

Mengdelære

Skrivemåter for mengder:

- listenotasjon: $\{1, 2, 3, 4, 5, 6\}$
 $\{1, 2, \dots\}$
- intervallnotasjon: $[0, 1]$
 $(0, \rightarrow)$

Begreper

- element: $w \in A$ (w element i A)
- delmengde: $A \subseteq B$ (A delmengde av B)
- tomme mengde: $\emptyset$ (mengde uten noen elementer)

Operasjoner

- Snitt: $A \cap B$ (består av elementer i A og B)
- Union: $A \cup B$ (består av elementer i A eller B)
- Komplement: A^c (består av elementer utenfor A)

Noen regneregler:

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

} de Morgans
lover

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

} distributiv
lov

Sum - notasjon:

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

Σ : Sigma, summetegn

x_i : uttrykket som skal summeres

$$\sum_{i=1}^n$$

indeksmengden er $\{1, 2, \dots, n\}$
og indeksvariabel er i , dvs

$$i=1, i=2, i=3, \dots, i=n$$

Skal settes inn i uttrykket x_i
for uttrykkene summeres

Ekse:

$$\sum_{i=1}^4 (2i-1) = \underbrace{(2 \cdot 1 - 1)}_{i=1} + \underbrace{(2 \cdot 2 - 1)}_{i=2} +$$

$$\underbrace{(2 \cdot 3 - 1)}_{i=3} + \underbrace{(2 \cdot 4 - 1)}_{i=4} = 1 + 3 + 5 + 7 = 16$$

Noen regneregler for summer:

$$i) \sum_{i=1}^n (a_i + b_i) = \sum_{i=1}^n a_i + \sum_{i=1}^n b_i$$

$$ii) \sum_{i=1}^n (c \cdot a_i) = c \cdot \sum_{i=1}^n a_i$$

$$iii) \sum_{i=1}^n c = n \cdot c$$

Husk: $\sum_{i=1}^n a_i^2 \neq \left(\sum_{i=1}^n a_i \right)^2$

A Sannsynlighetsregning

Stokastisk forsøk:

Utfallsrom $S = \{\omega_1, \omega_2, \dots\}$ er mengden av mulige utfall ω_i i forsøket

Sannsynlighetsmålet p er en funksjon

Som til hver hendelse A (dvs en delmengde $A \subseteq S$ av utfallsrommet)

tilordner en tallverdi

$$A \mapsto p(A)$$

slik at følgende krav er oppfylt:

a) $p(A) \geq 0$ for alle hendelser A

b) $p(S) = 1$

c) $p(A_1 \cup A_2 \cup \dots) = p(A_1) + p(A_2) + \dots$

for alle parvis disjunkte hendelser

(dvs $A_i \cap A_j = \emptyset$ for $i \neq j$)

Endelige utfallsrom (spesialtilfelle)

$$S = \{\omega_1, \omega_2, \dots, \omega_n\}$$

Da har vi: i) $p(A) = \sum_{\omega \in A} p(\omega)$

ii) Kravene a) - c) reduseres til følgende krav:

a) $p(\omega_i) \geq 0$ for alle ω_i

b) $p(\omega_1) + p(\omega_2) + \dots + p(\omega_n) = 1$

Uniform sannsynlighet:

Hvis alle utfall er like sannsynlige, dvs $p(\omega_1) = p(\omega_2) = \dots = p(\omega_n)$, har vi uniform sannsynlighet og $p(\omega_i) = \frac{1}{n}$.

Da har vi:

$$p(A) = \frac{\text{antall utfall i } A}{\text{antall utfall i } S} = \frac{\text{"antall gunstige"}}{\text{"antall mulige"}}$$

Kombinatorikk:

Kombinatorikk = å tegne
utførelse, eller andre ting

never
underestimate
a result
that counts
something!

