Mat. Stat. 31/3 Måndag LV1

TMA 321
Erik Broman
broman@math.chalmers.se

Sannolikhetsteori — "Regler för slumpens beteende"

Matematisk statistik — "Se mönster i data och dra slutsatser av dessa"

1) räkneregler
2) diskreta slumpvariabler
3) kontinuerliga SV.
4) fördelningar
5) gränsvärdessatser
6) uppskatta värden på en modell | ingående parameter.
7) Hypotesprövning (*)
8) Enkel linjär regression.

Ex: Typisk statistikfråga: Reduceras risken för hjärtattack av aspirin?
Studie: 11034 fick placebo, varav 189 hjärtattack.
11037 fick aspirin, varav 104 — " —.

(*) Hypotes: Nej, aspirin har ingen effekt.
Under vår hypotes: (kan vi se experimentet som):
Vi har 11034 svarta bollar i en urna } samma urna
" 11037 vita — " —
Vi drar 293 st bollar ur urnan, vad är sannolikheten att minst 189 är svarta? Om svaret blir "jätte-liten"

Räkne-regler Def: Utfallsrum (sample space)
Betecknas Ω := mängden av alla utfall i ett experiment

Def: Händelse $A \subset \Omega$

Ex: Tärningskast: $\Omega = \{1,2,3,4,5,6\}$
$A = \{1\}$ $A = \{2,4,6\}$ (jämnt antal)

Ex: Tiden till vi kommer fram till kassan på Willy's
(tid är kontin.)
$\Omega = \mathbb{R}$ $\Omega = [0,T]$ T = tid fram till stängning
$A = \{t\}$ $A = [t_1, t_2]$ $0 < t_1 < t_2 < T$

Ex: 36 personer, betrakta deras födelsedagar:

$\Omega = \{(i_1, i_2, \ldots, i_{36})\}$

$1 \leq i_1, i_2, \ldots, i_{36} \leq 365$

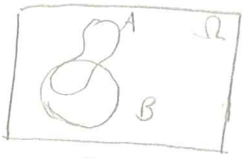
$A = \{$minst två har födelsedag samtidigt$\}$

Mängdlära: Venn-diagram:

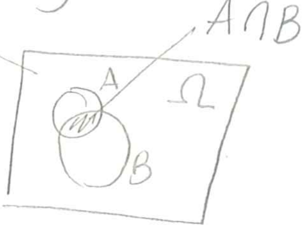
$A, B \subset \Omega$

i) $A \cap B :=$ $\{x \in \Omega :$ $x \in A$ och $x \in B\}$

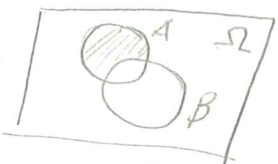
snitt — def. som

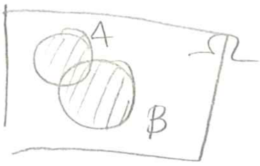


ii) $A \setminus B := \{x \in \Omega : x \in A$ och $x \notin B\}$



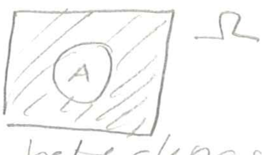
iii) $A \cup B := \{x \in \Omega :$ $x \in A$ och/eller $x \in B\}$





iv) Komplementet till A betecknas A^c. ($\overline{A}$) alt. bet.

$A^c := \{x \in \Omega \quad x \notin A\}$



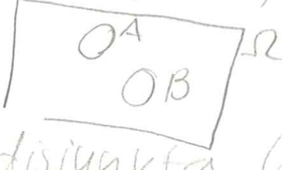
Def: Sannolikheten för $A \subset \Omega$ betecknas $\mathbb{P}(A)$ och är ett tal mellan 0 och 1.

$\mathbb{P}$ kallas ofta för ett sannolikhetsmått och uppfyller:

i) $\mathbb{P}(\Omega) = 1$

ii) $A \subset \Omega$ $\mathbb{P}(A) \geq 0$

iii) Om $A \cap B = \emptyset$ (disjunkta, särskilda) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$



t.ex: tärning: få en etta och få ett jämnt tal = disjunkta

Def: A, B är disjunkta (disjoint, mutually exclusive) om $A \cap B = \emptyset$

Några räkneregler

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

ii) $P(A) + P(A^c) = 1 \Rightarrow P(A) = 1 - P(A^c)$

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

Divisions-regeln: Om i ett experiment alla utfall är lika sannolika:

$$P(A) = \frac{\#\text{gynnsamma utfall}}{\#\text{utfall}} = \frac{\#A}{\#\Omega}$$

Ex: 5 familjer

P P P P P

F F F P P

$A = \{$den valda familjen har två pojkar$\}$

i) välj en familj slumpmässigt.

$$P(A) = \frac{2}{5}$$

ii) Välj en pojke och betrakta hans familj.

$$P(A) = \frac{4}{7}$$

Kombinatorik Ett experiment består av k steg.

$N_l = \#$utfall i steg l

$\#\Omega = N_1 \times N_2 \times \dots \times N_k$

Ex: Dra tre kort ur en kortlek:

$$\#\Omega = 52 \times 51 \times 50$$

Def en permutation av objekt = en ordning av objekten

Ex: 123 321 132 är tre permutationer av $\{1, 2, 3\}$

Vad blir antalet permutationer av n objekt som välj s ur N objekt? Svar: $N \times (N-1) \times \dots \times (N-n+1) = \frac{N!}{(N-n)!}$

"Bevis": 1 2 3 4 . . . N

1:a 3 — N st. sätt

2:a 17 — (N-1) st sätt.

Ex: 52 löpare, hur många sätt kan de första 5 bestämmas? $\frac{52!}{47!}$

Vad blir antalet kombinationer av n objekt som väljs ur N.

Svar: $\binom{N}{n} = \frac{N!}{(N-n)!\,n!}$

Talen $\binom{N}{k}$ kallas för Binomialkoefficienter:

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \dots + \binom{n}{k}a^{n-k}b^k + \dots + b^n$$

$$(a+b)^n = \underbrace{(a+b)(a+b)\cdots(a+b)}_{n \text{ st.}}$$

Ex: 36 personer $A = \{$minst två av 36 har samma födelsedag$\}$

$\# \Omega = 365^{36}$

$\# A^c = 365 \times 364 \times \dots \times 330$

$\mathbb{P}(A^c) = \dots \approx 0{,}17 \Rightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c) \approx 0{,}83$

Multinomial-koefficienter

sätt att dela in n objekt i k grupper av storlek $n_1, n_2, \dots, n_k$ med $n_1 + n_2 + \dots + n_k = n$

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!}$$

"B" # sätt att välja ut n_1 ur n (till grupp 1) $\binom{n}{n_1}$, när detta är gjort återstår $n-n_1$ objekt.

Vi skall då välja ut n_2 objekt ur $n-n_1$

sätt: $\binom{n-n_1}{n_2}$.

Vi får $\binom{n}{n_1, n_2 \ldots, n_k} = \binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \binom{n-n_1-n_2}{n_3} \times \ldots \times \binom{n-n_1-\ldots-n_{k-1}}{n_k}$

$$(\#) = \frac{n!}{n_1!\cancel{(n-n_1)!}} \times \frac{\cancel{(n-n_1)!}}{n_2!\cancel{(n-n_1-n_2)!}} \times \ldots \times \frac{\cancel{(n-n_1-n_2-\ldots-n_{k-1})!}}{n_k!\underbrace{(n-\underbrace{n_1-\ldots-n_k}_{=n})!}_{=0!=1}}$$

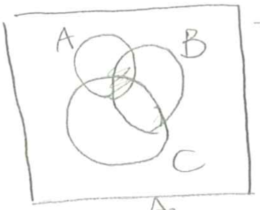
$$= \frac{n!}{n_1!\ldots n_k!} \quad \square$$

Räknestuga LV1

1.8

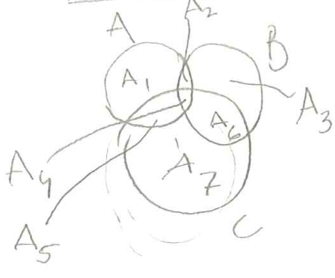
2) Visa $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

a) Motivera mha Venn-diagram



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

b)



$A_1 := A \setminus \{B \cup C\}$

$A_2 := A \cap B \setminus C$

$A_3 := B \setminus \{A \cup C\}$

$A_4 := A \cap B \cap C$

$A_5 := A \cap C \setminus B$

$A_6 := B \cap C \setminus A$

$A_7 := C \setminus \{A \cup B\}$

(sannolikhetsmått är additivt)

$$P(A) + P(B) + P(C) = P(A_1) + \cancel{P(A_2)} + P(A_4) + P(A_5) + P(A_2) + \cancel{P(A_3)} + \cancel{P(A_4)} + P(A_6) + P(A_4) + \cancel{P(A_5)} + \cancel{P(A_6)} + P(A_7)$$

$$= \sum_{i=1}^{7} P(A_i) + (P(A_2) + P(A_4)) + (P(A_4) + P(A_5)) + (P(A_6) + P(A_4)) - P(A_4) = P(A \cup B \cup C) + P(A \cap B) + P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)$$

13. Vad är sannolikheten att en pokerhand (utan byten) är en stege? $= A$

Hur många pokerhänder finns det? $\binom{52}{5}$ totalt antal möjliga pokerhänder.

Hur många finns det som är stege?

$$10 \text{ st.} \begin{cases} A2345 & 4^5-4 = \# \text{ stegar. A till 5 ej alla i samma färg.} \\ 23456 & 4^5-4 \\ \vdots & \\ TJQKA & 4^5-4 \end{cases}$$

totala # stegar $= 10 \cdot (4^5 - 4)$

$$P(A) = \frac{10 \cdot (4^5-4)}{\binom{52}{5}}$$

b) Vad är sannolikheten för $B = \{\text{vi får ett 4-tal}\}$?

AAAA * ← femte kortet — 48 st händer har 4A

2222 * — 48 st — '' — 4 2

⋮

KKKK* — '' '' '' 4 K

$$P(B) = \frac{13 \cdot 48}{\binom{52}{5}}$$

c) kåk: 3 kort samma valör + 2 kort i samma valör.

$\binom{4}{3}$ $\binom{4}{2}$

$$12 \begin{cases} AAA22 & 222AA & & KKKAA \\ AAA33 & 22233 & \dots & KKK22 \\ \vdots & \vdots & & \vdots \\ AAAKK & 222KK & & KKKQQ \end{cases}$$

13 st.

Varje kåk i listorna finns i $\binom{4}{3} \cdot \binom{4}{2}$ antal färg-

$C = \{\text{vi har en kåk}\}$

$$P(C) = \frac{12 \cdot 13 \cdot 4 \cdot 6}{\binom{52}{5}}$$

26. n objekt varav k är defekta, vi väljer m st.
A = {alla m fungerar}

Vad är P(A)?

$$P(A) = \frac{n-k}{n} \cdot \frac{n-1-k}{n-1} \cdot \frac{n-2-k}{n-2} \cdot \ldots \cdot \frac{n-m+1-k}{n-m+1}$$

Sannolikhet första ok.

a) Givet n = 1000 k = 10 vad ska m vara s.a

$P(A) \leq 0.1$?

---

21) Vi singlar slant 5 ggr
A = {vi får en sekvens av 3 H}

$$P(A) = \frac{2^3}{2^5} = \frac{1}{4}$$

$\# \Omega = 2^5$

HHHHH
HHHTH
HHHHT
HHHTT
THHHT
TTHHH
HTHHH

8 st.

18) Vi spelar mastermind. vi har 4 "pluppar" i rad. Varje "plupp" har en av 6 olika färger.
A = {motspelaren gissar rätt på första försöket}
Vad är P(A)?

$\# A = 1$
$\# \Omega = 6^4$

$\rightarrow P(A) = \frac{1}{6^4}$

35 a) Visa och motivera varför sätt att

$$\binom{n}{r} = \binom{n}{n-r}$$

Mot:

Att välja ut r objekt ur n är samma som att välja n-r objekt ur n.

O^n

r

O O^{n-r}

B: $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ $\binom{n}{n-r} = \frac{n!}{(n-r)!(n-(n-r))!} = \frac{n!}{(n-r)!r!}$

b) $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$

$1 \quad 2 \quad 3 \quad \cdots\cdots \quad n-1 \quad n$

r st

$= 1\ 2\ 3 \ \cdots\cdots\ n-1 \quad + \quad 1\ 2 \ \cdots\cdots\ n-1 = \binom{n-1}{r-1} + \binom{n-1}{r}$

r-1 st — fallen då objekt n väljs

r st obj — fallen då objekt n ej väljs

B: $\binom{n-1}{r-1} + \binom{n-1}{r} = \frac{(n-1)!}{(r-1)!\underbrace{(n-1-(r-1))!}_{=(n-r)!}} + \frac{(n-1)!}{r!(n-1-r)!}$

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

---

2 april, ons

Multinominal-koefficienter:

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!}$$

sätt att dela in n obj. i k grupper av storlek

$n_1, n_2, \ldots, n_k : n_1 + n_2 + \cdots + n_k = n$

$$(a_1 + a_2 + \cdots + a_k)^n = \sum_{n_1, n_2, \ldots, n_k \,:\, n_1+n_2+\cdots+n_k=n} \binom{n}{n_1, n_2, \ldots, n_k} a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k}$$

Ex: 36 studenter

$A = \{$3 st. är födda i jan, 3 st födda i feb, ..., 3 st i dec$\}$

$\# \Omega = 12^{36}$ $\quad \# A = \binom{36}{\underbrace{3, 3, \ldots, 3}_{12\text{ st}}}$

$P(A) = \frac{\# A}{\# \Omega} \approx 2.4 \cdot 10^{-7}$

Betingade sannolikheter (conditional probability)

Def: $P(A|B) := \dfrac{P(A\cap B)}{P(B)}$ ← givet

Enkel konsekvens:

$$P(A\cap B\cap C) = P(A|B\cap C)P(B\cap C) =$$
$$= P(A|B\cap C)P(B|C)P(C)$$

(Vad är sannolikheten att vi var någonstans i A givet att vi var någonstans i B.)

Ex: Igen har vi 36 studenter.

$A_i := \{$person nr i har ej samma födelsedag som person j för $1\le j\le i-1\}$

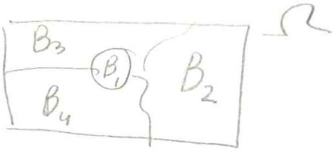
$A = \{$minst två personer har samma födelsedag$\}$

$\Rightarrow A^c = A_1\cap A_2\cap A_3\cap \dots \cap A_{36}$

$$P(A^c) = P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\cdots P(A_{36}|A_1\cap A_2\cap\dots\cap A_{35})$$

1:a pers: $= 1 \cdot \dfrac{364}{365}\cdot\dfrac{363}{365}\cdot \dots \cdot \dfrac{330}{365}$ ← givet

Def: En partition av Ω är mängder $B_1, B_2, \dots, B_k$ s.a. $B_i\cap B_j = \emptyset$ $i\neq j$ och $\Omega = \bigcup_{i=1}^{k} B_i$



LTP: "Law of total probability": Om $B_1,\dots,B_k$ är en partition av Ω och $A\subset\Omega$

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

$\left(A_1\cup A_2\cup\dots\cup A_k = \bigcup_{i=1}^{k} A_i\right)$

B: $\sum_{i=1}^{k} P(A|B_i)P(B_i) = \sum_{i=1}^{k} P(A\cap B_i) = P\left(\bigcup_{i=1}^{k} A\cap B_i\right) =$

$= P(A\cap\Omega) = P(A)$ $\square$

Ex: Vi har ett spel där vi 1) Vi singlar slant: $P(H) = P(T) = \frac{1}{2}$

2) Om H inträffar slår vi en tärning

Om T inträffar — " — två tärningar.

Låt D = totala tärningssumman. $\#\Omega = 6+36 = 42$

$P(D=5|H) = \dfrac{1}{6}$ $\quad P(D=5|T) = \dfrac{4}{36} = \dfrac{1}{9}$

$P(D=5) = P(D=5|H)P(H) + P(D=5|T)P(T) = \frac{1}{6}\cdot\frac{1}{2} + \frac{4}{36}\cdot\frac{1}{2} =$

$= \dfrac{5}{36}$ //

Ex: Spel: 3 dörrar, bakom en finns en bil bakom de andra finns ingenting.
Dörrarna har (osynliga) nummer 1,2,3.
bilen finns bakom nummer 1.

1) Välj en dörr slumpmässigt X_1 $P(X_1=1)=P(X_1=2)=P(X_1=3)=\frac{1}{3}$
2) Martin T. väljer en annan dörr som inte har en bil bakom sig. (kalla dörren X_2)
3) Välj X_1 eller byt dörr till X_3. och vi får vad som är bakom.

Vad är bäst strategi?

Bayes formel:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

B: $P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$ $\square$

Låt $B_1, B_2, \dots, B_k$ vara en partition av Ω

$$P(B_i|A) \underset{\uparrow}{=} \frac{P(A|B_i)P(B_i)}{P(A)} \underset{\uparrow}{=} \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}$$

Bayes formel LTP

Def: Apriori-sannolikheten för B är $P(B)$
Aposteriori-sannolikheten " " givet A $P(B|A)$

Om vi har ett experiment är $P(B)$ sann för B. innan experimentet är utfört.
$P(B|A)$ är sann för B givet att experimentet utföll A.

Ex: Apriori: $P(H) = P(T) = \frac{1}{2}$
Aposteriori (givet $A = \{D=5\}$)

$$P(H|D=5) = \frac{P(H\cap D=5)}{P(D=5)} = \frac{P(D=5|H)P(H)}{P(D=5)} = \frac{\frac{1}{6}\cdot\frac{1}{2}}{\frac{5}{36}} = \frac{3}{5}$$

$$P(T|D=5) = \frac{P(D=5|T)P(T)}{P(D=5)} = \frac{\frac{1}{9}\cdot\frac{1}{2}}{\frac{5}{36}} = \frac{2}{5}$$

Oberoende: Intuitivt: A och B är oberoende om $P(A|B) = P(A)$ symmetriskäl $P(B|A) = P(B)$
$B \neq 0$
Nackdelen är att $P(A|B)$ måste vara väldefinierad dvs $P(B) > 0$

Istället: A och B är oberoende ($A \perp B$) omm $P(A \cap B) = P(A)P(B)$

om $P(B) > 0 \Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$ //

Def: A, B, C är invördes oberoende om A är oberoende av B, $B \perp C$, $A \perp C$ och $P(A \cap B \cap C) = P(A)P(B)P(C)$

Ex: En krona kastas 2 ggr. $A = \{1\text{:a kastet H}\}$ $B = \{2\text{:a kastet H}\}$ $C = \{\text{exakt ett kast } \cancel{\text{är}} \text{ H}\}$

i) Uppenbart att A och B oberoende.

ii) $P(A \cap C) = P(C|A)P(A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(C) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$\Rightarrow P(A)P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \Rightarrow$ A och C oberoende

iii) pss. B och C oberoende.

$P(A \cap B \cap C) = 0 \neq \frac{1}{8} = P(A)P(B)P(C)$ (parvis ober. men inte inbördes oberoende)

Ex: Frågar n pers. om deras födelsedag

$B_n = \{\text{ingen av dem fyller år på samma dag som jag}\}$

$A_i = \{\text{person } i \text{ fyller } \underline{\text{ej}} \text{ år samma dag som jag}\}$

$B_n = A_1 \cap A_2 \cap \ldots \cap A_n$

$P(B_n) = P(A_1 \cap \ldots \cap A_n) = P(A_1)^n = \left(\frac{364}{365}\right)^n$ (ty oberoende)

Linda 31 år ensamstående, hon är aktiv (engagerad) i frågor som nedrustning, miljö och jämställdhet.
Vilka av dessa påståenden är mest troliga:
1) Linda är bankkassörska
) — " — bankkassörska och aktiv feminist

om cancer 80% rätt
om ej cancer 90% rätt
1% har cancer.

Räkneövn. 2

46) Urna A: 3 röda bollar och 2 vita
" B: 2 röda bollar " 5 vita

Spel 1) vi singlar slant och
2) om H så drar vi en boll från A
" T " " B

a) X är en s.v. X = färgen på bollen $\in \{r, v\}$

$$P(X=r) = P(X=r \mid H)P(H) + P(X=r \mid T)P(T) =$$

$$= \frac{3}{5} \cdot \frac{1}{2} + \frac{2}{7} \cdot \frac{1}{2} = \frac{31}{70}$$

b) Om $X=r$, vad är sann att den kom från A?

$$P(H \mid X=r) = \frac{P(H \cap X=r)}{P(X=r)} = \frac{P(X=r \mid H)P(H)}{P(X=r)} \quad \frac{\frac{3}{5} \cdot \frac{1}{2}}{\frac{31}{70}} =$$

$$= \frac{21}{31}$$

53) Vi har ett brandförsäkringsbolag. Sann att vi får ett krav:

	H_r	M_r	L_r
krav:	0,02	0.01	0.0025
proportion av # klienter	0.10	0.20	0.70

Vilken proportion av alla krav kommer från H_r?

N = # klienter   # klienter från H_r med krav: $N \cdot 0{,}10 \cdot 0{,}02$
" " M_r $N \cdot 0{,}2 \cdot 0{,}01$
" " " L_r $N \cdot 0{,}70 \cdot 0{,}0025$

$$\text{prop. krav från } H_r = \frac{N \cdot 0{,}002}{N \cdot 0{,}002 + N \cdot 0{,}002 + N \cdot 0{,}00175} = \frac{2}{5{,}75} \approx 0{,}35$$

61) En defekt komponent upptäcks vara defekt i 95% av fallen.
En hel komponent upptäcks —"— hel i 97% av fallen.
Vi antar att 0,5% av alla komp. är defekta.
Vad är sannolikheten att en komponenten som sägs vara felaktig är hel?

prop. av komp. som är defekta och sägs vara defekta = 0.005 · 0,95 prop. av komp som är hela men sägs vara defekta = 0,995 · 0,03

$$\mathbb{P}(\text{komponent som sägs vara defekt men är hel}) = \frac{0.03 \cdot 0{,}995}{0{,}005 \cdot 0{,}95 + 0{,}995 \cdot 0{,}03} \approx 0{,}86$$

69). $A \cap B = \emptyset$ Kan A och B vara oberoende?
Ja: om $\mathbb{P}(A)$ eller $\mathbb{P}(B) = 0$.

73) Vi har ett system med n komponenter (oberoende av varandra). Varje komp. går sönder med sann P.
Systemet är sönder om k eller fler av komponenterna är sönder. Vad är sann att systemet är sönder.
X = # komp. som är sönder

$$\mathbb{P}(X=k) = \underbrace{p^k}_{k\ \text{första}} \underbrace{(1-p)^{n-k}}_{\text{resten hela}} \binom{n}{k}$$

antal sätt som k st kan vara sönder på

1 2 3 . . . n

k st
$p^k(1-p)^{n-k}$
att exakt dessa är sönder

$$\mathbb{P}(\text{systemet sönder}) = \sum_{l=k}^{n} \mathbb{P}(X=l) = \sum_{l=k}^{n} p^l (1-p)^{n-l} \binom{n}{l}$$

78) Enkel genetik modell: En organism har genotyper
AA Aa aa
a) Antar att $F_1 = AA$, $F_2 = Aa$ Vad kan B = "barnets" genotyp vara och med vilken sannolikhet?
B = AA B = Aa

$$\mathbb{P}(B = AA) = \frac{1}{2}$$
$$\mathbb{P}(B = Aa) = \frac{1}{2}$$

b) Antag: $P(F_1=AA)=p \quad P(F_1=Aa)=2q \quad P(F_1=aa)=r$

Vi skall beräkna $P(B=AA) \quad P(B=Aa) \quad P(B=aa)$

C = genotypen för 3:e gen.

$P(C=AA) \quad P(C=Aa) \quad P(C=aa)$

kommer vara samma.

$$
\begin{aligned}
P(B=AA) = {} & P(B=AA \mid (F_1,F_2)=(AA,AA))\,P((F_1,F_2)=(AA,AA)) \\
& + \quad AA \qquad AA,Aa \qquad AA,Aa \\
& + \quad AA \qquad Aa,AA \qquad Aa,AA \\
& + \quad AA \qquad Aa,Aa \qquad Aa,Aa \\
& + \quad AA \qquad AA,aa \qquad AA,aa \\
& + \quad AA \qquad aa,AA \qquad aa,AA
\end{aligned}
$$

$$= 1p^2 + \tfrac{1}{2}\cdot p \cdot 2q + \tfrac{1}{2} p\, 2q + \tfrac{1}{4}(2q)^2 = p^2 + 2pq + q^2$$

$$
\begin{aligned}
P(B=Aa) = {} & P(B=Aa \mid (F_1,F_2)=(AA,aa))\,P((F_1,F_2)=(AA,aa)) \\
& + \qquad aa,AA \qquad aa,AA \\
& \qquad\quad AA,Aa \qquad AA,Aa \\
& \qquad\quad Aa,AA \qquad Aa,AA \\
& \qquad\quad Aa,aa \qquad Aa,aa \\
& \qquad\quad aa,Aa \qquad aa,Aa \\
& \qquad\quad Aa,Aa \qquad Aa,Aa
\end{aligned}
$$

$$= 1\cdot pr + 1\cdot p\cdot r + \tfrac{1}{2}\cdot p\, 2q + \tfrac{1}{2} p\, 2q + \tfrac{1}{2}\cdot 2q\cdot r + \tfrac{1}{2} r\cdot 2q + \tfrac{1}{2}(2q)^2 = 2pr + 2pq + 2qr + 2q^2$$

$$P(B=aa) = r^2 + 2rq + q^2$$

Vad med C?

$$P(C=AA) = \bar{p}^2 + 2\bar{p}\bar{q} + \bar{q}^2 = (\bar{p}+\bar{q})^2$$

$$P(C=Aa) = 2\bar{p}\bar{r} + 2\bar{p}\bar{q} + 2\bar{q}\bar{r} + 2\bar{q}^2$$

$$P(C=aa) = \bar{r}^2 + 2\bar{q}\bar{r} + \bar{q}^2$$

subst

$$\left\{\begin{aligned}
\bar{p} &= p^2 + 2pq + q^2 = P(B=AA) = (p+q)^2 \\
2\bar{q} &= 2pr + 2pq + 2qr + 2q^2 = P(B=Aa) \\
\bar{r} &= r^2 + 2qr + q^2 = P(B=aa)
\end{aligned}\right.$$

Har kvar: visa $\bar{p}=p \quad \bar{q}=q \quad \bar{r}=r$

$\bar{p}+\bar{q}=p+q \Rightarrow \bar{p}=p!$

$\bar{p}+\bar{q}=\cancel{p^2}+\cancel{2pq}+q^2+\cancel{pr}+pq+qr+q^2=p\underbrace{(p+2q+r)}_{1}+q\underbrace{(p+2q+r)}_{=1}=p+q$ usv.

