetsfunktionen

Två parametrar: μ, σ
X ~ N(μ, σ)

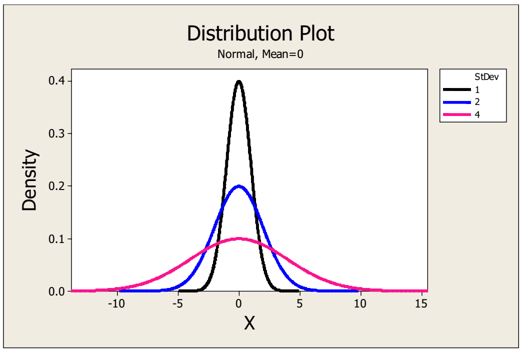
$$f_X(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Väntevärdet μ



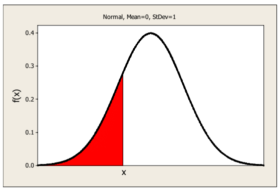
Graph / Probability Distribution Plot / Vary Parameters / Normal / Means = 0 0.6 1.8

Standardavvikelsen σ



Graph / Probability Distribution Plot / Vary Parameters / Normal / Standard deviations = 1 2 4

Fördelningsfunktionen (cdf)



$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

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

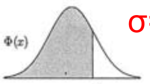
Det finns inget analytiskt uttryck för $F_X(x)$

Standard normalfördelning: Φ(x)

μ = 0
σ= 1

Blom s. 397

Tabell 1. Standard normalfördelning



$\Phi(x) = P(X \leq x)$, där $X \in N(0,1)$.
För negativa x, utnyttja att $\Phi(-x) = 1 - \Phi(x)$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

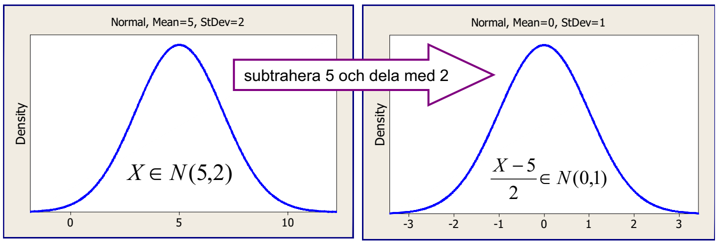
tabellerad från 0 .. 4.0 i 0.01-steg

Allmän normalfördelning

$$X \in N(\mu,\sigma) \quad \Rightarrow \quad \frac{X-\mu}{\sigma} \in N(0,1)$$

Standardiserad normalvariabel

Om X är normalfördelad, så är (X-μ)/σ standard-normalfördelad.



Allmän normalfördelning

$$X \in N(\mu,\sigma) \quad \Rightarrow$$

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

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

Ur tabellen vet vi Φ för ”alla” argument

Φ är fördelningsfunktionen för standard-normalfördelningen, dvs. när μ = 0 och σ = 1 (tabell).

Med hjälp av denna formel och en tabell över Φ kan man beräkna fördelningsfunktionen för alla värden av x, för alla μ och σ (med en viss precision).

Fördelningsfunktion för allmän normalfördelning

$$X \in N(\mu,\sigma)$$

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

Exempel

Exempel 1: $X \in N(4,2)$ Vad är $P(X \leq 3)$??

$$\boxed{P(X \leq 3) = F_X(3) = \Phi\left(\frac{3-4}{2}\right)} = \Phi(-0.5)$$
$$= 1 - \Phi(0.5) = 1 - 0.6915 = \mathbf{0.3085}$$

Exempel 2: $X \in N(4,2)$ Vad är $P(X \geq 5)$??

$$\begin{aligned} P(X \geq 5) &= 1 - P(X < 5) = 1 - P(X \leq 5) \\ &= 1 - F_X(5) = 1 - \Phi\left(\frac{5-4}{2}\right) \\ &= 1 - \Phi(0.5) = 1 - 0.6915 = \mathbf{0.3085} \end{aligned}$$