Binomialkoeffisienter

$n! = n(n-1) \cdots 2 \cdot 1$ (n faktoriell)

husk:

$$0! = 1! = 1$$

$$\binom{n}{r} = \frac{n \cdot (n-1) \cdots (n-r+1)}{r \cdot (r-1) \cdots 1} = \frac{n!}{r! \cdot (n-r)!}$$

(binomialkoeffisient)

husk: $\binom{n}{0} = 1$ $\binom{n}{1} = n$ $\binom{n}{n-1} = n$ $\binom{n}{n} = 1$
 $\binom{n}{i} = \binom{n}{n-i}$

Pascals trekant:

		1			
		1	1		
	1	2	1		
	1	3	3	1	
	1	4	6	4	1
1	5	10	10	5	1

$n=0$
 $n=1$
 $n=2$
 $n=3$
 $n=4$
 $n=5$

Hver rad
består av
binomialkoeff.
 $\binom{n}{i}$ for

$i=0, 1, 2, \dots, n$

Ex:

$$\binom{4}{2} = 6$$

Newtons binomialformel

$$(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^{n-i} b^i$$

$$= a^n + n \cdot a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots \\ \dots + n a b^{n-1} + b^n$$

Ex: $(a+b)^4 = a^4 + 4a^3b + \binom{4}{2}a^2b^2 + 4ab^3 + b^4$
 $= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$

Urmodell

En urne inneholder n kuler, nummerert fra 1 til n , og vi trekker r kuler. Antall mulige måter å gjøre dette på er:

	med tilbakelegging	uten tilbakelegging
ordnet	n^r	$\frac{n!}{(n-r)!}$
uordnet	$\binom{n+r-1}{r}$	$\binom{n}{r}$

Generelle sannsynlighetsmå

Ikke alle sannsynlighetsmå er urifenne;
det er ikke alltid nok å telle opp antall utfall.

Følgende gjelder for alle sannsynlighetsmå:

i) $P(A^c) = 1 - P(A)$

ii) $P(\emptyset) = 0$

iii) $P(A) \leq P(B)$ når $A \subseteq B$

iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Tolkning av $P(A)$:

Vi tolker $P(A)$ som frekvensen (hvor ofte A vil forekomme) dersom vi kunne gjentatt forsøket uenige ganger. Vi kaller $P(A)$ for sannsynligheten til A .

Betinget sannsynlighet

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

Sannsynligheten
for A, gitt at
B forekommer

Bayes lov

Anta at $B_1, B_2, \dots, B_r$ er pervis disjunkte
slik at $B_1 \cup B_2 \cup \dots \cup B_r = S$. Da har vi:

$$p(B_i|A) = \frac{p(A \cap B_i)}{p(A)} = \frac{p(A|B_i) \cdot p(B_i)}{\sum_{j=1}^r p(A|B_j) \cdot p(B_j)}$$

Uavhengighet:

A og B er uavhengige hendelser hvis

$$p(A \cap B) = p(A) \cdot p(B). \text{ Dette er det samme}$$

$$\text{Som at: } p(A|B) = p(A)$$

Stokastiske variable

En stokastisk variabel X tar verdier $X(\omega)$ som avhenger av utfallet ω i et stokastisk forsøk.

Utfallsrom:

$$S = \{\omega_1, \omega_2, \dots\}$$

Mulige verdier for X :

$$X(\omega_1), X(\omega_2), \dots$$

To hovedtilfeller:

i) Diskret stokastisk variabel:

$X(S) = \{X(\omega_1), X(\omega_2), \dots\}$ er diskret mengde

Typiske mengder: $\{1, 2, 3, 4, 5, 6\}$ (endelig)

$\{1, 2, 3, \dots\}$ (uendelig diskret)

ii) Kontinuerlig stokastisk variabel

$X(S)$ er kontinuerlig mengde

Typiske mengder: $[0, \infty)$ } intervall
 $[0, 10]$

Regning med diskrete stokastiske variable

$$X(S) = \{X(\omega) : \omega \in S\} \quad \text{mulige verdier for } X$$

$$p(x) = p(X=x) : \text{sannsynligheten for at } X \text{ har verdien } x$$

$$X=x \text{ definerer en hendelse} \\ \{\omega : X(\omega)=x\} \subseteq S$$

Spesifikasjon av X :

x	$p(x)$
x_1	$p(x_1)$
x_2	$p(x_2)$
x_3	$p(x_3)$
$\vdots$	$\vdots$

Angir alle mulige verdier x i $X(S)$ og deres sannsynligheter.

Krav:

- i) $p(x_i) \geq 0$ for alle x_i
- ii) $\sum_{x_i} p(x_i) = 1$

Kumulativ fordelingsfunksjon:

$$F(x) = p(X \leq x) \\ = \sum_{x_i \leq x} p(x_i)$$

kalles noen ganger CDF
(cumulative distribution fn.)

