Matematisk statistik F/Kf

Föreläsningar 1999

68 sidor

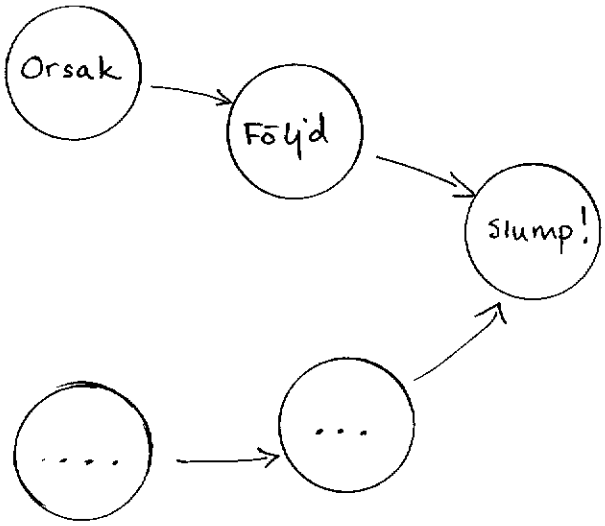
~~20:-~~ ~~30:-~~ 35:-

MATEMATISK STATISTIK:

Slumpen! Sannolikhet!

slumpmässighet och regelbundenhet!

August Cournot:



Experiment:

- En uppsättning villkor
- Kan upprepas många gånger.
- Unika experiment.
- Slumpmässigt urval. (stickprov)

Sannolikhet:

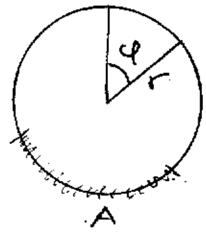
1) Subjektiv sannolikhet.
2) Frekvenstolkning, $\frac{f}{n}$ - relativa frekvensen,
3) Klassisk, $P[A] = \frac{n(A)}{n(S)}$ ← antalet utfall som leder till A; ← totala antalet utfall.

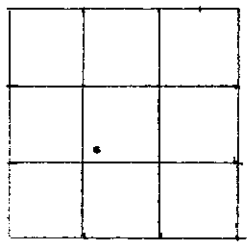
(2.)

4) Geometrisk sannolikhet

$0 < \varphi < 2\pi$

$\dfrac{\text{Längd}(A)}{2\pi r}$





$$P(A) = \frac{\text{area}(A)}{\text{area}(S)}$$

(1.2) UTFALLSRUM OCH HÄNDELSE

Ex. Experiment: Två myntkast.

$S = \{$(klave, klave) (klave, krona) (krona, klave) (krona, krona)$\}$

Utfall

S – utfallsrum.

0 – klave, 1 – krona $\Rightarrow$ $S = \{00, 01, 10, 11\}$

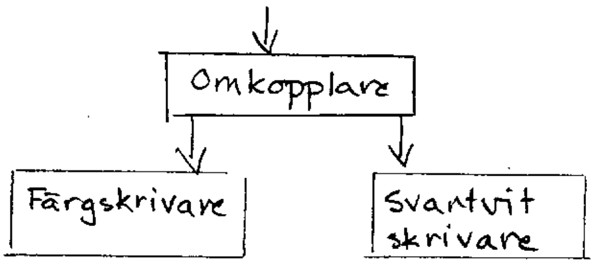
Händelse A; en delmängd av S. $A \subset S$

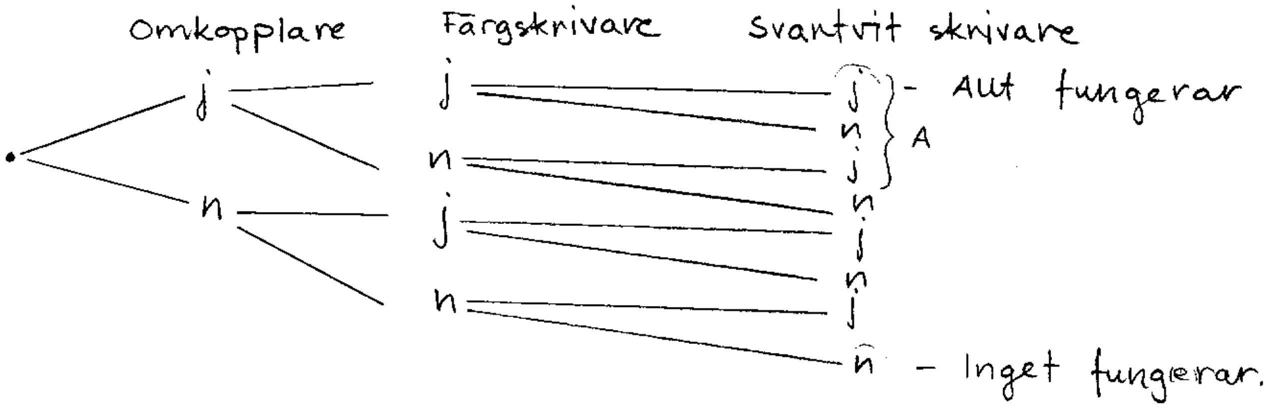
Experiment: Vi får utfall u.

Om $u \in A$, A inträffade.

$u \notin A$, A inträffade inte.

Träddiagram:





$S = \{jjj, jjn, jnj, jnn, njj, njn, nnj, nnn\}$

A - "utskriftsmöjlighet finns". $A = \{jjj, jjn, jnj\}$

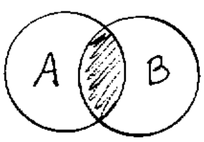
B_1 - "färgskrivare fungerar". $B_1 = \{jjj, jjn, njj, njn\}$

B_2 - "svartvit skrivare fungerar". $B_2 = \{jjj, jnj, njj, nnj\}$

C - "omkopplare fungerar". $C = \{jjj, jnj, jjn, jnn\}$

$$A = \underset{\text{"och"}}{C \cap} (B_1 \underset{\text{"eller"}}{\cup} B_2)$$

Mängdteoretiska operationer med händelser:



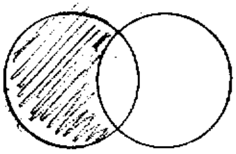
$A \cap B$ - <u>snitt</u> av händelse A och B.

$A \cap B$ inträffar omm både A och B inträffar.

4.

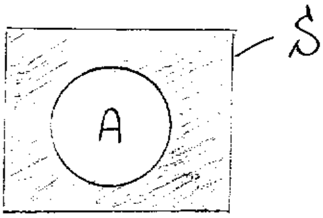
$A \cup B$ – <u>union</u>

$A \cup B$ inträffar om åtminstone en av händelserna A och B inträffar.



$A \setminus B$ – <u>differens</u>

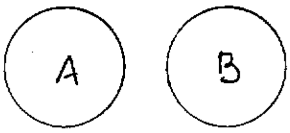
$A \setminus B$ inträffar om A inträffar men inte B.



A' – <u>komplement</u>händelse

A^c

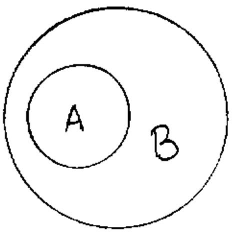
A' inträffar omm A inte inträffar.



$A \cap B = \emptyset$

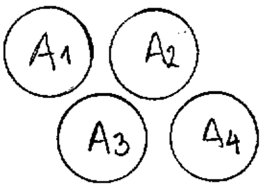
$\emptyset$ – tomma mängden (omöjlig händelse)

A och B oförenliga.



$A \subset B$ – A innehålls i B.

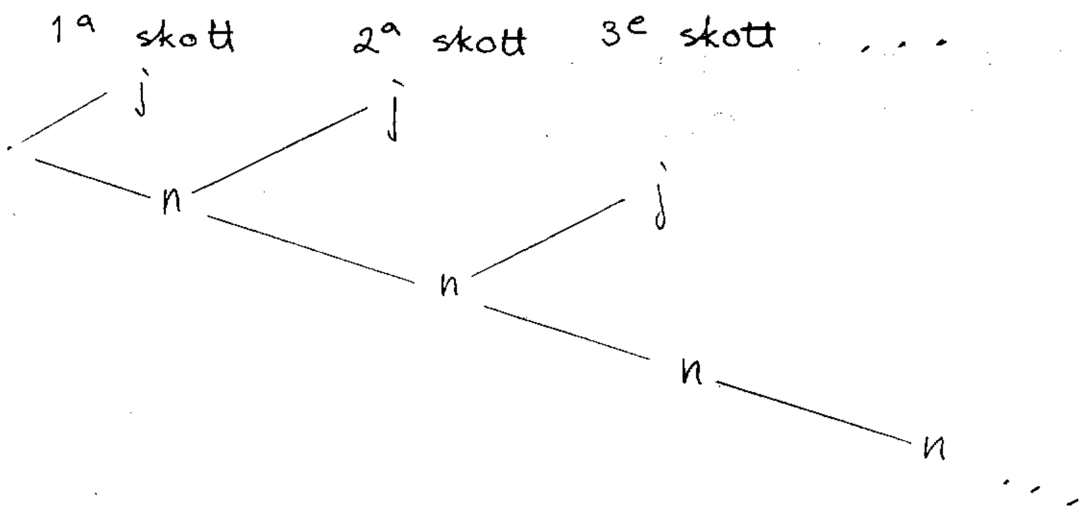
A medför B.



A_1, A_2, A_3, A_4 är oförenliga.

$A_i \cap A_j = \emptyset$ om $i \neq j$.

Ex. Skjutning tills första träff.



(1.3) PERMUTATIONER OCH KOMBINATIONER:

Klassiska formeln: $P[A] = \frac{n(A)}{n(S)}$

Multiplikationsprincipen:

Ett experiment sker i k steg. För steg nr. j finns n_j möjligheter hur kan det gå, $j = 1, 2, \ldots, n$. Då finns det totala $\prod_{j=1}^{n} = n_1 \cdot n_2 \cdot n_3 \ldots n_j$ möjligheter till hur hela experimentet skall gå.

19/3

Antalet permutationer:

Ex. 5 st. tävlande i ett sprinterlopp.
Hur många platsordningar finns det?

$\frac{5}{1.a\ pl.} \ \frac{4}{2.a\ pl.} \ \frac{3}{3{:}e\ pl.} \ \frac{2}{4{:}e\ pl.} \ \frac{1}{5{:}e\ pl.} = \frac{120}{\text{Totalt!}}$

6.

Allmänt: $n! = n(n-1)(n-2)(n-3)\ldots 2\cdot 1$

Antalet ordnande placeringar hos r st. objekt, utvalda ur en mängd med n objekt.

$$\underset{1:a}{n}\cdot\underset{2:e}{(n-1)}\ldots\underset{r:te}{n-(r-1)} = \underset{\text{Totalt!}}{n(n-1)\ldots(n-r+1)}$$

$$\Rightarrow \quad {}_nP_r = n(n-1)\ldots(n-r+1)$$

<u>Antalet kombinationer</u> (Antalet delmängder)

En mängd med n objekt. Vi väljer ut en delmängd med r objekt. Hur många möjligheter finns?

$$\Rightarrow \quad {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} =$$

$$= \frac{n(n-1)\ldots(n-r+1)\cancel{(n-r)}\cancel{\ldots}1}{r!\,\cancel{(n-1)}\cancel{\ldots}1} = \frac{{}_nP_r}{r!}$$

<u>Ex.</u> ${}_9C_7 = \binom{9}{7} = \dfrac{9\cdot 8\cdot\cancel{7}\cdot\cancel{6}\cdot\cancel{5}\cdot\cancel{4}\cdot\cancel{3}}{\cancel{7}\cdot\cancel{6}\cdot\cancel{5}\cdot\cancel{4}\cdot\cancel{3}\cdot 2\cdot 1} = \binom{9}{2} = \dfrac{9\cdot 8}{2} = 36$

$$\Rightarrow \quad {}_nC_r = \frac{n!}{r!(n-r)!} = {}_nC_{(n-r)}$$

<u>Permutationer av oskiljbara objekt:</u>

n objekt av k typer

$$\left.\begin{array}{llll} n_1 & \text{stycken av} & \text{typ} & 1 \\ n_2 & \text{"} \quad \text{"} & \text{"} & 2 \\ \vdots & & & \\ n_k & \text{"} \quad \text{"} & \text{"} & k \end{array}\right\} \quad n_1+n_2+\ldots+n_k = n$$

Hur många ordnade finns det?

Svar: $\dfrac{n!}{n_1!\,n_2!\cdots n_k!}$

Bevis: Det sökta antalet (placeringar hos oskiljbara objekt) sättes till x.

Vi namnger våra objekt så att alla går att skilja från varandra. Det finns $n!$ ordnade placeringar hos dessa skiljbara objekt.

