FÖRELÄSNING 23/3

FINNS SLUMP? Ja, i kvantvärlden tycks det vara så, men det vi betraktar som slump är oftast osäkerhet; skeenden som är för komplexa för att man skall kunna ha kontroll över dem, men som i princip kan vara deterministiska.

EXEMPEL: - Hur kommer ett tärningsslag utfalla. I princip förutsägbart men i praktiken oförutsägbart.
- Hur kommer svenska folket rösta i valet 2022.
- Finns liv på andra planeter?
- Dog farao Tutankhamon av en genetisk sjukdom?
- Härstammar fåglar från dinosurier?

VAD ÄR SANNOLIKHET?

Förslag: Vad relativa frekvensen konvergerar mot. T.ex. singla slant n gånger. Låt X_n vara antalet gånger du får klave. Rimligt att anse att sannolikheten att få klave vid kast är:

$$\lim_{n \to \infty} \frac{X_n}{n}$$

Två problem: Existerar gränsvärdet? Kvicker upprepbarhet.

En del slumpproblem/fenomen är inte ens i princip upprepbara, men vi vill ändå applicera sannolikhetsteori på dem.

EXEMPEL: • Vad är sannolikheten att Barcelona slår Chelsea?
• Vad är sannolikheten att det regnar på fredag?

Di är sannolikhet snarare en grad av tro, som man vill sätta siffror på. Den kan uppdateras med ny information ("Det regnar idag") Viktigt inom AI.

Sannolikhetsteori säger inget om vad sannolikhet är, bara hur den måste uppföra sig. Man utgår alltid från en modell, precis som i annan matematik. Modellen har man goda skäl att tro eller intressera sig för (sannolikhetsteori) eller vill testa (statistikteori)

SANNOLIKHETSTEORI

Ett försök vars utfall beror på slumpen kallas för ett <u>slumpförsök</u>.

Mängden S av alla möjliga utfall, u, kallas för försökets <u>utfallsrum</u>

Mer bildligt: Om slumpförsöket är ett lotteri, så är S tunnan med lotter.

EXEMPEL: Singla slant. De möjliga utfallen är krona/klave. Boken skriver T, H för Tails, Heads. Då får vi:

$$S = \{H, T\}$$

EXEMPEL: Välj ett tal på måfå mellan 0 och 1.

$$S = [0, 1]$$

EXEMPEL: Hur länge håller LED-lampan?

$$S = [0, \infty)$$

DEFINITION: En delmängd A till S kallas för en händelse.

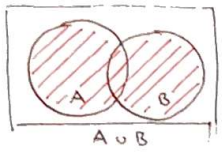
EXEMPEL: Betrakta ett tärningskast. $S = \{1, 2, 3, 4, 5, 6\}$

$$A = \{1, 3, 5\} = \{u \in S : u \text{ udda}\} = \{\text{udda utfall}\}$$

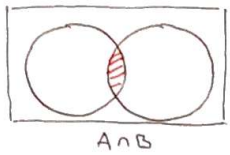
$$B = \{1, 2\} = \{u \in S : u \leq 2\} = \{\text{Vi får högst en tvåa}\}$$

Om slumpförsöket får ett utfall $u \in A$, säger vi att "A inträffar"

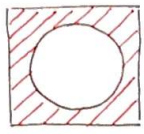
Att säga att $A \cup B$ inträffar är di att säga att A eller B inträffar
Att säga att $A \cap B$ inträffar är samma sak som att säga både A och B inträffar
A^c inträffar är att säga A inte inträffar.



$A \cup B$



$A \cap B$



A^c

SANNOLIKHETSMÅTT

Man skriver $\mathcal{P}(S)$ för mängden av alla delmängder till S.

DEFINITION: Om $\mathbb{P} : \mathcal{P}(S) \longrightarrow \mathbb{R}$ uppfyller:

(I) $0 \leq \mathbb{P}(A) \leq 1 \quad \forall A$

(II) $\mathbb{P}(S) = 1$

(III) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ för alla parvis disjunkta mängder $A_1, A_2, \ldots, \; (A_i \cap A_j = 0 \text{ om } i \neq j)$

kallas $\mathbb{P}$ sannolikhetsmått.

Tag $A_1 = S$ och $A_2 = A_3 = \ldots = \phi$ i (III) och få

$$\mathbb{P}(S) = \mathbb{P}(S) + \mathbb{P}(\phi) + \mathbb{P}(\phi) + \ldots \longrightarrow \mathbb{P}(\phi) = 0$$

PROPOSITION: För alla händelser A och B gäller att:

a) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

b) $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$

c) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

d) $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

BEVIS: a) Eftersom A och A^c är disjunkta och $A \cup A^c = S$ gäller enligt (II) och (III) att $1 = \mathbb{P}(S) = \mathbb{P}(A) + \mathbb{P}(A^c)$

b) Enligt (III) gäller

$$\mathbb{P}(B) = \mathbb{P}((B \setminus A) \cup (A \cap B)) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$$

c) Enligt (III) och sedan (b) gäller

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

d) Om $A \subseteq B$ gäller $A \cap B = A$ så b)

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A) \geq \mathbb{P}(A) \quad \text{enligt (I)}$$

EXEMPEL: Risken för regn på lördag är 50%, risken för regn på söndag är 40%. Risken för regn båda dagarna är 30%. Vad är chansen till uppehåll i helgen?

Vad är S? Inte så mycket, men vi kan specifiera:

$$S = \{(\text{regn}, \text{regn}), (\text{regn}, \text{uppehåll}), (\text{uppehåll}, \text{regn}), (\text{uppehåll}, \text{uppehåll})\}$$

Låt $A = \{\text{regn på lördag}\}$

$B = \{\text{regn på söndag}\}$

Vi söker $\mathbb{P}((A \cup B)^c)$

$$\mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) = 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B))$$
$$= 1 - (0.5 + 0.4 - 0.3) = 0.4$$

För tre händelser gäller det att A, B och C.