Forventningsverdi (diskret tilfelle)

$$E(X) = \sum_{x \in X(s)} x \cdot p(x) = x_1 \cdot p(x_1) + x_2 p(x_2) + \dots$$

forventningsverdien til X

$$E[g(X)] = \sum_{x \in X(s)} g(x) \cdot p(x) = g(x_1) p(x_1) + g(x_2) p(x_2) + \dots$$

forventningsverdien til $g(X)$, en
funksjon av X (for eksempel $g(x) = x^2$)

Skrivemåte:

$$\mu = E(X) \quad (\mu = \text{mu})$$

Tolkning:

$E(X)$ er gjennomsnittsverdien til X om
det statistiske forsøket gjentas
mange ganger

Regneregler for forventningsværdi

i) $E(ax+b) = a \cdot E(x) + b$ (a, b konst.)

ii) $E(g(x)+h(x)) = E(g(x)) + E(h(x))$

Husk: $E(x^2) \neq E(x)^2$

Varians:

$$\text{Var}(x) = E[(x-\mu)^2] \quad \text{hvor } \mu = E(x)$$

variansen til x

Vi har: $\boxed{\text{Var}(x) \geq 0}$, og $\text{Var}(x) = 0$ kun hvis x er en konstant

Standard-deviation:

Skruenåle: σ ($\sigma = \text{sigma}$)

$\sigma = \sqrt{\text{Var}(x)}$, dvs

$\boxed{\text{Var}(x) = \sigma^2}$

Regneregler for varians:

i) $\text{Var}(aX + b) = a^2 \text{Var}(X)$

ii) $\text{Var}(X) = E(X^2) - E(X)^2$

(a, b konst.)

Noen sentrale diskrete fordelinger

Binomisk fordeling

En binomisk forsøksrekke karakteriseres ved at vi gjentar et stokastisk forsøk n ganger, i hver gjentakelse er sannsynligheten $p = p(A)$ for en bestemt hendelse A den samme, og forsøkene er uavhengige. Da kalles

$X =$ antall ganger A forekommer

en binomisk fordelt stokastisk variabel med parametre n og p .

$$X(S) = \{0, 1, 2, \dots, n\}$$

$$p(X=i) = \binom{n}{i} p^i q^{n-i} \quad \text{med } q = 1-p$$

$$E(X) = np$$

$$\text{Var}(X) = npq$$

Poisson fordeling

$$\lambda = \text{lambd}$$

En stokastisk variabel X kaldes
Poisson-fordelt med parameter λ
dersom

$X(S) = \{0, 1, 2, \dots\}$ er de mulige
og verdier til X

$$p(X=i) = e^{-\lambda} \cdot \frac{\lambda^i}{i!} \quad \text{for } i = 0, 1, 2, \dots$$

Da har vi:

$$\begin{cases} E(X) = \lambda \\ \text{Var}(X) = \lambda \end{cases}$$

Binomisk tilnærmelse:

Dersom X er binomisk
fordelt, med n stor og p
liten, er X tilnærmelse
Poisson-fordelt med $\lambda = np$.

$$\left. \begin{array}{l} n \text{ stor} \\ p \text{ liten} \end{array} \right\} \left\{ \begin{array}{l} n > 10 \\ p < 0.10 \end{array} \right.$$

Typisk anvendelse:

X = antall
forekomster
av en hendelse
per tidsintervall,

λ = forventet antall
forekomster
per tidsintervall

Regning med kontinuerlige variable

Spesifikasjon av X :

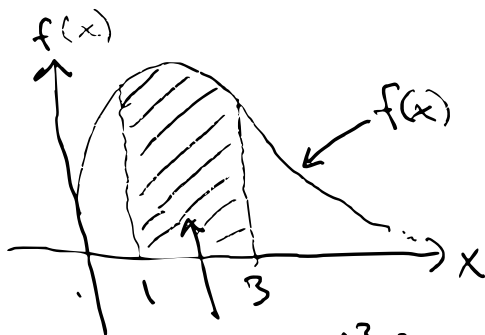
En kontinuerlig tetthetsfunksjon $f(x)$
(som er null utenfor $X(S)$) slik at

$$p(a \leq X \leq b) = \int_a^b f(x) dx$$

Krao t.l. tetthetsfunksjoner:

i) $f(x) \geq 0$ for alle x

ii) $\int_{X(S)} f(x) dx = 1$



Eksempel der
 $X(S) = [0, \infty)$

$$\text{Areal} = \int_1^3 f(x) dx = p(1 \leq X \leq 3)$$

Kumulativ fordelingsfunksjon:

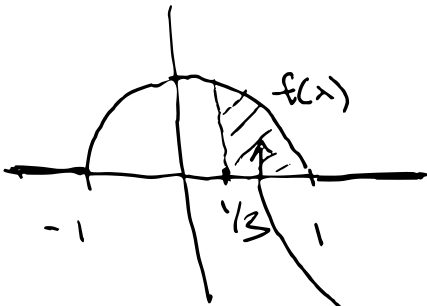
$$F(x) = p(X \leq x) = \int_{-\infty}^x f(z) dz$$

kalles
også CDF

Ekso: X kontinuerlig stokastisk variabel med nullte verdier $X(S) = [-1, 1]$

og

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2), & -1 \leq x \leq 1 \\ 0, & \text{ellers} \end{cases}$$



$$P(X \geq 1/3)$$

$$= \int_{1/3}^1 \frac{3}{4}(1-x^2) dx$$

$$= \frac{3}{4} \left[x - \frac{x^3}{3} \right]_{1/3}^1$$

$$= \frac{3}{4} \left[\left(1 - \frac{1}{3}\right) - \left(\frac{1}{3} - \frac{1}{81}\right) \right]$$

$$= \frac{3}{4} \left(\frac{1}{3} + \frac{1}{81} \right) = \frac{3}{4} \cdot \frac{27+1}{81} = \frac{7}{27} \approx 0.26$$

Stille kravene:

$f(x) \geq 0$ i $[-1, 1]$: ok.

$$\int_{-1}^1 \frac{3}{4}(1-x^2) dx$$

$$= \frac{3}{4} \left[x - \frac{x^3}{3} \right]_{-1}^1$$

$$= \frac{3}{4} \left(\left(1 - \frac{1}{3}\right) - \left(-1 + \frac{1}{3}\right) \right)$$

$$= \frac{3}{4} \left(2 - \frac{2}{3} \right)$$

$$= \frac{3}{4} \cdot \frac{4}{3} = 1 \quad \underline{\text{ok.}}$$

Forventningsverdi (kontinuerlig tilfælde)

$$E(x) = \int_{x(s)} x \cdot f(x) dx \quad (\text{forventningsverdi for } X)$$

$$E(g(x)) = \int_{x(s)} g(x) f(x) dx \quad (\text{forventningsverdi for } g(x), \text{ en funktion af } x)$$

Skriveemåle: $\mu = E(x)$ ($\mu = m_\mu$)

Tolkning: $E(x)$ er gennemsnitsverdi til X om forsøket gentages mange gange

Regneregler for forventningsverdi:

i) $E(ax+b) = aE(x) + b$ (a, b konst.)

ii) $E(g(x) + h(x)) = E(g(x)) + E(h(x))$

Husk: $E(x^2) \neq E(x)^2$

Varians:

$$\text{Var}(x) = E[(x - \mu)^2]$$

variansen til x

hvor $\mu = E(x)$

Vi har: $\boxed{\text{Var}(x) \geq 0}$

og $\text{Var}(x) = 0$
kun hvis
 x er konstant

Standardafvik:

Stokvariabel: σ ($\sigma = \text{sigma}$)

$$\sigma = \sqrt{\text{Var}(x)} \quad \text{dvs} \quad \boxed{\text{Var}(x) = \sigma^2}$$

Regneregler for varians:

i) $\text{Var}(ax+b) = a^2 \text{Var}(x)$

(a, b konst.)

ii) $\text{Var}(x) = E(x^2) - E(x)^2$

Noen sentrale kontinuerlige fordelinger

Normalfordeling

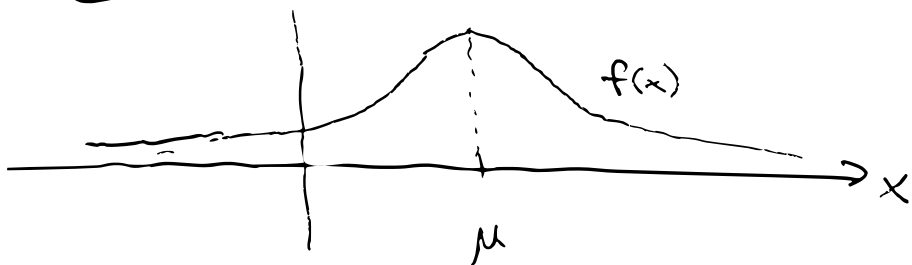
La μ, σ være gitte tall, med $\sigma > 0$.