Vi kan få en ordnad placering på följande sätt:

a) Vi bygger en ordnad placering med de oskiljbara objekten.

b) Därefter ordnar vi objekten av samma typ i en viss ordning, för typ $1, \dots, k$.

$$n! = \underbrace{x}_{a)} \cdot \underbrace{n_1!}_{b)\,typ\,1} \cdot \underbrace{n_2!}_{b)\,typ\,2} \cdots \underbrace{n_k!}_{b)\,typ\,k} = \underbrace{x \cdot n_1!\,n_2! \cdots n_k!}_{Totalt!}$$

Uttrycker $x = \dfrac{n!}{n_1!\,n_2!\cdots n_k!}$

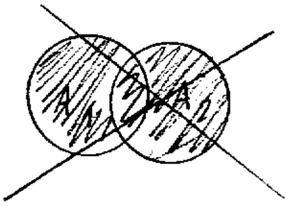
(2.1) SANNOLIKHETSTEORINS AXIOM:

Utfallsrum S

1. $P[S] = 1$.
2. $P[A] \geq 0$
3. Om $A_1, A_2, \dots$ är en ändlig eller oändlig följd av parvis oförenliga händelser, dvs $A_i \cap A_j = \emptyset$ för $i \neq j$,

så gäller att

$$P[A_1 \cup A_2 \cup \ldots] = P[A_1] + P[A_2] + \ldots$$



utan

<u>Sats:</u> $P[\emptyset] = 0$

Bevis: $S = S \cup \emptyset$ Vi ser också att S och tom mängd är oförenliga.

Axiom 3 $\Rightarrow$ $\cancel{P[S]} = \cancel{P[S]} + P[\emptyset]$

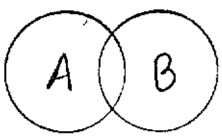
$$P[\emptyset] = 0 \quad \blacksquare$$

<u>Sats:</u> $P[A^c] = 1 - P[A]$

Bevis: $A \cup A^c = S$, $A \cap A^c \cap \emptyset$

$1 = P[S] = P[A] + P[A^c]$ $\blacksquare$

<u>Adderingsregel:</u> $P[A \cup B] = P[A] + P[B] - P[A \cap B]$.

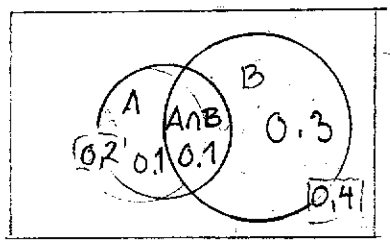


<u>(2.2) BETINGAD SANNOLIKHET:</u>

<u>Ex.</u> A: Betyg 5
B: Teknisk linje

$P[A] = 0{,}2$ $P[B] = 0.4$

$P[A \cap B] = 0{,}1$



Hur stor är sannolikheten att en student som har gått på teknisk linje får betyg 5?

Svar: $\frac{0.1}{0.4} = 0.25$, $P[A|B] = \frac{P[A \cap B]}{P[B]}$

Betingad sannolikhet av A givet B.

Def. $P[A|B] = \frac{P[A \cap B]}{P[B]}$ (*) , då $P[B] \neq 0$.

Om A och B inte har något samband så är $P[A|B] = P[A]$

Vidare ger (*) multiplikationsregeln:

$$P[A \cap B] = P[B] \cdot P[A|B]$$

(2.3) OBEROENDE HÄNDELSER:

Det finns inga helt oberoende händelser, men de som har så svaga samband, att dessa är försumbara, kan behandlas så.

Ex. Två tärningskast.

A: På första kastet < 3. $P(A) = \frac{2 \cdot \cancel{6}}{6 \cdot \cancel{6}} = \frac{1}{3}$

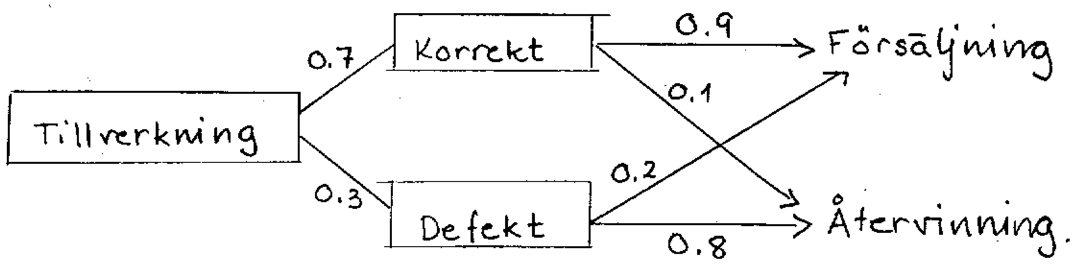
B: På andra kastet > 3. $P(B) = \frac{\cancel{6} \cdot 3}{\cancel{6} \cdot 6} = \frac{1}{2}$

$A \cap B$: $P(A \cap B) = \frac{2}{6} \frac{3}{6} = P(A) \cdot P(B)$

Def. Händelser A och B kallas oberoende om
$$P[A \cap B] = P[A] \cdot P[B]$$

(2.4.) BAYES SATS:

Ex.



Hur stor andel i försäljningen är defekt?

Lösn.: A_1 : föremålet är korrekt
A_2 : " " defekt
B : " klassificeras till försäljning.

Söker $P[A_2 | B]$.

$$P[A_2 | B] = \frac{P[A_2 \cap B]}{P[B]} = \frac{P[B | A_2] \cdot P[A_2]}{P[B]}$$

$$B = \underbrace{(A_1 \cap B) \cup (A_2 \cap B)}_{\text{oförenliga}} \Rightarrow P[B] = P[A_1 \cap B] + P[A_2 \cap B] =$$

$$= P[B | A_1] \cdot P[A_1] + P[B | A_2] \cdot P[A_2]$$

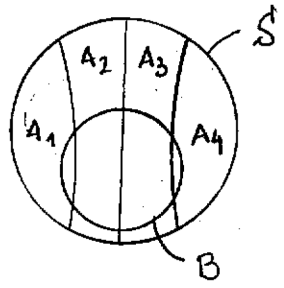
$$\Rightarrow P[A_2 | B] = \frac{P[B|A_2] \cdot P[A_2]}{P[B|A_1] \cdot P[A_1] + P[B|A_2] \cdot P[A_2]} =$$

$$= \frac{0{,}2 \cdot 0{,}3}{0{,}9 \cdot 0{,}7 + 0{,}2 \cdot 0{,}3} = \frac{0.06}{0.69} = \ldots$$

Allmänt: En uppsättning händelser $A_1, A_2, \ldots, A_n$ kallas partition av S om

1) $A_i \cap A_j = \emptyset$ då $i \neq j$

2) $A_1 \cup A_2 \cup \ldots \cup A_n = S$



Sats 2.4.1: (Bayes sats)
Låt $A_1, A_2, \dots, A_n$ utgöra en partition av S och låt B vara en händelse. Då är

$$P[A_i \mid B] = \frac{P[B \mid A_i] \cdot P[A_i]}{\sum_{k=1}^{n} P[B \mid A_k] \cdot P[A_k]}$$

- Terminologi:

$A_1, A_2, \dots, A_n$ – hypotes

$P[A_1], \dots, P[A_n]$ – apriori sannolikheter

$P[A_1 \mid B], \dots, P[A_n \mid B]$ – apostoriori sannolikheter.

(3.1–3.9) DISKRETA STOKASTISKA VARIABLER

Stokastisk variabel – Ett talvärde som vi observerar i ett experiments resultat.

En stokastisk variabel kallas diskret om mängden av dess möjliga värden är ändlig eller uppräknelig.

Sannolikhetstäthet och Geometrisk fördelning:

Def. 3.2.1 Låt X vara en diskret stok. var.

12.

Funktionen

$$f(x) = P[X = a]$$

kallas (sannolikhets)<u>täthet</u> för X.

[Ibland säger man: frekvensfunktion]

<u>Ex</u>. Myntkast tills krona kommer upp första gången.

X – antalet gånger som behövs.

$f(x) = ?$

$$\underset{1}{\overset{\text{kl.}}{\bullet}} \quad \underset{2}{\overset{\text{kl.}}{\bullet}} \quad \bullet \quad \cdots \quad \underset{x-1}{\overset{\text{kl.}}{\bullet}} \quad \underset{x}{\overset{\text{kr}}{\bullet}} \quad \bullet \quad \cdots$$

$$f(x) = P[X = x] = \underbrace{\underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2}}_{x-1} \cdot \frac{1}{2}}_{x} = \frac{1}{2^x} \quad \text{om } x \text{ heltal.}$$

x	1	2	3	...	x
$f(x)$	$1/2$	$1/4$	$1/8$	...	$1/2^x$

• Experiment: A, $P[A] = p$

X – antalet gånger tills A förekommer första gången.

$$f(x) = P[X = x] = \underbrace{(1-p)\cdots(1-p)}_{x-1}\,p = (1-p)^{x-1}p$$

$x = 1, 2, 3, \ldots$

Vi betecknar $q = 1 - p$

$$\boxed{f(x) = q^{x-1} \cdot p, \quad x = 1, 2, 3, \ldots}$$

<u>Geometrisk fördelning</u>!

• $P[X \leq 5]$?

$$P[X \leq 5] = f(1) + f(2) + f(3) + f(4) + f(5) = \sum_{k=1}^{5} q^{k-1} p = \frac{p - pq^5}{1-q}$$

$$= [q = 1-p] = \frac{p(1-q^5)}{p} = 1 - q^5$$

• $P[4 \leq X \leq 8]$??? (forts. längre ner)

<u>Kummulativ fördelning:</u> (Fördelningsfunktion)

<u>Def.</u> $F(x) = P[X \leq x]$ för alla reella x.

<u>Ex.</u> Geometrisk fördelning.

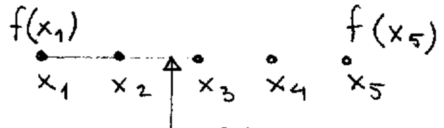
$F(x) = 1 - q^{[x]}$, $[x]$ - heltalsdel av x.

$\Rightarrow P[4 \leq X \leq 8] = F(8) - F(3)$

<u>Väntevärde:</u>

<u>Def.</u> $E[X] = \sum_x x f(x)$

<u>Ex.</u>



$f(x)$ - massa

$E(x)$ - tyngdpunkt

Om $Y = H(X)$ gäller

$$E[H(X)] = \sum_x H(x) f(x)$$

<u>Sats 3.3.1:</u>

1) $E[c] = c$

2) $E[cX] = c\,E[X]$

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

(14.)

Varians:

<u>Def.</u> $\operatorname{Var} X = E\left[(X-\mu)^2\right]$ där $\mu = E[X]$.

- Formel: $\operatorname{Var} X = E[X^2] - \left(E[X^2]\right)^2$

Bevis: $\operatorname{Var} X = E\left[(X-\mu)^2\right] = E\left[X^2 - 2X\mu + \mu^2\right] =$
$= E[X^2] - E[2X\mu] + E[\mu^2] = E[X^2] - 2\mu \underbrace{E[X]}_{\mu}$
$+\mu^2 = E[X^2] - \mu^2$

- Beteckning: $\sigma^2 = \operatorname{Var} X$
$\sigma = \sqrt{\operatorname{Var} X}$

- Egenskaper:
  1) $\operatorname{Var} c = 0$
  2) $\operatorname{Var} cX = c^2 \operatorname{Var} X$
  3) Om X och Y är <u>oberoende</u> stok. var, så är

$$\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y$$

OBS! Oberoende stok. var. ej ännu def.!

Momentgenererande funktion:

Momentgenererande funktion av en stokastisk variabel X definieras;

$$m_X(t) = E\left[e^{tX}\right]$$

<u>Ex.</u> $X \sim \operatorname{Geom}(p)$
↑ är fördelad

$$f(x) = P[X = x] = (1-p)^{x-1} \cdot p = q^{x-1} \cdot p \quad , \quad q = (1-p)$$

$$m_X = E\left[e^{tX}\right] = \sum_{x=1}^{\infty} e^{tx} q^{x-1} p = \sum_{1}^{\infty} (e^t q)^x \cdot p \cdot q^{-1} =$$

$$= \frac{e^t p}{1 - e^t q} = \frac{pe^t}{1 - qe^t}$$

men $e^t q < 1$, dvs $t < \ln \frac{1}{q}$.