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

KONTINUITET HOS SANNOLIKHETER

Antag att $A_1 \subseteq A_2 \subseteq \dots$ och skriv $A = \bigcup_{n=1}^{\infty} A_n$. Då gäller enligt (III) att ($A_0 = \emptyset$)

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A_n \setminus A_{n-1}) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})$$
$$= \lim_{N \to \infty} \sum_{n=1}^{N} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1}) = \lim_{N \to \infty} \mathbb{P}(A_N)$$

Kortfattat: Om $A_1 \subseteq A_2 \subseteq \dots$ och $A = \bigcup_{n=1}^{\infty} A_n$ gäller att $\mathbb{P}(A_n) \longrightarrow \mathbb{P}(A)$ då $n \to \infty$

Om istället $A_1 \supseteq A_2 \supseteq \dots$ och $A = \bigcap_{n=1}^{\infty} A_n$ gäller enligt det ovanstående att $\mathbb{P}(A_n^c) \longrightarrow \mathbb{P}(A^c)$ och därmed att $\mathbb{P}(A_n) \longrightarrow \mathbb{P}(A)$ då $n \to \infty$

EXEMPEL: Välj ett tal på mäfi $[0,1]$. Detta betyder att $S = [0,1]$ och $\mathbb{P}(A) = l(A)$ = längden av A. Detta är ett sannolikhetsmått. (kräus mätteori)

Låt $A_n = [0, \frac{1}{2} + \frac{1}{n}]$. Då är $A = \bigcap_n A_n = [0, \frac{1}{2}]$ och $\mathbb{P}(A_n) \longrightarrow \mathbb{P}(A)$.

SANNOLIKHET OCH STATISTIK 24/3

UPPRÄKNELIGA OCH ÄNDLIGA UTFALLSRUM.

$S = \{u_1, u_2, \dots\}$

Exempelvis om vi singlar slant till första klaven är det naturligt att ta $S = \{1, 2, 3, \dots\}$

Antag att $p_1, p_2, \dots$ är sådana att $p_i \geq 0$ för alla i och $\sum_{i=1}^{\infty} p_i = 1$. Låt för alla $A \in P(A)$

$$\mathbb{P}(A) = \sum_{i: u_i \in A} p_i$$

Då är $\mathbb{P}$ sannolikhetsmått och $\mathbb{P}(\{u_i\}) = p_i$

EXEMPEL: Slant singling till första klaven, $S = \{1, 2, 3, \dots\}$
Med en rättvis slant har man $p_n = \frac{1}{2^n}$, ds

$p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$, ...

Vi får exempelvis

$$\mathbb{P}(\{\text{Högst tre kast krävs}\}) = \mathbb{P}(\{1,2,3\}) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$$

DET KLASSISKA SANNOLIKHETSMÅTTET

Här är $S = \{u_1, \dots, u_n\}$ och $p_i = \frac{1}{n}$. Då blir:

$$\mathbb{P}(A) = \frac{\text{antalet utfall } : A}{\text{totala antalet utfall}} = \frac{|A|}{n} = \frac{|A|}{|S|}$$

Vanlig situation

EXEMPEL: Om man slår en tärning tre gånger, vad är sannolikheten att få en sexa?

$S = \{(x,y,z) \mid x, y, z \in \{1,2,3,4,5,6\}\}$. Då

Det gäller att $|S| = 6^3 = 216$. Vi söker $\mathbb{P}(A)$ för

$A = \{(x,y,z) \in S \mid \text{exakt en av } x, y, z \text{ är } 6\}$

Vi ser $|A| = 1 \cdot 5 \cdot 5 + 5 \cdot 1 \cdot 5 + 5 \cdot 5 \cdot 1 = 75$. Alltså

$$\mathbb{P}(A) = \frac{75}{216} = \frac{25}{72}$$

EXEMPEL: I en hatt ligger tre kort: rött-rött, rött-svart, svart-svart
Ett kort dras på måfå, och läggs med måfå sidan upp. Vad är sannolikheten att den andra sidan har samma färg?

kalla korten a, b, c.

$S = \{a \text{ sida } 1, a \text{ sida } 2, b \text{ sida } 1, b \text{ sida } 2, c \text{ sida } 1, c \text{ sida } 2\}$

Vi söker $\mathbb{P}(A)$ där $A = \{a \text{ sida } 1, a \text{ sida } 2, c \text{ sida } 1, c \text{ sida } 2\}$

Alltså blir svaret $\mathbb{P}(A) = \frac{4}{6}$

Vi behöver alltså kunna räkna antal element i en mängd

KOMBINATORIK

MULTIPLIKATIONSPRINCIPEN: Om man utför r stycken experiment i tur och ordning och experimenten kan utfalla på $n_1, n_2, \ldots$ respektive n_r sätt, så är det totala antalet utfall för alla experiment tillsammans

$$n_1 \cdot n_2 \cdot \ldots \cdot n_r$$

EXEMPEL: Födelsedagsproblemet. En skolklass har r elever. Vad är sannolikheten att alla har olika födelsedagar?

$$S = \{(x_1, x_2, \ldots, x_r) \mid \forall i : x_i \in \{1, 2, \ldots, 365\}\}$$

$$A = \{(x_1, x_2, \ldots, x_r) \in S \mid \forall i \neq j : x_i \neq x_j\}$$

$$\mathbb{P}(A) = \frac{365 \cdot 364 \cdot \ldots \cdot (365 - r + 1)}{365^r}$$

Detta är mindre är $\frac{1}{2}$ då $r \geq 23$

Detta var ett exempel på att: Antalet sätt att välja r element ur en mängd med n element utan återläggning med hänsyn till ordning är:

$$n(n-1)(n-2)\ldots(n-r+1) = \frac{n!}{(n-r)!} =: (n)_r$$

Detta följer av multiplikationsprincipen.
Antalet sätt att välja r element ur en mängd av n element utan återläggning utan hänsyn till ordning betecknas

$$\binom{n}{r}$$

Enligt multiplikationsprincipen gäller

$$(n)_r = \binom{n}{r} \cdot r!$$

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

EXEMPEL: Att välja två kort ur en kortlek kan göras, med hänsyn till ordning på $52 \cdot 51 \cdot 50 = 132\,600$ sätt
Utan hänsyn till ordning kan det väljas på:

$$\binom{52}{3} = \frac{52 \cdot 51 \cdot 50}{1 \cdot 2 \cdot 3} = 52 \cdot 17 \cdot 25 = 22100 \text{ sätt}$$