En stokastisk variabel X kalles normalfordelt med parametre μ, σ dersom

$$X(S) = (-\infty, \infty) = \mathbb{R}$$

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

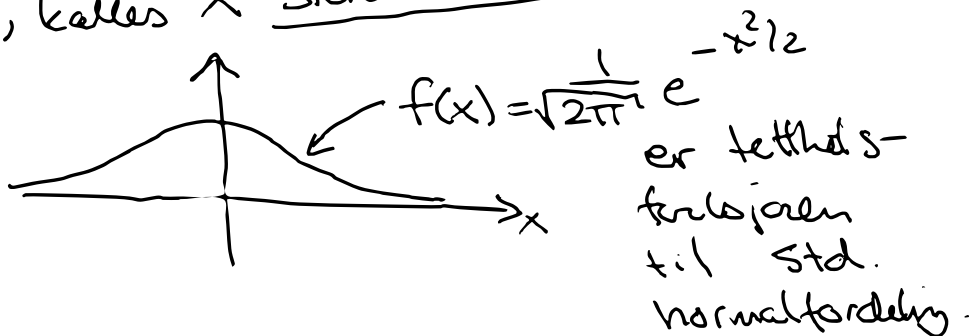
Skrivemåte: $X \sim N(\mu, \sigma)$ betyr at X er normalfordelt med $E(X) = \mu$, $\text{Var}(X) = \sigma^2$



Tetthetskurven er symmetrisk om $x = \mu$ og har en "klokkeform". Grafen kalles en Gauss-kurve.

Standard-normalfordeling:

Når X er normalfordelt med parametre $\mu=0$, $\sigma=1$, kaldes X standard normalfordelt.



Förventning og varians:

Når $X \sim N(\mu, \sigma)$ er normalfordelt med parametre μ, σ , har vi:

$$\begin{aligned} E(X) &= \mu \\ \text{Var}(X) &= \sigma^2 \end{aligned}$$

Viktig egenskab:

Dersom X er normalfordelt, er også $aX + b$ normalfordelt når a, b er konstanter

Regning med normalfordeling

Det er uensartet at regne ut sannsynligheter direkte (ved integrasjon).

Standardisering:

Hvis X er stokastisk variabel med $E(X) = \mu$, $\text{Var}(X) = \sigma^2$, så kalles

$$Z = \frac{X - \mu}{\sigma}$$

standardiseringen av X . Det er en

Stokastisk variabel med $E(Z) = 0$, $\text{Var}(Z) = 1$.

Hvis X er normalfordelt, så er også Z normal fordelt, dvs standard normalfordelt.

Tolkning av Z : antall std. avvik fra μ

Regning med X , når X er normalfordelt:

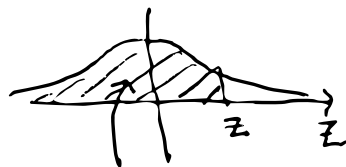
$$P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = G\left(\frac{x - \mu}{\sigma}\right)$$

hvor G er kumulativ fordelingsfunksjon for std. normalfordeling.

$G(z)$ = kumulativ
fordelingsfunksjon
for std. normalfordeling

$$= P(Z \leq z)$$

med $Z \sim N(0,1)$



$$G(z) = P(Z \leq z)$$

Kalk:

$$Z \Leftrightarrow P$$

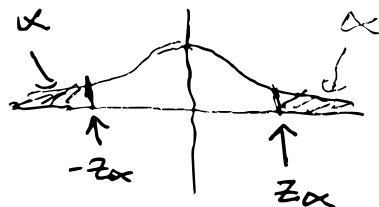
Kvantiler:

Før $0 < \alpha < 1$ definer vi

$z_\alpha = \alpha$ -kvantilet til

Standard-normalfordelingen

ved



$$P(Z > z_\alpha) = \alpha$$

Merke: arealet $= \alpha$ er på høyresiden

av z_α pga $Z > z_\alpha$ mens $-z_\alpha$

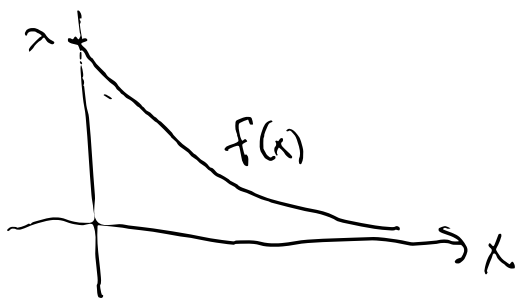
har areal $= \alpha$ på venstresiden
pga symmetri

Ekspontentialfordeling:

Vi kaller X ekspontentialfordelt med parameter $\lambda > 0$ dersom $X(s) = [0, \infty)$

og

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , \text{ellers} \end{cases}$$



Typisk anvendelse:

X = tid til første ankomst (ventetid) i en Poisson-proces, med forventet antall forekomster per tidsenhet lik λ .

Forventning og varians:

$$E(X) = 1/\lambda \quad \text{Var}(X) = 1/\lambda^2$$

Simultant fordelte stokastiske variabler

To variabler X og Y , som begge tar verdier som avhenger av samme stokastiske prosess, kalles simultant fordelte.

Diskret tilfelle

For hvert tallpar (x, y) av mulige verdier, dvs verdier slik at $x = X(\omega)$ og $y = Y(\omega)$ for et utfall ω , skriver vi:

$p(x, y) = p(X=x, Y=y)$: Sannsynligheten for at $X=x$ og $Y=y$

Krav til p :

- i) $p(x, y) \geq 0$ for alle x, y
- ii) $\sum_{x, y} p(x, y) = 1$

Kumulativ fordelingsfunksjon: $F(x, y) = p(X \leq x, Y \leq y)$

Marginale fordelinger:

$$p_X(x) = \sum_y p(x,y)$$

$$p_Y(y) = \sum_x p(x,y)$$

	y_1	y_2	...
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	...
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	...
$\vdots$	$\vdots$	$\vdots$	

Fra sannsynlighetstabellen finner vi:

$$p_X(x) = \text{radsum} \quad (\text{raden for } x)$$

$$p_Y(y) = \text{kolonnesum} \quad (\text{kolonnen for } y)$$

Forventningsverdier:

$$E[g(x,y)] = \sum_{x,y} g(x,y) \cdot p(x,y)$$

$$E[g(x)] = \sum_x g(x) \cdot p_X(x)$$

$$E[g(y)] = \sum_y g(y) \cdot p_Y(y)$$

Ex:

$$E(X) = \sum_x x \cdot p_X(x)$$

$$E(XY) = \sum_{x,y} xy \cdot p(x,y)$$

Kovarians:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$\text{hvor } \mu_X = E(X), \mu_Y = E(Y)$$

(kovarians til X og Y)

Korrelationskoeffisient:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

$$\text{hvor } \begin{cases} \sigma_X = \sqrt{\text{Var}(X)} \\ \sigma_Y = \sqrt{\text{Var}(Y)} \end{cases}$$

(korrelationskoeffisient
til X og Y , skrives
også ρ_{XY})

(std. dev. for X og Y)

Tolkning:

$$\text{Cov}(X, Y), \rho > 0$$

: { større verdier av X har
en tendens til å si større
verdier av Y

$$\text{Cov}(X, Y), \rho < 0$$

: { større verdier av X har en
tendens til å si mindre verdier
av Y (og omvendt)

Regneregler for forventningsverdier:

$$i) E(X+Y) = E(X) + E(Y)$$

$$ii) \text{Cov}(Y, X) = \text{Cov}(X, Y)$$

$$\text{Cov}(X, X) = \text{Var}(X)$$

$$\text{Cov}(aX+bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$iii) \text{Var}(aX+bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$$

Resultater for korrelationskoefficienten:

$$i) \rho_{XY} \text{ er ubestemt (bestemt)}$$

$$ii) -1 \leq \rho_{XY} \leq 1 \text{ for alle } X, Y$$

Dessuden har vi:

$$\rho_{XY} = 1 \iff Y = aX + b \text{ med } a > 0$$

$$\rho_{XY} = -1 \iff Y = aX + b \text{ med } a < 0$$

(a, b er konstanter)

Uafhængighed:

X og Y er uafhængige dersom

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

for alle x, y .