Har visat $P(B=AA)=P(C=AA)$

Symmetri ger $P(B=aa)=P(C=aa)$

2 samma $\Rightarrow$ 3:e samma

$P(B=Aa)=1-P(B=AA)-P(B=aa)=1-P(C=AA)$
$-P(C=aa)=P(C=Aa)$ $\square$

---

7 april Mån LV2 Diskreta slumpvariabler:

$\mathbb{N}_0=\{0,1,2,\ldots\}$ (SV: stokastisk variabler (slump) (process))

$\mathbb{N}=\{1,2,3,\ldots\}$

Def: Diskret s.v. $X:\Omega\to\{x_1,x_2,\ldots\}$

om möjligt $x_1<x_2<x_3\ldots$

X är resultatet av ett slumpexperiment. Med antingen ändligt eller uppräkneligt oändligt antal möjliga värden.

uppräkneligt oändligt: ex $\mathbb{N}=\{1,2,3\ldots\}$

överuppräkneligt oändligt: ex $\mathbb{R}$ eller $[0,1]$

Om vi försöker lista alla tal i $[0,1]$:

0.137859001.... vår lista innehåller alla tal
0.295310789..... i $[0,1]$
0.3795

Konstruera $x=0.290\ldots$ x ej lika med något tal i listan. Dvs det finns ej någon lista!

~~Ex~~: Antar vi drar ett kort ur en kortlek. om det ej är ett A så lägger jag tillbaka kortet, blandar och drar ett nytt kort

X = # kort jag måste dra innan jag får ett A.

$X=\{1,2,3,\ldots\}=\mathbb{N}$

Slumpräkning: $X:\Omega\to\mathbb{N}_0$ X är resultatet av en <u>slumpräkning</u> i ett experiment.

<u>Ex</u>: Kastar tärning 4 ggr. X = # 6:or vi får

$X \in \{0,1,2,3,4\}$ (X <u>räknar</u> det <u>slumpmässiga</u> # 6:or)

Observera: X inducerar en partition av Ω på följande sätt: $B_i := \{x \in \Omega : X(x) = i\}$

Vi har: $B_i \cap B_j = \emptyset$ $i \neq j$ och $\Omega = \bigcup_{i=0}^{\infty} B_i$

i vårt exempel. om $x = \{5,5,6,1\}$

$X(x) = 1$ $x \in B_1$

<u>Def</u>: <u>Sannolikhetsfördelningen</u> av en s.v X är mängden av alla sannolikheter för alla möjliga värden på X.

<u>Probability mass function</u> (pmf) (kallas också frekvens funktion)

P $\{p_k\}_{k \in \mathbb{N}}$ $\{p(k)\}_{k \in \mathbb{N}}$ $p_k = p(k) := P(X = x_k)$

<u>Vi har</u>: $p_1 + p_2 + p_3 + \dots = 1$

<u>Ex</u>: slant - tärning

1) Singlar slant (H, T)

2) Om H kastar vi en tärning.

Om T " två tärningar.

D = totala tärningssumman ger en <u>slumpräkning</u>

$p_k = P(D = k)$ $1 \leq k \leq 12$

$p_1 = \frac{1}{12}, p_2 = \frac{7}{72},$

X = (resultat av slant-sing, D) $\in \{(H,1), (H,2), \dots, (H,6), (T,2), \dots, (T,12)\}$ x_1 x_2 x_7

ty vektorer

<u>Def</u>: <u>Cumulative distribution function</u> (cdf) Om X är en reellvärd ($x \in \mathbb{R}$) diskret s.v

$F(x) := P(X \leq x)$

Om X är en slumpräkning $k \in \mathbb{N}_0$

$F(k) = P(X \leq k) = \sum_{l=0}^{k} P(X=l) = p_0 + p_1 + \dots + p_k$

Några enkla egenskaper för cdf av en slumpräkning: $P(X > k) = 1 - F(k) = P(X \geq k+1)$

$P(k_1 < X \leq k_2) = F(k_2) - F(k_1)$

$p_k = F(k) - F(k-1)$

Def: Väntevärde (medelvärde)

X är en diskret slumpvariabel med $X \in \mathbb{R}$

så är ~~väntevärdet~~ av X

$$\underset{\text{expectation}}{\mathbb{E}[X]} := \sum_{k=0}^{\infty} \underset{\text{reella tal}}{x_k} p_k = \sum_{k=0}^{\infty} x_k P(X = x_k)$$

Ibland betecknas $\mathbb{E}[X]$ med μ

(slumpräkning specialfall av diskreta sv)

För en slumpräkning: $\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot p_k = \sum_{k=1}^{\infty} k P(X=k)$

Def: Variansen av X betecknas med $Var(X)$ eller σ^2 och definieras av: $Var(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$

σ kallas för standardavvikelsen.

Egenskaper: $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

$\mathbb{E}[cX] = c\mathbb{E}[X]$ } visas på övning.

$\mathbb{E}[g(X)] = \sum_{k=0}^{\infty} g(x_k) p_k$ vanligast $g(x) = x^n$ $n \in \mathbb{N}$

$\mathbb{E}[X] = \sum_{k=0}^{\infty} P(X \geq k)$ $Var(cX) = c^2 Var(X)$

(är inte linjär)

Observera att: $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 + \mathbb{E}[X]^2 - 2X\mathbb{E}[X]]$

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

$= \mathbb{E}[X^2] + \mathbb{E}[X]^2 \mathbb{E}[1] - 2\mathbb{E}[X]\mathbb{E}[X]$

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

$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ om X och Y oberoende.

$Var(X+Y) = Var(X) + Var(Y)$ — " —

Ex: "Deal or no deal" Vi är i slutet av spelet och har 3 väskor kvar. I en väska finns 1 000 000 i den andra 500 000 och i den sista 1kr.

1) Du kan acceptera 360 000

2) Du går vidare varvid en slumpmässigt vald väska slängs bort.

Bör vi acceptera budet?

Fråga dig, vad händer om vi aldrig ~~tar~~ /

X = pengarna i sista väskan $\mathbb{E}[X] = 1 \cdot \frac{1}{3} + 500000 \cdot \frac{1}{3} + \frac{1}{3} \cdot 1000\,000 =$

$\approx 500\,000$ DVS den förväntade förtjänsten är 500 000.

Def Likformig fordelning $X \in \{x_1, x_2, \ldots, x_n\}$

om $p_k = \mathbb{P}(X = x_k) = \frac{1}{n}$ $\forall k$ så är X likformigt fordelad på $\{x_1, x_2, \ldots x_n\}$

Ex: Tärning. X är likf ~~för~~ förd. på $\{1, 2, \ldots, 6\}$

$\mathbb{E}[X] = 3.5$

Def: Bernoulli: $X \in \{0,1\}$ $p = \mathbb{P}(X=1)$

Def: Binomialfördelningen $X \sim Bin(n,p)$ $n \in \mathbb{N}$ $0<p<1$

$\mathbb{P}(X=k) = p_k = \binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \ldots, n$

har fördeln. el. fordelat som

$(q = 1-p)$

$\mathbb{E}[X] = np$ $Var(X) = np(1-p)$

Ex: Vi upprepar ett exp. n ggr (exp är ober.) Varje exp. lyckas med s.p X = #lyckade exp. $\Rightarrow X \sim Bin(n,p)$

$p^k (1-p)^{n-k} \binom{n}{k}$

Exp: 1 2 3 ... n

1) välj k st — k st

2) Vad är sannolikheten att exakt dessa k var lyckade? (kompl. misslyckade)

För att få den totala sann för $\{X=k\}$ måste summera över # sätt att göra steg 1.

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

2.2 En slant singlas 4 ggr. Hitta pmf och cdf för

8/4 - tisdag
Räkneövn.

a) $X = \# H$ inför första T

pmf. $p_0 = P(X=0) = \frac{1}{2}$

$p_1 = P(X=1) = P(HT) = \frac{1}{4}$

$p_2 = P(X=2) = P(HHT) = \frac{1}{8}$

$p_3 = P(X=3) = P(HHHT) = \frac{1}{16}$

$p_4 = P(X=4) = P(HHHH) = \frac{1}{16}$

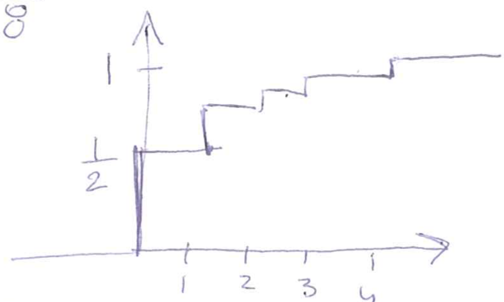
~~b)~~ $F(0) = p_0 = \frac{1}{2}$

$F(1) = p_0 + p_1 = \frac{3}{4}$

$F(2) = F(1) + p_2 = \frac{3}{4} + \frac{1}{8} = \frac{7}{8}$

$F(3) = F(2) + p_3 = \frac{7}{8} + \frac{1}{16} = \frac{15}{16}$

$F(4) = F(3) + p_4 = \frac{15}{16} + \frac{1}{16} = 1$



b) $X = \# H$ i rad efter första T

$p_0 = P(X=0) = P(TT**) + P(HTT*) + P(HHTT) + P(HHHT) + P(HHHH) = \frac{9}{16}$

$p_1 = P(X=1) = P(THT*) + P(HTHT) + P(HHTH) = \frac{4}{16} = \frac{1}{4}$

$p_2 = P(X=2) = P(THHT) + P(HTHH) \quad \frac{2}{16} = \frac{1}{8}$

$p_3 = P(X=3) = P(THHH) = \frac{1}{16}$

cdf. $F(0) = p_0 = \frac{9}{16}$

$F(1) = \frac{13}{16}$, $F(2) = \frac{15}{16}$, $F(3) = 1$

c) $X = \# H - \# T \qquad X \in \{-4, -2, 0, 2, 4\}$

$p_{-4} = P(X=-4) = \frac{1}{16} \qquad p_2 = \frac{4}{16} \qquad p_4 = \frac{1}{16}$

$p_{-2} = P(X=-2) = \binom{4}{1}\frac{1}{16} = \frac{4}{16}$

$p_0 = P(X=0) = \binom{4}{2}\frac{1}{16} = \frac{6}{16}$

d) $X = \#H \times \#T \qquad X \in \{0, 3, 4\}$

$p_0 = P(X=0) = P(HHHH) + P(TTTT) = \frac{2}{16}$

$p_3 = P(\text{exakt } 3H) + P(\text{exakt } 3T) = \binom{4}{1}\frac{1}{16} + \binom{4}{1}\cdot\frac{1}{16} =$

$= \frac{8}{16}$

$p_4 = P(X=4) = P(\text{exakt } 2H) = \binom{4}{2}\cdot\frac{1}{16} = \frac{6}{16}$ //

---

4) $X \in \mathbb{Z}$ s.v. visa att $p_k = F(k) - F(k-1)$

B: $F(k) = P(X \leq k) = \sum_{l=-\infty}^{k} P(X=l) = P(X=k) + \sum_{l=-\infty}^{k-1} P(X=l)$

$= p_k + P(X \leq k-1) = p_k + F(k-1)$ $\square$

---

14. Spelare A och B kastar straffkast i basket. A börjar och den som först träffar vinner,

a) X_A = # kast A behöver ~~innan~~ för träff.

$p_1 = P(X=1) = p_A$

p_A = sann att A träffar.

$p_k = P(k-1 \text{ missar följt av träff}) =$

X_A är Geom(p_A)-fördelad. $= (1-p_A)^{k-1}p_A$

b) Vad är sann att A vinner?

$P(A \text{ vinner}) = \sum_{k=1}^{\infty} P(A \text{ vinner i } \overset{\text{omgång}}{\text{kast}}\ k) = \sum_{k=1}^{\infty}(1-p_A)^{k-1}(1-p_B)^{k-1}p_A =$

$= p_A \sum_{k=0}^{\infty}\left((1-p_A)(1-p_B)\right)^k = p_A \frac{1}{1-(1-p_A)(1-p_B)}$ //

Alt: Väntar till den omgång l då minst en träffade

$P(A \text{ vinner}) = P(A \text{ träffade} \mid A \text{ eller } B \text{ träffade}) =$

$= \frac{P(A \text{ tr.})}{P(A \text{ el. } B \text{ tr})} = \frac{p_A}{1-(1-p_A)(1-p_B)}$ //

---

17) Vi har en sekvens av oberoende Bernoulli-försök som lyckas med sann. p.

X = # misslyckade innan första lyckade, bestäm pmf:

$p_0 = p \qquad p_k = (1-p)^k p$

4.7 $X \in \{0,1,2\}$ $p_0 = \frac{1}{2}$, $p_1 = \frac{3}{8}$ $p_2 = \frac{1}{8}$

a) Bestäm $\mathbb{E}[X]$ $\mathbb{E}[X] = \sum_{k=0}^{\infty} k p_k = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 =$

$= \frac{3}{8} + \frac{2}{8} = \frac{5}{8}$ //

b) Låt $\cancel{X}\ Y = X^2$ Hitta pmf och $\mathbb{E}[Y]$ $Y \in \{0,1,4\}$

$p_0 = \frac{1}{2}$, $p_1 = \frac{3}{8}$ $p_4 = \frac{1}{8}$

$\mathbb{E}(Y] = 0 \cdot p_0 + 1 \cdot p_1 + 4 p_4 = \frac{3}{8} + \frac{4}{8} = \frac{7}{8}$

c) Använd: $\mathbb{E}\{g(X)] = \sum g(x) p(x)$

$p(x) = \mathbb{P}(X = x)$

med $g(x) = x^2$

Vi får: $\mathbb{E}[X^2] = g(0)p(0) + g(1)p(1) + g(2)p(2) =$

$= 0 \cdot \frac{1}{2} + 1 \cdot \frac{3}{8} + 4 \cdot \frac{1}{8} = \frac{7}{8}$ //

d) Hitta variansen $Var(X)$ mha 1) definitionen.

2) $Var(X) = \cancel{\mathbb{E}[(X}\ \mathbb{E}[X^2] - \mathbb{E}[X]^2$

2): $Var(X) = \frac{7}{8} - \left(\frac{5}{8}\right)^2 = \frac{56-25}{64} = \frac{31}{64}$

$g(x) = \left(x - \frac{5}{8}\right)^2 = (x - \mathbb{E}(X])^2$

$Var(X) = \mathbb{E}[g(X)] = g(0)p_0 + g(1)p_1 + g(2)p_2 =$

$= \left(\frac{5}{8}\right)^2 \cdot \frac{1}{2} + \left(\frac{3}{8}\right)^2 \frac{3}{8} + \left(\frac{11}{8}\right)^2 \cdot \frac{1}{8} = \dots = \frac{31}{64}$ //

---

g) $X \in \mathbb{N}$ Visa att $\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$

B: Använd def: $\mathbb{E}[X] = \sum_{k=1}^{\infty} k \mathbb{P}(X = k) = \mathbb{P}(X=1) + 2\mathbb{P}(X=2) + 3\mathbb{P}(X=3) + \dots$

$= (\mathbb{P}(X=1) + \mathbb{P}(X=2) + \mathbb{P}(X=3) + \dots) + (\mathbb{P}(X=2 + \mathbb{P}(X=3) + \mathbb{P}(X=4) + \dots) + (\mathbb{P}(X=3) + \mathbb{P}(X=4) + \dots) + (\mathbb{P}(X=k) + \mathbb{P}(X=k+1) + \dots)$

$= \mathbb{P}(X \geq 1) + \mathbb{P}(X \geq 3) + \dots \mathbb{P}(X \geq k) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$ $\boxtimes$

Några fler fördelningar (diskreta) 8/4 ons.

~~Hypergeometrisk fördelning~~ med parametrar
N, n, p där $N, n, (Np) \in \mathbb{N}$ $n \le N$ $0<p<1$
$(q = 1-p)$
har pmf: $p_k = \dfrac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}$ för $\max(n-Nq, 0) \le k$
$k \le \min(n, Np)$

Om $X \sim Hg(N, n, p)$ då har vi $\mathbb{E}[X] = np$
$Var(X) = npq\left(1 - \frac{n-1}{N-1}\right)$

Ex: N = # bollar i en urna (svarta eller vita)
p = initiala proportionen av svarta bollar
X = #svarta bollar i en grupp av n bollar som vi plockat ur urnan.

Om $I_j = \begin{cases} 1 & \text{om boll nr } j \text{ är svart} \\ 0 & \text{'' '' vit} \end{cases}$
då har vi att $X = I_1 + I_2 + \ldots + I_n$
Vi vill bestämma pmf för X. $p_k = \mathbb{P}(X = k)$
1) Antal sätt att välja k svarta bollar ur totalt Np svarta bollar är $\binom{Np}{k}$
2) Antal sätt att välja $n-k$ vita bollar ur totalt $N(1-p) = Nq$ vita bollar är $\binom{Nq}{n-k}$
3) Antal sätt att välja n bollar ur N är $\binom{N}{n}$

$\mathbb{P}(X = k) = \dfrac{\text{\# sätt att välja } k \text{ svarta och } n-k \text{ vita}}{\text{\# sätt att välja } n \text{ bollar}}$

$= \dfrac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}} \Rightarrow X \sim Hg(N, n, p)$

Ex: Aspirin-problemet: 11034 fick placebo, 189 hjärtattacker
11037 '' aspirin, 104 —''—
Hypotes: Ingen effekt.
X = # hjärtattacker i placebogruppen.

Under vår hypotes är $X \sim Hg(22071, 293, \frac{11034}{22071})$

Vad är $P(X = 189) = \frac{\binom{11034}{189}\binom{11037}{104}}{\binom{22071}{293}} = 0.00000015$

Kan vi dra en slutsats? Nej för $P(X=k) \leq 0.0468$ för varje k.

Snarare beräknar vi $P(X \geq 189)$

<u>Geometrisk fördelning</u>: $X \sim Geom(p)$ $0<p<1$ om pmf för X är $p_k = (1-p)^{k-1}p$, $k \in \mathbb{N}$

<u>cdf</u>: $F(k) = 1-(1-p)^k$

$$\mathbb{E}[X] = \frac{1}{p} \qquad \sigma^2 = Var(X) = \frac{q}{p^2} = \frac{1-p}{p^2}$$

<u>Ex</u>: Vi har ett försök som upprepas (oberoende) och lyckas med sannolikhet p.

X = # försök tills vi lyckas första gången.

$P(X=k) = P(\{\text{vi misslyckas } k-1 \text{ gånger}\} \cap \{\text{vi lyckas i försök } k\}) = (1-p)^{k-1}p \Rightarrow X \sim Geom(p)$

Användbar egenskap: $P(X>n+k \mid X>n) =$

$$= \frac{P(X>n+k)}{P(X>n)} = \frac{(1-p)^{n+k}}{(1-p)^n} = (1-p)^k = P(X>k)$$

"glömske-egenskap"
"loss of memory property"

<u>Ex</u>: Vi spelar rött på Roulette

X = # gånger vi måste spela tills vi får rött 1:a gången.

$$P(X>11 \mid X>10) = P(X>1)$$

<u>Ex</u>: X = # människor jag frågar innan jag hittar någon med "min" födelsedag.

$X \sim Geom(1/365)$ $P(X>253) = (1-\frac{1}{365})^{253} \approx 0.5$

$$\mathbb{E}[X] = 365$$

Negativa binomialfördeln. med parametrar r, p.

$r \in \mathbb{N}$ $\quad 0 < p < 1$ $\quad$ Om $X \sim Nb(r,p)$ så har X pmf: $\quad p_k = \binom{k-1}{r-1} p^r (1-p)^{k-r}$ $\quad (k = r, r+1, r+2 \ldots)$. $\quad k \geq r$

$\mathbb{E}[X] = \frac{r}{p}$ $\quad \sigma^2 = Var(X) = \frac{rq}{p^2}$ $\quad q = 1-p$

Ex: Bernoulli-försök (oberoende) med sannolikhet p att lyckas

X = # försök tills vi har r lyckade försök

Observera att: $X = Y_1 + Y_2 + \ldots + Y_r$ där $Y_i \sim Geom(p)$ och dom är oberoende.

$$\mathbb{E}[X] = \sum_{i=1}^{r} \mathbb{E}[Y_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

Beräkna pmf:en för X $\quad p_k = \mathbb{P}(X=k)$

Varje sekvens (som ger $\{X=k\}$) av försök som består av k försök varav r lyckade (inkluderat försök k) har sannolikhet $p^r(1-p)^{k-r}$

Hur många sådana sekvenser finns det?

Den sista är lyckad och $r-1$ av dom $k-1$ första försöken är lyckade $\Rightarrow \binom{k-1}{r-1}$ = #sekv.

$$\mathbb{P}(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

Poisson-fördeln (ej Poisson-process) med parameter λ. $\quad \lambda > 0$ $\quad X \sim Poi(\lambda)$ om X har pmf

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

term k i Maclaurin $\quad k \in \mathbb{N}_0$

$$\mathbb{E}[X] = Var(X) = \lambda$$

Poisson-fördelningen är "lite mer konstruerad"

Betrakta $Bin(n,p)$ n är stort och p litet.

Antag att $np = \lambda$ (för varje n)

$X_n \sim Bin(n,p)$ $\quad p_k = P(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} =$

$$= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \approx \frac{n^k}{k!} p^k e^{-pn} = \frac{(np)^k}{k!} e^{-np} =$$

$$= \frac{\lambda^k}{k!} e^{-\lambda}$$

$1-p \approx e^{-p}$ Maclaurin

Ex: Radioaktivt material, vi betraktar sönderfall p = sannolikheten för minst ett sönderfall per millisekund

X_t = # sönderfall under t timmar

$X_t \sim Bin(n,p)$ $\quad n = 3\,600\,000t$ "dvs"

$X_t \sim Poi(\lambda t)$ där $\lambda = 3\,600\,000\,p$

Vi säger att X_t räknar # sönderfall som förekommer med intensitet λ: sönderfall / timme

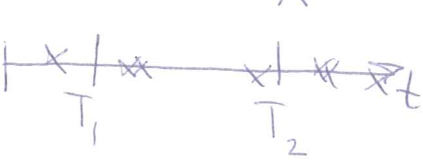
Poisson-process i planet är en punkt-process med intensitet λ = #punkter / area-enhet

X_A = # punkter i A

$X_A \sim Poi(\lambda \cdot Area(A))$ egenskap 1

$X_B \sim Poi(\lambda \cdot Area(B))$

Egenskap 2: X_A och X_B är oberoende (om $A \cap B = \emptyset$)



X = # punkter i $[T_1, T_2]$

$\sim Poi(\lambda |T_2 - T_1|)$

Räkneövn. ons LV2

19) Bestäm cdf för en Geom(p)-fördeln.

pmf: $p_k = P(X = k) = (1-p)^{k-1} p \qquad k \geq 1$

cdf: $F(n) = P(X \leq n) = \sum_{l=1}^{n} P(X = l) = \sum_{l=1}^{n} (1-p)^{l-1} p =$

$= p \sum_{l=0}^{n-1} (1-p)^l = p \dfrac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n$

---

31 a) Diana får telefonsamtal enligt en Poisson-process med intensitet $\lambda = 2$ samtal/timme.

a) Om Diana duschar i 10 min.

Hur stor är sannolikheten att hon missar ett samtal?

X = # samtal mellan $t=0$ och $t = \frac{1}{6}$ (t i timmar) 10 min

$X \sim Poi(\lambda \cdot \frac{1}{6}) = Poi(\frac{1}{3}) \qquad P(X \geq 1) = 1 - P(X = 0) =$ $\left[pmf = \frac{\lambda^k}{k!} e^{-\lambda} \right.$

$= 1 - e^{-1/3} \approx 0{,}28$

b) Hur länge kan Diana duscha om hon kan tänkas missa ett telefonsamtal med sannolikhet $1/2$?

X_T = antal samtal som ankommer i tidsintervallet $[0, T]$

$X_T \sim Poi(\lambda T) \qquad P(X_T \geq 1) = 1 - P(X_T = 0) = 1 - e^{-\lambda T} = \frac{1}{2}$

$\Rightarrow e^{-\lambda T} = \frac{1}{2} \Rightarrow -\lambda T = -\ln 2 \Rightarrow T = \dfrac{\ln 2}{\lambda}$ //

4.15 Vi har två lotterier med n lotter och samma utbetalningar. Vad är bäst (vi maximerar väntevärdet) av att 1) Köp två lotter från olika lotterier 2) Köper —"— samma lotteri?

Det spelar ingen roll! X = Vinsten från lott 1.

Y = " " 2.

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

4.24) $X_1, X_2, \ldots, X_n$ är diskreta sv. och $Y = a + \sum_{i=1}^{n} b_i X_i$, visa att $\mathbb{E}[Y] = a + \sum_{i=1}^{n} b_i \mathbb{E}[X_i]$

(dvs $\mathbb{E}[\;]$ är en linjär operator)

$$\mathbb{E}[Y] = \sum_{(x_1, x_2, \ldots, x_n)} \left(a + \sum_{i=1}^{n} b_i x_i\right) \underbrace{P((X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots, x_n))}_{= p(x_1, x_2, \ldots, x_n)}$$

$$= \sum a\, p(x_1, x_2, \ldots, x_n) + b_1 \sum_{(x_1, x_2, \ldots, x_n)} x_1\, p(x_1, \ldots, x_n) + \ldots +$$

$$+ b_n \sum_{(x_1, \ldots, x_n)} x_n\, p(x_1, \ldots, x_n) = a + b_1 \mathbb{E}[X_1] + \ldots + b_n \mathbb{E}[X_n]$$

$$\left(a \underbrace{\sum_{(x_1 \ldots x_n)} p(x_1, x_2, \ldots, x_n)}_{1 \text{ (tot sann)}} = a \right)$$

$$b_1 \sum_{(x_1, \ldots, x_n)} x_1\, p(x_1, \ldots, x_n) = b_1 \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} x_1\, p(x_1, x_2, \ldots, x_n) =$$

$$= \sum_{x_1} x_1 \underbrace{\sum_{(x_2, \ldots x_n)} p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)}_{= P(X_1 = x_1)} = \sum_{x_1} x_1 P(X_1 = x_1) =$$

$$= \mathbb{E}[X_1]$$

Slant-tärning (X, D) X = res. slant D = tärningssumman

om $x_1 = H$

$$\sum_{x_2} p(x_1, x_2) = P(X = H, D = 1) + P(X = H, D = 2) + \ldots P(X = H, D = 6)$$

$$= P(X = H)$$

---

30) En person samlar på idolkort på IFK's spelartrupp. Om IFK har n spelare och personen nöjer sig med r unika kort ($r < n$), vad är det förväntade # kort personen måste köpa?