Antalet sätt att välja r element ur en mängd med n element med återläggning och med hänsyn till ordning är förstås $(n)_r$

Antalet sätt att välja r element med återläggning men utan hänsyn till ordningen är lika med antalet icke-negativa lösningar till

$$x_1 + x_2 + \ldots + x_n = r$$

Detta är antalet sätt att placera ut $n-1$ vägar mellan r kollar

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

BETINGAD SANNOLIKHET OCH OBEROENDE

EXEMPEL: Två tärningar slås. Vad är sannolikheten för två sexor?
Tärningarna påverkar inte varandra, så om A_i är händelsen att tärning i visar sexa $(i = 1,2)$ får vi:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$$

EXEMPEL: Vad är sannolikheten att det regnar både den 1 och 2 augusti i år? Lät R_i vara händelsen att det regnar i:te augusti. Eventuell korrelation om att det regnar 1/8 påverkar hur sannolikt vi tror ~~att det~~ att det regnar 2/8

$$\mathbb{P}(R_1 \cap R_2) \neq \mathbb{P}(R_1)\mathbb{P}(R_2)$$

Utan svarare $\mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_1) \cdot (\text{sannolikheten för } R_2 \text{ givet } R_1)$

DEFINITION: ~~Givet~~ För två händelser A och B ges sannolikheten för B givet A av:

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

DEFINITION: Två händelser sägs vara oberoende om

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Notera att om $\mathbb{P}(A) > 0$ så är A och B oberoende om och endast om $\mathbb{P}(B|A) = \mathbb{P}(B)$

EXEMPEL: Slå en blå tärning och en gul tärning. Låt:

$A = \{\text{blå tärning sexa}\} = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$

$B = \{\text{gul tärning sexa}\} = \{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\}$

$C = \{\text{summan är } 7\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

$D = \{\text{summan är } 8\} = \{(2,6),(3,5),(4,4),(5,3),(2,6)\}$

Rimlig modell att anta att alla 36 utfall är lika sannolika.

$\Rightarrow \mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{6}, \quad \mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{36}$

Så A och B är oberoende, A och C oberoende, B och C oberoende.
$\mathbb{P}(A \cap B \cap C) = 0$ ej rimligt att säga A, B, C oberoende

Vidare har vi:

$\mathbb{P}(D) = \frac{5}{36}, \quad \mathbb{P}(A \cap D) = \frac{1}{36}$

vilket ger

$$\mathbb{P}(D|A) = \frac{\mathbb{P}(A \cap D)}{\mathbb{P}(A)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} \neq \mathbb{P}(D)$$

$\Longrightarrow$ A och D ej oberoende

PROPOSITION: Om A och B är oberoende, så är A och B^c oberoende

BEVIS: $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B)$
$= \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c)$

$\square$

PROPOSITION: Fixera A och låt för varje $B \in \mathcal{P}(S)$
$Q(B) = \mathbb{P}(B|A)$
Då är Q ett sannolikhetsmått.

BEVIS: (i) Eftersom $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(A)$ gäller att $0 \leq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \leq 1$
dvs $0 \leq Q(B) \leq 1$

(ii) $Q(S) = \mathbb{P}(S|A) = \frac{\mathbb{P}(A \cap S)}{\mathbb{P}(A)} = 1$

(iii) Låt $B_1, B_2, \ldots$ vara disjunkta. Då gäller att:

$$Q(\cup_n B_n) = \mathbb{P}(\cup_n B_n | A) = \frac{\mathbb{P}(\cup_n (A \cap B_n))}{\mathbb{P}(A)}$$
$$= \frac{\sum_n \mathbb{P}(A \cap B_n)}{\mathbb{P}(A)} = \sum_n Q(B_n)$$

Alltså är betingad sannolikhet verkligen en sannolikhet, så att bex

$$\mathbb{P}(B_1 \setminus B_2 | A) = \mathbb{P}(B_1|A) - \mathbb{P}(B_1 \cap B_2 | A)$$

och

$$\mathbb{P}(B_1 \cup B_2 | A) = \mathbb{P}(B_1|A) + \mathbb{P}(B_2|A) - \mathbb{P}(B_1 \cap B_2 | A)$$

En tankebild är att om A betingas, så sätter man bort A^c från utfallsrummet utan att förändra de inbördes sannolikheterna av delmängder till A.

SANNOLIKHET OCH STATISTIK. 26/3

OBEROENDE AV FLERA HÄNDELSER.

Man säger att händelserna A, B och C är oberoende om de är parvis oberoende och

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)$$

Mer generellt:

DEFINITION: Man säger att händelserna $A_1, A_2, A_3 \dots$ är oberoende om det för alla n, och indexmängder $i_1, \dots, i_n$ gäller att

$$\mathbb{P}\Big(\bigcap_{k=1}^{n} A_{i_k}\Big) = \prod_{k=1}^{n} \mathbb{P}(A_{i_k})$$

Att vissa händelser är oberoende är oftast något man antar i sin modell, snarare än att man räknar ut det. Exempelvis, upprepade tärningskast är rimligt att anta att händelser som har med olika kast att göra är oberoende.
Om A_n är händelsen att man får 6 i kast n, så antar vi att $A_1, A_2, \dots$ är oberoende

EXEMPEL: $\mathbb{P}(\text{första sexan i kast nr } n) = \mathbb{P}(A_1^c \cap A_2^c \cap \dots \cap A_{n-1}^c \cap A_n) = \left(\frac{5}{6}\right)^{n-1} = \frac{1}{6}$

Vi har också, enligt kontinuitet hos sannolikheter:

$$\mathbb{P}(\text{man får aldrig sexa}) = \mathbb{P}(A_1^c \cap A_2^c \cap \dots)$$
$$= \lim_{n\to\infty} \mathbb{P}(A_1^c \cap \dots \cap A_n^c) = \lim_{n\to\infty} \left(\frac{5}{6}\right)^n = 0$$

$\Rightarrow$ Om en händelse har positiv sannolikhet och försöket upprepas oändligt, kommer det garanterat vara.