Resultat: $\{ X, Y \text{ uafhængige} \Rightarrow \text{Cor}(X, Y) = 0$

Husk: implikationen går kun not højre

Konsekvens:

$X_1, X_2, \dots, X_n$ uafhængige stokastiske variable

$\Downarrow$

$$\text{Var}(a_1 X_1 + a_2 X_2 + \dots + a_n X_n)$$

$$= a_1^2 \cdot \text{Var}(X_1) + a_2^2 \cdot \text{Var}(X_2) + \dots + a_n^2 \cdot \text{Var}(X_n)$$

(alle kovarianser $\text{Cor}(X_i, X_j) = 0$ når $i \neq j$)

Kontinuerlig tilfelle

For å regne ut sannsynligheter og forventningsverdier, må vi bruke en kontinuerlig simultan tetthetsfunksjon $f(x,y)$ og regne med dobbelteintegraler. Vi går ikke inn på hvordan dette gjøres i kurset.

Når vi regner med forventning, varians og kovarians, gjelder samme regneregler som i det diskrete tilfellet. Dette bruker vi i kurset.

Exo: La X, Y være normalfordelte stokastiske variabler med $\text{Cor}(X, Y) = 1$ så er:

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) = \mu_X + \mu_Y \\ \text{Var}(X+Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cor}(X, Y) \\ &= \sigma_X^2 + \sigma_Y^2 + 2 \end{aligned}$$

når $X \sim N(\mu_X, \sigma_X)$, $Y \sim N(\mu_Y, \sigma_Y)$.

Noen viktige resultater:

i) Chebyshev's ulikhet:

For enhver stokastisk variabel X med $E(X) = \mu$, $\text{Var}(X) = \sigma^2$ gjelder

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2 \quad \text{for } k > 0$$

Eksp: $k=2$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \geq 3/4 = 75\%$$

(innenfor ± 2 Std. avvik) ($1 - 1/k^2$)

ii) Store talls lov:

$X_1, X_2, \dots$ uavhengige stokastiske variable
med $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$ for alle i

$\Downarrow$

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \mu \quad \text{når } n \rightarrow \infty$$

Merk: $E(\bar{X}_n) = \mu$
 $\text{Var}(\bar{X}_n) = \sigma^2/n$

Varians / std. avvik
blir mindre og
mindre når $n \rightarrow \infty$

iii) Sentralgrenseteoremet

$x_1, x_2, \dots$ uavhengige, identiske fordelte
stokastiske variable med
 $E(x_i) = \mu$ og $Var(x_i) = \sigma^2$ for
alle i

$\Downarrow$

$$\bar{X}_n = \frac{x_1 + \dots + x_n}{n} \rightarrow N(\mu, \sigma/\sqrt{n})$$

(normalfordeling)

for $n \rightarrow \infty$

Merke:

- i) Sentralgrenseteoremet gir ofte god tilnærming for $n \geq 30$
- ii) Hvis $x_1, x_2, \dots$ er normalfordelte, er $\bar{X}_n$ eksakt normalfordelt for alle n siden vi har

$x_1, x_2, \dots, x_n$ normalfordelt $\Rightarrow a_1 x_1 + a_2 x_2 + \dots + a_n x_n + b$
eksakt normalfordelt
($a_1, a_2, \dots, a_n, b$ konst.)

iii) Hvis X er binomisk fordelt med parametre (n, p) så er X tilnærmet normalfordelt $N(\mu, \sigma)$ med $\mu = np$ og $\sigma^2 = npq = np(1-p)$ dersom $npq \geq 5$.

Skal vi regne ut kumulative sannsynligheter for en binomisk variabel uha normaltilnærming, blir tilnærmingen mye bedre om vi bruker

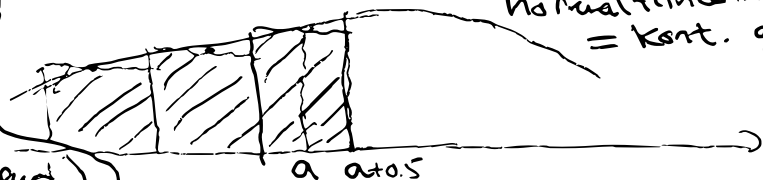
heltalls korrigering:

Ek:
$$p(X \leq a) \approx p(Y \leq a + 0.5)$$
$$\approx p\left(Z \leq \frac{a + 0.5 - \mu}{\sigma}\right) = \Phi\left(\frac{a + 0.5 - \mu}{\sigma}\right)$$

hvor X binomisk fordelt (n, p)
 Y normal fordelt med $\mu = np, \sigma^2 = npq$
 Z std. normal fordelt.