X_1 = # kort tills första unika bilden = 1

X_2 = "fler" " andra "

X_3 = tredje

$\vdots$

X_r = " ... r:te "

Om X = # kort personen måste köpa.

$$X = X_1 + X_2 + \dots + X_r$$

Vad har X_2 för fördelning?

$$X_2 \sim \text{Geom}\left(\frac{n-1}{n}\right)$$

Vad har X_k " " $\quad X_k \sim \text{Geom}\left(\frac{n-k+1}{n}\right)$

$$\mathbb{E}[X] = \sum_{i=1}^{r} \mathbb{E}[X_i] = \sum_{i=1}^{r} \frac{1}{\frac{n-i+1}{n}} = n\sum_{i=1}^{r} \frac{1}{n-i+1}$$

---

20) Beräkna $\mathbb{E}\left[\frac{1}{1+X}\right]$ där $X \sim \text{Poi}(\lambda)$

Vi har $\mathbb{E}\left[\frac{1}{1+X}\right] = \sum_{k=0}^{\infty} \frac{1}{1+k} P(X=k) =$ $\left[\mathbb{E}[g(X)] = \dots\right]$

$$= \sum_{k=0}^{\infty} \frac{1}{1+k} \frac{\lambda^k}{k!} e^{-\lambda} = \frac{1}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} e^{-\lambda} = \frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} =$$

$$= \frac{1}{\lambda}\left(\underbrace{\underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{e^{\lambda}\ (\text{MacL.})} e^{-\lambda}}_{1} - e^{-\lambda}\right) = \frac{1-e^{-\lambda}}{\lambda}$$

$\boxed{14/4}$

Kontinuerliga slumpvariabler $\quad X : \Omega \to (-\infty, \infty) = \mathbb{R}$

X är ett reellt tal som är resultatet av ett slumpexperiment

cdf: $F(x) := P(X \le x)$

$F(x)$ är en kontinuerlig funktion som är icke-avtagande

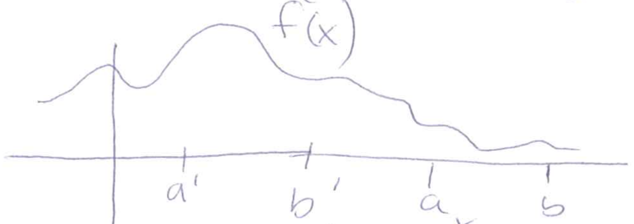
"från 0 till 1" dvs $\lim_{x \to -\infty} F(x) = 0$ och $\lim_{x \to \infty} F(x) = 1$

Vi har också att för varje $x \in \mathbb{R}$ gäller att $P(X = x) = 0$

~~Def~~: Probability density function (pdf) betecknas: $f(x) = F'(x)$

$f(x)$ beskriver sann.-dens. enligt: om X är en s.v. med pdf $f(x)$ så gäller:

$$\mathbb{P}(a < X < b) = \int_a^b f(x)dx$$



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

Vi def väntevärdet av X enligt:

(mek. tyngdpunkten) $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)dx$ (betecknas även med μ)

Vi kan också definiera variansen enligt:

Variansen: $Var(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$

Man kan bevisa att för $g(X)$ gäller: $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)dx$

Vi får att: $Var(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x)dx$ (bet. även σ^2)

σ kallas för standardavvikelsen.

Def: Kontinuerligt likformig fördelning med parametrar a, b: $-\infty < a < b < \infty$

$X \sim U(a,b)$ om $f(x) = \begin{cases} \frac{1}{b-a} & \text{om } a \leq x \leq b \\ 0 & \text{annars} \end{cases}$

$\mathbb{E}[X] = \frac{a+b}{2}$ $\quad \sigma^2 = Var(X) = \frac{(b-a)^2}{12} \Rightarrow$

$\Rightarrow \sigma = \frac{(b-a)}{\sqrt{12}}$

Ex: Vi har ett lås med 10 000 kombinationer dvs 4 siffror mellan 0 och 9, vi börjar med 0000, 0001, osv. X = # försök tills vi hittar rätt. $\sim U(\{0, 1, \ldots, 10\,000\})$ (diskret)

$T = \frac{X}{1000}$ söktiden i timmar (om vi hinner 1000 försök/timme), med god approximation är $T \sim U(0,10)$

$\mathbb{E}[T] = 5$ timmar, $\sigma = 2$ timmar 53 min.

Notera: $\mathbb{P}(T>3) = 0{,}7 = 1 - 0{,}3$, medans:

$$\mathbb{P}(T>5 \mid T>2) = \frac{\mathbb{P}(T>5)}{\mathbb{P}(T>2)} = \frac{\frac{10-5}{10}}{\frac{10-2}{10}} = \frac{0.5}{0.8} = 0.625$$

<u>Def</u>: ~~Exponential fördelningen~~ med parameter λ. Vi skriver $X \sim Exp(\lambda)$, med $\lambda > 0$ och pdf: $f(x) = \lambda e^{-\lambda x}$ för $x > 0$.

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x} \lambda e^{-\lambda y}\,dy = \left[-e^{-\lambda y}\right]_0^x = 1 - e^{-\lambda x}$$

$$\mathbb{E}[X] = Var[X] = \frac{1}{\lambda}$$

Uppkommer t.ex. som gränsvärde till den geometriska fördelningen.

Antag att $X \sim Geom(p)$ med p "litet"

Låt $T_n = \frac{X}{n}$ så gäller $\mathbb{P}(T_n > t) = \mathbb{P}(X > nt) =$

$= (1-p)^{nt} \approx e^{-pnt} = e^{-\lambda t}$ om $\lambda = np$

Om $T \sim Exp(\lambda)$: $\mathbb{P}(T>t) = e^{-\lambda t} \approx e^{-pnt} = e^{-\lambda t}$ om $\lambda = np$.

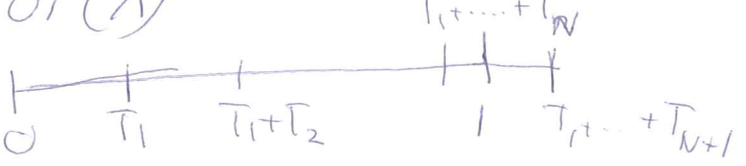
<u>Man kan visa att</u>: Låt $X_n \sim Geom(p_n)$ där $p_n = \frac{\lambda}{n}$

Låt $T_n := \frac{X_n}{n}$ så gäller $T_n \to T \sim Exp(\lambda)$

Viktig egenskap: $T_1, T_2, T_3, \ldots \sim Exp(\lambda)$ (oberoende)

Låt $N := \max_n T_1 + T_2 + \ldots + T_n \leq 1$

$\Rightarrow N \sim Poi(\lambda)$



"Mellanrummen mellan punkter i Poisson-processen (med par. λ) är exponentialfördelade (med par λ)!"

Ex: När vi modellerar radioaktivt sönderfall så är tiden till sönderfall exponentialfördelad

Parametern λ kallas för skalningsparameter

Om $T \sim Exp(\lambda)$ och $T_1 \sim Exp(1)$

$$\Rightarrow P\left(\frac{T_1}{\lambda} > t\right) = P(T_1 > \lambda t) = 1 - P(T_1 \leq \lambda t)$$

$$= 1 - \int_0^{\lambda t} e^{-x} dx = 1 - \left[-e^{-x}\right]_0^{\lambda t} = e^{-\lambda t} = \ldots = P(T > t)$$

dvs $Exp(\lambda) \sim \frac{1}{\lambda} Exp(1)$

Def: Medianvärde X är en s.v och vi väljer M s.a. $P(X < M) = P(X > M) = \frac{1}{2}$ kallas M för medianen.

Exponentialfördeln: Ex: $X \sim Exp(\lambda)$

$$P(X < M) = \int_0^M \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^M = 1 - e^{-\lambda M} = \frac{1}{2}$$

$$\Rightarrow -\lambda M = -\log 2 \Rightarrow M = \frac{\log 2}{\lambda}$$

M är halveringstiden i exemplet ovan.

Ex: Vi provar slumpmässigt en kod (av 10000 möjliga) och glömmer bort den om den ej fungerar.

X = # försök tills vi lyckas $\sim Geom\left(\frac{1}{10000}\right)$

$$P(X > 10\,000) = (0.9999)^{10000} = e^{-1}$$

Låt T vara söktiden i timmar (antar 1000 försök/timme) Då är $T \sim Exp(0.1)$ med god approximation

$$\underset{n=1000}{np} = \frac{1}{10000} = 0.1$$

Egenskap: $T \sim Exp(\lambda)$ och $0 < a < b < \infty$

$$P(T>b \mid T>a) = \frac{P(T>b)}{P(T>a)} = \frac{e^{-\lambda b}}{e^{-\lambda a}} = e^{-\lambda(b-a)} =$$

$P(T>b-a)$ med $b=5$, $a=2$

$P(T>5 \mid T>2) = P(T>3)$ "glömske-egenskap"

Egenskap ~~II~~: $T_1 \sim Exp(\lambda_1)$ $T_2 \sim Exp(\lambda_2)$ oberoende

$T = \min(T_1, T_2)$ Vi har att $T \sim Exp(\lambda_1+\lambda_2)$

"B:" $P(T>t) = P(T_1>t \cap T_2>t) \underset{\text{oberoende}}{=} P(T_1>t)\,P(T_2>t)$

$= e^{-\lambda_1 t} e^{-\lambda_2 t} = e^{-(\lambda_1+\lambda_2)t}$ $\square$

Gammafördelningen med parametrar α och λ

$X \sim Gamma(\alpha, \lambda)$ med pdf: $f(x) = \frac{1}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} \lambda e^{-\lambda x}$

för $x>0$.

$\Gamma(\alpha)$ den sk. gamma-funktionen och definieras av

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$$

För $k \in \mathbb{N}$ $\Gamma(k) = (k-1)!$

$E[X] = \frac{\alpha}{\lambda}$, $Var(X) = \frac{\alpha}{\lambda^2}$ Räkna ut genom att lösa integralen.

Observera att: $\alpha = 1 \Rightarrow \Gamma(1) = 0! = 1$

$f(x) = \frac{1}{\Gamma(1)} (\lambda x)^{1-1} \lambda e^{-\lambda x} = \lambda e^{-\lambda x}$ pdf för $Exp(\lambda)$

I själva verket: ~~att~~ om vi låter $Y_1, Y_2, \ldots, Y_k \sim Exp(\lambda)$ oberoende så är: $Y := Y_1 + Y_2 + \ldots + Y_k \sim Gamma(k, \lambda)$ //

15/4 R.öv. 2.34) pdf: $f(x) = \begin{cases} \frac{1+\alpha x}{2} & -1 \le x \le 1 \\ 0 & \text{annars} \end{cases}$
och $-1 \le \alpha \le 1$

cdf: $F(x) = \int_{-1}^{x} f(y)dy = \int_{-1}^{x} \frac{1+\alpha y}{2} dy = \left[\frac{1}{2}y + \frac{\alpha}{4}y^2\right]_{-1}^{x}$

$= \frac{1}{2}x + \frac{\alpha}{4}x^2 - \left(-\frac{1}{2} + \frac{\alpha}{4}\right) = \frac{1}{2}(x+1) + \frac{\alpha}{4}(x^2-1)$

$F(1) = 1 \Rightarrow f(x)$ är en pdf □

M def. av att $\underbrace{P(X < M)}_{=F(M)} = \frac{1}{2}$

$F(M) = \frac{M+1}{2} + \frac{\alpha}{4}(M^2-1) = \frac{1}{2} \Rightarrow \dots M = -\frac{1}{\alpha} \pm \sqrt{\left(\frac{1}{\alpha}\right)^2 + 1}$ två st, endast en är rätt

i) $0 \le \alpha \le 1$ $\quad -\frac{1}{\alpha} \le -1 \Rightarrow -\frac{1}{\alpha} - \sqrt{\left(\frac{1}{\alpha}\right)^2+1} < -1$

$\Rightarrow M = -\frac{1}{\alpha} + \sqrt{\left(\frac{1}{\alpha}\right)^2+1}$

ii) $-1 \le \alpha \le 0$ $\quad -\frac{1}{\alpha} \ge 1 \Rightarrow -\frac{1}{\alpha} + \sqrt{\left(\frac{1}{\alpha}\right)^2+1} > 1$

$\Rightarrow M = -\frac{1}{\alpha} - \sqrt{\left(\frac{1}{\alpha}\right)^2+1}$ //

$Q_{1/4}$: $P(X < Q_{1/4}) = \frac{1}{4}$

2.40) X har pdf: $f(x) = cx^2 \quad 0 \le x \le 1$

a) Bestäm c $\quad \int_0^1 f(x)dx = \int_0^1 cx^2 dx = c\left[\frac{x^3}{3}\right]_0^1 = \frac{c}{3} = 1$

$\Rightarrow c = 3$

b) Bestäm cdf $\quad F(x) \overset{def:}{=} \int_0^x f(y)dy = \int_0^x 3y^2 dy = \left[y^3\right]_0^x = x^3$ ty pdf

c) Vad är $P(0.1 \le X \le 0.5) = \int_{0.1}^{0.5} f(x)dx = F(0.5) - F(0.1) =$
$= \left(\frac{1}{2}\right)^3 - \left(\frac{1}{10}\right)^3$

2.42) Hitta pdf för : X = avståndet mellan en händelse och dess närmaste granne för en Poisson(λ)-process i planet.

Y_x = # händelser i en bestämd disk med radie x.

Vi får att $Y_x \sim Poi(\lambda \pi x^2)$

$P(X \geq x) = P($  $) = P(Y_x = 1 \mid Y_x \geq 1) =$

$= P(Y_x = 0) = e^{-\lambda \pi x^2}$

$F(x) = P(X \leq x) = 1 - P(X \geq x) = 1 - e^{-\lambda \pi x^2}$

$\Rightarrow \underbrace{f(x)}_{\text{pdf}} = \underbrace{F'(x)}_{\text{derivatan av cdf}} = 2\pi\lambda x e^{-\lambda \pi x^2}$

[Om $Y \sim Poi(\lambda)$

$P(Y = k_2 \mid Y \geq k_1) =$

$= P(Y \geq k_2 - k_1)$]

2.48) $T \sim Exp(\lambda)$ och $P(T < 1) = 0.05$

$P(T<1) = 1 - P(T \geq 1) = 1 - e^{-\lambda} = 0.05 \Rightarrow$

$e^{-\lambda} = 0.95 \Rightarrow -\lambda = \log 0.95 \Rightarrow \lambda = -\log 0.95$!

koll, λ pos. ok

2.61) $X \sim Gamma(\alpha, \lambda)$ Bestäm pdf för cX

pdf för X : $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ $x \geq 0$

(Bestäm pdf genom att hitta cdf och derivera) $F(x) = P(X \leq x) = \int_0^x f(y)dy$

Låt $F_c(x)$ vara cdf för cX

$F_c(x) = P(cX \leq x) = P(X \leq \frac{x}{c}) = F(\frac{x}{c})$

Låt $f_c(x)$ vara pdf för cX

Vi har $f_c(x) = F_c'(x) = \frac{d}{dx}\left(F\left(\frac{x}{c}\right)\right) = \frac{1}{c} F'\left(\frac{x}{c}\right) =$

$= \frac{1}{c} f\left(\frac{x}{c}\right)$ Vi får: $f_c(x) = \frac{1}{c} f\left(\frac{x}{c}\right) = \frac{1}{c} \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{x}{c}\right)^{\alpha-1} e^{-\frac{\lambda x}{c}}$

$= \left(\frac{\lambda}{c}\right)^{\alpha} \cdot \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\frac{\lambda}{c}} \Rightarrow f_c(x)$ är pdf för Gamma$(\alpha, \frac{\lambda}{c})$

Dvs multiplikation med c ger en ny Gamma-fördelning. s.a. variera $\lambda \Leftrightarrow$ variera en skalnings-koefficient framför X

4.2) X har cdf $F(x) = 1 - x^{-\alpha}$, $x \geq 1$

a) Hitta $\mathbb{E}[X]$ för dom α som det $\exists$.
dvs hitta α s.a. $\mathbb{E}[X] < \infty$.

pdf för X: $f(x) = F'(x) = \alpha x^{-\alpha-1}$

$$\Rightarrow \mathbb{E}[X] = \int_1^{\infty} x f(x)dx = \int_1^{\infty} \alpha x \cdot x^{-\alpha-1} dx = \alpha \int_1^{\infty} x^{-\alpha} dx =$$

$$= \alpha \left[+ \frac{x^{-\alpha+1}}{1-\alpha} \right]_1^{\infty} = \begin{cases} \infty & \text{om } \alpha \leq 1 \\ \alpha\left(0 - \left(+\frac{1}{1-\alpha}\right)\right) = -\frac{\alpha}{1-\alpha} = \frac{\alpha}{\alpha-1} & \alpha > 1 \end{cases}$$

b) Hitta $Var(X)$ för dom α s.a. $Var(X) < \infty$

Vi beräknar: $\mathbb{E}[X^2] = \int_1^{\infty} x^2 \alpha x^{-\alpha-1} dx = \alpha \int_1^{\infty} x^{1-\alpha} dx =$

$$= \alpha \left[\frac{x^{2-\alpha}}{2-\alpha} \right]_1^{\infty} = \begin{cases} \infty & \text{om } \alpha \leq 2 \\ \alpha\left(0 - \frac{1}{2-\alpha}\right) = \frac{\alpha}{\alpha-2} & \text{om } \alpha > 2 \end{cases}$$

$$Var(X) = E[X^2] - (\mathbb{E}[X])^2 = \frac{\alpha}{\alpha-2} - \left(\frac{\alpha}{\alpha-1}\right)^2 \text{ för } \alpha > 2$$

16/4 Normalfördelningen med parametrar μ och σ^2

$X \sim N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$

$\mathbb{E}[X] = \mu \quad Var(X) = \sigma^2$

Alla pdf:er för normalfördelningarna ser ungefär lika dana ut, där μ translaterar kurvan och σ^2 bestämmer formen (flack/spetsig "puckel")

Om $\mu = 0$ och $\sigma^2 = 1$ så kallas $(N(0,1))$ för en standardiserad normalfördelning.

Om $X \sim N(\mu, \sigma^2)$ så är pdf: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Vidare gäller att om $Z := \frac{X-\mu}{\sigma}$ så är $Z \sim N(0,1)$

pdf för $Z \sim N(0,1)$ betecknas $\phi(x)$ och

$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $(\sigma = 1, \mu = 0)$

Observera att $f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$

cdf för $Z \sim N(0,1)$ betecknas $\Phi(x)$

$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \phi(x)dx$

<u>Egenskaper</u>: $P(Z > z) = 1 - \Phi(z)$

Symmetri: $P(Z < -z) = 1 - \Phi(z)$

Vi får att: $P(|Z| > z) = 2(1 - \Phi(z))$

$\Rightarrow P(-z < Z < z) = 2\phi(z) - 1$

$\boxed{P(Z > z) = P(Z \ge z)}$

ty int. ~~öve~~ fr. en pkt till samma punkt $= 0$.

<u>Tre-sigma regeln</u> Ungefär 99,74% av alla $N(\mu, \sigma^2)$ utfall ligger inom $(\mu - 3\sigma, \mu + 3\sigma)$

vi behöver i genomsnitt $\frac{1}{1-0.9974} \approx 385$ obsenationer för att få ett utfall utanför $(\mu, 3\sigma, \mu + 3\sigma)$

[Varför? Jo # gånger tills vi är utanför $(\mu - 3\sigma, \mu + 3\sigma)$ är Geom$(1 - 0.9974)$]

För stora z så gäller $1 - \Phi(z) \approx \frac{1}{z\sqrt{2\pi}} e^{-z^2/2}$, användbar approximation för vissa räkningar.

<u>Ex</u>: Låt IQ vara "ett mått på intelligens" för en viss individ.

Det gäller att $IQ \sim N(100, 15^2)$, dvs $P(IQ < 85) \approx$
≈ 0.16 $P(IQ > 130) \approx 0.024$ enl. tabell

$P(IQ > 175) \approx 3 \cdot 10^{-7}$

<u>Transformationer av slumpvariabler</u> Givet en pdf $f(x)$ för s.v. X och en icke-avtagande funktion $h(x)$, vill vi beräkna pdf $g(x)$ för $Y := h(X)$.

Vi får: $G(y) = P(Y \le y) = P(h(X) \le y) = P(X \le h^{-1}(y)) =$
$= F(h^{-1}(y))$

h är icke avtagande

$g(y) = G'(y) = \frac{d}{dy} F(h^{-1}(y)) = (h^{-1}(y))' \cdot F'(h^{-1}(y)) =$

$= \frac{1}{h'(h^{-1}(y))} f(h^{-1}(y))$

inversa fkn satsen

Om $h(x)$ är en icke-växande funktion
får vi $g(y) = \frac{1}{|h'(h^{-1}(y))|} f(h^{-1}(y))$

Ex: Linjär transformation: dvs $h(x) = ax + b \Rightarrow$
$Y = h(X) = aX + b$

$h'(x) = a$ för att beräkna $h^{-1}(y)$:

$y = h(x) = ax + b \Rightarrow x = \frac{y-b}{a} = h^{-1}(y)$

$\Rightarrow g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right)$

Ex: $Z \sim N(0,1)$ $Y = Z^2$ (kvadrerad stand. Norm.)

$h(y)$ ej monoton men: $G(y) = P(Y \le y) = P(Z^2 \le y) =$
$= P(-\sqrt{y} \le Z \le \sqrt{y}) = 2\Phi(\sqrt{y}) - 1 \Rightarrow g(y) = G'(y) =$
$= 2 \cdot \frac{1}{2\sqrt{y}} \phi(\sqrt{y}) =$

derivatan av Φ

$= \frac{1}{\sqrt{2\pi y}} e^{-y/2}$

Övning: $Y \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$? Kolla hemma...

Ex $U \sim U(0,1)$ $Y = \frac{1}{U}$ dvs $h(x) = \frac{1}{x} \Rightarrow h'(x) = -\frac{1}{x^2}$

och $h^{-1}(y) = \frac{1}{y} \Rightarrow (h^{-1}(y))' = -\frac{1}{y^2}$

$\Rightarrow g(y) = |(h^{-1}(y))'| \underbrace{f(h^{-1}(y))}_{} = \frac{1}{y^2}$ för $y > 1$

$E[Y] = \int_1^\infty y \cdot \frac{1}{y^2} dy = \int_1^\infty \frac{1}{y} dy \text{ "}= \infty\text{"}$ för $f(x)$ är pdf för $U(0,1)$ (E finns inte)

Vad är nyttan? Antag att vi vill generera X med cdf $F(x)$, låt $U \sim U(0,1)$ (Matlab), låt då $X = F^{-1}(U)$, räkningar som ovan (övning) ger att X har cdf $F(x)$.

Ex: Antag att vi vill ha en $Exp(\lambda)$-sv.

$F(x) = 1 - e^{-\lambda x}$, sätt $y = 1 - e^{-\lambda x} \Rightarrow e^{-\lambda x} = 1 - y$

$\Rightarrow -\lambda x = \log(1-y) \Longrightarrow x = -\frac{\log(1-y)}{\lambda} = F^{-1}(y)$

Simulera en $U \sim U(0,1)$, då är $X = -\frac{\log(1-U)}{\lambda} =$

$= -\frac{\log(U)}{\lambda} \sim Exp(\lambda)$

Väntevärde och Varians av transformerade sv

Om $Y = h(X) = aX + b \Rightarrow E[\cancel{X}] = aE[X] + b$

$Var(Y) = a^2 Var(X)$, Om $Y = h(X) = aX + b + c(X - E[X])^2$

$\Longrightarrow E[Y] = aE[X] + b + c\,Var(X)$

I allmänhet kan vi använda Taylor-utveckling på $h(x)$.

$h(x) \approx h(E[X]) + (x - E[X])h'(E[X]) + \frac{1}{2}(X - E[X])^2 h''(E[X])$

$E[h(X)] = h(E[X]) + 0 + \frac{1}{2}Var(X)h''(E[X])$

Fundera på hemma: Vad är fördelningen för

1) Väntetiden innan bussen kommer?

2) Dagliga utgifter.

3) Hur många barn innan första dotter/son.

4) Fördeln. för mänsklig livslängd.

2.52) En population X = längden hos en individ

Vi antar $X \sim N(\mu, \sigma^2)$ $\mu = 70$, $\sigma = 3$

Räkne-stuga

Vilken proportion har längd minst 6 fot?