I många sammanhang är det lättare att förstå $\mathbb{P}(B|A)$ snarare än $\mathbb{P}(A \cap B)$, därför är formeln:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B|A)$$

lika användbar som definitionen av betingad sannolikhet.
Antag att $A_1, A_2, \dots, A_n$ är en partition av S, dvs parvis disjuncta mängder sådana att $A_1 \cup A_2 \dots \cup A_n = S$. Då gäller

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{P}(B|A_i)$$

Detta är det totala sannolikhetslagar.

Special fall: $\mathbb{P}(B) = \mathbb{P}(A)\,\mathbb{P}(B|A) + \mathbb{P}(A^c)\,\mathbb{P}(B|A^c)$

EXEMPEL: Tre procent av svenskarnas befolkning dödas av lungcancer. Bland rökare dödas 10%. Rökare utgör 20% av befolkningen. Vad är risken för icke-rökare?

Välj en person NN, låt $A = \{\text{NN är rökare}\}$, $B = \{\text{NN får lungcancer}\}$

Vi söker $\mathbb{P}(B|A^c)$. Givet är $\mathbb{P}(A) = 0{,}2$, $\mathbb{P}(B|A) = 0{,}1$, $\mathbb{P}(B) = 0{,}03$

$$\mathbb{P}(B) = \mathbb{P}(A)\,\mathbb{P}(B|A) + \mathbb{P}(A^c)\,\mathbb{P}(B|A^c)$$

$$\Rightarrow \mathbb{P}(B|A^c) = \frac{\mathbb{P}(B) - \mathbb{P}(B|A)}{\mathbb{P}(A^c)} = \frac{0{,}03 - 0.2 \cdot 0.1}{0.8} = 0{,}0125$$

Hur stor andel av dem är rökare? Hur sökt vi $\mathbb{P}(A|B)$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{0.1 \cdot 0.2}{0.03} = \frac{2}{3}$$

I slutet på exemplet vände vi på betingning. Enligt TSL kan vi skriva om nämnaren och få:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}$$

Detta är <u>BAYES FORMEL</u>

Mer generellt gäller att om $A_1, A_2, \dots, A_n$ är en partition så att

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{k=1}^{n} \mathbb{P}(B|A_k)\mathbb{P}(A_k)}$$

EXEMPEL: I en hatt ligg <u>tre</u> mynt

mynt 1 ger klave med sannolikhet 0.3

mynt 2 ger klave med sannolikhet 0.5

mynt 3 ger klave med sannolikhet 0.8

Ett mynt valdes, kastades tre gånger och blev klave en gång. Vad är den betingade sannolikheten att mynt i valdes?

Låt M_i vara händelsen att mynt $i = 1,2,3$ valdes, och låt A vara händelsen att det blir klave exakt en gång på tre kast. För alla i gäller $\mathbb{P}(M_i) = \frac{1}{3}$. Vidare gäller

$$\mathbb{P}(A|M_1) = 3 \cdot 0.3 \cdot 0.7^2 = 0.441$$

$$\mathbb{P}(A|M_2) = 3 \cdot 0.5 \cdot 0.5^2 = 0.375$$

$$\mathbb{P}(A|M_3) = 3 \cdot 0.8 \cdot 0.2^2 = 0.096$$

Enligt Bayes är

$$\mathbb{P}(M_i|A) = \frac{\mathbb{P}(A|M_i)\mathbb{P}(M_i)}{\sum_{j=1}^{3} \mathbb{P}(A|M_j)\mathbb{P}(M_j)}$$

Nämnaren är

$$\sum_{j=1}^{3} \mathbb{P}(A|M_j)\mathbb{P}(M_j) = \frac{1}{3}\left(0.441 + 0.375 + 0.096\right) = 0.304$$

Detta ger:

$$\mathbb{P}(M_1|A) = \frac{\frac{1}{3} \cdot 0.441}{0.304} \approx 0.48$$

$$\mathbb{P}(M_2|A) = \frac{\frac{1}{3} \cdot 0.0375}{0.304} \approx 0.41$$

$$\mathbb{P}(M_3|A) = \frac{\frac{1}{3} \cdot 0.096}{0.304} \approx 0.11$$

EXEMPEL: Sjukdomen S. Av de patienter som testas för S, är det 1% som har S. Om patienten har S visar testet positivt med sannolikhet 0.95. (sensitivitet) (recall)
Om patienten inte har S, så visar testet negativt utslag med sannolikhet 0.9 (specificitet)
Om patient får positivt utslag, vad är den betingade sannolikheten att den har S? (precision)

Låt A vara händelsen att patienten har S, och låt T vara händelsen att testet ger ett positivt utslag.
Enligt uppgift gäller:

$$\mathbb{P}(A) = 0.01, \quad \mathbb{P}(T|A) = 0.95, \quad \mathbb{P}(T|A^c) = 0.1$$

Vi söker $\mathbb{P}(A|T)$

$$\mathbb{P}(A|T) = \frac{\mathbb{P}(T|A)\mathbb{P}(A)}{\mathbb{P}(T|A)\mathbb{P}(A) + \mathbb{P}(T|A^c)\mathbb{P}(A^c)} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.1 \cdot 0.99} = 0.095$$

Om man testar en person från gatan (screening), om S förekommer hos 0.01% av befolkningen $\longrightarrow \mathbb{P}(A) = 0.0001$

$$\Longrightarrow \mathbb{P}(A|T) = \frac{0.95 \cdot 0.0001}{0.95 \cdot 0.0001 + 0.1 \cdot 0.9999} = 0.00095$$

Det är väldigt vanligt i maskinlärning att utvärdera en klassificeringsmetod med precision - recall. Men, vi ser att precision beror väldigt mycket på fördelningen mellan klasserna, vilket inte egentligen har med testet att göra. Alltså ingen bra metod om man har stor obalans mellan klasser.