Sannolikhet att slumpvariabeln hamnar mellan a och b

$$P(a < X \leq b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

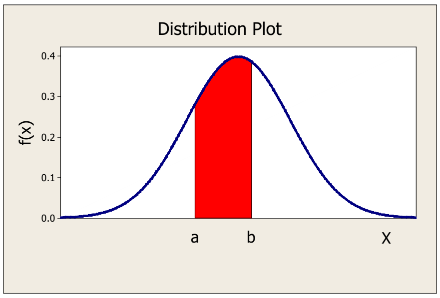
gäller allmänt

gäller om X är normalfördelad

s.v.= slumpvariabel

Sannolikhet att s.v. hamnar mellan a och b

$$P(a < X \leq b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$



Exempel

$$\boxed{P(a < X \leq b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

Exempel: $X \in N(4,2)$ Vad är $P(3 \leq X \leq 5)$??

$$\begin{aligned} P(3 \leq X \leq 5) &= F_X(5) - F_X(3) \\ &= \Phi\left(\frac{5-4}{2}\right) - \Phi\left(\frac{3-4}{2}\right) \\ &= \Phi(0.5) - \Phi(-0.5) \\ &= \Phi(0.5) - [1 - \Phi(0.5)] \\ &= 2 \cdot \Phi(0.5) - 1 \\ &= 2 \cdot 0.6915 - 1 = 0.383 \end{aligned}$$

Sannolikheten inom σ-intervaller

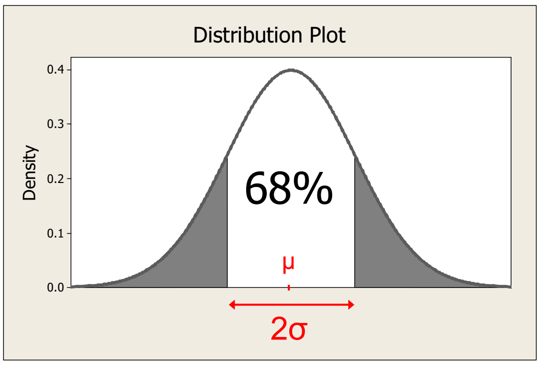
$$\boxed{P(a < X \leq b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

$$\begin{aligned} P(\mu - 2\cdot\sigma < X \leq \mu + 2\cdot\sigma) &= F_X(\mu + 2\cdot\sigma) - F_X(\mu - 2\cdot\sigma) \\ &= \Phi\left(\frac{\mu + 2\cdot\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - 2\cdot\sigma - \mu}{\sigma}\right) \\ &= \Phi(2) - \Phi(-2) = 0.954 \end{aligned}$$

Samma med 1 och 3:

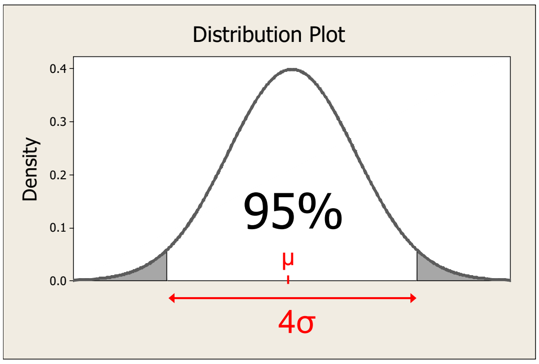
$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= 0.682 \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 0.954 \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 0.997 \end{aligned}$$

$$\begin{aligned} \boxed{P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.682} \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.954 \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997 \end{aligned}$$

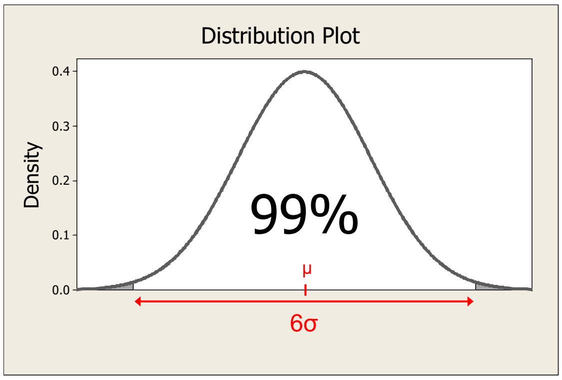


nsigma.MPJ

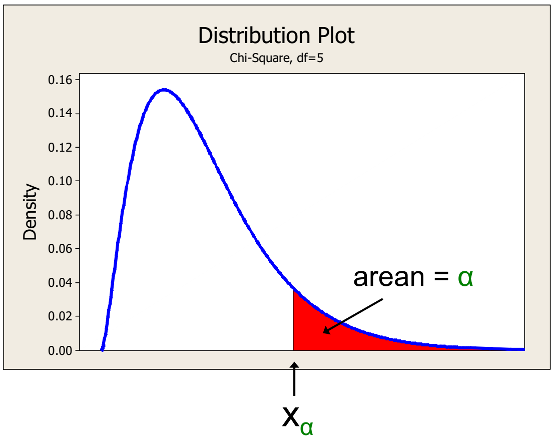
$$
\begin{aligned}
P(\mu-\sigma \le X \le \mu+\sigma) &= 0.682\\
\boxed{P(\mu-2\sigma \le X \le \mu+2\sigma) = 0.954}\\
P(\mu-3\sigma \le X \le \mu+3\sigma) &= 0.997
\end{aligned}
$$