6 fot = 12 · 6 = 72 inch

$P(X \geq 72)$? Sätt $Z = \frac{X - 70}{3} \sim N(0,1)$

$= P(Z \geq \frac{2}{3}) = 1 - \Phi(\frac{2}{3}) = 1 - 0{,}75 = 0.25$

enl. tabell

55) $X \sim N(\mu, \sigma^2)$ hitta c s.a. $P(\mu - c \leq X \leq \mu + c) =$
$= 0.95$

$P(\mu - c \leq X \leq \mu + c) = P(-c \leq X - \mu \leq c) = P(-\frac{c}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{c}{\sigma})$

$P(|Z| \geq z) = 0.95 \Rightarrow z = 1.96$ $\Rightarrow \frac{c}{\sigma} = 1{,}96$

$\Rightarrow c = 1{,}96\sigma$

4.6) X är en sv med pdf $f(x) = 2x$ $0 \leq x \leq 1$

a) $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \, 2x dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$

b) Låt $Y = X^2$, hitta pdf för Y

Kalla cdf för Y $F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y})$

$= \int_0^{\sqrt{y}} 2x dx = y \Rightarrow f_Y(y) = 1$ //

$\Rightarrow \mathbb{E}[Y] = \int_0^1 y dy = \frac{1}{2}$ //

c) Använd $\mathbb{E}[g[X]] = \int_{-\infty}^{\infty} g(x) f(x) dx$, vi har $g(x) = x^2$

$\Rightarrow \mathbb{E}[X^2] = \int_0^1 x^2 \cdot 2x dx = 2\left[\frac{x^4}{4}\right]_0^1 = \frac{1}{2}$

d) $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{9-8}{18} = \frac{1}{18}$ //

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

$= \int_0^1 (x - \mathbb{E}[X])^2 2x dx = \ldots = \frac{1}{18}$

13) $X \geq 0$ är en kont. s.v. Visa $\mathbb{E}[X] = \int_0^\infty [1 - F(x)]dx$

B: $\mathbb{E}[X] = \int_0^\infty y f(y) dy = \int_0^\infty f(y) \int_0^y 1 dx\, dy = \int_0^\infty f(y) \int_0^\infty \mathbb{1}_{\{x<y\}} dx$

$\mathbb{1}_{\{x<y\}}(x) = \begin{cases} 1 & x<y \\ 0 & \text{annars} \end{cases}$

$\begin{cases} 1 & \text{om händelsen sann} \\ 0 & \text{annars} \end{cases}$

$= \int_0^\infty \int_0^\infty f(y) \mathbb{1}_{\{x<y\}} dx dy = \int_0^\infty \int_0^\infty f(y) \mathbb{1}_{\{x<y\}} dy dx =$

$= \int_0^\infty \int_x^\infty f(y) dy dx = \int_0^\infty \left(1 - \int_0^x f(y) dy\right) dx = \int_0^\infty [1 - F(x)] dx$ $\square$

Låt $T \sim Exp(\lambda)$ $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$

$\Rightarrow 1 - F(t) = e^{-\lambda t}$ $\Rightarrow \mathbb{E}[T] = \int_0^\infty e^{-\lambda t} dt = \left[\frac{-e^{-\lambda t}}{\lambda}\right]_0^\infty =$

$= \left(0 - \left(-\frac{1}{\lambda}\right)\right) = \frac{1}{\lambda}$ //

4.26 Vi bryter en sticka, hitta väntevärdet av kvoten mellan den långa och den korta delen.
Vi antar att den bryts av på ett likformigt fördelat ställe. Kvoten blir $(U \sim U(0,1))$

$\frac{\max(U, 1-U)}{\min(U, 1-U)}$ $\mathbb{E}\left[\frac{\max(U,1-U)}{\min(U,1-U)}\right] = \int_0^1 \frac{\max(u,1-u)}{\min(u,1-u)} 1 du$

$\mathbb{E}[g(X)] = \int_0^1 g(x) f(x) dx$ $= \int_0^{1/2} \frac{1-u}{u} du + \int_{1/2}^1 \frac{u}{1-u} du =$

$= 2 \int_{1/2}^1 \frac{u}{1-u} du = 2 \int_{1/2}^1 \frac{u-1}{1-u} du + 2 \int_{1/2}^1 \frac{1}{1-u} du = -1 + 2\left[-\log(1-u)\right]_{1/2}^1$

$" = \infty "$ $\Rightarrow$ väntevärdet $\nexists$

Gemensamma fördelningar (joint distributions)

21/4 Mån
LV4

(diskret) <u>Def</u>: X och Y är två diskreta s.v. så betecknar vi den <u>gemensamma pmf:en</u>:

$p_{XY}(x,y) := \mathbb{P}(X=x, Y=y)$ och dom två <u>marginal-fördelningarna</u> betecknas $p_X(x) = \mathbb{P}(X=x)$ och $p_Y(y) = \mathbb{P}(Y=y)$

observera att: $p_X(x) = \sum_y p_{XY}(x,y)$

<u>Def</u>: <u>Betingade pmf</u>: Om X och Y är <u>slumpvariabler</u> så definierar vi: i) Den betingade pmf:en för Y givet X: $p_{Y|X}(k|n) := \dfrac{p_{XY}(n,k)}{p_X(n)} = \dfrac{\mathbb{P}(X=k, Y=n)}{\mathbb{P}(X=n)}$

ii) Den beting. pmf:en för X givet Y:

$$p_{X|Y}(n|k) := \frac{p_{XY}(n,k)}{p_Y(k)}$$

<u>Def</u>: slumpvariablerna X och Y är <u>oberoende</u> om

$$p_{XY}(x,y) = p_X(x)p_Y(y) \quad \forall\, x,y$$

[jämför: $A = \{X=x\}$ $B=\{Y=y\}$ oberoende om

$p_{XY}(A,B) = p_X(A)p_Y(B)$]

<u>Ex</u>: Tärning/slant singling: i) kastar tärning X = värdet ii) Vi singlar slant X ggr. $Y = \#H$

Vad är $p_X(n)$? $p_X(n) = \mathbb{P}(X=n) = \frac{1}{6}$ $n=1,2,\dots,6$

$X \sim U(\{1,2,\dots,6\})$

$$p_{Y|X}(k|n) = \mathbb{P}(Y=k|X=n) = \binom{n}{k}2^{-k}2^{-n+k} = \binom{n}{k}2^{-n}$$

$k=0,1,\dots,n$

$$\Rightarrow p_{XY}(n,k) = p_X(n)\,p_{Y|X}(k|n) = \frac{1}{6}\binom{n}{k}2^{-n}$$

$1 \le n \le 6$

$0 \le k \le n$

$n,k \in \mathbb{N}$

Även om $X \sim U(\{1, \dots, 6\})$ så är X givet Y helt annorlunda!

<u>Kontinuerliga fallet</u> <u>Def</u>: Gemensamma fördelningen av X, Y (pdf:en) betecknas $f_{XY}(x,y)$ och definieras av: $P(a<X<b, c<Y<d) = \int_a^b \int_c^d f_{XY}(x,y)\,dx\,dy$

Vi kan visualisera $f_{XY}(x,y)$ som en yta över $\mathbb{R}^2$, Denna yta beskriver den gemensamma densiteten.

<u>Def</u>: Marginal-fördelningarna (densiteterna) för X, Y betecknas $f_X(x)$, $f_Y(y)$ och definieras av:

$$f_X(x) := \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

(jämför: $P_X(x) = \sum_y P_{XY}(x,y)$ i det diskreta fallet)

<u>Betingade pdf</u>: Betecknas $f_{X|Y}(x|y)$ och $f_{Y|X}(y|x)$ (I:a. Densiteten av X givet Y) och definieras av: $f_{X|Y}(x|y) := \dfrac{f_{XY}(x,y)}{f_Y(y)}$ resp. $f_{Y|X}(y|x) := \dfrac{f_{XY}(x,y)}{f_X(x)}$

Om $f_{XY}(x,y)$ är en yta så är $f_{Y|X}(y|x)$ en profil av ytan längs snittet $\{X = x\}$

<u>Def</u>: X och Y är <u>oberoende</u> om $f_{XY}(x,y) = f_X(x) f_Y(y)$

Vi får: $P(a<X<b, c<Y<d) = \int_a^b \int_c^d f_{XY}(x,y)\,dy\,dx \overset{\text{def. ober.}}{=}$

$$\int_a^b \int_c^d f_X(x) f_Y(y)\,dy\,dx = \int_a^b f_X(x)\,dx \int_c^d f_Y(y)\,dy = P(a<X<b) \cdot P(c<Y<d)$$

Ex: (X, Y) är likformigt fördelade på enhetskv. $[0,1]^2$

Om densiteten $f_{XY}(x,y) = 1$

Marginalfördelning för X: $f_X(x) = \int_0^1 f_{XY}(x,y)\,dy = \int_0^1 1 \cdot dy = 1$

pss $f_Y(y) = 1 \Rightarrow X \sim U(0,1)$ (marginalen för X). Dessutom är X, Y oberoende då $f_{XY}(x,y) = 1$

$= 1 \cdot 1 = f_X(x) f_Y(y)$

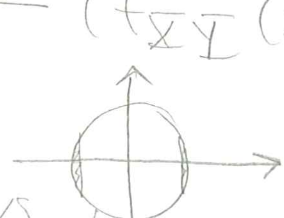
Ex: (X, Y) är likformigt fördelade på $D := \{x^2 + y^2 \leq 1\}$

Om $f_{XY}(x,y) = \frac{1}{\pi}$ (ty arean är π); vi får: $f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy =$

$= \frac{2\sqrt{1-x^2}}{\pi}$ pss. $f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi}$

X och Y är beroende $(f_{XY}(x,y) \neq f_X(x) f_Y(y))$:

"om $|X|$ är nära 1



$\Rightarrow$ $|Y|$ är nära 0." dvs beroende.

Vi hade kunnat betrakta (R, θ) istället, då hade R, θ blivit oberoende $\theta \sim U([0, 2\pi])$
$R \sim$?

Betingade väntevärden och betingad varians. Det betingade väntevärdet av Y givet X betecknas $\mathbb{E}[Y|X]$. Den –"– variansen –"– $Var(Y|X)$. Dessa är slumpvariabler. Vi har också $\mathbb{E}[Y|X=k]$

$= \sum_y y\, p_{Y|X}(y|k)$ (diskret)

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

Law of Total Expectation och L. of T. Variance

LTE: $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$

"B": $\mathbb{E}[\mathbb{E}[Y \mid X]] = \sum_x \mathbb{E}[Y \mid X = x] p_X(x) =$

$= \sum_x \sum_y y\, p_{Y|X}(y|x)\, p_X(x) = \sum_y y \sum_x p_{Y|X}(y|x)\, p_X(x) =$

$= \sum_y y\, p_Y(y) = \sum_y y\, P(Y = y) = \mathbb{E}[Y] \quad \square$

LTV: $\mathrm{Var}(Y) = \mathrm{Var}[\mathbb{E}[Y \mid X]] + \mathbb{E}[\mathrm{Var}(Y \mid X)]$

Ex: Lås med 10 000 kombinationer.
Tiden i timmar innan vi hittar rätt är
1) $S \sim U([0,10])$ om vi letar systematiskt.
2) $R \sim Exp(0,1)$ — " — ej — " — $\left[P(A) = \mathbb{E}[1_{\{A\}}]\right]$
Vad är $P(R > S)$? $P(R>S) = \mathbb{E}[P(R > S \mid S)] = \mathbb{E}[e^{-0.1S}]$
$= \int_0^{10} e^{-0.1x} \frac{1}{10} dx = \int_0^1 e^{-x} dx = 1 - e^{-1} = 0.632$

Optimal prediktor: Vi har ett experiment och vill avgöra det (slumpmässiga) värdet på Y. Vi kan dock ej mäta Y men vi vet att (X, Y) har en viss gemensam fördelning. Antag att vi kallar uppskattningen av Y för $\hat{Y}$, vad är bästa valet av $\hat{Y}$?
Låter $\hat{Y} = h(X)$, hur skall jag välja h s.a. $\mathbb{E}[(Y - \hat{Y})^2]$ blir minimalt? $\mathbb{E}[(Y-\hat{Y})^2] = \mathbb{E}[(Y - h(X))^2] =$

$= \mathbb{E}[\mathbb{E}[(Y - h(X))^2 \mid X]] \overset{\text{(lägger till och drar ifrån)}}{=} \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2 \mid X]]$
$+ \mathbb{E}[\mathbb{E}[(\mathbb{E}[Y \mid X] - h(X))^2 \mid X]] + 2\mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y \mid X])(\mathbb{E}[Y \mid X]$
$- h(X)) \mid X]] \Rightarrow \mathbb{E}[(Y - \hat{Y})^2]$ minimeras om $\hat{Y} = h(X) = \mathbb{E}[Y \mid X]$

Räkneövn 22/4 LV4

3.8/ X och Y har den gemensamma pdf:en $f_{XY}(x,y) =$

$$= \frac{6}{7}(x+y)^2, \quad 0 \le x \le 1, \quad 0 \le y \le 1$$

a) i) Hitta $P(X > Y)$ $\quad P(X>Y) = \mathbb{E}\left[1_{\{X>Y\}}\right] =$

$$= \int_0^1\int_0^1 1_{\{x>y\}} f(x,y)\,dx\,dy = \int_0^1\int_y^1 \frac{6}{7}(x+y)^2\,dx\,dy = \int_0^1 \frac{6}{7}\left[\frac{(x+y)^3}{3}\right]_y^1 dy =$$

(0 annars)

$$= \frac{6}{7}\int_0^1 \frac{(1+y)^3}{3} - \frac{(2y)^3}{3}\,dy = \frac{2}{7}\left[\frac{(1+y)^4}{4} - 8\frac{y^4}{4}\right]_0^1 =$$

$$= \frac{2}{7}\left(\frac{2^4}{4} - 2\cdot 1^4 - \left(\frac{1}{4}\right)\right) = \dots = \frac{1}{2} //$$

ii) $P(X+Y \le 1) = \mathbb{E}\left[1_{\{X+Y\le 1\}}\right] = \int_0^1\int_0^1 1_{\{x+y\le 1\}} f(x,y)\,dx\,dy$

$$= \int_0^1\int_0^{1-y} \frac{6}{7}(x+y)^2\,dx\,dy = \int_0^1 \frac{6}{7}\left[\frac{(x+y)^3}{3}\right]_0^{1-y} dy = \int_0^1 \frac{6}{7}\left(\frac{1}{3} - \frac{y^3}{3}\right) dy =$$

$$= \frac{2}{7}\left[y - \frac{y^4}{4}\right]_0^1 = \frac{2}{7}\left(1 - \frac{1}{4}\right) = \frac{2}{7}\cdot\frac{3}{4} = \frac{3}{14} //$$

(måste kolla att svaret ligger mellan 0 och 1)

iii) $P(X \le \frac{1}{2}) = \int_0^1\int_0^1 1_{\{x\le\frac{1}{2}\}} f(x,y)\,dx\,dy =$

$$= \int_0^1\int_0^{1/2} \frac{6}{7}(x+y)^2\,dx\,dy = \int_0^1 \frac{6}{7}\left[\frac{(x+y)^3}{3}\right]_0^{1/2} dy = \int_0^1 \frac{2}{7}\left(\left(\frac{1}{2}+y\right)^3 - y^3\right) dy$$

$$= \dots = \frac{2}{7} //$$

b/ Hitta marginalfördelningarna för X och Y

$$f_X(x) = \int_0^1 f(x,y)\,dy = \int_0^1 \frac{6}{7}\left[\frac{(x+y)^3}{3}\right]_0^1 = \frac{2}{7}\left((1+x)^3 - x^3\right)$$

p.s.s. $f_Y(y) = \frac{2}{7}\left((1+y)^3 - y^3\right)$

c) Hitta $f_{X|Y}(x,y)$ och $f_{Y|X}(y|x)$

$$f_{X|Y}(x|y) \overset{\text{jmfr diskreta}}{=} \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{\frac{6}{7}(x+y)^2}{\frac{2}{7}\left((1+y)^3 - y^3\right)} = \frac{3(x+y)^2}{\left((1+y)^3 - y^3\right)}$$

3.14/ En punkt är likformigt fördelad i enhetssfären i $\mathbb{R}^3$.

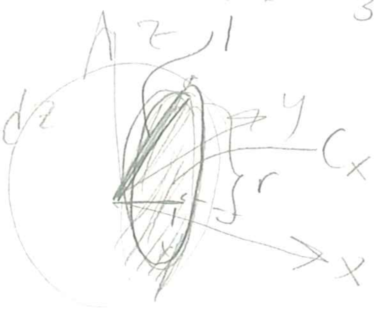
a) Hitta marginalfördelningarna för x, y, z -koord.

Vad är $f_{XYZ}(x,y,z)$? $\Rightarrow f_{XYZ}(x,y,z) = \frac{1}{Vol(B_1)} = \frac{1}{\frac{4\pi}{3}} = \frac{3}{4\pi}$ //

$\left(\frac{1}{\text{volymen}}\right)$

$f_X(x) = \text{Area}(C_x) \cdot \frac{3}{4\pi} = \iint_{C_x} \frac{3}{4\pi} \, dy dz$

dens.



$= \frac{3}{4\pi} \cdot \pi r^2 = \frac{3}{4}(1-x^2)$ //

$1 = x^2 + r^2$

Pyth.

$\Rightarrow r^2 = 1 - x^2$

b/ Hitta $f_{XY}(x,y)$ Vi har: $x^2+y^2+z^2 \leq 1$

$\Rightarrow -\sqrt{1-x^2-y^2} \leq z \leq \sqrt{1-x^2-y^2}$

$f_{XY}(x,y) = \int_{-\infty}^{\infty} f_{XYZ}(x,y,z) dz = \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} \frac{3}{4\pi} dz = \frac{3}{2\pi}\sqrt{1-x^2-y^2}$ //

c) Hitta fördelningen av X, Y betinget på $Z = 0$.

$f_{XY|Z}(x,y|0) = \frac{f_{XYZ}(x,y,0)}{f_Z(0)} = \frac{\frac{3}{4\pi}}{\frac{3}{4}(1-0^2)} = \frac{1}{\pi}$ //

3.10 Antag att $f_{XY}(x,y) = xe^{-x(y+1)}$, $0 \leq x < \infty$, $0 \leq y < \infty$

a/ Hitta $f_X(x)$, $f_Y(y)$: $f_X(x) = \int_0^\infty xe^{-x(y+1)} dy =$

$= \left[-e^{-x(y+1)}\right]_0^\infty = 0 - (-e^{-x}) = e^{-x}$ //

$f_Y(y) = \int_0^\infty xe^{-x(y+1)} dx \overset{p.i}{=} \left[-x\frac{e^{-x(y+1)}}{y+1}\right]_0^\infty - \int_0^\infty -\frac{e^{-x(y+1)}}{y+1} dx =$

$= \left[-\frac{e^{-x(y+1)}}{(1+y)^2}\right]_0^\infty = 0 + \frac{1}{(1+y)^2} = \frac{1}{(1+y)^2}$ //

$$P(A) = \int_A f_{XY}(x,y)\,dx\,dy$$

$$= \int f_{XY}(x,y) \underbrace{1_A}_{g(x,y)}\,dx\,dy = E[g(X,Y)]$$

$$E[g(X,Y)] := \int g(x,y) f(x,y)\,dx\,dy$$

$P(X > Y) = E[1_{\{X>Y\}}]$

---

b) $f_{X|Y}(x|y) = \dfrac{f_{XY}(x,y)}{f_Y(y)} = \dfrac{xe^{-x(y+1)}}{\frac{1}{(1+y)^2}} = (1+y)^2 xe^{-x(y+1)}$

$f_{Y|X}(y|x) = \dfrac{xe^{-x(y+1)}}{e^{-x}} = xe^{-xy}$

3.23 X givet N är $Bin(N,p)$ $N \sim Bin(m,r)$

vad har X för fördelning?

$$P(X=k) = \sum_{n=0}^{m} P(X=k \mid N=n)\,P(N=n) = \sum_{n=k}^{m} P(X=k \mid N=n)\,P(N=n)$$

$$= \sum_{n=k}^{m} \binom{n}{k} p^k (1-p)^{n-k} \binom{m}{n} r^n (1-r)^{m-n} = \sum_{n=k}^{m} \frac{\cancel{n!}}{k!(n-k)!} \frac{m!}{\cancel{n!}(m-n)!} p^k (1-p)^{n-k} r^n (1-r)^{m-n}$$

$$\overset{n \to n+k}{=} \sum_{n=0}^{m-k} \frac{m!}{k!\,n!\,(m-k-n)!} p^k (1-p)^n r^{n+k} (1-r)^{m-k-n} =$$

$$= (pr)^k \frac{m!}{k!(m-k)!} \sum_{n=0}^{m-k} \frac{(m-k)!}{n!(m-k-n)!} (r(1-p))^n (1-r)^{m-k-n} =$$

$$= (pr)^k \binom{m}{k} \sum_{n=0}^{m-k} \binom{m-k}{n} (r(1-p))^n (1-r)^{m-k-n} =$$

$$= (pr)^k \binom{m}{k} (r(1-p) + 1 - r)^{m-k} = \binom{m}{k} (rp)^k (1-rp)^{m-k} \Longrightarrow$$

$$\Longrightarrow X \sim Bin(m, rp)$$

$$\left[\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n\right]$$

Experiment  m st.

sannolikhet för $\boxtimes$ är rp.

dvs vi har m oberoende experiment som lyckas med sann. $rp \Rightarrow Bin(m, rp)$

/ indikerar lyckas (map r) $\sim Bin(m,r)$

$\underline{3.48}$ $N_1 \sim Poi(\lambda_1)$ $N_2 \sim Poi(\lambda_2)$ oberoende

Visa $N_1 + N_2 \sim Poi(\lambda_1 + \lambda_2)$

$\underline{B:}$ $P(N_1 + N_2 = k) = \sum_{n=0}^{k} P(N_1 = k-n \mid N_2 = n) P(N_2 = n) =$

ober.

$$\stackrel{\text{ober.}}{=} \sum_{n=0}^{k} P(N_1 = k-n)\, P(N_2 = n) = \sum_{n=0}^{k} \frac{\lambda_1^{k-n}}{(k-n)!} e^{-\lambda_1} \frac{\lambda_2^{n}}{n!} e^{-\lambda_2} =$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{n=0}^{k} \frac{k!}{n!(n-k)!} \lambda_1^{k-n} \lambda_2^{n} = \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{n=0}^{k} \binom{k}{n} \lambda_1^{k-n} \lambda_2^{n} =$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k \quad \square$$

3.65) $T_1, T_2, \ldots, T_n$ är ober. $T_i \sim Exp(\lambda_i)$

$T := \min(T_1, T_2, \ldots, T_n)$ $\quad T \sim Exp\left(\sum_{i=1}^{n} \lambda_i\right)$

$$P(T > t) = P(\min(T_1, T_2, \ldots, T_n) > t) = P(T_1 > t, T_2 > t, \ldots, T_n > t) = \prod_{i=1}^{n} P(T_i > t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-\sum_{i=1}^{n} \lambda_i t} =$$

$$= e^{-t \sum_{i=1}^{n} \lambda_i} \Rightarrow T \sim Exp\left(\sum_{i=1}^{n} \lambda_i\right) \quad \square$$

$\underline{23/4}$ $\underline{\text{Covarians och correlation}}$

$\underline{\text{Def}}$: X och Y sv då betecknas covariansen $Cov(X,Y)$ och definieras av: $COV(X,Y) := E[(X - E[X])(Y - E[Y])]$

Vi får: $Var(X+Y) = Var(X) + Var(Y) + 2COV(X,Y)$ och

$COV(X,Y) = E[XY] - E[X]E[Y]$

(kolla räkn. hemma)

$\underline{\text{Def}}$: Korrelationskoefficienten för X, Y betecknas ρ och definieras av: $\rho := \frac{COV(X,Y)}{\sqrt{Var(X)Var(Y)}}$, $-1 \leq \rho \leq 1$

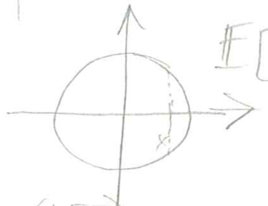
Vi säger att X, Y är <u>okorrelerade</u> om $\rho = 0$

Då gäller: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ (oberoende $\Rightarrow$ okorrelerade) men ej $\Leftarrow$

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

<u>Ex</u>: (X,Y) är likformigt fördelade på enhetsdisken. (i $\mathbb{R}^2$)

$\mathbb{E}[Y|X] = 0$ [för varje x: $\mathbb{E}[Y|X=x] = 0$ pga symmetri.]



$\mathbb{E}[XY] = \mathbb{E}[X\,\mathbb{E}[Y|X]] \overset{(LTE)}{=} \mathbb{E}[X \cdot 0] = 0 \Rightarrow Cov(X,Y) = 0$

$\mathbb{E}[X] = \mathbb{E}[Y] = 0 \Rightarrow \rho = 0$

$$\boxed{X \text{ och } Y \text{ oberoende} \Rightarrow \rho = 0 \quad \not\Leftarrow}$$

Vi känner (eller kan mäta X) men vill veta Y. Det bästa vi kan göra är att uppskatta Y mha $\hat{Y} = \mathbb{E}[Y|X]$

Problem, vi måste veta den gemensamma fördelningen för (X, Y). Om vi ej vet denna testar vi $\hat{Y} = \alpha + \beta X$,

Optimering ger: $\alpha = \mathbb{E}[Y] - \beta\,\mathbb{E}[X]$, $\beta = \frac{\mathbb{E}[Y]}{\mathbb{E}[X]}\rho$

$$\Rightarrow \hat{Y} = \mathbb{E}[Y] + \frac{\mathbb{E}[Y]}{\mathbb{E}[X]}\rho(X - \mathbb{E}[X])$$

<u>Multinomialfördelningen</u> med parametrar $n, p_1, \dots, p_r$

$n \in \mathbb{N}$ $\quad p_1 + \dots + p_r = 1$

Vi skriver $(X_1, X_2, \dots, X_r) \sim M_n(n, p_1, \dots, p_r)$

$$P(X_1 = k_1, X_2 = k_2, \dots X_r = k_r) = \binom{n}{k_1, k_2, \dots k_r} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

multinomialkoeff.

Uppkommer när vi upprepar ett exper. n ggr (oberoende) med $r = \#$ möjliga utfall. $\quad k_1 + k_2 + \dots + k_r = n$

($r = 2$ ger spec. fall Binomialfförd.)

$X_i = \#$ utfall av typ i.

Marginalfördelning: $X_i \sim Bin(n, p_i)$ $\quad \mathbb{E}[X_i] = np_i$

, $Var(X_i) = n \cdot p_i q_i$ $\quad (q_i = 1 - p_i)$ $\quad X_i + X_j \sim Bin(n, p_i + p_j)$

$\mathrm{Var}(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j)$

$\mathrm{Cov}(X_i, X_j) = \{\text{kontrollera!! (ehradare)}\} = -np_ip_j$

X_i och X_j är negativt korrelerade med $\rho_{ij} = -\sqrt{\frac{p_ip_j}{q_i q_j}}$

intuitivt: "om utfall i inträffar många gånger $\Rightarrow$ det finns mindre utrymme för utfall j"

$i = 1,2,3,4,5,6$

Ex: Yatzy. Vi kastar fem tärningar $X_i = \#i$

$(X_1, X_2, X_3, X_4, X_5, X_6) \sim Mn(5, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$

$\rho_{ij} = \ldots = -0{,}2$

"Varje fix sekvens av utfall med k_i utfall av typ i så har denna sekvens sannolikhet $p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$. Antalet sådana sekvenser är $\binom{n}{k_1, k_2, \ldots, k_r}$"

Bivariat normalfördelning (kontin.)

$(X, Y) \sim N(\underbrace{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho}_{5 \text{ parametrar}})$

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right)$$

Man kan visa att marginal-fördeln. $X \sim N(\mu_1, \sigma_1^2)$
$Y \sim N(\mu_2, \sigma_2^2)$

Vi har att: $\mathbb{E}[Y \mid X] = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(X - \mu_1)$

Moment-genererande funktion (mgf) Def: För en s.v. X definieras mgf: en (betecknas $M_X(t)$ eller $M(t)$)

$M_X(t) := \mathbb{E}[e^{tX}]$ (om väntevärdet $\exists$)

1) Om $M(t)$ $\exists$ i ett intervall $t \in (-\varepsilon, \varepsilon)$ så bestämmer $M(t)$ fördelningen för X unikt dvs om X, Y är s.v. och $M_X(t) = M_Y(t)$

$\forall\, t \in (-\varepsilon, \varepsilon) \iff X$ och Y har samma fördelning $(X \overset{d}{=} Y)$

2) $M^{(n)}(0)$ = n:te derivatan av $M(t)$ i $t=0$ $n \in \mathbb{N}$

$\rightarrow \mathbb{E}[X^n] = M^{(n)}(0)$ $\mathbb{E}[X^n]$ = n:te <u>momentet</u> av X.