- $m_X'(t)\Big|_{t=0} = \frac{d}{dt} E\left[e^{tX}\right]\Big|_{t=0} = E\left[\frac{d}{dt} e^{tX}\right]\Big|_{t=0} = E\left[e^{tX} X\right]\Big|_{t=0} = E[X]$

$$\Rightarrow \quad \boxed{E[X] = \frac{d}{dt} m_X(t)\Big|_{t=0}}$$

<u>Ex.</u> $X \sim \text{Geom}(p)$

$$m_X(t) = \frac{pe^t}{1 - qe^t}$$

$$m_X'(t)\Big|_{t=0} = \frac{pe^t(1-qe^t) - (-qe^t)\,pe^t}{(1-qe^t)^2}\Bigg|_{t=0} = \frac{p(1-q) + qp}{(1-q)^2} =$$

$$= \frac{\cancel{pp} + p - \cancel{pp}}{p^2} = \frac{1}{p}$$

- $m_X''(t)\Big|_{t=0} = \ldots = E[X^2]$

$$\Longrightarrow \quad \boxed{\text{Var}(X) = \frac{q}{p^2}}$$

(18.)

- Geometrisk fördelning är minneslös!

Vi har en produktionsprocess.

P – sannolikheten för ett defekt föremål.

Vi observerar processen tills dess att vi för första gången stöter på ett defekt föremål.

X – antalet föremål.

Vi antar att 15 föremål redan är observerade och alla har varit korrekta.
Hur stor är sannolikheten att vi måste observera åtminstone fem till?

Antagandet kan tolkas som att händelse $\{X > 15\} = \{X \geqslant 16\}$ inträffar.

Frågan: $P\{X \geqslant 20 \mid X \geqslant 16\} = ?$

meningslöst!

$$P\{X \geqslant 20 \mid X \geqslant 16\} = \frac{P\{X \geqslant 20 \text{ och } \cancel{X \geqslant 16}\}}{P\{X \geqslant 16\}} = \frac{q^{19}}{q^{15}} =$$

$$= q^4 = P\{X \geqslant 5\}$$

Binomialfördelning:

Ex. Tre myntkast!

X – antalet "kronor"

Krona – 1, Klave – 0

Detta ger följande utfallsrum:

$$S = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

$$P\{000\} = \frac{1}{2^3} = \frac{1}{8} \quad , \quad P\{111\} = \frac{1}{2^3} = \frac{1}{8}$$

- Utfallens relation till stokastiska variabeln:

$$\{X=3\} = \{111\}$$
$$\{X=2\} = \{011, 101, 110\}$$
$$\{X=1\} = \{001, 010, 100\}$$
$$\{X=0\} = \{000\}$$

Detta ger följande sannolikheter:

$$P\{X=3\} = \frac{1}{2^3} = \frac{1}{8}$$
$$P\{X=2\} = 3 \cdot \frac{1}{8} = \frac{3}{8}$$
$$P\{X=1\} = \frac{3}{8}$$
$$P\{X=0\} = \frac{1}{8}$$

x	0	1	2	3
$f(x)$	$1/8$	$3/8$	$3/8$	$1/8$

Binomialfördelningen med $n=3$, $p=1/2$

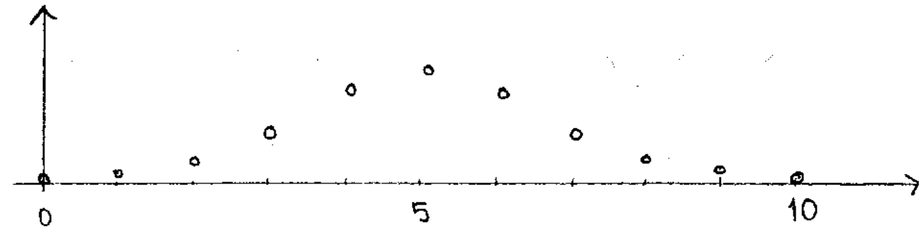
<u>Def.</u> Vi säger att $X \sim Bin(n,p)$ om

$$f(x) = P\{X=x\} = \binom{n}{p} \cdot p^k (1-p)^{n-k}$$

(16.)

$p = 0{,}5$

$n = 10$



Kurvan ser som "en liten kulle", med sin topp i punkten $n \cdot p$.

- Väntevärdet hos binomialfördelningen.

Om $X \sim Bin(n, p)$, så är $E[X] = n \cdot p$,

dvs toppen på "kullen".

Ex. Antal pojkar av 100 nyfödda.

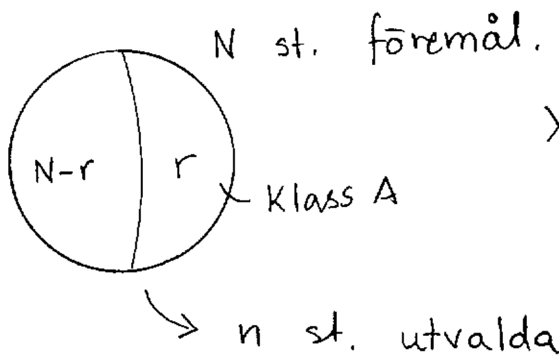
$p = 1/2$, $n = 100$ $\therefore$ Väntevärdet $= 50$.

- Variansen hos binomialfördelningen:

$Var(X) = n \cdot p \cdot q$, $q = 1 - p$

$\sigma = \sqrt{npq}$

Hypergeometrisk fördelning:



X - antalet objekt av klass A bland de utvalda.

$P\{X = x\} = ?$

Lösning: $N = \overbrace{\circ\circ\circ\ldots\circ}^{N-r} \mid \overbrace{\textcircled{A}\textcircled{A}\ldots\textcircled{A}}^{r}$

$$\overbrace{O\ldots O}^{N-k}\ \Big|\ \overbrace{Ⓐ\ldots Ⓐ}^{k}$$

Att välja n objekt ur N objekt! Det finns $\binom{N}{n}$ möjliga sätt att göra detta på.

För att få exakt k objekt av klass A måste dessa väljas ur r, vilket ger

$$\binom{r}{k}\cdot\binom{N-r}{n-k}$$

Antalet möjligheter som leder till exakt k objekt av klass A bland de utvalda blir alltså

$$P[X=x] = \frac{\binom{r}{k}\cdot\binom{N-r}{n-k}}{\binom{N}{n}}$$

- Väntevärdet: $E[X] = n\,\frac{r}{N}$
- Varians: $Var(X) = n\cdot\frac{r}{N}\cdot\frac{N-r}{N}\cdot\frac{N-n}{N-1}$

<u>Poissonfördelning:</u>

<u>Def.</u> Vi säger att X är Poissonfördelad med parametern λ om

$$P\{X=k\} = e^{-\lambda}\frac{\lambda^k}{k!}\quad,\quad k=0,1,2,3,\ldots$$

- Momentfunktionen: $m_X(t) = e^{\lambda(e^t-1)}$

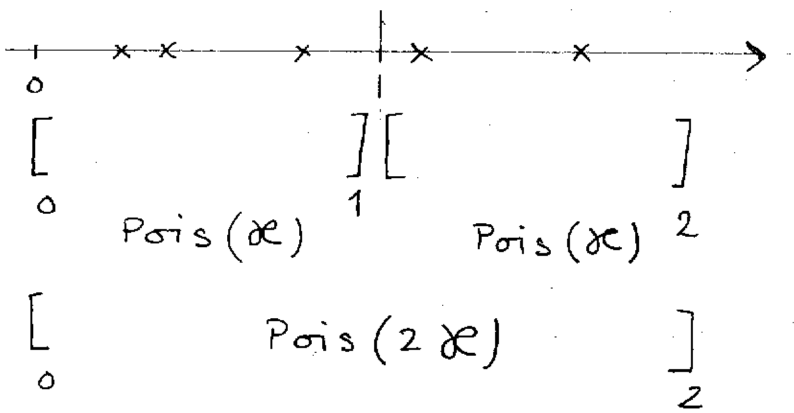
Bevis: $m_X(t) = \sum_{k=0}^{\infty} e^{tk}\cdot e^{-\lambda}\frac{\lambda^k}{k!} = \ldots$ (vänd)

$$\ldots = e^{-\varkappa} \sum_{k=0}^{\infty} \frac{(\varkappa \cdot e^t)^k}{k!} = e^{-\varkappa} \cdot e^{\varkappa e^t} = e^{\varkappa (e^t - 1)}$$

Detta ger:

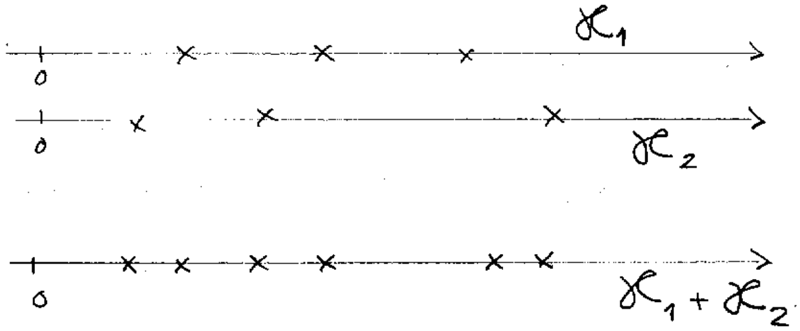
- Väntevärdet: $E[X] = \varkappa$
- Varians: $\text{Var } X = \varkappa$, dvs $\sigma_X = \sqrt{\varkappa}$

Poissonprocess:



Antalet impulser på $[0, t]$ är $\text{Pois}(t \cdot \varkappa)$

Sammansättning av Poissonprocesser:



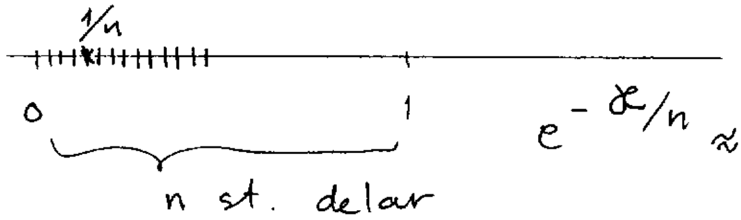
Poissonapproximation av binomialfördelningen:

$\underline{X} \sim \text{Bin}(n,p)$, n är stor

p är liten.

$np \approx \varkappa$

$$P[\underline{X}=k] \approx e^{-\varkappa} \frac{\varkappa^k}{k!}$$



$$e^{-\varkappa/n} \approx 1 - \frac{\varkappa}{n}$$

(4.1-4.9) KONTINUERLIGA FÖRDELNINGAR:

$\underline{X}$ är kontinuerlig. <u>Täthet</u> för $\underline{X}$
- en funktion $f(x)$ med egenskaperna:

1) $f(x) \geq 0$, $\forall x$

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

3) $P[\underline{X} \in [a,b]] = \int_a^b f(x)\,dx$, $\forall a,b$, $a>b$.

Tätheten benämns även som <u>frekvens-funktion</u>.

(22)

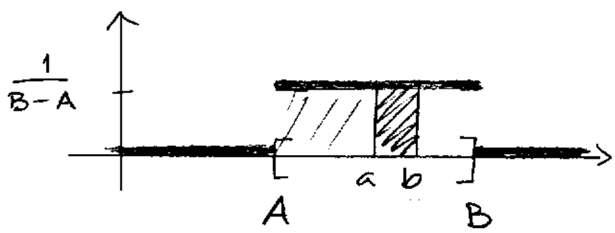
Kummulativ fördelning (Fördelningsfunktion):

$$F(x) = P[\mathbb{X} \leq x]$$

$$F(x) = \int_{-\infty}^{\infty} f(y)\,dy$$

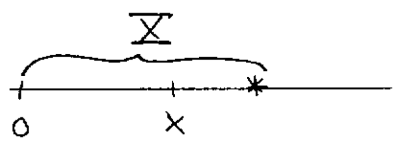
$$F'(x) = f(x)$$

Ex. Likformig fördelning



$$f(x) = \begin{cases} \frac{1}{B-A}, & \text{då } x \in [A,B] \\ 0, & \text{annars} \end{cases}, \quad P[\mathbb{X} \in [A,B]] = \frac{b-a}{B-A}$$

Ex.



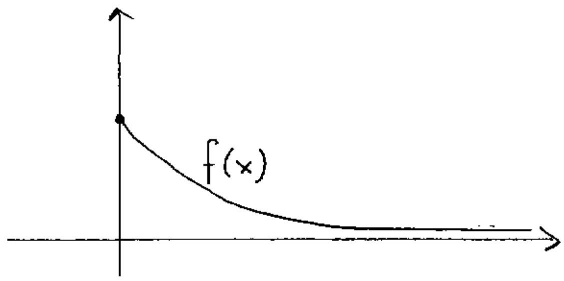
$P[\mathbb{X} \leq x] = 1 - P[\mathbb{X} > x] = 1 - P[$0 impulser på $[0,x]] = 1 - e^{-\varkappa x}$

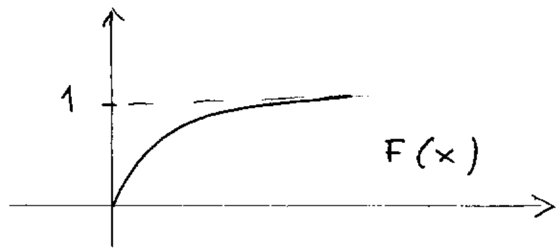
Exponentialfördelning par. med $\varkappa$

$$\beta = \frac{1}{\varkappa}, \quad F(x) = 1 - e^{-\varkappa/\beta} \quad \text{då } x \geq 0$$

$$f(x) = \frac{1}{\beta} e^{-\varkappa/\beta} \quad \text{då } x \geq 0$$

$E[X] = \beta$, exponent fördelning med väntevärdet β.