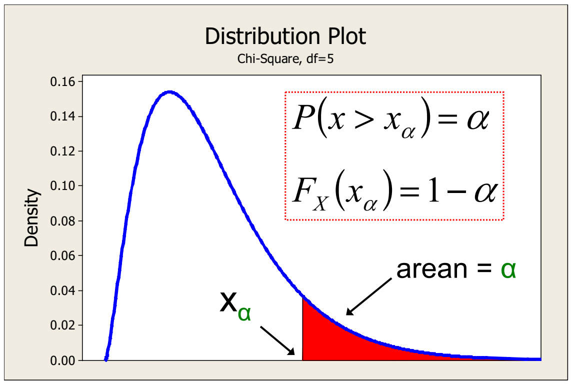
$$
\begin{aligned}
P(\mu-\sigma \le X \le \mu+\sigma) &= 0.682\\
P(\mu-2\sigma \le X \le \mu+2\sigma) &= 0.954\\
\boxed{P(\mu-3\sigma \le X \le \mu+3\sigma) = 0.997}
\end{aligned}
$$



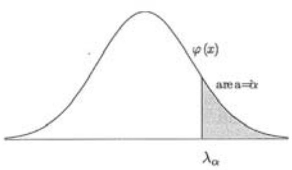
Definition av kvantilerna



Kvantiler



Kvantilerna för N(0,1) kallas λ_α



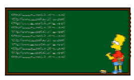
Figur: Kvantiler för $N(0,1)$

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

Sannolikhet inom $\mu \pm \lambda_{\alpha/2}\sigma$ för allmän normalfördelning

$$
\boxed{P(a < X \le b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}
$$

$$
\begin{aligned}
P(\alpha) &= P(\mu - \lambda_{\alpha/2}\sigma < X \le \mu + \lambda_{\alpha/2}\sigma)\\
&= F_X(\mu + \lambda_{\alpha/2}\sigma) - F_X(\mu - \lambda_{\alpha/2}\sigma)\\
&= \Phi\left(\frac{\mu + \lambda_{\alpha/2}\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - \lambda_{\alpha/2}\sigma - \mu}{\sigma}\right)\\
&= \Phi(\lambda_{\alpha/2}) - \Phi(-\lambda_{\alpha/2})\\
&= \Phi(\lambda_{\alpha/2}) - [1 - \Phi(\lambda_{\alpha/2})]\\
&= 2 \cdot \Phi(\lambda_{\alpha/2}) - 1\\
&= 2[1 - \alpha/2] - 1\\
&= 1 - \alpha
\end{aligned}
$$



Sannolikhet inom $\mu \pm \lambda_{\alpha/2}\sigma$ för allmän normalfördelning

$$\boxed{P\left(\mu - \lambda_{\alpha/2}\sigma < X \leq \mu + \lambda_{\alpha/2}\sigma\right) = 1 - \alpha}$$

Sätt $\alpha = 0.05,\ 0.001,\ 0.0001$

$$\begin{aligned} P(\mu - 1.96\sigma < X \leq \mu + 1.96\sigma) &= 0.95 \\ P(\mu - 2.58\sigma < X \leq \mu + 2.58\sigma) &= 0.99 \\ P(\mu - 3.29\sigma < X \leq \mu + 3.29\sigma) &= 0.999 \end{aligned}$$

Ex: $\alpha = 0.05,\ \alpha/2 = 0.025,\ \lambda_{\alpha/2} = 1.96,\ 1 - \alpha = 0.95$

Låt $X \in N(3, 1) \implies P(1 < X \leq 5) \approx 95\%$

Medelvärdet är en slumpvariabel !