3) $\mathbb{E}[e^{tX+sY}] =$ (X, Y gemensamt fördelade sv)

$= \mathbb{E}[e^{tX}]\mathbb{E}[e^{sY}]$ $\forall (s,t) \in (-\varepsilon, \varepsilon) \iff X$ och Y oberoende

Ex: 1) $X \sim$ Bernoulli med par. p $\begin{bmatrix} & \text{sann.} \\ 1 \to & p \\ 0 \to & 1-p \end{bmatrix}$

$\Rightarrow M(t) = \mathbb{E}[e^{tX}] = e^{t\cdot 0}\mathbb{P}(X=0) + e^{t\cdot 1}\mathbb{P}(X=1) = 1-p+pe^t$

2) $X \sim Bin(n,p)$ $M(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{tk}$

$\overset{\text{Bin.satsen}}{=} (pe^t + 1 - p)^n$

3) $X \sim Geom(p)$ $M(t) = \mathbb{E}[e^{tX}] = \sum_{k=1} e^{tk}(1-p)^{k-1}p =$

$= pe^t \sum_{k=1}^{\infty} e^{t(k-1)}(1-p)^{k-1}$ $pe^t \sum_{k=0}^{\infty} (e^t(1-p))^k = \dfrac{pe^t}{1-e^t(1-p)}$

om $e^t(1-p) < 1 \Rightarrow t < -\log(1-p)$

4) $Nb(r,p)$ $M(t) = \ldots = \left(\dfrac{pe^t}{1-e^t(1-p)}\right)^r$

(neg bin)

5) $Poi(\lambda)$: $M(t) = \sum_{k=0}^{\infty} e^{tk} \dfrac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} \dfrac{(\lambda e^t)^k}{k!} e^{-\lambda} = e^{\lambda e^t} e^{-\lambda} = e^{\lambda(e^t-1)}$

$= e^{-\lambda} e^{\lambda(e^t)} = e^{\lambda(e^t-1)}$?

Kontinuerliga förd:

1) $Exp(\lambda)$: $M(t) = \mathbb{E}[e^{tX}] = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{(t-\lambda)x} dx =$

$= \lambda \left[\dfrac{e^{(t-\lambda)x}}{t-\lambda}\right]_0^{\infty} = \lambda\left(0 - \dfrac{1}{t-\lambda}\right) = \dfrac{\lambda}{\lambda - t}$ om $t<\lambda$

2) $Gamma(\alpha, \lambda)$: $M(t) = \ldots = \left(\dfrac{\lambda}{\lambda - t}\right)^{\alpha}$ $t<\lambda$

3) $N(0,1)$: $M(t) = \ldots = e^{t^2/2}$

Ex: $X_1, X_2, \ldots, X_k$ ober. $Exp(\lambda)$ $Y := X_1 + \ldots + X_k$

$$M_Y(t) = \mathbb{E}\left[e^{t(X_1+\ldots+X_k)}\right] \underset{\text{ober.}}{=} \prod_{i=1}^{k} \mathbb{E}\left[e^{tX_i}\right] = \prod_{i=1}^{k} \frac{\lambda}{\lambda - t} =$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^k \Longrightarrow Y \text{ är Gamma}(k,\lambda)\text{-fördelad.}$$

Ex: $X \sim \text{Gamma}(r,\lambda)$ Y givet X är $Poi(X)$.

Vad har $Y + R$ för fördelning?

$$M_{Y+r}(t) = \mathbb{E}\left[e^{t(Y+r)}\right] = e^{tr}\mathbb{E}\left[e^{Yt}\right] \overset{\text{LTE}}{=} e^{tr}\mathbb{E}\left[\mathbb{E}\left[e^{tY} \mid X\right]\right]$$

$$= e^{tr}\mathbb{E}\left[e^{X(e^t-1)}\right] = e^{tr}\left(\frac{\lambda}{\lambda-(e^t-1)}\right)^r = \left(\frac{\frac{\lambda}{1+\lambda}e^t}{1-\frac{1}{1+\lambda}e^t}\right)^r$$

$$\Longrightarrow Y + r \sim Nb\left(r, \frac{\lambda}{1+\lambda}\right)$$

4.36 $X \sim U(0,1)$, $Y = \sqrt{X}$ Räknestuga 2 23/4

Hitta $\mathbb{E}[Y]$ mha a) pdf:en för Y.

Vi börjar med att hitta cdf:en.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\sqrt{X} \leq y) = \mathbb{P}(X \leq y^2) = \int_0^{y^2} 1\,dx = y^2$$

y^2 ty likf. fördelad

$$\Longrightarrow f_Y(y) = F_Y'(y) = 2y \qquad 0 \leq y \leq 1$$

$$\mathbb{E}[Y] = \int_0^1 y\, \underbrace{2y}_{\text{densiteten}}\, dy = \left[2\frac{y^3}{3}\right]_0^1 = \frac{2}{3}$$

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

$$\mathbb{E}[\sqrt{X}] = \int_0^1 x^{1/2} \cdot 1\,dx = \left[\frac{2}{3}x^{3/2}\right]_0^1 = \frac{2}{3} \;//$$

c) Visa: $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

B: $Var(X-Y) = \mathbb{E}\left[(X - Y - \mathbb{E}[X] + \mathbb{E}[Y])^2\right] =$

$= \mathbb{E}\left[(X-\mathbb{E}[X])^2 + (Y-\mathbb{E}[Y])^2 - 2(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right]$

$= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] + \mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] - 2\mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right]$

$= Var(X) + Var(Y) - 2Cov(X,Y)$

4.54) X, Y och Z är okorrelerade. Låt $U = Z + X$, $V = Z + Y$

Bestäm $Cov(U,V)$ och ρ_{UV}

$Cov(U,V) = \mathbb{E}[(U-\mathbb{E}[U])(V-\mathbb{E}[V])] = \mathbb{E}[(Z+X-\mathbb{E}[Z]-\mathbb{E}[X])(Z+Y-\mathbb{E}[Z]-\mathbb{E}[Y])] =$

$= \mathbb{E}\left[(Z-\mathbb{E}[Z])^2 + (Z-\mathbb{E}[Z])(Y-\mathbb{E}[Y]) + (X-\mathbb{E}[X])(Z-\mathbb{E}[Z]) + (X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] =$

$= Var(Z) + \underbrace{Cov(Z,Y)}_{0} + \underbrace{Cov(X,Z)}_{0} + \underbrace{Cov(X,Y)}_{0} =$

$= Var(Z)$ (ty okorrelerade, $\rho = 0$)

$\rho_{UV} = \dfrac{Cov(U,V)}{\sqrt{Var(U)Var(V)}} = \dfrac{Var(Z)}{\sqrt{(Var(Z)+Var(X))\cdot(Var(Z)+Var(Y))}}$

$Var(U) = Var(X+Z) = \mathbb{E}[(X+Z-\mathbb{E}[X]-\mathbb{E}[Z])^2]$

$= \mathbb{E}\left[(Z-\mathbb{E}[Z])^2 + (X-\mathbb{E}[X])^2 + 2(X-\mathbb{E}[X])(Z-\mathbb{E}[Z])\right]$

$= Var(Z) + Var(X) + \cancel{Cov(Z,X)} = Var(Z) + Var(X)$

$Var(V) = Var(Z) + Var(Y)$

4.60) X har pdf med egenskapen $f(x) = f(-x)$

$S \sim U(\{-1,1\})$. Låt $Y = SX$ X, S ober.

Visa att $Cov(X,Y) = 0$ men ej oberoende.

B: $Cov(X,Y) = \mathbb{E}[(X-\underbrace{\mathbb{E}[X]}_{=0})(SX-\underbrace{\mathbb{E}[SX]}_{=0})] =$

$= \mathbb{E}[SX^2] = \mathbb{E}[S]\mathbb{E}[X^2] = 0$ ty int. över udda intervall

för beroende, hitta $\varepsilon > 0$ s.a $0 < P(X > \varepsilon) < \frac{1}{2}$

Låt $p_\varepsilon = P(X > \varepsilon)$ $P(SX > \varepsilon) = P(X > \varepsilon \mid S=1)P(S=1) +$

$+ \mathbb{P}(\bar{X} < -\varepsilon \mid S = -1)\mathbb{P}(S=-1) \underset{\text{ober.}}{\Rightarrow} = \mathbb{P}(\bar{X} > \varepsilon)\mathbb{P}(S=1) +$
$+ \mathbb{P}(\bar{X} < -\varepsilon)\mathbb{P}(S=-1) = P_\varepsilon \cdot \frac{1}{2} + P_\varepsilon \cdot \frac{1}{2} = P_\varepsilon$

Men: $\mathbb{P}(S\bar{X} > \varepsilon \cap \bar{X} > \varepsilon) = \mathbb{P}(S\bar{X} > \varepsilon \mid \bar{X} > \varepsilon)\mathbb{P}(\bar{X} > \varepsilon)$
$= \mathbb{P}(S=1)\mathbb{P}(\bar{X} > \varepsilon) = \frac{1}{2}P_\varepsilon \neq P_\varepsilon^2 = \mathbb{P}(S\bar{X} > \varepsilon)\mathbb{P}(\bar{X} > \varepsilon)$ □

4.66 T_1, T_2 är två väntetider med $\mathbb{E}[T_1] = 3$ $\mathbb{E}[T_2] = 1$

$S \in \{0,1\}$ $\mathbb{P}(S=0) = \frac{1}{3}$ $\mathbb{P}(S=1) = \frac{2}{3}$

$S = \begin{cases} 1 & \text{tar vi } T_2\text{-hissen} \\ 0 & \text{— " — } T_1\text{-hissen.} \end{cases}$

$T = ST_2 + (1-S)T_1$ Vad är $\mathbb{E}[T]$?

$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T \mid S]] = \mathbb{E}[T \mid S=0]\mathbb{P}(S=0) + \mathbb{E}[T \mid S=1]\mathbb{P}(S=1) =$
$= \mathbb{E}[T_1]\frac{1}{3} + \mathbb{E}[T_2]\cdot\frac{2}{3} = 1 + \frac{2}{3} = \frac{5}{3}$

4.81) $X \sim \text{Bern}(p)$ dvs $X \in \{0,1\}$ $\mathbb{P}(X=1) = p$

$M_X(t) = \mathbb{E}[e^{tX}] = \sum_x e^{tx}p(x) = e^{t\cdot 0}p(0) + e^{t\cdot 1}p(1) =$
$= (1-p) + e^t p = 1 - p + pe^t$ //

Beräkna $\mathbb{E}[X^n]$ $M'_X(0) = \mathbb{E}[X]$ $M_X^{(n)}(0) = \mathbb{E}[X^n]$

$M'_X(t) = pe^t \Rightarrow \mathbb{E}[X] = M'_X(0) = p$

Vi ser att $M_X^{(n)}(t) = pe^t \Rightarrow \mathbb{E}[X^n] = p$

$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$ //

4.82) $X_1, \dots, X_n \sim \text{Bern}(p)$ ober. Låt $X = X_1 + \dots + X_n$

Visa $X \sim \text{Bin}(n,p)$ B: $M_X(t) = \mathbb{E}[e^{t(X_1 + \dots + X_n)}] =$
$= \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] = \mathbb{E}[e^{tX_1}]^n = (1-p+pe^t)^n$ - mgf för $\text{Bin}(n,p)$ □.

Beräkna $\mathbb{E}[X]$ och $Var[X]$:

$\mathbb{E}[X] = M_X'(0)$ $\quad M_X'(t) = pe^t n(1-p+pe^t)^{n-1}$

$\Rightarrow \mathbb{E}[X] = np$ $\quad M_X''(t) = np(1-p+pe^t)^{n-1}e^t +$

$\mathbb{E}[X^2] = M_X''(0) =$ $\quad + npe^t pe^t(n-1)(1-p+pe^t)^{n-2}$

$= np + n(n-1)p^2$

$\Rightarrow Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = np + n(n-1)p^2 - n^2p^2 =$

$= np - np^2 = np(1-p)$ //

$\mathbb{P}(A) = \iint_A f_{XY}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{A}(x,y) f_{XY}(x,y)\,dx\,dy$

$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\,dx\,dy$ $\quad 1_A(x,y) = \begin{cases} 1 & (x,y) \in A \\ 0 & (x,y) \notin A \end{cases}$

$= \mathbb{E}[1_A(X,Y)] = \mathbb{E}[1_A]$

Ex: $A = \{X \le \frac{1}{2}\}, \{X + Y \le 1\}$

28/4 gränsvärdessatser LLN (Law of Large Numbers el. stora talens lag)

Ex: Singlar slant n# ggr. $X = \#H \sim Bin(n, \frac{1}{2})$

LLN ger: $\frac{X}{n} \xrightarrow[n\to\infty]{} \frac{1}{2}$ (från LLN)

Anledn. är (informellt) att $\mathbb{E}[\frac{X}{n}] = \frac{1}{2}$ $\quad Var(\frac{X}{n}) = \frac{1}{4n} \to 0$

avvikelse fr. medel

Chebyschevs olikhet: $\mathbb{E}[X] = \mu$ och $Var(X) = \sigma^2$ så är

$\mathbb{P}(|X - \mu| > \varepsilon) \le \frac{\sigma^2}{\varepsilon^2}, \forall \varepsilon > 0$.

B: $\mathbb{P}(|X-\mu| > \varepsilon) = \mathbb{E}[1_{\{|X-\mu|>\varepsilon\}}] \le \mathbb{E}[\frac{(X-\mu)^2}{\varepsilon^2} \underbrace{1_{\{|X-\mu|>\varepsilon\}}}_{\le 1}]$

$\le \mathbb{E}[\frac{(X-\mu)^2}{\varepsilon^2}] = \frac{\sigma^2}{\varepsilon^2}$ $\square$

Sats (LLN): Låt $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$ där X_i är oberoende med $\mathbb{E}[X_i] = \mu$ och $Var(X_i) = \sigma^2 \ \forall i$.
(kräver ej samma fördelning bara samma medel & varians)

$\Rightarrow P(|\bar{X}_n - \mu| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0$

B: $P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{Var(\frac{1}{n}(X_1 + \ldots + X_n))}{\varepsilon^2} =$

$= \frac{\frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)}{\varepsilon^2} = \frac{\frac{1}{n}\sigma^2}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0 \quad \square$

"Sannolikheten att $\bar{X}_n$ avviker från $\mu = \mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1]$ mer än $\varepsilon > 0$ går mot 0"

Ex: Kastar tärning: X_i = resultat av kast i. $\Rightarrow$

$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n} \rightarrow 3.5 = \mathbb{E}[X_1]$

Ex: (Monte-Carlo simulering) Låt $X_1, \ldots, X_n \sim U(0,1)$ oberoende. Vi har $\mathbb{E}[e^{X_i^2}] = \int_0^1 e^{x^2} dx$

Vi har: $\bar{X}_n = \frac{\sum_{i=1}^{n} e^{X_i^2}}{n} \xrightarrow[n\to\infty]{LLN} \mathbb{E}[e^{X_1^2}] = \int_0^1 e^{x^2} dx$

$\Rightarrow$ Vi kan uppskatta $\int_0^1 e^{x^2} dx$ numeriskt genom att simulera $X_1, \ldots X_{1000} \sim U(0,1)$ och har:

$$\int_0^1 e^{x^2} dx \approx \frac{1}{1000}\sum_{i=1}^{1000} e^{X_i^2}$$

CLT (Central Limit Theorem) Centrala gränsvärdessatsen

Sats (CLT): Låt $S_n = X_1 + \ldots + X_n$ där $X_1, \ldots X_n$ är oberoende med $\mathbb{E}[X_i] = \mu$ och $Var(X_i) = \sigma^2$

$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \xrightarrow[n\to\infty]{} \Phi(x), \ \forall x$ där $\Phi(x)$ är cdf för en $N(0,1)$

Def: $\{X_n\}_{n\geq 1}$ är en sekvens av s.v

vi säger att $X_n \to X$ om $F_n(\cancel{x}) \to F(\cancel{x})$ där $F_n(x)$

är cdf för X_n, och $F(x)$ för X.

Kom ihåg att om $M_X(t) = M_Y(t) \quad \forall t \in (-\varepsilon, \varepsilon)$

$\Rightarrow X \stackrel{d}{=} Y$ dvs X och Y har samma fördeln.

Sats: $X_n \to X \iff M_{X_n}(t) \to M_X(t) \; \forall \; t \in (-\varepsilon, \varepsilon)$

"B (CLT)": Antag $\mu = 0$, $\sigma = 1$ (annars transformera)

om $X \sim N(0,1)$ har vi: $M_X(t) = \mathbb{E}[e^{tX}] = e^{t^2/2}$

Vi har: $M_X(t) = 1 + \frac{1}{2}t^2 + O(t^2)$ då $t \to 0$

menas $\frac{O(t^2)}{t^2} \xrightarrow[t\to 0]{} 0$

$M_n(t) = \mathbb{E}[e^{t\frac{S_n}{\sqrt{n}}}]$

$= \mathbb{E}[e^{\frac{t}{\sqrt{n}}(X_1 + \ldots + X_n)}] \stackrel{\text{ober.}}{=} \prod_{i=1}^{n} \mathbb{E}[e^{\frac{t}{\sqrt{n}}X_i}] = (M(\frac{t}{\sqrt{n}}))^n$

Vi har $M_{X_1}(s) \stackrel{\text{Taylor}}{=} \underbrace{M(0)}_{=1} + s\underbrace{M'(0)}_{=\mathbb{E}[X_1]=0} + \frac{s^2}{2}\underbrace{M''(0)}_{=\mathbb{E}[X_1^2] = Var(X_1) = 1} + O(s^2)$

$= 1 + \frac{s^2}{2} + O(s^2)$

$\Rightarrow M_n(t) = \left(1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + O\left(\left(\frac{t}{\sqrt{n}}\right)^2\right)\right)^n = \left(1 + \frac{1}{2}\frac{t^2}{n} + O\left(\frac{t^2}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{\frac{1}{2}t^2}$

dvs $M_n(t) \to M_X(t) \stackrel{\text{sats}}{\iff} \frac{S_n}{\sqrt{n}} \to X$

dvs $P\left(\frac{S_n}{\sqrt{n}} \leq x\right) \to \Phi(x) \; \forall x$. $\square$

$\left(1 + \frac{x}{n}\right)^n \to e^x$

Ofta används detta för approximation:

1) Vi gör många försök (oberoende) och tittar (t.ex. kastar en tärning) på medlet $\bar{X}_n \Rightarrow \bar{X}_n \approx N(\mu, \frac{\sigma^2}{n})$

2) $X \sim Bin(n,p) \Rightarrow X \approx N(np, npq)$ $\quad np \geq 5, nq \geq 5$ giltighetsområde

$q = 1-p$

3) $X \sim Poi(\lambda) \Rightarrow X \approx N(\lambda, \lambda)$ $\quad \lambda \geq 5$

4) $X \sim Hg(N, n, p)$ (hypergeometrisk fördeln.) $X \approx N(np, npq\frac{N-n}{N-1})$, $np \geq 5$, $nq \geq 5$

5) $X \sim Gamma(\alpha, \lambda)$ $X \approx N(\frac{\alpha}{\lambda}, \frac{\alpha}{\lambda^2})$, α stort

Ex

Spel med tre val.

1) Du får 4500 SEK

2) Singla slant, om H får du 10 000

— " — T — " — 0

3) Singlar 10 000 slantar, behåll alla H.

X = # H om vi väljer 3) $X \sim Bin(10\,000, \frac{1}{2})$

X är med god approx. $N(5000, 2500)$

3-sigma-regeln: Med sann 99.74% 0.9974 är

$X \in (5000 - 3\sqrt{Var(X)}, 5000 + 3\sqrt{Var(X)}) = (5000 - 3 \cdot 50, 5000 + 3 \cdot 50) =$

$= (4850, 5150)$

Dvs 3 bättre än 1!

~~Ex~~: χ^2 och t-fördeln. χ^2 om $Z \sim N(0,1) \Rightarrow$

$Z^2 \sim \chi^2$ dvs "chi-kvadrat med 1 frihetsgrad"

Def: χ^2-fördelning med k frihetsgrader betecknas χ^2_k och definieras av att fördelningen av χ^2_k är samma som $Z_1^2 + \ldots + Z_k^2$ $Z_i \sim N(0,1)$ Z_i oberoende.

mgf $M_{\chi^2_k}(t) = (1-2t)^{-k/2}$

pdf $f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$

Hur uppkommer den? Låt $X_1, \ldots, X_n$ vara oberoende

$N(\mu, \sigma^2)$ Vi har $\overline{X}_n = \frac{X_1 + \ldots + X_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$

Sats: $\bar{X}_n$ och $(X_1 - \bar{X}_n), \ldots, X_n - \bar{X}_n)$ är oberoende

"B:" Använd $\mathbb{E}[e^{sX+tY}] = \mathbb{E}[e^{sX}]\mathbb{E}[e^{tY}] \Leftrightarrow X, Y$ oberoende.

Sats: $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$

B: $\mu = 0$, $\sigma^2 = 1$

$$\underbrace{\sum_{i=1}^n X_i^2}_{=\chi^2_n} = \sum_{i=1}^n (X_i - \bar{X}_n + \bar{X}_n)^2 = \underbrace{\sum_{i=1}^n \bar{X}_n^2}_{=n\bar{X}_n^2 \sim \chi^2_1} + \underbrace{\sum_{i=1}^n (X_i - \bar{X}_n)^2}_{=?} + 2\bar{X}_n \underbrace{\sum_{i=1}^n (X_i - \bar{X}_n)}_{=0}$$

$$\Rightarrow M_{\chi^2_n}(t) = M_{\chi^2_1}(t)\, M_?(t)$$

$$\Rightarrow (1-2t)^{-n/2} = (1-2t)^{-1/2} M_?(t)$$

$$\Rightarrow M_?(t) = (1-2t)^{-\frac{(n-1)}{2}} = M_{\chi^2_{n-1}}(t)$$

$$\Rightarrow \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi^2_{n-1} \quad \square$$

Räknestuga Mat.stat. 29/4

4.88) $X \sim N(0,\sigma^2)$, beräkna $\mathbb{E}[X^n]$ $\forall n$ mha mgf

$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}[e^{\sigma t \frac{X}{\sigma}}] = M_Z(\sigma t)$

där $Z = \frac{X}{\sigma} \sim N(0,1) \Rightarrow M_Z(t) = e^{\frac{1}{2}t^2}$

$\Rightarrow M_X(t) = e^{\frac{1}{2}\sigma^2 t^2}$

$\mathbb{E}[X^n] = M_X^{(n)}(0)$

Testa att derivera $M_X(t)$, ser vi ett mönster?

$M_X(t) = e^{\frac{1}{2}\sigma^2 t^2} \Rightarrow M_X'(t) = \sigma^2 t e^{\frac{1}{2}\sigma^2 t}$

$M''_X(t) = \sigma^2 e^{\frac{1}{2}\sigma^2 t^2} + \sigma^2 t \cdot \sigma^2 t e^{\frac{1}{2}\sigma^2 t^2} = (\sigma^2 + (\sigma^2)^2 t^2) e^{\frac{1}{2}\sigma^2 t^2}$

$M_X^{(3)}(t) = \sigma^2 \cdot \sigma^2 t e^{\frac{1}{2}\sigma^2} + 2(\sigma^2)^2 t e^{\frac{1}{2}\sigma^2 t^2} + (\sigma^2)^2 \cdot t^2 \cdot \sigma^2 t e^{\frac{1}{2}\sigma^2 t^2}$

$\vdots$

Funkar ej! $e^t = 1 + t + \frac{t^2}{2} + \ldots = \sum_{k=0}^{\infty} \frac{t^k}{k!}$

$\Rightarrow e^{\frac{1}{2}\sigma^2 t^2} = \sum_{k=0} \frac{\sigma^{2k} t^{2k}}{2^k k!}$

$\Rightarrow M_X^{(2n)}(t) = \sum_{k=n}^{\infty} \frac{\sigma^{2k}}{2^k k!} t^{2k-2n} \cdot (2k)(2k-1)(2k-2)\cdots(2k-2n+1) =$

$= \sum_{k=n}^{\infty} \frac{\sigma^{2k}}{2^k k!} \frac{(2k)!}{(2k-2n)!} t^{2k-2n} \Rightarrow M_X^{(2n)}(0) = \frac{\sigma^{2n}(2n)!}{2^n n!(2n-2n)!} =$

$= \frac{\sigma^{2n}(2n)!}{2^n \cdot n!}$

(udda momenten blir 0 pga täthetsfunktionen jämn)

$M_X^{(2n+1)}(t) = \sum_{k=n+1}^{\infty} \frac{\sigma^{2k}}{2^k k!} t^{2k-2n-1} \frac{(2k)!}{(2k-2n-1)!} \Rightarrow M_X^{(2n+1)}(0) = 0$

5.4) $X_n \sim Bin(n, p_n)$ där $np_n \to \lambda$, $n \to \infty$

Visa att $X_n \to X$ där $X \sim Poi(\lambda)$

$$\frac{O(p_n^2)}{p_n^2} \to C < \infty$$

B: Vi visar $M_{X_n}(t) \to M_X(t)$

$$M_{X_n}(t) = (1 + p_n(e^t - 1))^n = \left(e^{p_n(e^t-1) + O(p_n^2)}\right)^n =$$

$$= e^{np_n(e^t-1) + np_n O(p_n)} \xrightarrow[n\to\infty]{} e^{\lambda(e^t-1)+0} = M_X(t)$$

$\square$

5.14 En full person går varje minut ett steg till höger med sann. $\frac{2}{3}$ och vänster $\frac{1}{3}$. (Diskret variant av Brownsk rörelse)

Bestäm den ungefärliga fördelningen för var personen befinner sig efter en timme.

$$X_i = \begin{cases} 1 & \text{om höger i steg } i \\ -1 & \text{om vänster i steg } i. \end{cases}$$

$S_n = \sum_{i=1}^{n} X_i$ är positionen vid tiden n.

$$\mathbb{E}[X_i] = \frac{2}{3} - \frac{1}{3} = \frac{1}{3} = \mu$$

$$Var(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = 1 - \frac{1}{9} = \frac{8}{9} = \sigma^2 \overset{CLT}{\Rightarrow} S_n \sim N(\mu n, n\sigma^2)$$ $n = 60$

för stora n. $\Rightarrow S_n \sim N(20, 20 \cdot \frac{8}{3})$ om steglängd $= 50$ cm

$\Rightarrow S_n \sim N(1000, 20 \cdot \frac{8}{3} \cdot 50^2)$ S_n i cm

---

5.16 $X_1, \ldots, X_{20}$ ober. med pdf $f(x) = 2x$ $0 \leq x \leq 1$

Låt $S = X_1 + \ldots + X_{20}$. Använd CLT för att approximera $\mathbb{P}(S \leq 10)$

$$\mathbb{E}[X_1] = \int_0^1 2x^2 dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$$

$$\mathbb{E}[X_1^2] = \int_0^1 2x^3 dx = \left[\frac{2x^4}{4}\right]_0^1 = \frac{1}{2} \Rightarrow Var(X_1) = \frac{1}{2} - \frac{4}{9} =$$

$$= \frac{9-8}{18} = \frac{1}{18}$$ //

CLT ger: $\frac{S-n\mu}{\sigma\sqrt{n}}$ är ca $N(0,1)$ $n=20$

$P(S \leq 10) =$

$Z \sim N(0,1)$

$= P\left(\frac{S-\frac{40}{3}}{\sqrt{\frac{1}{18}}\sqrt{20}} \leq \frac{10-\frac{40}{3}}{\sqrt{\frac{1}{18}\cdot 20}}\right) \underset{CLT}{\approx} P\left(Z \leq \frac{10-\frac{40}{3}}{\sqrt{\frac{20}{18}}}\right) \approx P(Z \leq -3.16)$

$= P(Z \geq 3.16) = 1 - P(Z \leq 3.16) \underset{\text{tabell}}{\approx} 1 - 0.9992 = 0.0008$ //

tabell s.A7 i App. B

5.18) X_i = vikten av paket i med $E[X_i] = 15$ kg

$\mathrm{Var}(X_i) = 100\ \mathrm{kg}^2 \Rightarrow \sigma = 10$ kg

$S_{100} = \sum_{i=1}^{100} X_i$ (vikten av 100 paket)

uppskatta $P(S_{100} \leq 1700)$ (*)

Vi använder CLT (centr. gränsv. satsen)

Vi har att $\frac{S_{100} - 100\cdot 15}{\underbrace{10}_{\sigma}\cdot\underbrace{10}_{\sqrt{n}}} \underset{CLT}{\approx} N(0,1) \overset{(*)}{\Rightarrow} P\left(\frac{S_{100}-1500}{100} \leq \frac{1700-1500}{100}\right)$

$Z \sim N(0,1)$

$\approx P(Z \leq 2) \underset{\text{tabell}}{\approx} 0.9772$

Om $X_1, \ldots, X_n$ är s.v. med $E[X_i] = \mu$ [5/5]

$\bar{X} = \bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$

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

$\boxed{\begin{array}{ll} \bar{X} \text{ motsvarar } \mu \\ S^2 \quad '' \quad \sigma^2 \end{array}}$

t-fördelningen: (med $k \geq 1$ grader av frihet)

Om $Z \sim N(0,1)$ och $X \sim \chi^2_k$ oberoende så är

$\frac{Z\sqrt{k}}{\sqrt{X}} \sim t_k$-fördelad pdf: $f(t) = \frac{\Gamma((k+1)/2)}{\sqrt{k\pi}\,\Gamma(k/2)} \left(1+\frac{t^2}{k}\right)^{-(k+1)/2}$

Konsekvens: $\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\left(\frac{(n-1)S^2}{\sigma^2}\right)}}\sqrt{n-1} \sim t_{n-1}$ fördelat

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Sampling Motivation:

Ex: 1) Vi vill veta medellängden av alla i Sverige

2) Vi vill veta konsekvenserna av Tjernobyl olyckan.