<u>Kontinuerligt väntevärde:</u>

<u>Def.</u> $E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$

En allmän formel: $E[H[X]] = \int_{-\infty}^{\infty} H(x) f(x) dx$

<u>Ex.</u> Beräknar momentgenererande funktion för exponentialfördelad X.

$$m_X(t) = E[e^{tX}] = \frac{1}{\beta} \int_0^{\infty} e^{tx} e^{-x/\beta} dx = \frac{1}{\beta} \int_0^{\infty} e^{-x(\frac{1}{\beta}-t)} dx$$

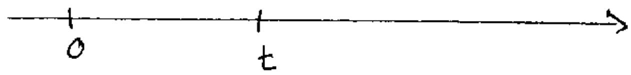
$$= \begin{bmatrix} y = x(\frac{1}{\beta}-t) \\ dx = \frac{dy}{1/\beta - t} \end{bmatrix} = \frac{1}{\beta} \cdot \frac{1}{\frac{1}{\beta}-t} \int_0^{\infty} e^{-y} dy = \frac{1}{\beta} \cdot \frac{1}{1/\beta - t} =$$

$$= \frac{1}{1 - t\beta} \quad \therefore \quad m_X(t) = \frac{1}{1 - t\beta}$$

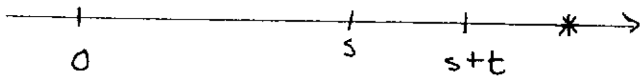
$$E[X] = m'_X(t)\Big|_{t=0} = -\frac{1 \cdot (-\beta)}{(1 - t\beta)^2}\Big|_{t=0} = \beta$$

$$\operatorname{Var} X = m''(t)\Big|_{t=0} = \ldots = \beta^2 \quad , \quad \sigma_X = \beta$$

<u>Minneslöshet hos exponentialfördelningen:</u>



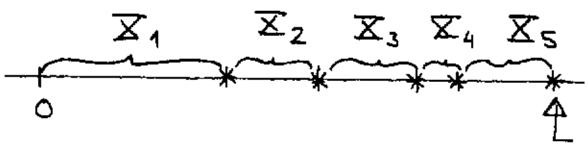
$$P[X \geq t] = e^{-x/\beta}$$



$$P[X > t+s \mid X \geq s] = \frac{P[X \geq t+s \cancel{\cap X \geq s}]}{P[X \geq s]} =$$

$$= \frac{e^{-(t+s)/\beta}}{e^{-s/\beta}} = e^{-t/\beta} = P[X \geq t]$$

<u>Ex.</u> När sker impuls nummer 5.



$X_1 + \ldots + X_5$.

Gammafördelning:

Gammafunktion: $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z}\, dz$

Om α är heltal $\Gamma(\alpha) = (\alpha-1)!$

Detta ger gammafördelningen

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad , x \geq 0.$$

- Momentgenererande funktion:

$$m_X(t) = \int_0^t e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\, dx = \dots =$$

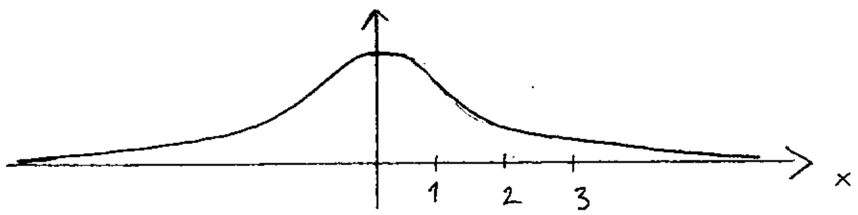
$$= \frac{1}{(1 - t_\beta)^\alpha}$$

- Väntevärdet: $E[X] = \alpha \cdot \beta$
- Varians: $\text{Var}\, \underline{X} = \alpha \cdot \beta^2$

Normalfördelning:

- Standardnormalfördelning:

Täthet – $f(x) = \varphi(x) \stackrel{\text{def.}}{=} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$



26

- Kummulativ fördelning: $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$

- Allmän normalfördelning:

$$f(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , -\infty < \mu < \infty \quad , \sigma > 0 .$$

- Momentfunktion:

$$m_X(t) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx = \ldots \quad (*)$$

$$\text{exponenten} = -\frac{1}{2\sigma^2}\left(-2\sigma^2 tx + x^2 - 2x\mu + \mu^2\right) =$$

$$= -\frac{1}{2\sigma^2}\left(x^2 - 2x(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^2 - (\mu^2 + \sigma^2 t) + \mu^2\right)$$

$$= -\frac{1}{2\sigma^2}\left[\left(x - (\mu + \sigma^2 t)\right)^2\right] - \frac{1}{2\sigma^2}\left[-(\mu + \sigma^2 t)^2 + \mu^2\right]$$

$(*) \quad \ldots = e^{\mu t + \frac{\sigma^2 t^2}{2}}$ (efter substituering).

- Väntevärde:

$$E[X] = m_X'(t)\Big|_{t=0} = e^{\mu t + \frac{\sigma^2 t^2}{2}} = (\mu t + \sigma^2 t)\Big|_{t=0} = \mu$$

- Varians: σ^2

Entydighetssats för momentgenererande funktioner:

Om X_1, $m_{X_1}(t)$, och X_2, $m_{X_2}(t)$ samt

$$m_{X_1}(t) \equiv m_{X_2}(t)$$

så sammanfaller fördelningarna $F_{X_1}(x)$ och $F_{X_2}(x)$.

Låt X vara normalfördelad med par. μ, σ $(X \sim N(\mu, \sigma))$ så $Z = \frac{X-\mu}{\sigma}$ är standard-normalfördelad. $(Z \sim N(0,1))$

Bevis:

$$m_Z(t) = E\left[e^{tZ}\right] = E\left[e^{\frac{t(X-\mu)}{\sigma}}\right] = e^{-\frac{t\mu}{\sigma}} E\left[e^{\frac{t}{\sigma}X}\right] =$$

$$= e^{-\frac{t\mu}{\sigma}} m_X\left(\frac{t}{\sigma}\right) = \cancel{e^{-\frac{t\mu}{\sigma}}} \cdot e^{\cancel{t\mu/\sigma} + \frac{\cancel{\sigma^2}}{2} \cdot \frac{t^2}{\cancel{\sigma^2}}} = e^{-t^2/2}$$

Normalapproximation av normalfördelningen:

Låt $X \sim Bin(n,p)$, $q = 1-p$

Vi vet att $E[X] = np$, $\mathrm{Var}\, X = npq$

$\sigma_x = \sqrt{npq}$.

$$Z = \frac{X - np}{\sqrt{npq}}$$

Om n är stor och p inte är för nära till noll eller till ett, så är Z approximativt $N(0,1)$. Villkor:

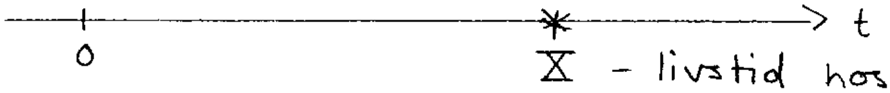
$[p < 0{,}5$ och $np > 5]$ eller $[p \geq 0.5$ och $n(1-p) > 5]$

28.

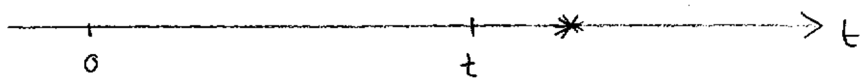
Weibull - fördelning:

Täthet $-$ $f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}$, $\alpha > 0$, $\beta > 0$.

- Väntevärde: $\mu = E[X] = \alpha^{-1/\beta}\,\Gamma(1+\frac{1}{\beta})$
- Variansen: $\sigma^2 = \mathrm{Var}\, X = \alpha^{-2/\beta}\,\Gamma(1+\frac{2}{\beta}) - \mu^2$

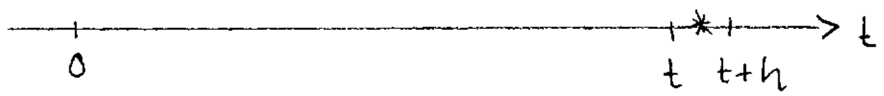


X - livstid hos en komponent eller tiden tills den kräver teknisk betjäning.



Sannolikheten att apparaten fungerar vid tidpunkten t $= P[X > t] = 1 - F(t) = R(t)$

$R(t)$ $-$ tillförlitlighetsfunktionen.



$$\rho(t) = \lim_{h \to 0} \frac{P[t < X < t+h \mid X > t]}{h}$$ felintensitet!

$$\rho(t) = \frac{f(t)}{R(t)}$$ $\leftarrow$ täthet $\leftarrow$ tillförlitlighetsfunktion.

<u>Bevis:</u> Omskrivning av definitionen ger

$$\frac{P(t < X < t+h)}{P(X > t)} = \frac{F(t+h) - F(t)}{R(t)\ h} \xrightarrow[h \to 0]{} \frac{f(t)}{R(t)}$$

$$R(t) = e^{-\int_0^t \varrho(x)\,dx}$$

<u>Bevis</u>: $\varrho(t) = \frac{f(t)}{R(t)} = \frac{F'(t)}{R(t)} = \frac{(1-R(t))'}{R(t)} = -\frac{R'(t)}{R(t)} =$

$$= -\left[\ln R(t)\right]'$$

$$\int_0^t \varrho(x)\,dx = -\ln R(t) - (-\ln R(0))$$

$$R(t) = e^{-\int_0^t \varrho(x)\,dx}$$

$\square$

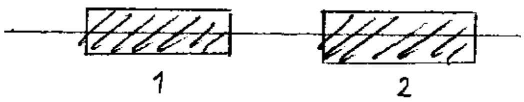
<u>Ex</u>. $\varrho(t) = \alpha \cdot \beta t^{\beta-1}$

$$R(t) = e^{-\int_0^t \alpha\beta x^{\beta-1}\,dx} = e^{-\alpha t^{\beta}}$$

$f(t) = \varrho(t)\,R(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}}$ ← Weibullfördelning!

<u>Tillförlitlighet hos serie- och parallellkopplade system:</u>

· Seriekoppling:

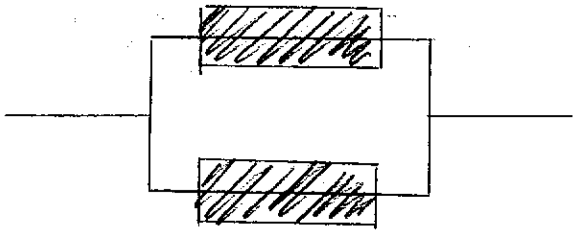


Livstiden hos komponenterna är $X_1, X_2, \ldots, X_n$
Livstiden hos systemet är X_s

$$[X_s > t] = \bigcap_{i=1}^{n} [X_i > t]$$

$$R_s(t) = P[X_s > t] = \prod_{i=1}^{n} R_i(t)$$

• Parallellkoppling:



$$[X_s < t] = \bigcap_{i=1}^{n} [X_i < t]$$

$$R_s(t) = 1 - \prod_{i=1}^{n} (1 - R_i(t))$$

(5.1-5.3) TVÅDIMENSIONELLA FÖRDELNINGAR:

X, Y DISKRET	X, Y KONTINUERLIG
Täthet:	Täthet:
$f_{X,Y}(x,y) = P[X=x, Y=y]$	1) $f_{X,Y}(x,y) \geq 0$
	2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = \underline{1}$
	3) $P[(X,Y) \in A] = \iint_A f_{X,Y}(x,y)\,dx\,dy$

(forts.)

(forts.)

X \ Y	1	2	3	
1	f(1,1)	f(1,2)	...	$f_y(1)$
2	...	...	...	
3	...	...	f(3,3)	
	$f_x(1)$		$f_x(3)$	1,0

$$f_x(1) = \sum_Y f_x(1,y)$$

$$f_y(1) = \sum_X f_y(x,1)$$

$$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dy$$

$$f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dx$$

Oberoende:

Def.: X och Y kallas oberoende om,
$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ för alla x och y.