EXEMPEL: Gambler's ruin.

Ann och Bo spelar på upprepade oberoende slantsinglar
Klave: Ann ger 1 kr till Bo
Krona: Bo ger 1 kr till Ann
Detta upprepas tills någon av dem har slut på pengar!
Om Ann vid starten har k kr och Bo har $n-k$ kr, vad är sannolikheten att Ann vinner?

Låt A vara händelsen att Ann vinner. Låt P_k vara sannolikhetsmåttet vid starten från k kr för Ann. (varje k ger en egen modell) Låt H vara händelsen att första kastet blir klave.

$P_k = \mathbb{P}_k(A)$. Vi har $P_0 = 0$, $P_n = 1$, ovanför k ger TSL

$$P_k = \mathbb{P}_k(A) = \mathbb{P}_k(A|H)\mathbb{P}_k(H) + \mathbb{P}_k(A|H^c)\mathbb{P}_k(H^c) = \frac{1}{2}(P_{k-1} + P_{k+1})$$

$$\Longrightarrow P_{k+1} = 2P_k - P_{k-1} \longrightarrow P_k = \frac{k}{n}$$

SANNOLIKHET OCH STATISTIK 30/3

STOKASTISKA VARIABLER

En stokastisk variabel är ett slumptal, dvs ett tal vars värde beror på slumpen, dvs vars värde bestäms av utfallet till ett slumpförsök.

DEFINITION: Låt S vara utfallsrummet till ett slumpförsök. En stokastisk variabel är en funktion

$$X : S \longrightarrow \mathbb{R}$$

EXEMPEL: Låt $S = \{\text{äpple, apelsin, banan}\}$. Exempel på stokastiska variabler:

$X(u)$ = vikten av u

$Y(u)$ = energiinnehållet i u

$Z(u)$ = sockerinnehållet i u

u	$X(u)$	$Y(u)$	$Z(u)$
äpple	130	80	15
apelsin	100	60	10
banan	160	120	29

EXEMPEL: Välj på måfå i cirkelskivan av radie 1.

$$S = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

EXEMPEL: Exempel på svar

$$X(u) = |u|, \quad Y(u) = \pi |u|^2, \quad Z(u) = \text{poäng vid pilkast.}$$

Låt X vara en stokastisk variabel. Det slumpförsöket är utfört vet vi alltså exakt vad X blev. Före försöket kan vi ytterst ofta om sannolikheterna för att X antar vissa värden. Talen

$$\mathbb{P}(X \in B), \quad B \subseteq \mathbb{R}$$

kallas (sannolikhets) fördelning.

Observera att $\mathbb{P}(X \in B) = \mathbb{P}(\{u \in S \mid X(u) \in B\})$

Det räcker med angav en fördelning för en sv X angav vad inte $\mathbb{P}(X \in B)$ för alla B, endast för en uppsättning B, sådant att dessa räcker för att räkna ut alla sannolikheter som har med X att göra.

EXEMPEL: ~~Låt~~ Singla slant två gånger

$$S = \{HH, HT, TH, TT\}.$$

Låt X vara klase, dvs:

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0$$

X:s fördelning ges av talen

$$\mathbb{P}(X = 0) = \frac{1}{4}, \quad \mathbb{P}(X = 1) = \frac{1}{2}, \quad \mathbb{P}(X = 2) = \frac{1}{4}$$

Låt $X : S \to \mathbb{R}$ vara en SV. Man säger att X är diskret om V_X är ändlig eller uppräkneligt:

$$V_X = \{x_1, x_2, x_3, \dots\}$$

Kom ihåg att $V_X = \{X(u) \mid u \in S\}$, mängden av alla värden som X kan anta. Skriv:

$$p(x_k) = \mathbb{P}(X = x_k) = \mathbb{P}(\{u \in S \mid X(u) = x_k\})$$

Då blir $p : V_X \to [0,1]$ och p kallas för X:s frekvensfunktion (pmf) (massfunktion)

EXEMPEL: Slant singling med rättvist mynt. $S = \{H, T\}$. Låt

$$X(H) = 0, \quad X(T) = 1$$

Då är X en SV med $V_X = \{0, 1\}$ om

$$p(0) = \frac{1}{2}, \quad p(1) = \frac{1}{2}$$

EXEMPEL: Följd av tärningskast. Låt X vara antal kast till första sexan.

$$V_X = \{1, 2, 3, \dots\}$$

Vi har tidigare räknat att

$$p(k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}$$

Notera att varje funktion $p : \{x_1, \dots\} \to [0,1]$ där att $\sum_{k=1}^{\infty} p(x_k) = 1$ är frekvensfunktion till en SV. nämligen den av X för vilken $\mathbb{P}(X = x_k) = p(x_k)$

DEFINITION: Funktionen $F : \mathbb{R} \to \mathbb{R}$ given av

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

kallas för X:s fördelningsfunktion.