Vi kan: 1) Vi mäter alla som bor i Sverige.

2) Undersök alla människor och se om dom har (en viss typ) av cancer.

Ej praktiskt möjligt!

I stället väljer vi ut ett antal personer och mäter/undersöker dom. Detta bör ske dumpmässigt för att få ett representativt urval och för att undvika bias/subjektivitet hos undersökaren.

Vi har en population $\{1,2,3,\dots,N\}$ med värdet $\{x_1, x_2, \dots, x_N\}$ (längd, ålt cancer/ej cancer)

PD = populations-fördelningen av $\{x_1, \dots, x_n\}$

Vad kan gå fel? 1) Vårt urval är ej representativt.

i) Vi frågar bara folk i Danderyd om deras politiska åsikter.

ii) Man skickar ut frivilliga enkäter. (bara engagerade svarar)

2) Vi mäter fel (svaren ej sanningsenliga)

i) Är du heroinist?

ii) Har du varit otrogen?

3) Systematiska fel

Populationsparametrar och deras skattningar Vi vill uppskatta medel och standardavvikelsen av någon egenskap (längd, andel av pop med cancer)

Ex: PD = (p_r, p_m, p_l) proportioner med rött/mörkt/ljust hår.
PD = $U(0,\theta)$, något är likformigt fördelat på ett intervall av okänd längd θ. (vänta på bussen)
PD = Exp(λ) Vad är sönderfallshastigheten λ?

Def: punktestimering (punktskattning) av parametern θ betecknas $\hat{\theta}$ och $\hat{\theta} = \theta(X_1, X_2, \ldots, X_n)$ där $(X_1, \ldots, X_n)$ är data.

Samplingsfördelningen av $\hat{\theta}$ = fördelningen av $\hat{\theta}$.
$\hat{\theta}$ är en s.v. som beror på $(X_1, \ldots, X_n)$

Ex: 1/ Vi vill skatta μ dvs ett medelvärde.

$$\hat{\mu} = \bar{X} = \frac{X_1 + \ldots + X_n}{n}$$

2/ Om $\theta = \sigma$ dvs vi vill skatta standardavvikelsen används $\hat{\theta}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Samplingsfördelningen av $\bar{X}$ antas ofta vara $N(\mu, \frac{\sigma^2}{n})$

Användbart: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 + \frac{n}{n-1}\bar{X}^2 - \frac{2}{n-1}\underbrace{\sum_{i=1}^{n} X_i \bar{X}}_{= n\bar{X}^2} =$

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

Proportions-sampling PD = proportionen p har en viss egenskap (t.ex. cancer)

$X_i \sim \text{Bern}(p)$ oberoende $X_i = \begin{cases} 1 & \text{om } i \text{ har egenskapen} \\ 0 & \text{annars} \end{cases}$

$X_1 + \ldots + X_n \sim \text{Bin}(n, p)$ Om $\theta = p$ dvs vi vill skatta p så kallas $\hat{\theta}$ för $\hat{p}$ och $\hat{p} = \bar{X}$ och har approximativt fördeln. $N(p, \frac{p(1-p)}{n})$

Väntesvärdes-riktiga skattare Def: $\hat{\theta}$ är en väntskatt av θ om $\mathbb{E}[\hat{\theta}] = \theta$ Dvs vi gör ej något systematiskt fel! Vi har: för $\bar{X}$: $\mathbb{E}[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n} n\mu = \mu$

2) $\mathbb{E}[\hat{p}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\cdot n\cdot p = p$ // var ty $\mathbb{E}[\bar{X}]=\mu=c$

3) $\mathbb{E}[S^2] \overset{\mu=0}{=} \frac{n}{n-1}\mathbb{E}[X_1^2] - \frac{n}{n-1}\mathbb{E}[\bar{X}^2] = \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\cdot\frac{\sigma^2}{n}$

$= \frac{n-1}{n-1}\cdot\sigma^2 = \sigma^2$

Därför är $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ istället för $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

Standardfelet ~~och dess skattare~~ Def: Standardfelet av skattaren $\hat{\theta}$ är dess standardavvikelse $\sigma_{\hat{\theta}}$.

Vi är ofta intresserade av $\sigma_{\hat{\theta}}$, men känner den ej (av naturliga skäl).

I stället skattar vi $\sigma_{\hat{\theta}}$ med $s_{\hat{\theta}}$

Ex: $\sigma_{\bar{X}}^2 = Var(\bar{X}) = \frac{\sigma^2}{n}$ där S är som innan

Vi skattar $\sigma_{\bar{X}}$ med $S_{\bar{X}} = \frac{S}{\sqrt{n}}$

Ex: $\hat{p}$ skattar p med $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Vi skattar $\sigma_{\hat{p}}$ med $s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$

Def: Konsistent skattare: Givet en samplings-storlek n för att skatta θ med $\hat{\theta}$, så är $\hat{\theta}$ konsistent om $\mathbb{E}[(\hat{\theta}-\theta)^2] \xrightarrow[n\to\infty]{} 0$. Vi har: $\mathbb{E}[(\hat{\theta}-\theta)^2] = \mathbb{E}[(\hat{\theta}-\mathbb{E}[\hat{\theta}])^2 +$

$+ (\mathbb{E}[\hat{\theta}]-\theta)^2 - 2(\hat{\theta}-\mathbb{E}[\hat{\theta}])(\theta-\mathbb{E}[\hat{\theta}])] = \underbrace{Var(\hat{\theta})}_{\text{osäkerhet från mätning}} + \underbrace{(\mathbb{E}[\hat{\theta}]-\theta)^2}_{\text{Systematiska felet (vill man ha till 0)}}$

6.3/ Vi tar 16 st. samplingar som antas vara oberoende och $N(0,1)$-fördelade. Hitta c s.a $P(|\bar{X}|<c) = 0.5$ (R.öv. 6/5)

där $\bar{X}$ är medelvärdet.

Vi har $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ $\quad X_i \sim N(0,1)$ oberoende

$M_{\bar{X}}(t) = \mathbb{E}[e^{t\bar{X}}] = \mathbb{E}[e^{\frac{t}{n}\sum_{i=1}^{n}X_i}] \overset{\text{oberoende}}{=} \prod_{i=1}^{n}\mathbb{E}[e^{\frac{t}{n}X_i}] =$

momentgen. fkn.

$= \mathbb{E}[e^{\frac{t}{n}X_1}]^n = (e^{\frac{1}{2}(\frac{t}{n})^2})^n = e^{\frac{1}{2}(\frac{t}{\sqrt{n}})^2}$

$(N(0,\sigma^2)) = e^{\frac{1}{2}t^2\sigma^2}$ $\Rightarrow \bar{X} \sim N(0,\frac{1}{n})$

$\bar{X} \sim N(0, \frac{1}{n}) \Rightarrow \frac{\bar{X}}{1/\sqrt{n}} \sim N(0,1)$

cdf för $N(0,1)$

med $n=16 \Rightarrow \frac{\bar{X}}{1/4} = 4\bar{X} \sim N(0,1)$

(cdf i 4c)

$\Rightarrow P(|\bar{X}| < c) = P(|4\bar{X}| < 4c) = 2(1 - \Phi(4c)) = 0.5$

$\Rightarrow 1 - \Phi(4c) = \frac{1}{4} \Rightarrow \Phi(4c) = \frac{3}{4} \xrightarrow{\text{(tabell)}} 4c \approx 0.675$

$\Rightarrow c \approx \frac{0.675}{4}$

6.10) Visa hur vi skall använda χ^2 för att beräkna $P(a < \frac{S^2}{\sigma^2} < b)$. Vi har att om $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

så gäller $\frac{S^2}{\sigma^2}(n-1) \sim \chi^2_{n-1}$

$$P(a < \frac{S^2}{\sigma^2} < b) = P\left(a(n-1) < \frac{S^2}{\sigma^2}(n-1) < b(n-1)\right) =$$

$$= P\left(\frac{S^2}{\sigma^2}(n-1) < b(n-1)\right) - P\left(\frac{S^2}{\sigma^2}(n-1) < a(n-1)\right) = F_{\chi^2_{n-1}}(b(n-1))$$

$$- F_{\chi^2_{n-1}}(a(n-1))$$

7.4) Populationer I och II, vi samplar n_1 och $n_2 = 2n_1$ från dessa. Populationerna har st-avvikelserna σ_1 resp. $\sigma_2 = 2\sigma_1$

I vilken av dessa undersökningarna kan vi förvänta oss att få den mest exakta skattningen av medlet?

$\bar{X}_1$ är skattaren av medlet för pop I

$\bar{X}_2$ är — " — II

Frågan blir, vilken av $\bar{X}_1$ och $\bar{X}_2$ har minst varians?

$$Var(\bar{X}_1) = Var\left(\frac{1}{n_1}\sum_{i=1}^{n_1} X_{1,i}\right) = \frac{1}{n_1^2}\sum_{i=1}^{n_1} Var(X_{1,i}) = \frac{\sigma_1^2}{n_1}$$

p.s.s. $Var(\bar{X}_2) = \frac{\sigma_2^2}{n_2} = 2\frac{\sigma_1^2}{n_1}$

(måste ta 4ggr så många prov för att få dubbla exaktheten)

(Om σ är 2ggr större $\Rightarrow$ behöver 4ggr så många n

7.14) Vi har $N = 393$ st sjukhus med proportionen $p = 0.654$ som skriver ut färre än 1000 patienter.
Låt $T = Np = \#$ sjukhus med ≤ 1000 utskrivningar.
Vi kollar $n = 25$ styck sjukhus och låter $X = \#$ som hade ≤ 1000 utskrivningar.

Skatta T mha $\hat{T}$ och ange fördelningen av $\hat{T}$.

Vi skattar $\hat{p}$ och sätter $\hat{T} = N\hat{p}$

Vi låter $\hat{p} = \frac{X}{n}$ och observerar att $X \sim Bin(n, p)$

$$\overset{CLT}{\Rightarrow} \hat{p} \sim N(E[\hat{p}], Var(\hat{p})) \Rightarrow \hat{T} \approx N(E[\hat{T}], Var(\hat{T}))$$

$$E[\hat{T}] = NE[\hat{p}] = \frac{N}{n} E[X] = \frac{N}{n} \cdot np = Np = 257$$

$$Var(\hat{T}) = N^2 Var(\hat{p}) = \cancel{N^2 Np(1-p)} = N^2 \cdot \frac{p(1-p)}{n} = 1398.$$

Konfidensintervall

Def: Ett konfidensintervall (CI) för θ med konfidensnivå $V\% =$ ett intervall runt skattaren $\hat{\theta}$ som innehåller θ med sannolikhet $\frac{V}{100}$

7/5

Om vi skattar θ 100 ggr och hittar ett 95% CI så kommer intervallet att innehålla θ i genomsnitt 95 gånger.

EX: Vi vill hitta ett $100(1-\alpha)\%$ CI för μ. Vi skattar μ mha $\bar{X}$.

Vi har $\frac{\bar{X} - \mu}{S_{\bar{X}}} \sim t_{n-1}$-fördelad

$\approx N(0,1)$ för n stort

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{S_{\bar{X}}} < z_{\alpha/2}\right) \approx P(-z_{\alpha/2} < Z < z_{\alpha/2}) \quad Z \sim N(0,1)$$

$= 1 - \alpha$ (approx.)

Där z_α s.a $\Phi(z_\alpha) = 1 - \alpha$

$\Rightarrow \mu \in (\bar{X} - z_{\alpha/2} S_{\bar{X}}, \bar{X} + z_{\alpha/2} S_{\bar{X}})$ med sannol. $1-\alpha$

Om jag byter ut $z_{\alpha/2}$ mot $t_{n-1,\alpha/2}$

där $\mathbb{P}(-t_{n-1,\alpha/2} < T < t_{n-1,\alpha/2}) = 1-\alpha$ $T \sim t_{n-1}$-fördelad

Ex: Ägg: 68% CI: $\mu_L \in 22.41 \pm 1 \cdot 0.069$

95% CI: $\mu_L \in 22.41 \pm 1.96 \cdot 0.069$

$\mu_L \in 22.41 \pm 3.30 \cdot 0.069$

$z_{\alpha/2}$ $S_{\bar{X}}$ enl. tabell

Ju högre konfidensnivå desto bredare CI

Ju fler samplingar desto smalare CI.

Exakt CI för medelvärdet:

Antag att PD $= N(\mu, \sigma^2)$ med okända μ och σ^2 då är exakta CI för μ: $\bar{X} \pm t_{n-1,\alpha/2} S_{\bar{X}}$

$t_{n-1,\alpha/2}$ hittas i tabell!

Det exakta CI är bredare än det approximativa!

Ex: Approximativt vs exakt CI Vi använder två olika metoder för att mäta fetthalten i 16 kor.

Vi får 16 skillnader i mätningarna med $\bar{X} = 0.53$

$S = 1.06$

$(S^2 = \frac{1}{n-1}\Sigma(\dots)^2)$ Exakt 95% CI: $\mu \in \bar{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{16}} =$

$= 0.53 \pm 2.13 \cdot \frac{1.06}{4} = (-0.03, 1.08)$

Approx 95% CI: $\mu \in 0.53 \pm 1.96 \cdot \frac{1.06}{4} = (0.01, 1.05)$

Två metoder för punktskattningar:

MME (Method of Moments Estimator)

Antag att $\mathbb{E}[X] = f_1(\theta_1, \theta_2)$ och $\mathbb{E}[X^2] = f_2(\theta_1, \theta_2)$

Vi ersätter $\mathbb{E}[X]$ med $\bar{X}$ och $\mathbb{E}[X^2]$ med $\overline{X^2} = \frac{1}{n}(X_1^2 + \dots + X_n^2)$

Vi löser sedan $\overline{X} = f_1(\hat{\theta}_1, \hat{\theta}_2)$

$\overline{X}^2 = f_2(\hat{\theta}_1, \hat{\theta}_2)$ för att få $\hat{\theta}_1 = g_1(\overline{X}, \overline{X}^2)$

$\hat{\theta}_2 = g_2(\overline{X}, \overline{X}^2)$

Ex: Vi låter X vara väntetiden på en buss
Vi mäter och får: 2, 7, 4, 11, 15 min.
Vi antar $X \sim U(0, \theta)$ där θ är den fixa väntetiden mellan två bussar

Vi får $\mathbb{E}[X] = \frac{\theta}{2} \Rightarrow \overline{X} \Rightarrow \overline{X} = \frac{\hat{\theta}}{2} \Rightarrow$

$\hat{\theta} = 2\overline{X}$

$\hat{\theta}$ är vår MME och $\hat{\theta} = 15.6$ min

MLE (Maximum likelihood Estimator) Vi vill hitta $\hat{\theta}$ som "bäst" stödjer data.

Def: Likelihood-funktionen betecknas $L(\theta)$ och definieras av $L(\theta) := f(x_1, \dots, x_n \mid \theta)$ och är den gemensamma pdf/pmf för datamängden givet θ. Datamängden $(X_1, \dots, X_n)$ har ~~realisation~~ realiserad observation $(x_1, \dots, x_n)$

$f_{X_1 X_2 \dots X_n}(x_1, \dots \mid \theta)$

Def: Vår MLE $\hat{\theta}$ är det värde av θ som maximerar $L(\theta)$

Om samplingen är i.i.d. (Identically, Independently Distributed) så gäller $f(x_1, \dots, x_n \mid \theta) = f(x_1 \mid \theta) \dots f(x_n \mid \theta) = L(\theta)$

Ex: (Buss) $L(\theta) = \prod_{i=1}^{5} f(x_i \mid \theta) = \prod_{i=1}^{5} \frac{1}{\theta} 1_{\{x_i \leq \theta\}} =$

$= \left(\frac{1}{\theta}\right)^5 1_{\{\max(x_1, \dots, x_5) \leq \theta\}}$ Vi ser att $L(\theta)$ maximeras av

$\hat{\theta} = \max(x_1, \dots, x_5) = 15$

Med n observationer fås $\mathbb{E}[\hat{\theta}] = \frac{n}{n+1}\theta$

$n=2$: $\mathbb{E}[\hat{\theta}] = \mathbb{E}[\max(X_1, X_2)] = \int_0^\theta \int_0^\theta \max(x_1, x_2) \frac{1}{\theta^2} dx_1 dx_2 =$

$= \frac{1}{\theta^2} \int_0^\theta \int_0^\theta x_1 \mathbb{1}_{\{x_1 \geq x_2\}} dx_1 dx_2 + \frac{1}{\theta^2} \int_0^\theta \int_0^\theta x_2 \mathbb{1}_{\{x_2 \geq x_1\}} dx_1 dx_2 =$

$= \frac{2}{\theta^2} \int_0^\theta \int_{x_2}^\theta x_1 dx_1 dx_2 = \frac{2}{\theta^2} \left(\frac{\theta^3}{2} - \frac{\theta^3}{6} \right) = \frac{2}{3} \theta$

Därför är $\hat{\theta} \cdot \frac{n+1}{n}$ den så kallade korrigerade MLE'n (väntevärdesriktiga)

Buss: $\hat{\theta} \cdot \frac{n+1}{n} = 18$ min

Ex: Metod för att uppskatta N där N är antalet individer i en vargstam 1) Vi fångar 100 djur och märker dem. 2) Vi släpper ut dem. 3) Vi fångar in 50 djur varav 20 är märkta, vad är en bra skattning av N?

$X = \#$ märkta djur av dom 50

$\Rightarrow X \sim Hg(N, 50, \frac{100}{N})$ $\quad \frac{100}{N}$ prop. märkta vargar/svarta bollar.

$L(N) = P(X = 20 | N) = \frac{\binom{100}{20}\binom{N-100}{30}}{\binom{N}{50}}$

Vi får:

$\frac{L(N)}{L(N-1)} = \ldots = 1 - \frac{20}{N} \frac{N-250}{N-130} \geq 1 \Leftrightarrow N \leq 250$

$\Rightarrow \hat{N} = 250$ maximerar $L(N)$ dvs MLE: $\hat{N} = 250$.

Randomiserad svarsmetod Ex: Fängelse med 500 fångar varav Np är heroinister.

$Nq = N(1-p)$ är ej heroinister.

"Är du heroinist?" Låt internen slå en tärning: om $\{1,2,3,4,5\}$ svara sanningsenligt om $\{6\}$ så ljug.

7/5
R.öv.

7.18) Två oberoende undersökningar har gett 90% CI för μ = popul.-medlet
Vad är sannolikheten att inget av de två CI innehåller μ? Sannolikheten att en undersökning med 90% CI "missar" μ är $1-0.9=0.1 \Rightarrow$ svar: sannolikheten $=0.1^2=0.01$

Sann. båda innehåller μ? $0.9^2=0.81$ //

7.33) Vi undersöker 2 popul. med samplingsstorlek n för att skatta p_1, p_2. Unders. är oberoende
Vi väntar oss att $p_1, p_2 \approx \frac{1}{2}$
Vad skall n vara s.a $S_{\hat{p}_1-\hat{p}_2} \leq 0.02$?

$S^2_{\hat{p}_1-\hat{p}_2} \leq 0.0004 \Rightarrow$

$\Rightarrow \sigma^2_{\hat{p}_1-\hat{p}_2} = Var(\hat{p}_1 - \hat{p}_2) \leq 0.0004$

$Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(\hat{p}_2) = Var\left(\frac{X_1}{n}\right) + Var\left(\frac{X_2}{n}\right)$

summan av alla sampl.

$= \frac{1}{n^2} Var(X_1) + \frac{1}{n^2} Var(X_2) = \frac{1}{n^2} n p_1(1-p_1) + \frac{1}{n^2} n p_2(1-p_2)$

Variansen för en Bern. variabel

$\approx \frac{2}{n} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2n} \leq 0.0004 \Rightarrow$

$\Rightarrow n = 1250.$

[$X_1 = \sum_{i=1}^{n} X_{1,i}$ $Var(X_1) = \sum_{i=1}^{n} Var(X_{1,i}) = n Var(X_{1,1})$

(ober. Bern: exp. beror inte av föregående exp.)

$X_{1,i}$ = res. av sampl. i av undersökn. I]

7.37) Två undersökn. (oberoende) av samma popul. genomförs för att skatta medlet μ. Beteckna skattningarna $\bar{X}_1$ och $\bar{X}_2$ och deras standardfel $\sigma_{\bar{X}_1}, \sigma_{\bar{X}_2}$

Bilda $\bar{X} = \alpha \bar{X}_1 + \beta \bar{X}_2$, hur skall vi välja α, β s.a. $\bar{X}$ blir väntevärdesriktig? [Def: $E[\bar{X}] = \mu \Leftrightarrow$ väntevärdesriktig skattare]

gäller även om beroende

$E[\bar{X}] = E[\alpha \bar{X}_1 + \beta \bar{X}_2] = \alpha E[\bar{X}_1] +$

$+\beta\mathbb{E}[\bar{X}_2] = \alpha\mu + \beta\mu = \mu \Rightarrow$ ~~β~~ (linjäritet för $\mathbb{E}$ (ej för Var))

$\Rightarrow \beta = 1-\alpha \Rightarrow \bar{X} = \alpha\bar{X}_1 + (1-\alpha)\bar{X}_2$

b) Hur välja α så $Var(\bar{X})$ blir minimal?