binomisk
fordelt $\Rightarrow$
tilnærmet
normal -
fordelt

(sentral-
grenseteorem)



normaltilnærming
= kont. grad

Noen flere fordelinger:

χ^2 -fordelingen:

($\chi = \text{chi} / \text{khi}$)
greek X

Hvis $Z_1, Z_2, \dots, Z_k$ er uavhengige, std. normalfordelte variabler, så kalles

$$Y = Z_1^2 + Z_2^2 + \dots + Z_k^2$$

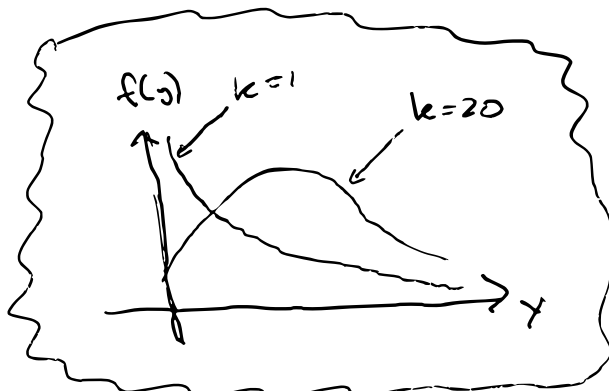
en χ^2 -fordelt stokastisk variabel med k frihetsgrader. Vi skriver $Y \sim \chi_k^2$.

Resultat:

i) $Y \geq 0$

ii) $E(Y) = k$

iii) $\text{Var}(Y) = 2k$



tetthets funksjon

Hovedresultat:

$X_1, X_2, \dots, X_n$ uavhengige, normalfordelte $N(\mu, \sigma^2)$

$\Downarrow$

$Y = \frac{(n-1)S^2}{\sigma^2}$ er χ_{n-1}^2 -fordelt, der $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2$

Student t-fordeling (= t-fordeling)

Hvis Z er std. normalfordelt og U er χ^2_{n-1} -fordelt, så kaldes

$$T = \frac{Z}{\sqrt{\frac{1}{n-1}U}} \quad \begin{array}{l} \text{t-fordelt med } n-1 \\ \text{frihedsgrader} \end{array}$$

Merk: T-fordelingen nærmer sig std. normal-fordeling når $n \rightarrow \infty$.

Viktig spæialtilfælle:

Anta $x_1, \dots, x_n$ uafhængige normalfordelte $N(\mu, \sigma)$.

Da er

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad \begin{array}{l} \text{t-fordelt med } n-1 \text{ df} \\ (\text{df} = \text{frihedsgrader}) \end{array}$$

$$\left(\text{sidn } T = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{S/\sigma} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1} \cdot \frac{(n-1)S^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{1}{n-1}U}} \right)$$

Siden $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ er std. normalfordelt
 mens $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ er t-fordelt med $n-1$ df.

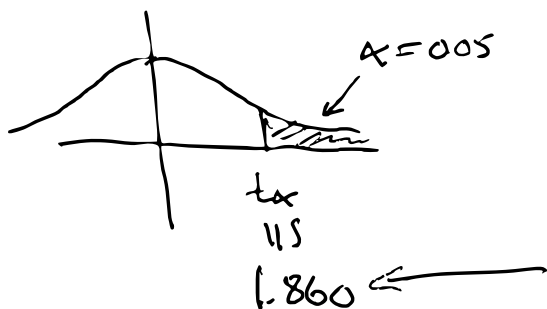
trenger vi ofte t-fordeling i stedet for normalfordeling når σ er ukendt og må estimeres ved hjælp af S .

Regning med t-fordeling:

Ex: T t-fordelt med $df=8$
 $P(T \leq 1) \approx 0.827 \leftarrow$

Kalk:
 $8 \text{ [df, t} \geq P \text{]}$
 $= \rightarrow 0.827$

Kvartil t_{α} når $\alpha=0.05$
 $df=8$,
 dvs $P(T \leq t_{\alpha}) = 1 - \alpha = 0.95$



Kalk:
 $8 \text{ [INV] [df, 1 - P]}$
 $0.95 \text{ [=]} \rightarrow 1.860$