Om X är diskret är

$$F(x) = \sum_{x_k \in V_X : x_k \leq x} p(x_k)$$

och $p(x_k) = F(x_k) - F(x_k -)$

EXEMPEL: Singla slant fyra gånger. Låt X vara klave (antal)

$$V_X = \{1, 2, 3, 4\}$$

$$p(k) = \binom{4}{k}\left(\frac{1}{2}\right)^4 = \binom{4}{k}\frac{1}{16}$$

k	$p(k)$	$F(k)$
0	$\frac{1}{16}$	$\frac{1}{16}$
0	$\frac{1}{4}$	$\frac{5}{16}$
0	$\frac{3}{8}$	$\frac{11}{16}$
0	$\frac{1}{4}$	$\frac{15}{16}$
0	$\frac{1}{16}$	1

Vänstra fördelningsfunktion F är vänsterkontinuerlig, $F(\infty) = 1$, $F(-\infty) = 0$ och

$$\lim_{n\to\infty} F\left(\frac{x+1}{n}\right) = \lim_{n\to\infty} \mathbb{P}\left(X \le \frac{x+1}{n}\right) = \mathbb{P}(X \le x) = F(x)$$

(ty $\{u \mid X(u) \le x\} = \bigcap_n \{u \mid X(u) \le \frac{x+1}{n}\}$) dvs F är högerkontinuerlig

Annan formula

$$\mathbb{P}(a < X \le b) = F(b) - F(a)$$
$$\mathbb{P}(X > x) = 1 - F(x)$$

Fördelningsfunktionen är inte så intressant för diskreta sv, eftersom frekvensfunktionen säger allt vi vill veta. Mer intressant för kontinuerliga stokastiska variabler.

DEFINITION: En sv X kallas kontinuerlig om det finns en funktion $f: \mathbb{R} \to \mathbb{R}$ sådan att $\forall x \in \mathbb{R}$:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Funktionen f kallas för X:s (sannolikhets) täthet (function)

Definitionen är ekvivalent med att säga att F är deriverbar med f som derivata. Def 2.5 i boken är inte korrekt; Det räcker inte att F är kontinuerlig.

Om X är kontinuerlig gäller

$$\mathbb{P}(X \in B) = \int_B f(t)\,dt$$

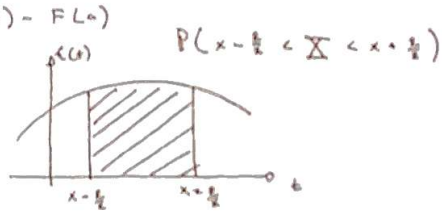
för alla $B \subseteq \mathbb{R}$

OBS: För alla x gäller $\mathbb{P}(X = x) = \int_x^x f(t)\,dt = 0$. Följdaktigen

$$\mathbb{P}(a \le X \le b) = \mathbb{P}(a < X < b) = F(b) - F(a)$$

Tolkning av f:

$$\mathbb{P}\left(x - \tfrac{\varepsilon}{2} < X < x + \tfrac{\varepsilon}{2}\right) \approx \varepsilon f(x)$$



LIKFORMIG FÖRDELNING

En kontinuerligt sv. sägs vara likformigt fördelat på $[a,b]$ om

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$$

dvs om:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

Täthet

Fördelningsfunktion

Kort skriversätt: $X \sim \text{likfo}[a,b]$

OBS: $\mathbb{P}(X \in [c,d]) = \frac{1}{b-a}\int_c^d dx = \frac{d-c}{b-a}$

PROPOSITION: Om $X \sim \text{likf}[0,1]$ och $Y = a + (b-a)X$ gäller att $Y \sim \text{likf}[a,b]$

BEVIS: Eftersom $F_X(x) = x$ för alla $x \in [0,1]$ gäller för alla $y \in [a,b]$ att

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(a + (b-a)X \le y)$$
$$= \mathbb{P}\left(X \le \frac{y-a}{b-a}\right) = \frac{y-a}{b-a}$$

Analogt gäller att $X \sim \text{likf}[a,b] \Rightarrow \frac{X-a}{b-a} \sim \text{likf}[0,1]$

FÖRDELNING FÖR FUNKTION AV STOKASTISK VARIABEL.

Låt X vara en sv och $g: \mathbb{R} \to \mathbb{R}$ en funktion och låt $Y = g(X)$. Vad är fördelningen för Y?

Fallet då X är diskret: $V_X = \{x_1, x_2, \ldots\}$

$$V_Y = \{ g(x_j) \mid x_j \in V_X \}$$

För varje $y_k \in V_Y$, låt $B_k = \{ x_j \in V_X \mid g(x_j) = y_k \}$. Då gäller:

$$p_Y(y_k) = \mathbb{P}(Y = y_k) = \mathbb{P}(X \in B_k) = \sum_{x_j \in B_k} p_X(x_j)$$

Fallet då X är kontinuerlig: Y kan bli diskret, kontinuerlig eller något annat beroende på hur g ser ut.

EXEMPEL: Låt $X \sim \text{likf}[0,1]$ och

$$g(x) = \begin{cases} 0 & x < \frac{2}{3} \\ 1 & x \geq \frac{2}{3} \end{cases}$$

Med $Y = g(X)$ blir $V_Y = \{0,1\}$ och $p_Y(1) = \frac{1}{3}$

Man jobbar oftast fall till fall

EXEMPEL: Låt $X \sim \text{likf}[0,2]$, dvs $f(x) = \frac{1}{2}$ och $F(x) = \frac{x}{2}$

Låt $Y = X(2-X)$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X(2-X) \leq y)$$

$$X(2-X) = y \iff X^2 - 2X + y = 0 \iff X = 1 \pm \sqrt{1-y}$$

Alltså $F_Y(x) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq 1 - \sqrt{1-y}) + \mathbb{P}(X > 1 + \sqrt{1-y})$

$$= \frac{1 - \sqrt{1-y}}{2} + 1 - \frac{1 + \sqrt{1-y}}{2} = 1 - \sqrt{1-y}$$

Derivering ger även

$$f_Y(y) = \frac{d}{dy}\left(1 - \sqrt{1-y}\right) = \frac{1}{2\sqrt{1-y}}$$

Ett generellt resultat: Antag att g är strängt växande. Då finns g^{-1} och är växande:

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F(g^{-1}(y))$$

Alltså $f_Y(y) = f_X(g^{-1}(y))(g^{-1})'(y)$

EXEMPEL: Låt $X \sim \text{likf}[0,1]$ och $Y = X^3$. Då är $g(x) = x^3$, så $g^{-1}(y) = y^{\frac{1}{3}}$ och $(g^{-1})'(y) = \frac{1}{3} y^{-\frac{2}{3}}$. Eftersom $f(x) = 1$ får vi

$$f_Y(y) = \frac{1}{3} y^{-\frac{2}{3}}, \quad y \in [0,1]$$

VÄNTEVÄRDE

Låt $X_1, X_2, X_3, \ldots$ vara stokastiska variabler definierade på samma utfallsrum.