$Var(\bar{X}) = Var(\alpha\bar{X}_1 + (1-\alpha)\bar{X}_2) \overset{\text{oberoende!}}{=} Var(\alpha\bar{X}_1) + Var((1-\alpha)\bar{X}_2)$

$= \alpha^2 Var(\bar{X}_1) + (1-\alpha)^2 Var(\bar{X}_2) = \alpha^2\sigma^2_{\bar{X}_1} + (1-\alpha)^2\sigma^2_{\bar{X}_2} =$

$= f(\alpha) \Rightarrow$ {minimera: som vanligt: derivera map α}

$f'(\alpha) = 2\alpha\sigma^2_{\bar{X}_1} - 2(1-\alpha)\sigma^2_{\bar{X}_2} = 0 \Rightarrow \alpha = \dfrac{\sigma^2_{\bar{X}_2}}{\sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}}$

8.3) 400 prover

# celler	# prover
0	213
1	128
2	37
3	18
4	3
5	1

Antag #celler i ett prov är Poisson(λ)

X_i = #celler i prov i.

a) Skatta λ $\quad \mathbb{E}[X_i] = \lambda \overset{\text{MME}}{\Rightarrow} \hat{\lambda} = \bar{X}$

$\hat{\lambda} = \bar{X} = \dfrac{213\cdot 0 + 128 + 37\cdot 2 + 18\cdot 3 + 3\cdot 4 + 1\cdot 5}{213+128+37+18+3+1} = \dfrac{273}{400}$

≈ 0.6825

b) Hitta ett approximativt 95% CI (använd $N(0,1)$)

95% CI : $\hat{\lambda} \pm 1.96\, S_{\hat{\lambda}}$

Det bästa är nog $S_{\hat{\lambda}} = \sqrt{\dfrac{\hat{\lambda}}{n}}$ också ok: $S_{\hat{\lambda}} = \dfrac{S}{\sqrt{n}}$

S^2 skattar σ^2

$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \dfrac{n}{n-1}\bar{X}^2 = \dfrac{1}{399}(213\cdot 0^2 + 128\cdot 1^2 +$

$+ 37\cdot 2^2 + 18\cdot 3^2 + 3\cdot 4^2 + 1\cdot 5^2) - \dfrac{400}{399}\left(\dfrac{273}{400}\right)^2$

95% CI : $0{,}6825 \pm 1{,}96\dfrac{0.9021}{20} = 0.6825 \pm 0.088$

8.8) # Hops

# Hops	Freq
1	48
2	31
3	20
4	9
5	6
6	5
7	4
8	2
9	1
10	1
11	2
12	1

MME & MLE används för att hitta en bra skattare

X_i = # hops som fågel i gör. Antag att $X_i \sim Geom(p)$

a) Skatta p

Vi har $E[X_i] = \frac{1}{p} \overset{MME}{\Longrightarrow} \bar{X} = \frac{1}{\hat{p}} \Rightarrow \hat{p} = \frac{1}{\bar{X}}$

$$\bar{X} = \frac{48 + 2\cdot 31 + 3\cdot 20 + \ldots + 12\cdot 1}{48 + 31 + \ldots + 1} = \frac{363}{130} \Rightarrow \hat{p} = \frac{130}{363} \approx 0.358$$

b) Hitta approximativt 95% CI för $\hat{p}$.

$$\sigma^2 = Var(X_i) = \frac{1-p}{p^2}$$

$$\sigma^2_{\bar{X}} = Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n}\frac{1-p}{p^2}$$

$$\Rightarrow S^2_{\bar{X}} = \frac{1}{n}\frac{1-\hat{p}}{\hat{p}^2} \approx 0.0385 \qquad S_{\bar{X}} \approx 0.1963$$

Vi har: $\frac{\bar{X} - \mu}{S_{\bar{X}}} \overset{(CLT)}{\approx} N(0,1)$

$Z \sim N(0,1)$ $P(-1.96 \le Z \le 1.96) = 0.95 \approx P\left(-1.96 \le \frac{\bar{X}-\mu}{S_{\bar{X}}} \le 1.96\right) =$ $\left[\mu = E[X_i] = \frac{1}{p}\right]$

$= P(-1.96 S_{\bar{X}} \le \bar{X} - \mu \le 1.96 S_{\bar{X}}) = P(\bar{X} - 1.96 S_{\bar{X}} \le \mu \le \bar{X} + 1.96 S_{\bar{X}}) = P\left(\frac{1}{\bar{X} + 1.96 S_{\bar{X}}} \le \frac{1}{\mu} \le \frac{1}{\bar{X} - 1.96 S_{\bar{X}}}\right) \Rightarrow$

$\Rightarrow p \in \left(\frac{1}{\bar{X} + 1.96 S_{\bar{X}}}, \frac{1}{\bar{X} - 1.96 S_{\bar{X}}}\right)$ med 95% konfidens

Ex: Randomiserade svars-metod

Fängelse med $N=500$ fångar. $Np = \#$ heroinanvändare
$N(1-p) = \#$ ej h-anv.

Använder du heroin?
Kasta tärning: Om $\{6\}$ ljug, annars tala sanning.

$Y = \#$ som svarade ja blev $Y = 125$

Vi får: $Y = Y_p + Y_q$ där $Y_p \sim Bin(Np, \frac{5}{6})$ och
$Y_q \sim Bin(N(1-p), \frac{1}{6})$

Observerad andel ja: $\pi = \frac{Y}{N} = 0.25$

$$\mathbb{E}[\pi] = \frac{1}{N}\mathbb{E}[Y] = \frac{1}{N}(\mathbb{E}[Y_p] + \mathbb{E}[Y_q]) = \ldots = \frac{1+4p}{6}$$

$$Var(\pi) = \frac{1}{N^2}(Var(Y_q) + Var(Y_p)) = \ldots = \frac{1}{N}\cdot\frac{5}{6}\cdot\frac{1}{6} \Rightarrow$$

$$\Rightarrow \sigma_\pi = \frac{1}{60}$$

$$MME: \pi = \frac{1+4\hat{p}}{6} \Rightarrow \hat{p} = \frac{6\pi - 1}{4} = 0.125$$

$$\sigma_{\hat{p}} = \frac{6}{4}\sigma_\pi = 0.025$$

Vi får att ett approximativt 95% CI för p blir:
$\hat{p} \pm 1{,}96 \cdot 0.025$

Hypotestest Def: Nollhypotesen betecknas H_0, den alternativa hypotesen betecknas H_1, eller H_A.

Ex: Singlar slant och vill uppskatta p-proport. T
$H_0: p = \frac{1}{2}$ $H_1: p > \frac{1}{2}$ (manipulerad slant)

Ex: Aspirin/~~cancer~~ hjärtinfarkt

H_0: Aspirin har ej förebyggande effekt.

H_1 : " <u>har</u> " "

H_0 är en etablerad teori som måste motbevisas.
~~Bevis~~bördan ligger på H_1.

<u>Def</u>: Typ I och typ II fel
Typ I fel = Vi förkastar H_0 fastän H_0 är sann.
Typ II fel = Vi accepterar H_0 fastän H_1 är sann.

Verkligheten	Accepterar H_0	Förkastar H_0
H_0 är sann	Rikt acc. $1-\alpha = P(\text{acc}H_0 \mid H_0)$	$\alpha = P(\text{förkastar } H_0 \mid H_0)$ sann. att vi gör ett typ I fel
H_0 är falsk	Falskt accept. $\beta = P(\text{acc } H_0 \mid H_1)$	Riktigt förkast. $1-\beta = P(\text{förk. } H_0 \mid H_1)$

α är signifikansnivån. $1-\beta$ kallas för teststyrkan (power)
Ofta är typ I fel allvarligare än typ II.
Vi vill därför kontrollera α (mer än β)

<u>Def</u>: En <u>test-statistika</u> är en funktion av datamängden med olika värden (fördelningar) under H_0 och H_1.
<u>Ex</u>: Singlar slant 100 gånger och låter $\hat{p} = \frac{\#T}{100}$ vara vår test-statistika. Vi kan bestämma oss för att förkasta H_0 om $\hat{p} \geq 0.7$.
Relevanta frågor: Varför 0.7? Vad skall vi ta annars?
Vad blir α?
<u>Def</u>: Förkastningsområde (RR, Rejection Region) är den mängd av värden på vår test-statistika som gör att vi förkastar H_0.
I exemplet är $RR = \{\hat{p} \in [0.7, 1]\}$
Om RR ökar, så $\alpha \nearrow$ $\beta \searrow$ (vi förkastar mer)
Vi hittar ett "bra" α. Därefter hittar vi ett RR s.a.
$P(\text{test-statistikan} \in RR \mid H_0) = \alpha$

När sannolikheten beräknas antas test-statistikan ha fördelningen som antas under H_0.

Ex: Ok att felaktigt anklaga för fusk med 5% sannolikhet. Vi vill hitta c s.a $P(\hat{p} \geq c \mid H_0) = 0.05$
($RR = [c, 1]$)

Om H_0 sant $\mathbb{E}[\hat{p}] = \frac{1}{2}$ $Var(\hat{p}) = \frac{p(1-p)}{n} = \frac{1}{400}$

$\overset{CLT}{\Longrightarrow} \hat{p} \approx N(\frac{1}{2}, \frac{1}{400}) \Rightarrow z = \frac{\hat{p} - \frac{1}{2}}{\frac{1}{20}} = 20(\hat{p} - \frac{1}{2}) \approx N(0,1)$

$0.05 = P(\hat{p} \geq c \mid H_0) = P(20(\hat{p} - \frac{1}{2}) \geq 20(c - \frac{1}{2})) =$
$= P(z \geq 20(c - \frac{1}{2}))$

Ur $N(0,1)$ tabell fås $20(c - \frac{1}{2}) \approx 1{,}645 \Rightarrow$
$\Rightarrow c \approx \frac{1.645}{20} + \frac{1}{2} = 0.8523$

Dvs om $\hat{p} \geq 0.5823$ så förkastar vi H_0. Risken för typ I fel är då 5%

Likelihood ratio-test

Def: En (simpel) enkel hypotes H är en hypotes som fullständigt bestämmer fördelningen (för teststatistikan)

Ex: Vi har $X \sim N(0, \sigma^2)$ $H_0 : \sigma = \sigma_0$
H_0 är en enkel hypotes!

Def: En hypotes som ej är enkel kallas för sammansatt.

Ex: $X \sim N(0, \sigma^2)$ $H_1 : \sigma \neq \sigma_0$ H_1 är en sammansatt hypotes.

Antag att H_0 och H_1 är enkla hypoteser.

Vi bildar $f(X \mid H_0)$ och $f(X \mid H_1)$ där $f(X \mid H_i)$ är den gemensamma pdf/pmf:en för data $X = (X_1, \ldots, X_n)$ givet hypotesen H_i $i = 0,1$

$\frac{f(X \mid H_0)}{f(X \mid H_1)}$ kallas för (sannolikhetskvoten)

Likelihood ration

Vi hittar c s.a vi förkastar H_0 om $\frac{f(\bar{X}|H_0)}{f(\bar{X}|H_1)} \leq c$

annars accepterar vi H_0.

Hur bestäms c? Från fall till fall!

Liknande exemplet med slantsingling ~~xxxx~~ räknar vi # samplingar med en viss egenskap. och betecknar den s.v Y. Vi antar att $Y \sim Bin(n, p)$ där p är proportionen med den sökta egenskapen. Enkel noll-hypotes: $H_0: p = p_0$ Test-statistika: $Z = \frac{Y - np_0}{\sqrt{np_0(1-p_0)}}$

$\approx N(0,1)$ för stora n.

där $N(0,1)$ är den approximativa noll-fördelningen

clt för st.n. fördeln.

Tre typiska alternativa hypoteser:

$\Phi(z_\alpha) = 1 - \alpha$

~~xxxx~~ ensidigt: $H_1: p < p_0$ — $RR = \{Z \leq -z_\alpha\}$

" $H_1: p > p_0$ — $RR = \{Z \geq z_\alpha\}$

tvåsidigt $H_1: p \neq p_0$ — $RR = \{Z \geq z_{\alpha/2}$ el. $Z \leq -z_{\alpha/2}\}$

RR med signifikansnivå α

Def: p-värdet av test givet data $\bar{X}$.

p-värde = den minsta signifikansnivån α (minst möjl. α) för vilket vårt test förkastar H_0 givet observerad data.

Detta (p-värdet) visar hur signifikant observerad data motsäger H_0. Om $p \leq \alpha$, kan vi förkasta H_0 med signifikansnivå α. Ex: Aspirin/h-attack • placebo: 11034, 189 hjärtattacker • aspirin: 11037, 104 —"—.

H_0: ingen effekt. H_1: Aspirin motverkar h-att.

Test-statistika: $\bar{X}$ = # h-attacker i placebogruppen. Under H_0: $\bar{X} \sim Hg(22071, 293, 0.4999)$. CLT ger $\bar{X} \approx N(146.48, 72.78)$

$\Rightarrow p = P(\bar{X} \geq 189 | H_0) = \ldots = 0.000\,0003$

8.13/ Vinkeln θ med vilken elektroner sänds ut från ett myonsönderfall har pdf $f(x|\alpha) = \frac{1+\alpha x}{2}$ $\quad -1 \leq x \leq 1, \; -1 \leq \alpha \leq 1$

Räkneövn. 13/5

$x = \cos\theta$. ~~X har~~ X har pdf $f(x|\alpha)$

a) Skatta α mha MME. Vi har $\mathbb{E}[X] = \int x \frac{1+\alpha x}{2} dx =$

$$= \left[\frac{x^2}{4} + \alpha \frac{x^3}{6}\right]_{-1}^{1} = \frac{1}{4} + \frac{\alpha}{6} - \left(\frac{1}{4} - \frac{\alpha}{6}\right) = \frac{\alpha}{3}$$

MME: $\mathbb{E}[X] \to \bar{X} \quad \alpha \to \hat{\alpha}$

$\bar{X} = \frac{X_1 + \ldots + X_n}{n}$ $\quad X_i$ oberoende och X_i

pdf $f(x|\alpha) \Rightarrow \bar{X} = \frac{\hat{\alpha}}{3} \Rightarrow \hat{\alpha} = 3\bar{X}$ //

b) Visa $\mathbb{E}[\hat{\alpha}] = \alpha$ (dvs $\hat{\alpha}$ är en väntevärdesriktig skattare)

B: $\mathbb{E}[\hat{\alpha}] = 3\mathbb{E}[\bar{X}] = 3\mathbb{E}\left[\frac{X_1 + \ldots + X_n}{n}\right] = \frac{3}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]$

$= \frac{3}{n} \cdot n \cdot \frac{\alpha}{3} = \alpha \quad \square$

c) Visa $Var(\hat{\alpha}) = \frac{3-\alpha^2}{n}$

B: $Var(\hat{\alpha}) = Var(3\bar{X}) = 9 Var\left(\frac{X_1 + \ldots + X_n}{n}\right) \overset{\text{ober.}}{=} \frac{9}{n^2}\sum_{i=1}^{n} Var(X_i)$

$= \frac{9}{n} Var(X_1)$

$\mathbb{E}[X_1^2] = \int_{-1}^{1} x^2 \frac{1+\alpha x}{2} dx = \left[\frac{x^3}{6} + \alpha\frac{x^4}{8}\right]_{-1}^{1} = \frac{1}{6} + \frac{\alpha}{8} - \left(-\frac{1}{6} + \frac{\alpha}{8}\right) =$

$= 1/3$ //

$\Rightarrow Var(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \frac{1}{3} - \frac{\alpha^2}{9} = \frac{3-\alpha^2}{9}$

$\Rightarrow Var(\hat{\alpha}) = \frac{3-\alpha^2}{n} \quad \square$

d) Använd CLT för att bestämma fördeln. av $\hat{\alpha}$ (approx.) Använd denna med $\alpha=0, n=25$ för att bestämma $P(|\hat{\alpha}|>0.5)$. CLT ger $\hat{\alpha} \approx N(E[\hat{\alpha}], Var(\hat{\alpha}))$

$\boxed{\text{(ngt är} \approx \text{normalfördelat)}}$ $= N\left(0, \frac{3-\cancel{0}}{25}\right)$

$$\Rightarrow P\left(|\hat{\alpha}|>\frac{1}{2}\right) = 1-P\left(-\frac{1}{2}<\hat{\alpha}<\frac{1}{2}\right) = 1-P\left(-\frac{1}{2\sqrt{\frac{3}{25}}} < \frac{\hat{\alpha}}{\sqrt{\frac{3}{25}}} < \frac{1}{2\sqrt{\frac{3}{25}}}\right)$$

$$\approx 1-P\left(-\frac{1}{2\sqrt{\frac{3}{25}}} < Z < \frac{1}{2\sqrt{\frac{3}{25}}}\right): \quad Z \sim N(0,1)$$

$$= 1-\left(\Phi\left(\frac{1}{2\sqrt{\frac{3}{25}}}\right) - \Phi\left(-\frac{1}{2\sqrt{\frac{3}{25}}}\right)\right) = 1-\left(\Phi\left(\frac{1}{2\sqrt{\frac{3}{25}}}\right) - \left(1-\Phi\left(\frac{1}{2\sqrt{\frac{3}{25}}}\right)\right)\right)$$

$$= 2-2\Phi\left(\frac{1}{2\sqrt{\frac{3}{25}}}\right) \approx 0.149 \text{ mha tabell}$$

8.21) $X_1, \ldots, X_n$ är i.i.d. (oberoende och lika fördelade) med pdf: $f(x|\theta) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0 & \text{annars} \end{cases}$

a) Använd MME för att hitta en skattare $\hat{\theta}$ (är liksom MLE metod för att hitta skattare)

X har pdf $f(x|\theta)$. $E[X] = \int_\theta^\infty x e^{-(x-\theta)} dx \overset{p.i}{=}$

$$= \left[-xe^{-(x-\theta)}\right]_\theta^\infty - \int_\theta^\infty -e^{-(x-\theta)} dx = 0-(-\theta e^0) +$$

$$+ \left[-e^{-(x-\theta)}\right]_\theta^\infty = \theta - 0 - (-1) = \theta+1$$

MME ($E[X] \to \bar{X}$, $\theta \to \hat{\theta}$)

$$\Leftrightarrow \bar{X} = \hat{\theta}+1 \Rightarrow \hat{\theta} = \bar{X}-1$$

b) Hitta MLE-skattaren $\hat{\theta}$. $L(\theta) = f(X_1, \ldots, X_n|\theta) =$

$$\underset{\text{ober.}}{=} \prod_{i=1}^n f(X_i|\theta) = \prod_{i=1}^n e^{-(X_i-\theta)} 1_{\{X_i \geq \theta\}} = e^{-\sum_{i=1}^n X_i + n\theta} 1_{\{\min(X_1, X_2, \ldots, X_n) \geq \theta\}}$$

(Derivera, sätt till 0 för att hitta maximum, svårt i detta fall)

Vi ser att $L(\theta)$ maximeras av $\underline{\hat{\theta} = \min(X_1, \ldots, X_n)}$

MLE-skattaren

$\underline{9.3}$ $X \sim Bin(100, p)$ $H_0 : p = \frac{1}{2}$

$H_1 = H_A$

förkastn.-område, rejection region

$H_A : p \neq \frac{1}{2}$

$RR = \{ |X - 50| > 10 \}$

Använd CLT för att bestämma (approx.)

a) α, Vi ~~förkastar~~ har $\alpha = P(\text{förkasta } H_0 | H_0) =$

$= P_{p=\frac{1}{2}}(|X - 50| > 10)$ Under H_0 är $X \sim Bin(\underset{n}{100}, \underset{p}{\tfrac{1}{2}})$

$\overset{CLT}{\Longrightarrow} X \approx N(\mathbb{E}[X], Var(X)) = N(50, \underbrace{25}_{\sigma^2}) \Rightarrow \frac{X - 50}{\underset{\sigma}{5}} \approx N(0,1)$

Vi har $\alpha = P(|X - 50| > 10) = P\left(\frac{|X - 50|}{5} > 2\right) \approx P(|Z| > 2) =$

$Z \sim N(0,1)$

$= 1 - P(-2 < Z < 2) =$

$= 1 - (\Phi(2) - \Phi(-2)) = 1 - (\Phi(2) - (1 - \Phi(2))) =$

$= 2 - 2\Phi(2) \underset{\uparrow \text{ tabell}}{\approx} 0.0456$ m.h.a tabell

b) Skissa teststyrkan som funktion av p.

Givet H_1 sann (med par. p)

$1 - \beta = P_p(|X - 50| > 10)$

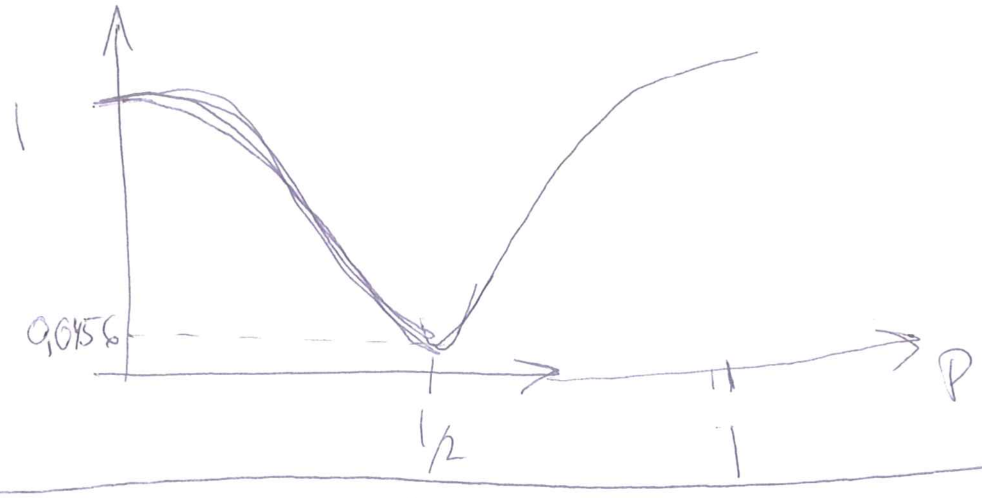
(Jag förkastar, och det var rätt) CLT ger $X \approx N(100p, 100p(1-p))$

$\Longrightarrow$ ~~$X \approx 100p$~~ $1 - \beta = P_p(|X - 50| > 10) = 1 - P(-10 < X - 50 < 10) =$

$= 1 - P(40 < X < 60) = 1 - P\left(\frac{40 - 100p}{10\sqrt{p(1-p)}} < \frac{X - 100p}{10\sqrt{p(1-p)}} < \frac{60 - 100p}{10\sqrt{p(1-p)}}\right)$

$= 1 - P\left(\frac{4 - 10p}{\sqrt{p(1-p)}} < Z < \frac{6 - 10p}{\sqrt{p(1-p)}}\right) = 1 - \left(\Phi\left(\frac{6 - 10p}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{4 - 10p}{\sqrt{p(1-p)}}\right)\right) =$

$= 1 - \Phi\left(\frac{6 - 10p}{\sqrt{p(1-p)}}\right) + \Phi\left(\frac{4 - 10p}{\sqrt{p(1-p)}}\right) = f(p)$



9.7 $X_1, \ldots, X_n$ ober. samplingar från en $Poi(\lambda)$

$H_0: \lambda = \lambda_0$ med $\lambda_1 > \lambda_0$

$H_1: \lambda = \lambda_1$

(Två enkla hypoteser, ska använda Likelihood ratio)

Bestäm likelihood ratio testet. $LR = \frac{f(X|H_0)}{f(X|H_1)}$

$$= \frac{f(X_1, \ldots, X_n | \lambda_0)}{f(X_1, \ldots, X_n | \lambda_1)} \overset{\text{ober.}}{=} \frac{\prod_{i=1}^{n} f(X_i | \lambda_0)}{\prod_{i=1}^{n} f(X_i | \lambda_1)} = \frac{\prod_{i=1}^{n} \frac{\lambda_0^{X_i}}{X_i!} e^{-\lambda_0}}{\prod_{i=1}^{n} \frac{\lambda_1^{X_i}}{X_i!} e^{-\lambda_1}} =$$

$f(X_i|\lambda_0) = P(X_i|\lambda_0)$

$$= \left(\frac{\lambda_0}{\lambda_1}\right)^{X_1 + \ldots + X_n} e^{n(\lambda_1 - \lambda_0)}$$

Vi förkastar om $\left(\frac{\lambda_0}{\lambda_1}\right)^{X_1 + \ldots + X_n} e^{n(\lambda_1 - \lambda_0)} \leq C$

Utnyttja att $X_1 + \ldots + X_n \sim Poi(n\lambda_0), Poi(n\lambda_1)$ under H_0 resp. H_1 för att förklara hur C skall väljas för givet α.

signifikansnivån

"LR litet" $\Longleftrightarrow$ $X_1 + \ldots + X_n$ "stort"

Vi kan hitta $c' = c'(\lambda_1, \lambda_0, n, c)$ s.a. $\{X_1 + \ldots + X_n \geq c'\}$

$\Longleftrightarrow \left(\frac{\lambda_0}{\lambda_1}\right)^{X_1 + \ldots + X_n} e^{n(\lambda_1 - \lambda_0)} \leq C$

Vi kan använda detta s.a $RR = \{X_1 + \ldots + X_n \geq c'\}$. Givet α kan vi under H_0 hitta c' s.a $\alpha = P(X_1 + \ldots + X_n \geq c')$

~~Proportionstest för små n~~ 14/4

Vi undersöker en population där proportionen p har en viss egenskap

Vi använder test-statistikan $Y \sim Bin(n,p)$ där n=# sampl.

Typiskt har vi $H_0: p = p_0$

Då är nollfördelningen (dvs fördelningen för Y under H_0) exakt $Bin(n, p_0)$. Om n litet är det ej ok. med $Y \sim N(np_0, np_0(1-p_0))$

(Villkor: $np > 5$, $nq > 5$)

Signifikanstest (på nivå α), ensidigt: $H_1: p > p_0$ RR= $\{Y \geq y_\alpha\}$

$H_1: p < p_0$ RR= $\{Y \leq y'_\alpha\}$

$RR = \{Y \leq y'_\alpha\}$

2-sidigt $H_1: p \neq p_0$ RR= $\{Y \leq y'_{\alpha/2}\} \cup \{Y \geq y_{\alpha/2}\}$

där $P(Y \geq y_\alpha) = 1 - F_{Bin(n,p_0)}(y_\alpha) = \alpha$

pss. $P(Y \leq y'_\alpha) = \alpha$

Ex: Övernaturlig perception Test: Man gissar färgen på 20 kort (s,h,r,k) som dras ur en kortlek. Varje kort som dragits blandas in i leken efter gissning.

Låt Y = # kort som gissas korrekt $\sim Bin(20, p)$

$H_0: p = \frac{1}{4}$ Y har nollfördelningen $Y \sim Bin(20, \frac{1}{4})$

$H_1: p > \frac{1}{4}$

y	8	9	10	11
$P_{1/4}(Y \geq y)$	0.101	0.041	0.014	0.004

5%

~~...~~: $RR = \{Y \geq 9\}$ exakt p-värde för RR är 0.041

dvs $P(Y \geq 9) = 0.041$

$P_W(p)$ = teststyrkan som funktion av p, dvs $P_W(p) = P(Y \geq 9$

$(Y \sim Bin(20, p))$

givet

p	0.27	0.3	0.4	0.5	0.6	0.7
$P_W(p)$	0.064	0.113	0.404	0.748	0.934	0.995

Dvs om personen har övernaturlig förmåga s.a $p=0.4$ så upptäcks det med sannolikhet 0.404

Det finns ett problem: Antag att 1000 personer testas med $\alpha=0.05 \Rightarrow Bin(1000, 0.05)$ kommer anses ha övernaturlig förmåga (även om H_0 sant)

Generaliserat likelihood-ratio test Vi har en parameter θ. Innan hade vi $H_0: \theta=\theta_0$

$H_1: \theta = \theta_1$

Låt $\Omega = \{$alla värden för $\theta\}$ Ofta har vi $\Omega = \Omega_0 \cup \Omega_1$ med $\Omega_0 \cap \Omega_1 = \emptyset$ (tomt snitt)

och $H_0: \theta_0 \in \Omega_0$

$H_1: \theta_1 \in \Omega_1$

Ex: $H_0: p=p_0$ $\quad X \sim Bin(100, p)$

$H_1: p \neq p_0$

Vi bildar då: $$\Lambda = \frac{\max_{\theta\in\Omega_0} f(X|\theta)}{\max_{\theta\in\Omega} f(X|\theta)} = \frac{\max_{\theta\in\Omega_0} L(\theta)}{\underbrace{\max_{\theta\in\Omega} L(\theta)}_{=L(\hat\theta) \text{ där } \hat\theta \text{ MLE}}}$$

$RR = \{\Lambda \leq \lambda_0\}$ där λ_0 väljs s.a $P(\Lambda \leq \lambda_0 | H_0) \leq \alpha$

där α är den apriori bestämda signifikansnivån.