Väntevärde: Givet $H(X,Y)$.

$$E[H(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y)\, f_{X,Y}(x,y)\,dx\,dy$$

$$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{X,Y}(x,y)\,dx\,dy$$

$$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{X,Y}(x,y)\,dx\,dy$$

Tyngdpunkt: $(E[X], E[Y])$

32.

Kovarians:

$$\mathrm{Cov}(X, Y) = E\left[(X-\mu_x)(Y-\mu_y)\right]$$

där $\mu_x = E[X]$, $\mu_y = E[Y]$

Korrelation: (Korrelationskoef.)

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot \mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \cdot \sigma_y}$$

Sats 5.2.2.:

Om X och Y är oberoende så är

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Bevis: Kontinuerligt fall

$$E[X \cdot Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \, f_{X,Y}(x,y)\,dx\,dy \underset{\text{oberoende}}{=}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \, f_X(x) \cdot f_Y(y)\,dx\,dy =$$

$$= \int_{-\infty}^{\infty} y\, f_Y(y)\,dy \int_{-\infty}^{\infty} x\, f_X(x)\,dy = E[Y] \cdot E[X]$$

Låt X och Y vara <u>oberoende</u>.

$$\mathrm{Cov}(X,Y) = E\left[(X-\mu_x)(Y-\mu_Y)\right] \underset{\text{ober.}}{=} E\left[X-\mu_X\right]\cdot E\left[Y-\mu_Y\right] =$$

$$= \left(E[X]-\mu_x\right)\cdot\left(E[Y]-\mu_Y\right) = 0\cdot 0 = 0$$

Vidare är även $\rho_{X,Y} = \dfrac{0}{\sigma_x\cdot\sigma_y} = 0$

då X och Y är oberoende.

<u>Sats 5,3.1. :</u>

Följande gäller alltid: $-1 \leq \rho_{X,Y} \leq 1$

<u>Bevis</u>:

a) Allmän olikhet:

$$|E[W\cdot Z]|^2 \leq E[W^2]\cdot E[Z^2]$$

Bevis av a):

$$0 \leq E\left[(\alpha W - Z)^2\right] = E\left[\alpha^2 W^2 - 2\alpha W Z + Z^2\right]$$

$$E[W^2]\alpha^2 - 2E[WZ]\alpha + E[Z^2] \geq 0, \quad \forall x.$$

$$\Rightarrow \left(2E[WZ]\right)^2 - 4E[W^2]\cdot E[Z^2] \leq 0$$

Om man inte minns kravet $b^2 - 4ac \leq 0$ så kan man även avsluta beviset med substitutionen:

$$\alpha = \frac{E[W\cdot Z]}{E[W^2]}$$

☒

34

Vi använder hjälpsatsen för $W = X - \mu_X$ och $Z = Y - \mu_Y$, och får därmed:

$$\left| E\left[(X-\mu_X)(Y-\mu_Y)\right] \right|^2 \leqslant E\left[(X-\mu_X)^2\right] E\left[(Y-\mu_Y)^2\right]$$

$$\left(\mathrm{Cov}(X,Y)\right)^2 \leqslant \mathrm{Var}(X) \cdot \mathrm{Var}(Y)$$

$$\Rightarrow \frac{\left(\mathrm{Cov}(X,Y)\right)^2}{\mathrm{Var}(X)\cdot \mathrm{Var}(Y)} \leqslant 1$$

$$\Rightarrow \rho^2_{X,Y} \leqslant 1 \Rightarrow \rho_{X,Y} \leqslant 1$$

Sats 5.3.2. :

$$|\rho_{X,Y}| = 1 \iff \begin{cases} \exists\ \beta_0, \beta_1 \neq 0, \text{ sådana att} \\ Y = \beta_0 + \beta_1 X \end{cases}$$

(5.4) BETINGAD TÄTHET:

(X, Y) $\qquad f_{X,Y}(x,y)$

Def. Betingad täthet av X givet $Y=y$:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Vad betyder t.ex. det diskreta fallet?

$$f_{X|Y}(x|y) = \frac{P[X=x, Y=y]}{P[Y=y]} = P[X=x \mid Y=y]$$

Ex.

$y \backslash x$	1	2	3	4	Σ
0	...	...	...	...	
1	...	$f_{X,Y}(2,1)$	...	...	$f_Y(y)$
2	...	...	...	...	

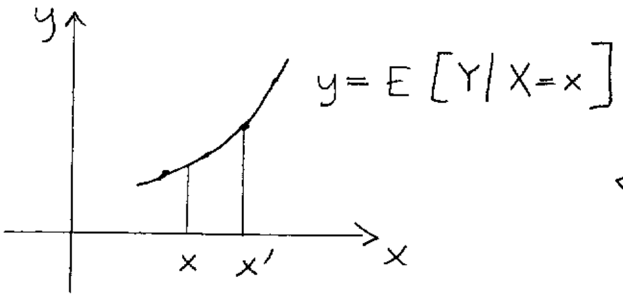
Analogt definieras $f_{Y|X}(y|x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$

Väntevärde:

Kontinuerliga fallet:

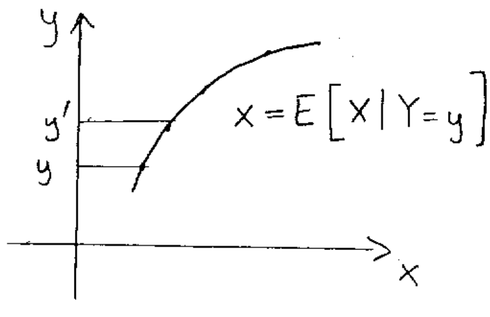
$$E[X|Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

$$E[Y|X=x] = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dy$$



Regressionskurva av Y m.a.p. X.

Regressionskurva av X m.a.p Y



(6.) DESKRIPTIV STATISTIK:

Population $\longrightarrow$ Stickprov

Experiment: Slumpmässigt val av en individ.

Stokastisk variabel X: Registrerade parametervärdet hos en individ.

Stickprov: N st, oberoende likafördelade stokastiska variabler, $X_1, X_2, \ldots, X_N$

Stickprovsvariabel:

Ett stickprov $X_1, X_2, \ldots, X_n$.

En funktion av n variabler: $f(x_1, x_2, \ldots, x_n)$

$$Y = f(X_1, X_2, \ldots, X_n)$$

kallas stickprovsvariabel. På engelska "statistics"!

- I övrigt hänvisas till självständiga studier av kapitel 6!

(7.1-8.7) SKATTNING:

Skattning av väntevärde och varians:

Vi antar att väntevärdet μ är okänt. Vi vill skatta μ utgående från stickprovet $X_1, X_2, \ldots, X_n$.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$, stickprovets medelvärde.

Visar att vi inte gör något systematiskt fel om vi skattar μ med $\bar{X}$.

Bevis: $E[\bar{X}] = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) =$

$= \frac{1}{n}\left(E[X_1] + \ldots + E[X_n]\right) = \frac{1}{n}(\mu + \ldots + \mu) = \frac{1}{n} n\mu = \mu$

$$\boxed{E[\bar{X}] = \mu}$$

$\bar{X}$ är en skattning av μ $(\hat{\mu} = \bar{X})$

$\bar{X}$ är en väntevärdesriktig skattning av μ.

$Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) =$

$= \frac{1}{n^2}(Var\, X_1 + \ldots + Var\, X_n) = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}$

Ober.

där σ^2 är populationens varians.

För att skatta σ^2 så tar vi

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ ← Stickprovsvarians.

OBS!

(38.)

Vi visar att s^2 är väntevärdesriktig uppskattning av σ^2.

Bevis:

$$E[s^2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}((X_i-\mu) - (\bar{X}-\mu))^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i-\mu)^2 - \sum_{i=1}^{n}2(X_i-\mu)(\bar{X}-\mu) + \sum_{i=1}^{n}(\bar{X}-\mu)^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}E\left[(X_i-\mu)^2\right] - 2E\left[\sum_{i=1}^{n}(X_i-\mu)(\bar{X}-\mu)\right] + E\left[\sum_{i=1}^{n}(\bar{X}-\mu)^2\right]\right] = \frac{1}{n-1}\cdot$$

$$\cdot\left(\sum_{i=1}^{n}\sigma^2 - 2E\left[(\bar{X}-\mu)\cdot n(\bar{X}-\mu)\right] + n\cdot \mathrm{Var}\,\bar{X}\right)$$

$$= \frac{1}{n-1}\left(n\sigma^2 - 2n\,\mathrm{Var}\,\bar{X} + n\,\mathrm{Var}\,\bar{X}\right) =$$

$$= \frac{1}{n-1}\left(n\sigma^2 - n\,\mathrm{Var}\,\bar{X}\right) = \frac{1}{n-1}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) =$$

$$= \frac{1}{\cancel{n-1}}\left(\cancel{(n-1)}\sigma^2\right) = \sigma^2$$

$\boxtimes$

Allmänt:

Vi antar att vår population karakteriseras med en fördelning

$$F_\theta(x)$$

som beror på en parameter θ, vars värde är okänt. Vi vill skatta θ utgående från stickprovet $X_1, \ldots, X_n$.

Vi tar en funktion $g(x_1, \dots, x_n)$ och sätter in $X_1, \dots, X_n$.

$$\hat{\theta} = g(X_1, \dots, X_n)$$

↑ skattning för θ

Skattningen kallas väntevärdesriktig om

$$E[\hat{\theta}] = \theta$$

(Väntevärdet beräknas enligt fördelningen $F_\theta(x)$).

Två metoder:

1) Momentmetoden

2) Maximum likelihood metoden

Momentmetoden:

$\alpha_k = E[X^k]$ moment av ordning k.

$M_k = \hat{\alpha}_k = \dfrac{X_1^k + X_2^k + \dots + X_n^k}{n}$ stickprovets moment av ordning k.

$$E[M_k] = \frac{1}{n}\left(E[X_1^k] + \dots + E[X_n^k]\right) = \frac{1}{n} n \cdot \alpha_k = \alpha_k$$

dvs M_k är väntevärdesriktig skattning för α_k.

Ex. $X \sim \text{Gamma}(\alpha, \beta)$

$$f_{\alpha,\beta}(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0$$

(40.) Vi antar att α och β är båda okända och vi vill skatta α och β.

Vi har visat att $\boxed{E[X] = \alpha \cdot \beta}$

$$\text{Var } X = \alpha\beta^2$$

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

Vi sätter in beteckningar för skattningar istället för parametrarna α, β och istället för momentet:

$$\begin{cases} M_1 = \hat{\alpha}\hat{\beta} \\ M_2 - M_1^2 = \hat{\alpha}\hat{\beta}^2 \end{cases} \qquad \begin{cases} \hat{\beta} = \dfrac{M_2 - M_1^2}{M_1} \\ \hat{\alpha} = \dfrac{M_1}{\beta} = \dfrac{M_1^2}{M_2 - M_1^2} \end{cases}$$

<u>Maximum - likelihood - metoden:</u>

$F_\theta(x)$ — Stickprov $X_1, \dots, X_n$

— Konkreta värden $x_1, \dots, x_n$

Maximum likelihood metoden består i att vi väljer det värde som ger största värdet för tätheten i punkten $(x_1, \dots, x_n)$ som skattning för θ.

$$L(\theta, x_1, \dots, x_n) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) =$$

$$= f_\theta(x_1) \cdot f_\theta(x_2) \cdot \dots \cdot f_\theta(x_n) \quad \text{där} \quad f_\theta(x) = F_\theta'(x)$$

Vi väljer $\hat{\theta}$ som sådant värde på θ, som maximerar uttrycket $L(\theta, x_1, \dots, x_n)$ för givna $x_1, \dots, x_n$.

$$L(\hat{\theta}, x_1, \dots, x_n) = \max_{\theta} L(\theta, x_1, \dots, x_n)$$

Ex. $X \sim N(\mu, \sigma)$, $X_1, \dots, X_n$ - stickprov.
Vi antar att både μ och σ är okända.

$$L(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x_1-\mu)^2/2\sigma^2} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x_n-\mu)^2/2\sigma^2} =$$

$$= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} e^{-\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma^2}.$$ Logaritmering!

$$\Rightarrow \ln L(\mu, \sigma) = \frac{n}{2}(-\log 2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\begin{cases} \frac{\partial L}{\partial \mu} = 0 \\ \frac{\partial L}{\partial \sigma} = 0 \end{cases} \qquad \begin{cases} \frac{1}{\sigma^2}\sum(x_i-\mu) = 0 & (i) \\ -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum(x_i-\mu)^2 = 0 & (ii) \end{cases}$$

(i) $\sum x_i - n\mu = 0 \Rightarrow \mu = \frac{1}{n}\sum x_i \Rightarrow \hat{\mu} = \bar{x}$

(ii) $\frac{1}{\sigma^2}\sum(x_i-\bar{x})^2 = \frac{n}{\sigma} \Rightarrow \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sigma^2$

$$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

(42.)