DEFINITION: Man säger att $X_1, X_2, \ldots$ är oberoende om det för alla $B_1, B_2, \ldots \subseteq \mathbb{R}$ gäller att händelserna $\{X_1 \in B_1\}, \{X_2 \in B_2\}$ är oberoende

Låt N vara stort och $X_1, \ldots, X_N$ vara oberoende och alla ha samma fördelning som X. Antag att $V_X = \{x_1, x_2, \ldots\}$.
Låt $N_k = |\{j \mid X_j = x_k\}|$
Vi förväntar oss att ha $N_k \approx N p(x_k)$, vilket medför:

medelvärdet av $X_j = \sum_{k=1}^{\infty} x_k \frac{N_k}{N} \approx \sum_{k=1}^{\infty} x_k p(x_k)$

(Detta är tyngdpunkten av tallinjen om massorna $p(x_k)$ läggs ut.

DEFINITION: Väntevärdet av en diskret stokastisk variabel ges av

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} x_k p(x_k)$$

För en kontinuerlig stokastisk variabel gäller

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

(För definition krävs att $\sum_{k=1}^{\infty} |x_k| p(x_k) < \infty$ resp $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$)

EXEMPEL: Satsa en marker på rött på roulette. Låt X vara vinsten.
Då gäller $\mathbb{P}(X = 1) = \frac{18}{37}$ och $\mathbb{P}(X = -1) = \frac{19}{37}$

$$\Longrightarrow \mathbb{E}[X] = (-1) \cdot \frac{19}{37} + 1 \cdot \frac{18}{37} = -\frac{1}{37}$$

EXEMPEL: Vv av diskret likformig: $V_X = \{1, \ldots, n\}$ och
$\mathbb{P}(X = k) = \frac{1}{n}$ för $k = 1, \ldots, n$

$$\mathbb{E}[X] = \sum_{k=1}^{n} k \frac{1}{n} = \frac{n+1}{2}$$

Exempelvis är väntevärdet av ett tärningskast $\frac{7}{2}$

EXEMPEL: $X \sim \text{likf}(a, b)$

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a}\, dx = \frac{b+a}{2}$$

EXEMPEL: $X \sim \text{likf}(0,1)$. Låt $Y = X^2$

$$F_Y(y) = \mathbb{P}(X \leq \sqrt{y}) = \sqrt{y} \quad \text{för } y \in (0,1)$$

, alltså $f_Y(y) = \frac{1}{2\sqrt{y}}$

$$\mathbb{E}[Y] = \int_0^1 y \frac{1}{2\sqrt{y}}\, dy = \frac{1}{3}\left[y^{3/2}\right]_0^1 = \frac{1}{3}$$

Notera att $\mathbb{E}[X^2] = \frac{1}{3}$, medan $\mathbb{E}[X]^2 = \frac{1}{4}$

PROPOSITION 2.9: Antag att X är kontinuerlig och $V_X = [0, \infty)$. Då gäller:

$$\mathbb{E}[X] = \int_0^\infty (1 - F(x))\,dx = \int_0^\infty \mathbb{P}(X > x)\,dx$$

BEVIS:

$$\int_0^\infty \mathbb{P}(X > x)\,dx = \int_0^\infty \int_x^\infty f(t)\,dt\,dx = \int_0^\infty \left(\int_0^t dx\right) f(t)\,dt$$

$$= \int_0^\infty t\,f(t)\,dt = \mathbb{E}[X]$$

DISKRET VARIANT: Antag att $V_X = \{0, 1, 2, \dots\}$. Då är

$$\mathbb{E}[X] = \sum_{n=1}^\infty \mathbb{P}(X \geq n) = \sum_{n=0}^\infty \mathbb{P}(X > n)$$

BEVIS:

$$\sum_{n=1}^\infty \mathbb{P}(X \geq n) = \sum_{n=1}^\infty \sum_{k=n}^\infty p(k) = \sum_{k=1}^\infty \sum_{n=1}^k p(k) = \sum_{k=1}^\infty k\,p(k)$$

EXEMPEL: I exemplet ovan hade vi $F_Y(y) = \sqrt{y}$, $y \in (0,1)$, så

$$\mathbb{E}[Y] = \int_0^1 (1 - \sqrt{y})\,dy = \frac{1}{3}$$

EXEMPEL: Låt X vara antal börjningsvärt till första sexan. Vi har:

$$\mathbb{P}(X > n) = \left(\frac{5}{6}\right)^n$$

Alltså

$$\mathbb{E}[X] = \sum_{n=0}^\infty \left(\frac{5}{6}\right)^n = \frac{1}{1 - \frac{5}{6}} = 6.$$

SANNOLIKHET OCH STATISTIK 13/4

IDAG: Väntevärden
Varians
Poisson

VÄNTEVÄRDEN AV FUNKTIONER AV STOKASTISKA VARIABLER.

Kom ihåg exempel med $X \sim \text{likf}(0,1)$ och $Y = X^2$. Vi fick $\mathbb{E}[Y] = \mathbb{E}[X^2] = \frac{1}{3}$. Nu var just:

$$\int_0^1 x^2 \, dx = \frac{1}{3}$$

dvs. $\mathbb{E}[X^2] = \int x^2 f_X(x) \, dx$, $\Longrightarrow$ för funktionen $g(x) = x^2$ gäller $\mathbb{E}[g(X)] = \int g(x) f_X(x) \, dx$. Tillfälligheter? Faktiskt inte.