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\cdot(X_1 + X_2 + \ldots + X_n)$$

Slumpvariabel

Slumpvariabler

Försök	X_i	medelvärde
1	10, 12, 11, 10, 9, 11, 11, 13, 9, 10	10,6
2	10, 13, 11, 9, 9, 11, 10, 13, 9, 14	10,9
3	12, 12, 13, 9, 10, 11, 12, 10, 11, 10	11,0
...	...	mindre spridning

$s_1 = 1.26 \quad s_2 = 1.85 \quad s_3 = 1.25 \qquad s_{\bar{x}} = 0.21$

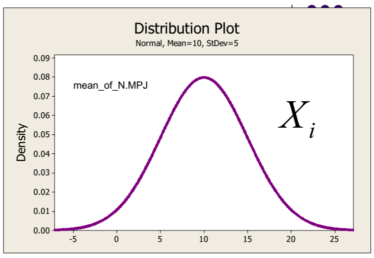
Flera N-fördelade variabler – fördelning av deras medelvärde

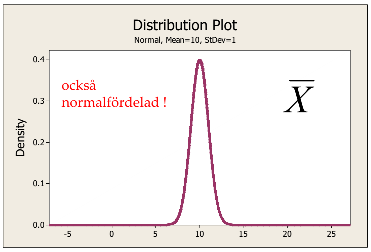
$X_i \in N(\mu, \sigma)$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\bar{X} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

SEM





Medelvärdet har en mindre spridning: det är just därför vi tar medelvärdet när vi ska mäta någonting!

Exempel: Fördelning för X_i

Exempel: Ett mätfel X är fördelad $N(0, 10)$.

Hur stor är slh. att felet är begränsad till $(-3, 3)$??

$$\begin{aligned} P(-3 < X < 3) &= F_X(3) - F_X(-3) \\ &= \Phi\left(\frac{3-0}{10}\right) - \Phi\left(\frac{-3-0}{10}\right) \\ &= \Phi(0.3) - \Phi(-0.3) = 2\cdot\Phi(0.3) - 1 \approx 0.236 \end{aligned}$$

Sannolikheten är 23% att felet hållas i gränserna mellan -3 och 3

Exempel: Fördelning för $\bar{X}$

Exempel: Ett mätfel X är fördelad $N(0, 10)$.

Hur stor är slh. att felets medelvärde (över 25 mätningar) är begränsad till $(-3, 3)$??

$$X \in N(0, 10) \ \Rightarrow\ \bar{X} \in N\left(0, \frac{10}{\sqrt{25}}\right) = N(0, 2)$$

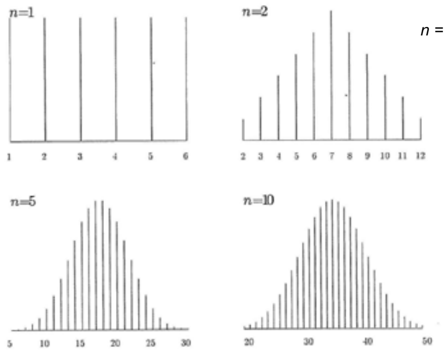
$$\begin{aligned} P(-3 < \bar{X} < 3) &= F_{\bar{X}}(3) - F_{\bar{X}}(-3) \\ &= \Phi\left(\frac{3-0}{2}\right) - \Phi\left(\frac{-3-0}{2}\right) \\ &= \Phi(1.5) - \Phi(-1.5) = 2\Phi(1.5) - 1 \approx 0.866 \end{aligned}$$

Sannolikheten är 87% att felets medelvärde hållas i gränserna mellan -3 och 3. Mycket bättre!

Centrala gränsvärdessatsen

- Vi har sett att summor av normalfördelade variabler är också normalfördelade
- **CGS:** även summor av *godtyckligt* fördelade slumpvariabler är (ungefär) normalfördelade, *bara de är*
  - många
  - oberoende
  - likafördelade

Nytt experiment: kasta flera tärningar och beräkna summan av ögontalet



n = antalet tärningar

Sammanfattning

Komponenterna måste vara oberoende och likafördelade (samma μ och σ).

	normalfördelad	godtycklig fördelad
$Y = \sum_{i=1}^{n} X_i$	$Y \in N\left(n\mu, \sigma\sqrt{n}\right) \ \forall n$	$Y \in N\left(n\mu, \sigma\sqrt{n}\right) \ \ n \to \infty$
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\bar{X} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \ \forall n$	$\boxed{\bar{X} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \ \ n \to \infty}$
	exakt	asymptotiskt