Intervallskattning:

Ett stickprov $X_1, X_2, \dots, X_n$ ur en population med fördelningen $F_\theta(x)$. θ - okänd
Vi skall konstruera

$$h_1 = h_1(X_1, \dots, X_n) \quad h_2 = h_2(X_1, \dots, X_n)$$

så att $P(h_1 \leqslant \theta \leqslant h_2) = 1 - \alpha$

där α är ett litet tal.

$[h_1, h_2]$ kallas konfidensintervall för θ med konfidensgrad $1-\alpha$.

Konfidensintervall för μ då σ är känd:

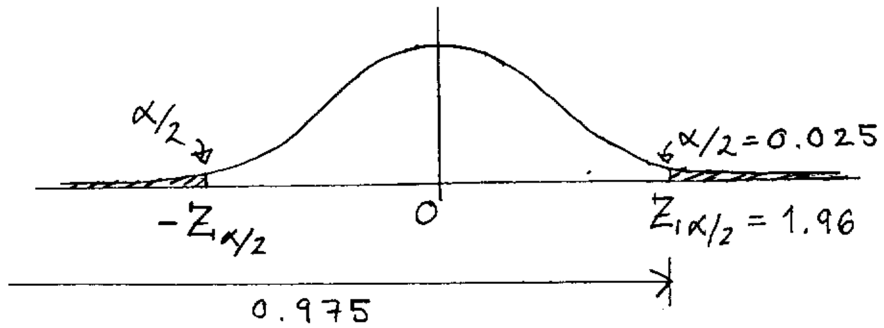
$X_1, X_2, \dots, X_n$ är ett stickprov ur en normalfördelad population med par. μ, σ.
μ är okänd. σ är känd.

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

Man kan välja att $\bar{X}$ är $N(\mu, \frac{\sigma}{\sqrt{n}})$.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{är} \quad N(0,1)$$

Ex. Vi fixerar α, $\alpha = 0.05$



$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha = 0.95$$

$$P\left(-Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right) = 0.95$$

$$P\left(\underbrace{\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{L_1} \leq \mu \leq \underbrace{\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{L_2}\right) = 0.95$$

$\left[\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}}\right]$ konfidensintervall för μ med konfidensgraden $1 - \alpha = 0.95$.

Centrala gränsvärdessatsen:

Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov ur en population med fördelningen F, med väntevärdet μ och variation σ^2. Då är, för stora n,

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

44.

Intervallskattning av variabiliteten:

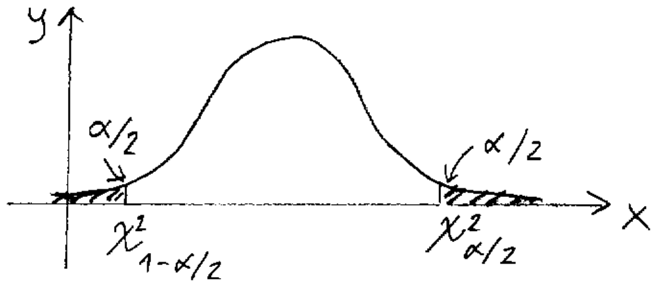
$X_1, X_2, \ldots, X_n$ är ett stickprov ur en normalfördelad $(N(\mu, \sigma))$ population.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Uttrycket $\frac{s^2}{\sigma^2}(n-1)$ är χ^2-fördelad med $n-1$ frihetsgrader.

χ^2-fördelning med r frihetsgrader definieras som fördelningen för $U_1^2 + U_2^2 + \ldots + U_r^2$ där $U_i \sim N(0,1)$ och $U_1, U_2, \ldots, U_r$ är ober.

Bl. a. χ^2-fördelning med r frihetsgrader sammanfaller med gammafördelningen med parametrar $\beta = 2$, $\alpha = \frac{r}{2}$



$$P\left(\chi^2_{1-\alpha/2} < \frac{s^2}{\sigma^2}(n-1) < \chi^2_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\frac{s^2}{\chi^2_{\alpha/2}}(n-1) < \sigma^2 < \frac{s^2}{\chi^2_{1-\alpha/2}}(n-1)\right) = 1 - \alpha$$

Skattning av medelvärdet då σ är okänd:

$X_1, X_2, \ldots, X_n$ är ett stickprov ur $N(\mu, \sigma)$-population. μ är okänd, σ är okänd. Vi skall skatta μ.

(Problemet löstes av Gosset, verksam på bryggeriet Guinnes i början av 1900-talet. Han använde pseudonymen "student".)

$$\frac{\bar{X}-\mu}{s/\sqrt{n}}$$ har s.k. T-fördelning med $n-1$ frihetsgrader.

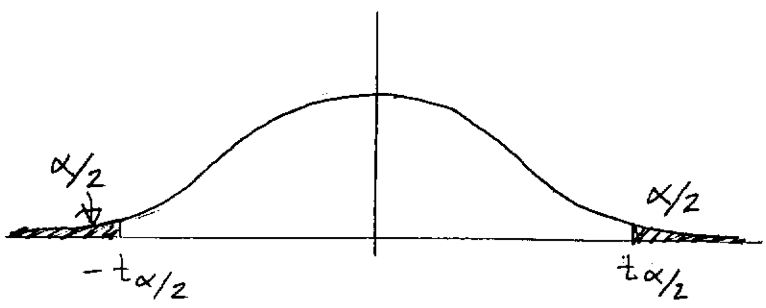
T-fördelning med r frihetsgrader:

Fördelning för en stor variabel av följande form:

$$T = \frac{Z}{\sqrt{X_r^2/r}}$$

Z är $N(0,1)$, X_r^2 är χ^2-fördelad med r frihetsgrader och Z och X_r^2 är ober.

Vi fixerar α;



$$P\left(-t_{\alpha/2} < \frac{\bar{X}-\mu}{s/\sqrt{n}} < t_{\alpha/2}\right) = 1-\alpha$$

$$P\left(\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

Hypotesprövning:

Ett stickprov $X_1, \dots, X_n$ ur en population med fördelningsfunktionen $F_\theta(x)$ och θ är okänd. Vi har en viss grund för att påstå att $\theta = \theta_0$.

$H_0: \theta = \theta_0$ nollhypotes

$H_1: \theta > \theta_0$ alternativhypotes

Om $\hat{\theta} > \hat{\theta}_{kr} \longrightarrow$ Vi förkastar H_0.

Vi fixerar ett litet tal, α - signifikansnivå.

Det är sannolikheten med vilken vi tillåter oss att fatta fel beslut (om H_0 verkligen är sant.)

Ex. $H_0: p = \frac{1}{2}$

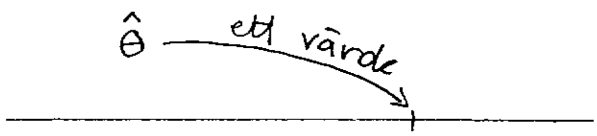
$H_1: p > \frac{1}{2}$

Beslut \ Verkligheten	H_0	H_1
H_0		Andra slags fel Sannolikhet $\beta = \beta(\theta)$
H_1	Första slags fel. Sannolikhet α	

$1-\beta$ - styrkefunktion!

Signifikansprövning:

Vi genomför liknande beräkningar som för hypotesprövning.



Hur stor är sannolikheten att ett sådant, eller ännu mer extremt värde skulle förekomma? Utgående från informationen om fördelningen av $\hat{\theta}$ får vi fram svaret.

$P = P(\theta)$, det s.k. P-värdet.

Om vi ser att P-värdet är litet, så förkastar vi H_0 och vid redovisning av resultatet säger vi: "Vi har förkastat H_0, och P-värdet var 'så och så' stort".

Hypotes och signifikansprövning av medelvärdet:

- $X_1, \ldots, X_n$ ett stickprov ur en normalfördelad population där μ och σ är okända. Vi har någon grund att tro att $\mu = \mu_0$ och vi vill kolla om det stämmer.

$$H_0 : \mu = \mu_0$$

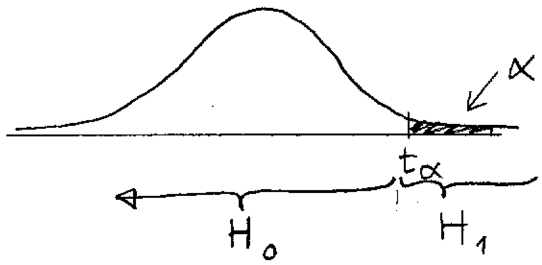
$$H_1 : \mu > \mu_0$$

Högersidig alternativhypotes!

Vi vet att, om H_0 är sant, så är

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Students T-fördelning med $n-1$ frihetsgrader. Vi fixerar signifikansnivån α, t.ex. $\alpha = 0.05$. Ur tabellen för T-fördelningen letar vi fram ett sådant tal t_α att $P(T > t_\alpha) = \alpha$



Beslutsregel: Om $T \geq t_\alpha$ så H_1
Om $T < t_\alpha$ så H_0

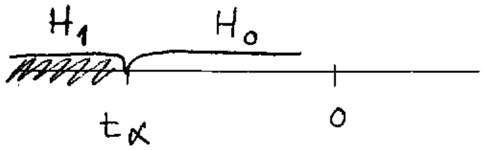
- Vi beräknar T enligt föregående formel. Vi letar fram P-värdet, P(t), ur tabellen; (m.h.a. interpolation) eller får P-värdet m.h.a. programvara (Minitab, Statgraphics, ...). Om P-värdet är litet så ger vi svaret:

"H_1 och P-värdet".

- Vänstersidig alternativhypotes:

$H_0: \mu = \mu_0$

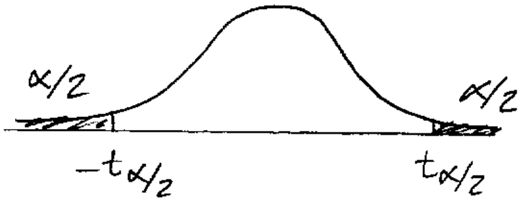
$H_1: \mu < \mu_0$

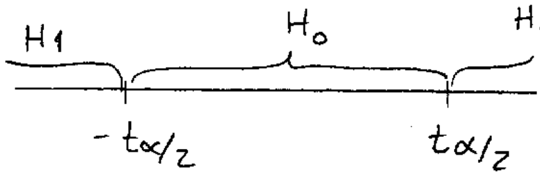


- Tvåsidig alternativhypotes:

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

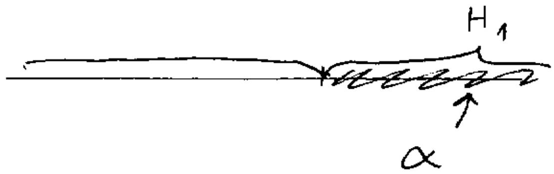




- Komplex 0-hypotes:

Ex $H_0: p \leq \frac{1}{2}$ — α bestämmer vi utgående från gränsfallet $p = \frac{1}{2}$

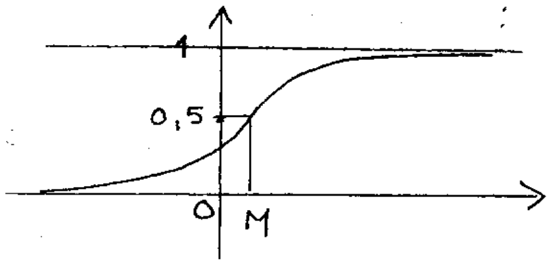
$H_1: p > \frac{1}{2}$



Tecken-test för medianen:

Medianen för en kontinuerlig stokastisk variabel X definieras som ett sådant tal M, att

$$P(X > M) = P(X < M) = \frac{1}{2}$$



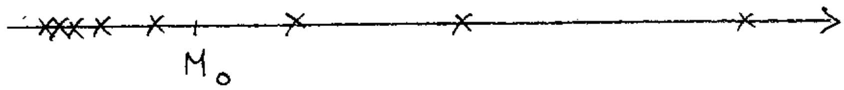
Ett stickprov $X_1, X_2, \ldots, X_n$ ur en population med helt okänd fördelning. Vi betecknar M - medianen (vilken också är okänd.)

$H_0: \quad M = M_0$

$H_1: \quad M > M_0$

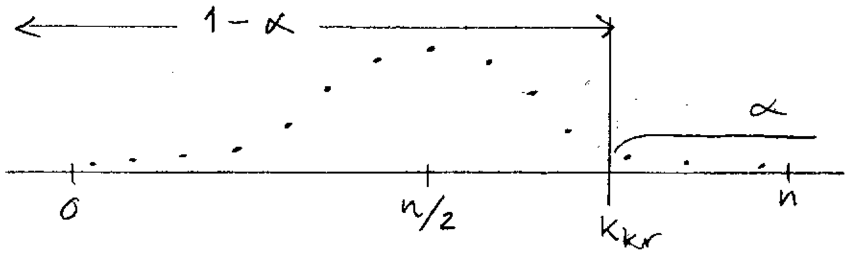
Låt Q_+ vara antalet positiva differenser $X_i - M_0$ bland alla n differenser.