Ex: $X_1, \dots, X_n$ är iid (oberoende lika fördelade)

$X_i \sim N(\mu, \sigma^2)$ där σ är känt. $H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

så $\Omega_0 = \{\mu_0\}$ $\quad \Omega = \{-\infty < \mu < \infty\}$

Vi får: $$\max_{\mu\in\Omega_0} L(\mu) = L(\mu_0) = f(X_1, \dots X_n | \mu_0) =$$

$$\underset{\text{ober.}}{=} \prod_{i=1}^n f(X_i | \mu_0) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(X_i-\mu_0)^2} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum(X_i-\mu_0)^2}$$

$$\max_{\mu\in\Omega} L(\mu) = L(\hat\mu) \text{ där } \hat\mu = \bar X = \frac{X_1 + \dots + X_n}{n}$$

är MLE'n för μ. $\Rightarrow \max_{\mu \in \Omega} L(\mu) = L(\bar{X}) = f(X_1, \dots, X_n | \bar{X}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(X_i - \bar{X})^2} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2}$

$\Rightarrow \Lambda = e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i - \mu_0)^2 - \sum_{i=1}^{n}(X_i - \bar{X})^2\right)}$

$\Rightarrow -2\log\Lambda = \underset{\text{övn.}}{\dots} = n\frac{(\bar{X}-\mu_0)^2}{\sigma^2}$

Vi förkastar H_0 om Λ "litet".

$\Rightarrow$ Vi förkastar H_0 om $-2\log\Lambda$ är "stort".

I själva verket är (övn.) (kolla def på χ_1^2)

$$-2\log\Lambda \sim \chi_1^2$$

Om vi vill ha signifikansnivå α förkastar vi alltså då $-2\log\Lambda > \chi_1^2(\alpha)$

Litet n igen Här antar vi PD $\sim N(\mu, \sigma^2)$ (okänt μ och σ). Om vi vill testa $H_0: \mu = \mu_0$ bildar vi test-statistikan $T = \frac{\bar{X} - \mu_0}{S_{\bar{X}}} \sim t_{n-1}$ -fördelad under H_0.

Ex: Fetthalt i korv. Skillnad i två mätningar av % fett i 16 korvar. $\bar{X} = 0.53$ $s = 1.06$

(med $\mu_0 = 0$)

Testa $H_0: \mu = 0$ $\quad S_{\bar{X}} = \frac{s}{\sqrt{16}} = 0.256 \Rightarrow T \approx 2.0$

$H_1: \mu \neq 0$

En-sidigt ($H_1: \mu > 0$) $P(T > 2 | T \sim t_{15}) = 0.032$

Två-sidigt ($H_1: \mu \neq 0$)

$$P(|T| > 2 | T \sim t_{15}) = 2 \cdot 0.0032 = 0.064$$

Vårt p-värde är 0,064 dvs vi kan ej förkasta H_0 på 5% nivån.

CI-metod: Accepterar $H_0: \mu = \mu_0$ på 5% nivån om ett 95% CI innehåller μ_0.

Förkastar H_0 om 95% **CI** ej innehåller μ_0

Ex: Fett-% Exakt 95% CI för μ var $(-0.03, 1.08)$

så vi kan ej förkasta H_0 på 5%-nivån.
Approximativa (m.h.a Normalapprox.) 95% CI för μ var $(0.01, 1.05)$ dvs vi hade förkastat $H_0: \mu = 0$ för $H_1: \mu \neq 0$ på 5%-nivån.

9.12) $X_1, \dots, X_n$ är iid sampl. från $Exp(\theta)$ med   R.öv. 14/5

pdf $f(x|\theta) = \theta e^{-\theta x}, \; x \geq 0$

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

Hitta likelihood-ratio test och visa att $RR = \{\bar{X} e^{-\theta_0 \bar{X}} \leq c\}$

Vi bildar $\Lambda = \dfrac{\max_{\theta \in \{\theta_0\}} f(X_1, \dots, X_n | \theta)}{\max_{\theta \geq 0} f(X_1, \dots, X_n | \theta)} = \dfrac{f(X_1, \dots, X_n | \theta_0)}{L(\hat{\theta})}$,

där $\hat{\theta}$ är vår MLE av θ

i) Hitta MLE

$L(\theta) = f(X_1, \dots, X_n | \theta) \overset{\text{ober.}}{=} \prod_{i=1}^{n} f(X_i | \theta) = \prod_{i=1}^{n} \theta e^{-X_i \theta} = \theta^n e^{-\sum_{i=1}^{n} \theta X_i} =$

$= \theta^n e^{-\theta n \bar{X}}$

$L'(\theta) = n \theta^{n-1} e^{-\theta n \bar{X}} - n \bar{X} \theta^n e^{-\theta n \bar{X}} = 0 \Rightarrow 1 = \bar{X} \theta$

$\Rightarrow \hat{\theta} = \frac{1}{\bar{X}}$ p.s.s. $f(X_1, \dots, X_n | \theta_0) = \theta_0^n e^{-\theta_0 n \bar{X}}$

$\Rightarrow \Lambda = \dfrac{\theta_0^n e^{-n \theta_0 \bar{X}}}{\left(\frac{1}{\bar{X}}\right)^n e^{-n \frac{1}{\bar{X}} \bar{X}}} = (\bar{X} \theta_0)^n e^{-n \theta_0 \bar{X} + n}$

Vi förkastar H_0 om $\Lambda \leq C \Rightarrow \bar{X} \theta_0 e^{-\theta_0 \bar{X} + 1} \leq \sqrt[n]{C} \Rightarrow$

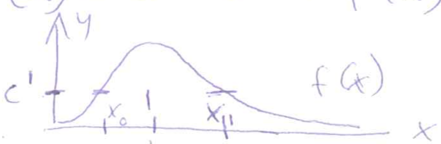
$\Rightarrow \bar{X} e^{-\theta_0 \bar{X}} \leq c' = c'(C, n, \theta_0)$ □

9.13) Antag i förra uppgiften $\theta_0 = 1$, $n = 10$, $\alpha = 0.05$ för att använda testet vill vi kunna hitta c.

a) Visa att RR kan skrivas som $\{\bar{X} \leq x_0\} \cup \{\bar{X} \geq x_1\}$ där x_0, x_1 bestäms av C. Vi hade $RR = \{\bar{X} e^{-\bar{X}} \leq c'\}$

Betrakta $f(x) = x e^{-x}$ i) $\lim_{x \to 0} f(x) = \lim_{x \to \infty} f(x) = 0$ och $f'(x) =$

$= e^{-x} - x e^{-x} = 0 \Rightarrow x = 1$ Typskiss:



dvs $f(x) \leq c' \iff x \leq x_0$ el. $x \geq x_1$ där $x_0 < x_1$ och
$f(x_0) = f(x_1) = c'$

b) Förklara varför c' skall väljas s.a. $\mathbb{P}(\bar{X}e^{-\bar{X}} \leq c') =$
$= 0.05$ då $\theta_0 = 1$

Vi förkastar H_0 även om H_0 är sann om $\bar{X} \in RR$
$\iff \bar{X}e^{-\bar{X}} \leq c'$. Detta får hända med sann. α

c) Förklara varför $\sum_{i=1}^{10} X_i$ och $\bar{X}$ är Gammafördelade

$X_i \sim Exp(\lambda) \quad 1 \leq i \leq 10$

$$M_{X_i}(t) = \mathbb{E}[e^{tX_i}] = \ldots = \frac{\lambda}{\lambda - t} \quad \text{(övning!)}$$

Om $X_1, \ldots, X_n$ är oberoende $\iff M_{X_1 + \ldots + X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t) =$

$$= (M_{X_i}(t))^n = \left(\frac{\lambda}{\lambda - t}\right)^n \leftarrow \text{Mgf för Gamma}(n, \lambda)$$

Vi använder: $M_X(t) = M_Y(t) \ \forall \ t \in (-\varepsilon, \varepsilon)$ (öppet område runt origo) $\iff X, Y$ har samma fördeln.

$$\Longrightarrow \sum_{i=1}^{n} X_i \sim Gamma(n, \lambda)$$

Med $\theta_0 = 1$ $n = 10 \Longrightarrow \sum_{i=1}^{10} X_i \sim Gamma(10, 1)$

Fördelning för $\bar{X}$: $M_{\bar{X}}(t) = \mathbb{E}\left[e^{t\frac{X_1 + \ldots + X_n}{n}}\right] = \mathbb{E}\left[e^{\frac{t}{n}(X_1 + \ldots + X_n)}\right]$

$$= M_{X_1 + \ldots + X_n}\left(\frac{t}{n}\right) = \left(\frac{\lambda}{\lambda - \frac{t}{n}}\right)^n = \left(\frac{\lambda n}{\lambda n - t}\right)^n \sim Gamma(n, \lambda n)$$

$$\Longrightarrow \theta_0 = 1 \quad n = 10 \quad \bar{X} \sim Gamma(10, 10)$$

Vi har $0.05 = \alpha = \mathbb{P}(\bar{X} \in RR \mid H_0) = \mathbb{P}(\bar{X} \leq x_0 \cup \bar{X} \geq x_1 \mid H_0) =$

$= \mathbb{P}(\bar{X} \leq x_0 \mid H_0) + \mathbb{P}(\bar{X} \geq x_1 \mid H_0) = F_{Gamma(10,10)}(x_0) + 1 - F_{Gamma(10,10)}(x_1)$

och vi har $f(x_0) = f(x_1)$

(2 villkor, 2 variabler)

$\Longrightarrow$ lös $\Longrightarrow f(x_0) = f(x_1) = c'$

$\Longrightarrow$ Vi kan bestämma även c

d) Om vi inte kommit på att $\bar{X} \sim Gamma(10, 10)$
Vad hade vi då gjort?

Vi kan generera (I MATLAB, exp-fördelade värden) $X_1, \dots, X_{10}$ 1000 ggr.
Bilda $\bar{X} \leq e^{-\bar{X}}$ alla 1000 ggr. Sedan väljer vi c s.a $\bar{X} e^{-\bar{X}} \leq c$ 50 ggr.

9.21 / $X \sim U[0, \theta]$ $H_0 : \theta = 1$ $H_1 : \theta = 2$ Hitta ett test med $\alpha = 0$ (dvs $P(\text{förk. } H_0 | H_0) = 0$) (aldrig förkasta H_0 felaktigt)

i) $RR = \{X = 1\}$ $P(X \in RR | H_0) = 0$

ii) $RR = \{X > 1\}$ $P(X > 1 | H_0) = 0$

Bestäm teststyrkan: $1 - \beta = P(\text{förkastar } H_0 | H_1)$

i) $P(X = 1 | H_1) = 0$

ii) $P(X > 1 | H_1) = \frac{1}{2}$

b) Låt $0 < \alpha < 1$. Antag att $RR = [0, \alpha]$. Vad är signifikansnivån? Teststyrkan?

$P(\text{förk. } H_0 | H_0) = P(X \in [0, \alpha] | \theta = 1) = \alpha$

$P(\text{förk. } H_0 | H_1) = P(X \in [0, \alpha] | \theta = 2) = \frac{\alpha}{2}$.

c) $RR = \{X \in [1-\alpha, 1]\}$ $P(\text{förk. } H_0 | H_0) = P(X \in [1-\alpha, 1] | \theta = 1) = \alpha$

$P(\text{förk. } H_0 | H_1) = P(X \in [1-\alpha, 1] | \theta = 2) = \frac{\alpha}{2}$

f) Låt $H_0 : \theta = 2$
$H_1 : \theta = 1$

b') $P(\text{förk } H_0 | H_0) = P(X \in [0, \alpha] | \theta = 2) = \frac{\alpha}{2}$

$P(\text{förk } H_0 | H_1) = P(X \in [0, \alpha] | \theta = 1) = \alpha$

c') $P(\text{förk. } H_0 | H_0) = P(X \in [1-\alpha, 1] | \theta = 2) = \frac{\alpha}{2}$

$P(\text{förk. } H_0 | H_1) = P(X \in [1-\alpha, 1] | \theta = 1) = \alpha$

Enkel linjär regression

Mån LV8, 19/5

Används för att beskriva sambandet mellan två (kontinuerliga) s.v.

X = den förklarande variabeln.

Y = den beroende variabeln.
(ej samma beroende som i sann.hetsteori)

Vi har att tillgå n st parvis observerade data (x_i, y_i)

Marginella samplingsvarianserna:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\cancel{X}_i - \cancel{\bar{X}})^2 \qquad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

medelvärdet av

Samplingens kovarians: $c_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) =$

$$= \ldots = \frac{n}{n-1}(\overline{xy} - \bar{x}\,\bar{y})$$

Samplingens korrelationskoefficient: $r = \dfrac{c_{xy}}{s_x s_y}$

<u>Minsta kvadratmetoden</u> Vi antar att sambandet mellan X och Y kan skrivas som: $Y = \beta_0 + \beta_1 X + \varepsilon$ där $\varepsilon \sim N(0, \sigma^2)$ är en oberoende störning.

Vi vill ur data bestämma det "sanna" sambandet (dvs utan störningsterm) $y = \beta_0 + \beta_1 x$ på "bästa" sätt mha våra n datapunkter.

Vi anpassar en linje $y = b_0 + b_1 x$ s.a

$$S(b_0, b_1) := \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \text{ minimeras.}$$

Här är $\hat{y}_i = b_0 + b_1 x_i$

(lättast att räkna med kvadratiskt)

$S(b_0, b_1)$ kallas för SSE (Summed Squared Error)

För att minimera $S(b_0, b_1)$ bildar vi $\frac{\partial S}{\partial b_0} = 0$ och $\frac{\partial S}{\partial b_1} = 0$.

Räkningar ger: lutn. (slope) $b_1 = \dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} =$

$$= r\frac{s_y}{s_x}$$

intercept (skärning) $b_0 = \bar{y} - b_1\bar{x}$

Funktionen $f(x) = b_0 + b_1 x$, kallas för den <u>linjära regressionen av y på x</u>.

Vi har alltså att b_0 och b_1 skattar β_0 resp β_1 (b_0, b_1 bet. också $\hat{\beta}_0$ och $\hat{\beta}_1$). Man kan visa att b_0, b_1 är väntevärdesriktiga och konsistenta.

<u>CI och hypotestest för b_0, b_1</u> Vi har (antar)

$$b_1 \sim N\left(\beta_1, \frac{\sigma_1^2}{N-1}\right), \quad \sigma_1^2 = \frac{\sigma^2}{S_x^2}$$

$$b_0 \sim N\left(\beta_0, \frac{\sigma_0^2}{N-1}\right) \quad \sigma_0^2 = \sigma_1^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Negativ kovarians

$$Cov(b_0, b_1) = -\frac{\sigma^2 \bar{x}}{(n-1)S_x^2}$$

Skattade standardfelen: $S_{b_1} = \frac{S}{S_x\sqrt{n-1}}$ $\quad S_{b_0} = S_{b_1}\sqrt{\frac{1}{n}\sum_{n=1}^{n} x_i^2}$

Exakt $100(1-\alpha)\%$ CI för β_1: $b_1 \pm t_{\alpha/2, n-1} S_{b_1}$

Typiskt hypotestest

$H_0 \quad \beta_1 = 0$

$H_1 \quad \beta_1 > 0$

Använd teststatistiskan $T = \frac{b_1 (-\beta_1)}{S_{b_1}} \sim t_{n-2}$

Repetition (urplock)

V1: Utfallsrum, slumphändelser. (diskreta sv)

Divisionsregeln: $\mathbb{P}(A) = \frac{\#\text{gynsamma utfall}}{\text{totala } \#\text{utfall}}$

Betingade sannolikheten (för A givet B)

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

A och B är oberoende om $\mathbb{P}(A \cap B) = \mathbb{P}(A)P(B)$

V2: Diskret s.v. pmf (sannolikhetsfördelningen) är $\{p_k\}$ $p_k = \mathbb{P}(X = x_k)$ $\{x_0, x_1, \ldots\}$ dom värden som X kan anta.

Vi introducerade cdf:en $F(x) = \mathbb{P}(X \leq x)$ om X antar värden i $\mathbb{R}$.

Vi definierade $\mathbb{E}[X]$, $\mathbb{E}[g(X)]$, $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Fördelningar: 1) Likformig 2) Bern(p) 3) Bin(n,p) 4) Hg(N,n,p) 5) Geom(p) 6) Nb(r,p) 7) Poi(λ) (Ta med på A4-papper)
- Hur ser det ut?
- När uppkommer de?

V3: Kontinuerliga s.v. Vi def. cdf:en $F(x) = P(X \leq x)$ med pdf:en (probab. density function) (sannolikhetsfördeln.) $f(x) = F'(x)$ (antar deriverbar)

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

$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Fördelningar: 1) $U(a,b)$ 2) $Exp(\lambda)$ (visst släktskap m. Poi) 3) $Gamma(\alpha, \lambda)$ 4) $N(\mu, \sigma^2)$

Transformationer av s.v.

Vy: Gemensamma fördelningar, marginal- och betingade fördelningar.

Diskret: 1) $P_{X,Y}(n,k) = \mathbb{P}(X=n, Y=k)$ $\forall$ n, k som X, Y kan antá. (gemensam fördeln.)

2) marginal förd: $P_X(n) = \mathbb{P}(X=n) = \sum_k \mathbb{P}(X=n, Y=k)$

3) betingade-förd: $P_{Y|X}(k|n) = \frac{P_{XY}(n,k)}{P_X(n)} = \frac{\mathbb{P}(X=n, Y=k)}{\mathbb{P}(X=n)}$

X och Y oberoende om $P_{XY}(n,k) = P_X(n) P_Y(k)$ $\forall$ n, k

Kontinuerliga fallet:

1) gemensamma fördeln: $\mathbb{P}(a < X < b, c < Y < d) =$

$= \int_a^b \int_c^d \underbrace{f_{XY}(x,y)}_{\text{gemensam pdf:en}} dx dy$

2) (motsv. diskreta: integrerar (ist för summera) bort Y-beroendet) marg-fördeln: $\cancel{\iint f_{XY}(x,y)}$

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

3) betingade pdf:en $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$

X och Y oberoende om $f_{XY}(x,y) = f_X(x) f_Y(y)$

Vi def. kovarians och korrelation mellan X och Y.

Fördelningar: 1) $Mn(n, p_1, \dots, p_r)$ 2) $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$
(Bivariat) korr. koeff.

Momentgenererande funktion (mgf)

Def: $M(t) = M_X(t) := \mathbb{E}[e^{tX}]$ om den $\exists$.

Egenskaper: 1) $X \overset{d}{=} Y$ (dvs X och Y har samma fördeln.) $\iff M_X(t) = M_Y(t)$ $\forall t \in (-\varepsilon, \varepsilon)$
har samma fördeln.

2) $\mathbb{E}[X^n] = M^{(n)}(0) \quad \forall\, n \in \mathbb{N}$

3) $\mathbb{E}[e^{tX+sY}] = \mathbb{E}[e^{tX}]\,\mathbb{E}[e^{sY}] \quad \forall\, (t,s) \in (-\varepsilon,\varepsilon)$

$\Leftrightarrow X, Y$ oberoende

4) $\{X_n\}_{n\in\mathbb{N}}$ en sekvens av s.v så gäller

$$X_n \to X \iff M_{X_n}(t) \to M_X(t) \quad \forall\, t \in (-\varepsilon,\varepsilon)$$

V5: LLN $\bar{X}_n = \dfrac{X_1 + \dots + X_n}{n}$ och $\mathbb{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$ och X_i är oberoende. Fixera $\varepsilon > 0$.

$$P(|\bar{X}_n - \mu| > \varepsilon) \xrightarrow[n\to\infty]{} 0$$

CLT: Låt $S_n = X_1 + \dots + X_n$ (X_i ober) $\mathbb{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2 < \infty$) $\Rightarrow$ $P\left(\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \xrightarrow[n\to\infty]{} \Phi(x)$ $\forall x$

$\Phi(x)$ cdf för $N(0,1)$

Visas gnm mgf.

Använd till (normal-) approximation.

1) $X \sim Bin(n,p) \Rightarrow X \approx N(\underbrace{np}_{\text{väntevärdet}}, \underbrace{np(1-p)}_{\text{variansen.}})$ $\quad np, n(1-p) \geq 5$

2) $X \sim Hg(N,n,p) \Rightarrow X \approx N\left(np, np(1-p)\dfrac{N-n}{N-1}\right)$

$np, n(1-p) \geq 5$

14.2) Data: Tis 20/5

LV8

x	0.34	1.38	-0.65	0.68	1.40	-0.88	-0.30	-1.18
Y	0.27	1.34	-0.53	0.35	1.28	-0.98	-0.72	-1.75

0.50
-0.81
0.64
-1.59

a) Använd minsta kvadratmetoden för att anpassa en linje $y = b_0 + b_1 x$ från dessa data.

Dvs hitta b_0, b_1 som minimerar SSE:n :

$$S(b_0, b_1) = \sum_{i=1}^{10} (y_i - b_0 - b_1 x_i)^2$$

Vi fick: $b_1 = \dfrac{10 \sum_{i=1}^{10} x_i y_i - \sum_{i=1}^{10} x_i \sum y_i}{10 \sum_{i=1}^{10} x_i^2 - (\sum_{i=1}^{10} x_i)^2}$; $b_0 = \bar{x} - b_1 \bar{y}$

Med våra data (använd räknare) så får vi $b_1 \approx 0.9044$

$b_0 \approx -0.0334$

b) Använd minsta kvadratmetoden för att anpassa en linje. $x = c_0 + c_1 y$. Dvs minimera $S(c_0, c_1) = \sum_{i=1}^{10} (x_i - c_0 - c_1 y_i)^2$. vi har: $c_1 = \dfrac{\sum_{i=1}^{10} x_i y_i - \sum_{i=1}^{10} x_i \sum_{i=1}^{10} y_i}{10 \sum_{i=1}^{10} y_i^2 - (\sum_{i=1}^{10} y_i)^2}$

$c_0 = \bar{y} - c_1 \bar{x} \implies c_1 \approx 1.055$

$c_0 \approx 0.0331$

(Linjerna blir ej ~~samma~~ likadana ty här har vi minimerat avstånden i höjdled ist för i sidled.)

c) Får vi samma linje? Nej, därför att i a) minimerar vi det vertikala (kvadratiska) felet och medans vi i b) min. det horisontella (kvadratiska) felet.

14.12) Visa att den linje som fås ur minsta kvadrat anpassning går igenom $(\bar{x}, \bar{y})$

B: $b_0 + b_1 \bar{x} = \bar{y} - b_1 \bar{x} + b_1 \bar{x} = \bar{y}$ $\square$

14.7) Vi har X, Y två s.v med $\mathbb{E}[X] = \mu_x$ $\mathbb{E}[Y] = \mu_y$

$Var(X) = \sigma_x^2$ $Var(Y) = \sigma_y^2$ $cov(X, Y) = \sigma_{xy}$

Vi skattar Y från X mha $\hat{Y} = \alpha + \beta X$

a) Hitta α, β som minimerar $\mathbb{E}[(Y - \hat{Y})^2]$

$\mathbb{E}[(Y - \hat{Y})^2] = \mathbb{E}[(Y - \alpha - \beta X)^2] = \mathbb{E}[(Y - \hat{Y} - \mathbb{E}[Y - \hat{Y}] + \mathbb{E}[Y - \hat{Y}])^2]$

$\mathbb{E}[((Y - \hat{Y}) - \mathbb{E}[Y - \hat{Y}])^2] + \mathbb{E}[Y - \hat{Y}]^2 +$

$+2(Y-\hat{Y}-\mathbb{E}[Y-\hat{Y}])\mathbb{E}[Y-\hat{Y}]] = Var(Y-\hat{Y}) +$
$+\mathbb{E}[Y-\hat{Y}]^2 + 2\mathbb{E}[Y-\hat{Y}]\mathbb{E}(Y-\hat{Y}-\mathbb{E}[Y-\hat{Y}]]$

Dvs vi har $\mathbb{E}[(Y-\hat{Y})^2] = Var(Y-\hat{Y}) + \mathbb{E}[Y-\hat{Y}]^2$

$\mathbb{E}[Y-\hat{Y}]^2 = \mathbb{E}[Y-\alpha-\beta X]^2 = (\mu_y-\alpha-\beta\mu_x)^2$

$Var(Y-\hat{Y}) = Var(Y) + Var(\hat{Y}) - 2Cov(Y,\hat{Y}) =$

$= \sigma_y^2 + Var(\beta X) - 2Cov(Y,\beta X) = \sigma_y^2 + \beta^2\sigma_x^2 - 2\beta\sigma_{xy}$

$[Cov(Y,\hat{Y}) = \mathbb{E}[Y-\mathbb{E}[Y]](\cancel{\alpha}+\beta X-\cancel{\alpha}-\beta\mathbb{E}[X]])$

$= Cov(Y,\beta X) = \beta Cov(Y,X)]$

Vi har: $\mathbb{E}[(Y-\hat{Y})^2] = \sigma_y^2 + \beta^2\sigma_x^2 - 2\beta\sigma_{xy} + (\mu_y - \alpha -$
$-\beta\mu_x)^2$

Låt $f(\beta) = \sigma_y^2 + \beta^2\sigma_x^2 - 2\beta\sigma_{xy}$

$\Rightarrow f'(\beta) = 2\beta\sigma_x^2 - 2\sigma_{xy} = 0$

$\beta = \frac{\sigma_{xy}}{\sigma_x^2}$

$f''(\beta) = 2\sigma_x^2 > 0 \Rightarrow \min$

sedan låt $\mu_y - \alpha - \beta\mu_x = 0 \Rightarrow \alpha = -\beta\mu_x + \mu_y$

Dessa värden minimerar $\mathbb{E}[(Y-\hat{Y})^2]$

---

b) Visa att med α, β som ovan gäller:

$$\frac{Var(Y) - Var(Y-\hat{Y})}{Var(Y)} = \rho^2$$ där ρ är korrelationskoeff.

B: $Var(Y) = \sigma_y^2$; $Var(Y-\hat{Y}) = \sigma_y^2 + \beta^2\sigma_x^2 - 2\beta\sigma_{xy}$

$= \sigma_y^2 + \frac{\sigma_{xy}^2}{\sigma_x^2} - 2\frac{\sigma_{xy}^2}{\sigma_x^2} = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}$

$$\Rightarrow \frac{Var(Y) - Var(Y - \hat{Y})}{Var(Y)} = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \rho^2$$