Ex.



I ex. $Q_+ = 3$.

Q_+ är $Bin\left(n, \frac{1}{2}\right)$. Vi fixerar α. Tabell för normalfördelning ger:



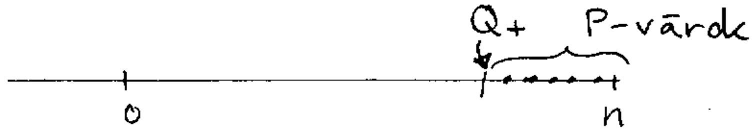
Beslutsregel: Om $Q_+ \geq K_{kr}$ H_1

Om $Q_+ < K_{kr}$ H_0

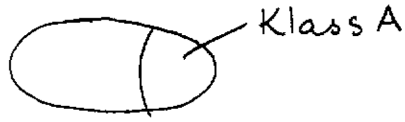
för stora n är

$$Z = \frac{Q_+ - n/2}{\sqrt{\frac{n}{4}}} \quad \text{, approx. } N(0,1)$$

- signifikansprövning:



(9.1–9.4) HYPOTESPRÖVNING AV PROPORTIONER:



$X_1, X_2, \dots, X_n$

$$X_i = \begin{cases} 1 & \text{om individen till klass A.} \\ 0 & \text{annars} \end{cases}$$

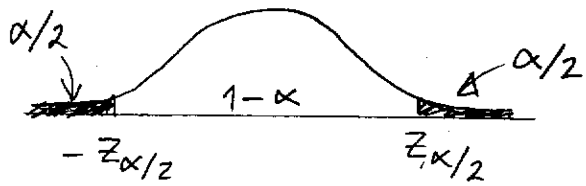
$X = X_1 + \dots + X_n$, antalet individer i stickprovet som tillhör klass A.

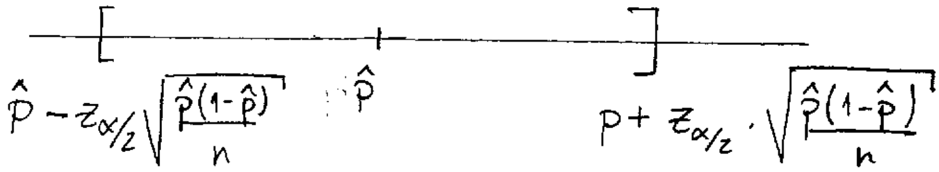
$\hat{p} = \frac{X}{n}$ skattning för proportionen (sannolikheten) p.

Vi vet att $X \sim Bin(n,p)$. Om n är stor så är $Z = \frac{X - np}{\sqrt{np(1-p)}}$, approx. $N(0,1)$.

Vi skriver $\hat{p}$ istället för p i nämnaren.
Konfidensintervall med konfidensgraden $1-\alpha$

blir då





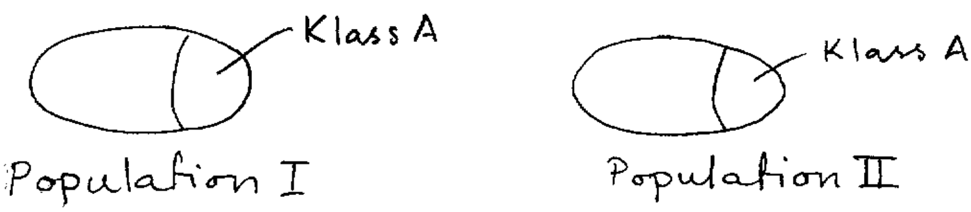
ty $$P\left(-z_{\alpha/2} \leqslant \frac{X-np}{\sqrt{n\hat{p}(1-\hat{p})}} \leqslant z_{\alpha/2}\right) \approx \alpha$$

$$np - z_{\alpha/2}\sqrt{n\hat{p}(1-\hat{p})} \leqslant X$$

$$p \leqslant \frac{X}{n} + z_{\alpha/2}\,\frac{\sqrt{n\hat{p}(1-\hat{p})}}{n}$$

$$\Longrightarrow \quad \boxed{\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Jämförelse av två proportioner:



$X_1^{(1)}, X_2^{(1)}, \dots, X_n^{(1)}$ $\qquad$ $X_1^{(2)}, X_2^{(2)}, \dots, X_n^{(2)}$

$X_1 = \sum_{i=1}^{n_1} X_i^{(1)}$ $\qquad$ $X_2 = \sum_{i=1}^{n_2} X_i^{(2)}$

$$p_1 - p_2 = ?$$

Punktskattning:

$$\hat{p}_1 = \frac{X_1}{n_1}, \qquad \hat{p}_2 = \frac{X_2}{n_2}$$

$$\widehat{p_1 - p_2} = \frac{X_1}{n_1} - \frac{X_2}{n_2} = \hat{p}_1 - \hat{p}_2$$

- $E[\hat{p}_1 - \hat{p}_2] = E[\hat{p}_1] - E[\hat{p}_2] = p_1 - p_2$
- $Var[\hat{p}_1 - \hat{p}_2] = Var(\hat{p}_1) + Var(\hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

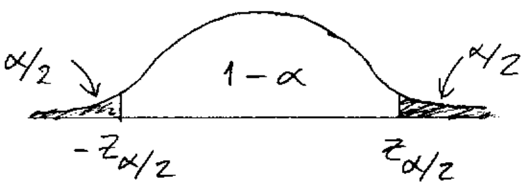
  $Var[\hat{p}_1 - \hat{p}_2] \approx \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$

Centrala gränsvärdessatsen:

$$Z = \frac{\widehat{p_1 - p_2} - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

är approximativt $N(0,1)$.

- Konfidensintervall för $p_1 - p_2$ med konf. graden $1-\alpha$ blir:



$$\widehat{p_1 - p_2} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

54

Hypotesprövning vid jämförelse av två proportioner:

$$\begin{cases} H_0: & p_1 = p_2 \\ H_1: & p_1 > p_2 \end{cases}$$

Om H_0 är sant, så har $\widehat{p_1 - p_2}$ väntevärdet 0 och variansen $p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$ där $p = p_1 = p_2$.

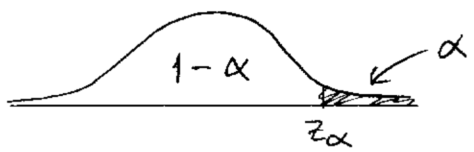
Den sammanvägda skattningen för p:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

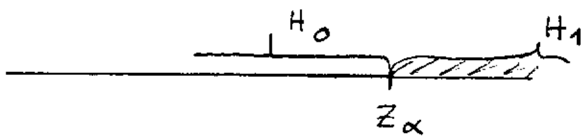
Om H_0 är sant så är

$$Z = \frac{\widehat{p_1 - p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

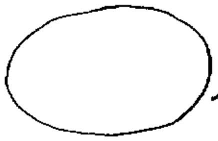
Vi fixerar α. Ur tabellen för $N(0,1)$



Beslutsregel:



(10.1 - 10.5) JÄMFÖRELSE AV TVÅ MEDELVÄRDEN.

 μ_1, σ_1

Population I

$X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}$

"Stickprov I"

$\hat{\mu}_1 = \overline{X^{(1)}}$

 μ_2, σ_2

Population II

$X_1^{(2)}, X_2^{(2)}, \ldots, X_n^{(2)}$

"Stickprov II"

$\hat{\mu}_2 = \overline{X^{(2)}}$

$$\mu_1 - \mu_2 = ?$$

Punktskattning:

$$\widehat{\mu_1 - \mu_2} = \hat{\mu}_1 - \hat{\mu}_2 = \overline{X^{(1)}} - \overline{X^{(2)}}$$

Vi antar att stickproven är <u>oberoende</u>

$$\widehat{\mu_1 - \mu_2} \text{ är } N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}\right)$$

<u>Jämförelse av två medelvärden då varianserna är lika:</u>

$\mu_1 - \mu_2 = ?$

Vi antar att $\sigma_1 = \sigma_2 = \sigma$ men σ är okänd.

Den <u>sammanvägda</u> stickprovsvariansen är

$$\boxed{s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}}$$

Beteckningen $s_p = s$ använd i boken, där p:et står för "pooled".

$(n_1+n_2-2)\frac{s^2}{\sigma^2}$ är χ^2-fördelad med (n_1+n_2-2) frihetsgrader.

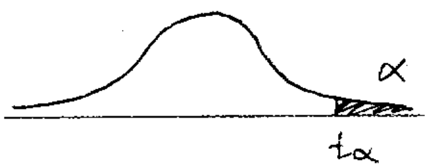
$$T = \frac{(\overline{X}^{(1)} - \overline{X}^{(2)}) - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_2} + \frac{1}{n_2}\right)}},$$

T-fördelad med n_1+n_2-2 frihetsgrader.

Hypotesprövning:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$$

Vi fixerar signifikansnivå α. Vi letar i tabellen av $T(n_1+n_2-2)$.



Vi beräknar
$$T = \frac{\overline{X}^{(1)} - \overline{X}^{(2)}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Beslutsregel: Om $T \geqslant t_\alpha$ så H_1.

Om $T < t_\alpha$ så H_0.

Jämförelse av två medelvärden då varianserna är olika:

$$T = \frac{(\overline{X}^{(1)} - \overline{X}^{(2)}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \overset{\text{approx.}}{\sim} T(\gamma)$$

där $\gamma \approx \dfrac{\left[s_1^2/n_1 + s_2^2/n_2\right]}{\frac{\left[s_1^2/n_1\right]^2}{n_1 - 1} + \frac{\left[s_2^2/n_2\right]^2}{n_2 - 1}}$ avrundat nedåt!

Kopplade data. Jämförelse av två medelvärden:

Ex.

X, Y - blodtryck hos patienter.

Före en viss behandling: Efter ...:

$$\begin{array}{ccc} X_1 & \longleftrightarrow & Y_1 \\ X_2 & \longleftrightarrow & Y_2 \\ \vdots & & \vdots \\ X_n & \longleftrightarrow & Y_n \end{array}$$

Stickproven "hänger ihop", dvs de är inte oberoende. Studera istället differensen.

$D = X - Y$

n	X_i	Y_i	D_i
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_n	Y_n	$D_n = X_n - Y_n$

(58.)

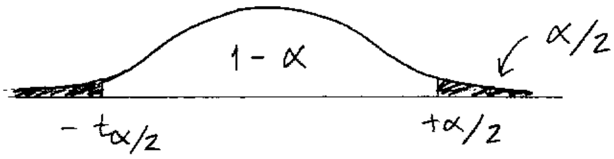
$\mu_D = E[D]$

- Punktskattning: $\hat{\mu}_D = \bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = [\bar{X} - \bar{Y}]$

Uttrycket $\dfrac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$ är Student T-fördelat med $(n-1)$ frihetsgrader. $S_D = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$

- Konf. intervall med konf. grad $1-\alpha$:

Ur $T(n-1)$-tabellen fås



$$\bar{D} \pm t_{\alpha/2}\frac{S_D}{\sqrt{n}}$$

- Hypotesprövning:

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D > 0 \end{cases}$$

Vi väljer signifikansnivå α:

Ur tabellen för $T(n-1)$ fås

Beslutsregel: Om $\dfrac{\bar{D}}{S_D/\sqrt{n}} \geq t_\alpha$ så H_1, annars H_0.

Jämförelse av två varianser:

Två oberoende stickprov:

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)} \qquad\qquad X_1^{(2)}, X_2^{(2)}, \ldots, X_n^{(2)}$$

Ur $N(\mu_1, \sigma_1)$ Ur $N(\mu_2, \sigma_2)$

$\mu_1, \mu_2, \sigma_1, \sigma_2$ är alla okända.

Hypotesprövning: $\begin{cases} H_0 : & \sigma_1 = \sigma_2 \\ H_1 : & \sigma_1 > \sigma_2 \end{cases}$

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(X_i^{(1)} - \bar{X}^{(1)} \right) \qquad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} \left(X_i^{(2)} - \bar{X}^{(2)} \right)$$

$$\frac{s_1^2}{s_2^2} = \frac{s_1^2/\sigma^2}{s_2^2/\sigma^2} = \left[\frac{\dfrac{(\chi^2\text{-fördelad med } n_1-1 \text{ frihetgr.})}{(n_1-1)}}{\dfrac{(\chi^2\text{-fördelad med } n_2-1 \text{ frihetsgr.})}{(n_2-1)}} \right]$$

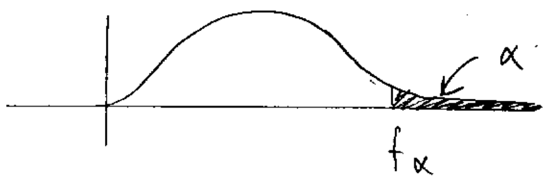
$$\frac{X^2_{\gamma_1}/\gamma_1}{X^2_{\gamma_2}/\gamma_2}$$

där $X^2_{\gamma_1}$ är χ^2-fördelad med γ_1 frihetsgrader, $X^2_{\gamma_2}$ är χ^2-fördelad med γ_2 frihetsgrader. $X^2_{\gamma_1}$ och $X^2_{\gamma_2}$ är oberoende. γ_1 och γ_2 kallas frihetsgrader. Uttrycket säges vara <u>F-fördelat</u>.

$\frac{S_1^2}{S_2^2}$ är F-fördelat med $n_1 - 1$ och $n_2 - 1$ frihetsgrader, om H_0 är sant.

Vi fixerar signifikansnivån α. Ur tabellen $F(n_1 - 1, n_2 - 1)$ får vi f_α.

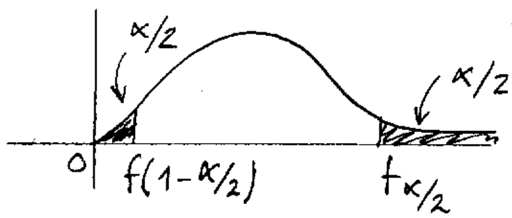
Beslutsregel: $\frac{S_1^2}{S_2^2} \geq f_\alpha$ så H_1 annars H_0.



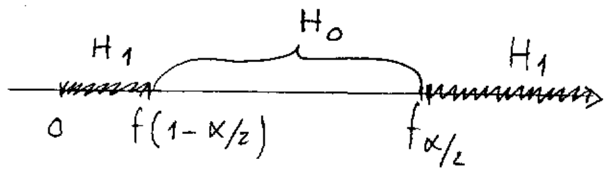
• Tvåsidig alternativhypotes:

$$\begin{cases} H_0: & \sigma_1 = \sigma_2 \\ H_1: & \sigma_1 \neq \sigma_2 \end{cases}$$

signifikansnivå α:



Beslutsregel:

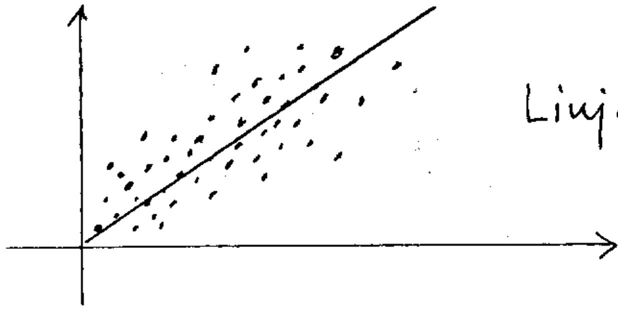


(11) LINJÄR REGRESSION:

Observationer: $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

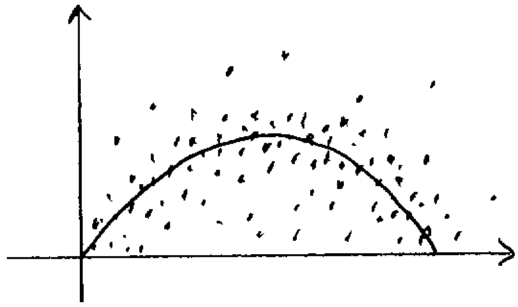
Y - någon kvalitetskarakteristik.

X - någon styrningsparameter vid tillverkning.



Linjär regression!

"Moln" av punkter $(x_1, y_1), (x_2, y_2), \ldots$



Icke-linjär-regression!

<u>Matematisk modell:</u>

X, Y stokastiska vektorer.

$$\mu_{Y|x} = E[Y \mid X = x] \quad - \text{regressionskurva.}$$

Vi antar att kurvan är en rak linje:

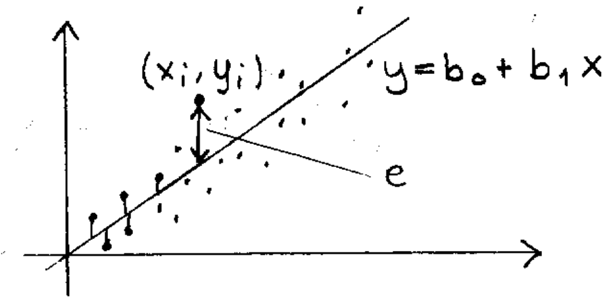
$$\mu_{Y|x} = \beta_0 + \beta_1 x \quad - \text{linjär regression.}$$

$y = \beta_0 + \beta_1 x$ kallas regressionlinje.

- Skattning: $\hat{\mu}_{Y|x} = b_0 + b_1 x$

Vi skall använda <u>minsta kvadratmetoden</u>.

62.



$$Y_i = b_0 + b_1 X_i + e_i$$, e_i – residual, avvikelse.

$$SSE = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

↑
"Sum of squared errors"

$$SSE \xrightarrow[b_0, b_1]{} \min.$$

Lösning:

$$\frac{\partial SSE}{\partial b_0} = \sum_{i=1}^{n} 2(Y_i - b_0 - b_1 X_i)(-1) = 0$$

$$\frac{\partial SSE}{\partial b_1} = \sum_{i=1}^{n} 2(Y_i - b_0 - b_1 X_i)(-X_i) = 0$$

$$\Rightarrow \begin{cases} n b_0 + b_1 \sum X_i = \sum Y_i \\ b_0 \sum X_i + b_1 \sum X_i = \sum X_i Y_i \end{cases}$$

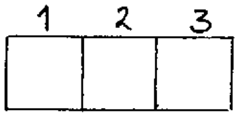
$$\begin{cases} b_1 = \dfrac{n \cdot \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n \cdot \sum X_i^2 - (\sum X_i)^2} \\ b_0 = \overline{Y} - b_1 \overline{X} \end{cases}$$

$y = b_0 + b_1 x$ – den skattade regressionslinjen.

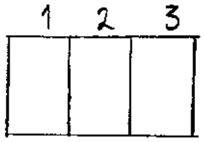
(13) VARIANSANALYS:

Enkel variansanalys:

Ex. Olika typer av gödsel: 1, 2, 3

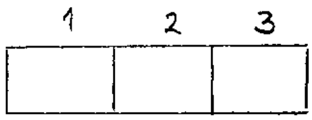


$y_{1,1}$, $y_{1,2}$, $y_{1,3}$



$y_{2,1}$, $y_{2,2}$, $y_{2,3}$

$\vdots$



$y_{n,1}$, $y_{n,2}$, $y_{n,3}$

Beskrivning av modell: $\boxed{Y_{ij} = \mu_j + E_{ij}}$

Väntevärde, beror på behandlingssätt

Totalt slumpmässig avvikelse

$$Y_{ij} = \mu + (\mu_i - \mu) + (Y_{ij} - \mu_i)$$

Totalt medelvärde

Slumpmässig avvikelse.

Avvikelse orsakad av behandlingen.

Har olika behandlingar verkligen olika effekt?

Hypotesprövning:
$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_k \\ H_1: \text{för några } i \text{ och } j,\ \mu_i \neq \mu_j \end{cases}$$

Vi antar att vi har k olika behandlingsmetoder.

Behandling				Σ	
1	Y_{11}	Y_{12}	...	$T_{1\cdot}$	$\bar{Y}_{1\cdot}$
2	Y_{21}	Y_{22}	...	$T_{2\cdot}$	$\bar{Y}_{2\cdot}$
⋮	–	–	–		
k					
Σ	$T_{\cdot 1}$	$T_{\cdot 2}$		$T_{\cdot\cdot}$	
	$\bar{Y}_{\cdot 1}$	$\bar{Y}_{\cdot 2}$			

$$\bar{Y}_{\cdot\cdot} = \frac{T_{\cdot\cdot}}{n}$$

$$\bar{Y}_{\cdot j} = \frac{T_{\cdot j}}{k}$$

$$\bar{Y}_{i\cdot} = \frac{T_{i\cdot}}{n_i}$$

$$SS_{TOT} = \sum_{i,j}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = \underbrace{\sum_{i=1}^{k} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}_{SS_{Tr}\ (\text{Treatment})\ (\text{Mellan behandlingar})} + \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2}_{SS_E\ (\text{Error})\ (\text{Inom behandlingar})}$$

SS_{Tr} innehåller $(k-1)$ termer.
SS_E — " — $(n-k)$ — " — .

F-fördelningen: $\dfrac{\chi^2_{\gamma_1}/\gamma_1}{\chi^2_{\gamma_2}/\gamma_2}$

ANOVA-tabellen:

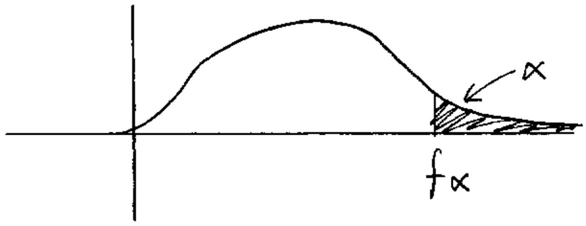
	VARIATIONS-KÄLLA	FRIHETS-GRADER	SS	MS	VÄNTEVÄRDE HOS MS	F
Behandling eller nivå		$k-1$	SS_{Tr}	$\dfrac{SS_{Tr}}{k-1} =$	•	$\dfrac{MS_{tr}}{M_{SE}}$
Fel eller residual.		$n-k$	SS_E	$\dfrac{SS_E}{n-k}$	•	
	Total	$n-1$	SS_{TOT}		•	

$$\frac{MS_{TR}}{M_{SE}}$$

SS - kvadratsummor, MS - medelkvadratsummor

$F = \dfrac{MS_{Tr}}{MS_E}$ är F-fördelad med $k-1$

och $n-k$ frihetsgrader.



Beslutsregel: Om $F \geqslant f_\alpha$, så H_1

annars H_0

(15.) KVALITATIVA DATA (KATEGORIALA DATA):

En egenskap $\mathcal{E}$ med värdena $a_1, a_2, \dots, a_k$.
Låt $p_1, p_2, \dots, p_k$ vara motsvarande sannolikheter.
Vi genomför experimentet n gånger (d.v.s. vi tar stickprov bestående av n individer)

n_1 - antalet individer med a_1.

\- - - - - - - - - - -

n_k - " " " a_k.

Naturligtvis $n_1 + n_2 + \dots + n_k = n$

Sannolikheten att en viss uppsättning $n_1, n_2, \dots, n_k$ förekommer är

$$\frac{n!}{n_1!\, n_2! \cdot \ldots \cdot n_k!}$$

66.

Det är s.k. multinomial fördelning.

Multinomial stokastisk variabel:

$$(X_1, \dots, X_k)$$

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1!\, n_2! \cdot \ldots \cdot n_k!}$$

Hypoteseprövning:

H_0: $p_1 = p_1^{(0)}, p_2 = p_2^{(0)}, \dots, p_n = p_n^{(0)}$

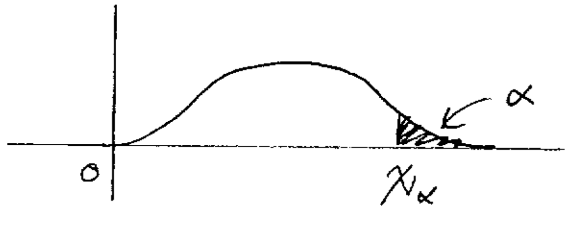
H_1: Det finns sådant i att $p_i \neq p_i^{(0)}$

Grovt: Om $n_i \approx n p_i^{(0)}$ för $i = 1, \dots, k$ så är H_0 troligtvis sant.

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n p_i^{(0)})^2}{n p_i}$$

Om n är stor så uttrycket χ^2 approximativt χ^2-fördelat med $k-1$ frihetsgrader.

Vi fixerar signifikansnivån α. Ur tabellen $\chi(k-1)$ fås



Om $\chi^2 \geq \chi_\alpha$, så H_1

Om $\chi^2 < \chi_\alpha$, så H_0

Annorlunda beteckningar:

n_i – observerad frekvens O_i

$np_i^{(0)}$ – förväntad frekvens E_i (expected)

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Kravet för att approximationen skall kunna användas:

$$E_i \geq 5 \qquad i = 1, 2, \ldots, k$$

Lite svagare krav:

a) Alla $E_i \geq 1$

b) Åtminstone 80 % av E_i satisfierar $E_i \geq 5$.

SLUT

A.N – 99