SATS: Den omedvetne statistikerns lag
Låt $g: \mathbb{R} \longrightarrow \mathbb{R}$ och X en stokastisk variabel. Om X diskret gäller:

$$\mathbb{E}[g(X)] = \sum_{k=1}^{\infty} g(x_k) p(x_k)$$

Om X kontinuerlig gäller:

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

BEVIS: Diskreta fallet:
Skriv $Y = g(X)$ och $V_Y = \{ y_1, y_2, \dots \}$

$$\mathbb{E}[Y] = \sum_j y_j \mathbb{P}(Y = y_j) = \sum_j y_j \sum_{k: g(x_k) = y_j} p(x_k)$$

$$= \sum_j \sum_{k: g(x_k) = y_j} g(x_k) p(x_k) = \sum_{k=1}^{\infty} g(x_k) p(x_k)$$

EXEMPEL: Låt a och b vara konstanter och X stokastisk variabel. Då är:

$$\mathbb{E}[aX + b] = \int_{-\infty}^{\infty} (ax + b) f(x) \, dx = a \int_{-\infty}^{\infty} x f(x) \, dx + b = a \mathbb{E}[X] + b$$

EXEMPEL: Låt $X \sim \text{likf}(0,1)$ och

$$g(x) = \begin{cases} 0 & x < \frac{2}{3} \\ 1 & x \geq \frac{2}{3} \end{cases}$$

Vi får

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

OÄNDLIGA VÄNTEVÄRDEN

I allmänhet kräver vi att $\int_{-\infty}^{\infty} |x| f(x)\,dx < \infty$ för att över huvudtaget definiera $\mathbb{E}[X]$. Detta då vi riskerar att få:

$$\int_{-\infty}^{\infty} x f(x)\,dx = \int_{0}^{\infty} x f(x)\,dx - \int_{-\infty}^{0} (-x) f(x)\,dx = \infty - \infty = ?$$

Men om $X \geq 0$ försvinner den andra termen och uttrycket blir entydigt oändligt om den första termen är oändlig. I sådana fall säger vi att $\mathbb{E}[X] = \infty$. Analogt i det diskreta fallet.

EXEMPEL: St. Petersburg. Låt X vara antalet kast t.o.m. den första klaven. Sammanlagd förlust före första klaven blir $2^X - 1$. Det gäller att $\mathbb{P}(X = k) = \left(\frac{1}{2}\right)^k$, $k = 1, 2, \ldots$

$$\mathbb{E}[2^X - 1] = \sum_{k=1}^{\infty} (2^k - 1)\left(\frac{1}{2}\right)^k = -1 + \sum_{k=1}^{\infty} 1 = \infty$$

VARIANS

Låt X vara en stokastisk variabel med $\mathbb{E}[X] = \mu < \infty$

DEFINITION: $\mathbb{V}ar(X) = \mathbb{E}[(X-\mu)^2]$

Man kan ha $\mathbb{V}ar(X) = \infty$. Man skriver $Std(X) = \sqrt{\mathbb{V}ar(X)}$
Vanlig betackning $\sigma^2 = \mathbb{V}ar(X)$

STEINERS FORMEL

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

BEVIS: Kontinuerliga fallet:

$$\mathbb{V}ar(X) = \int (x-\mu)^2 f(x)\,dx = \int x^2 f(x)\,dx - 2\mu \underbrace{\int x f(x)\,dx}_{\mu} + \mu^2 \int f(x)\,dx$$
$$= \mathbb{E}[X^2] - \mu^2$$

EXEMPEL: $X \sim$ likf $[a,b]$. Kom ihåg $\mathbb{E}[X] = \frac{a+b}{2}$

$$\mathbb{E}[X^2] = \int_a^b x^2 \frac{1}{b-a}\,dx = \frac{b^3 - a^3}{3(b-a)}$$

Alltså:

$$\mathbb{V}ar(X) = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

PROPOSITION: $\mathbb{V}ar(aX + b) = a^2\,\mathbb{V}ar(X)$

BEVIS:
$$\mathbb{V}ar(aX+b) = \mathbb{E}[(aX + b - \mathbb{E}[aX+b])^2]$$
$$= \mathbb{E}[(aX + b - (a\mu + b))^2]$$
$$= a^2\,\mathbb{E}[(X-\mu)^2] = a^2\,\mathbb{V}ar(X)$$

TEORETISKT VIKTIGA OLIKHETER.

Markovs olikhet:

Antag att $X \geq 0$. Då gäller för alla $a > 0$ att

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

BEVIS: Låt $g(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$

Då är $g(X) \leq X$, så $\mathbb{E}[g(X)] \leq \mathbb{E}[X]$. Eftersom $g(X) = a$ med sannolikhet $\mathbb{P}(X \geq a)$ och 0 annars, så:

$$\mathbb{E}[g(X)] = a\,\mathbb{P}(X \geq a)$$

Chebyshevs olikhet:

Skriv $\mu = \mathbb{E}[X]$ och $\sigma^2 = \mathrm{Var}(X)$. Det gäller för alla $\varepsilon > 0$ att

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

BEVIS: Enligt Markovs olikhet gäller att

$$\mathbb{P}(\underbrace{|X-\mu|}_{\geq \varepsilon}) = \mathbb{P}\left((X-\mu)^2 \geq \varepsilon^2\right) \leq \frac{\mathbb{E}[(X-\mu)^2]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

EXEMPEL: Längden X av en på måfå vald svensk man är en stokastisk variabel med väntevärde 180 cm och standardavvikelse 5 cm. Sätt en gräns för andelen svenskar som är över 210 cm.

Vi har $\mu = 180$ cm och $\sigma^2 = 5^2 = 25$

$$\Rightarrow \mathbb{P}(X \geq 210) \leq \mathbb{P}(|X - \mu| \geq 30) \leq \frac{25}{30^2} = \frac{1}{36}$$

INDIKATORFÖRDELNING

Låt A vara en händelse. Låt $X(u) = \begin{cases} 1 & u \in A \\ 0, & u \notin A \end{cases}$

Då kallas X för indikatorn för händelsen A.

Skriv $p = \mathbb{P}(X = 1)$. Då gäller $\mathbb{E}[X] = p$

Eftersom $\mathbb{E}[X^2] = p$, fås vi $\mathrm{Var}(X) = p - p^2 = p(1-p)$

Man skriver ofta $X = I_A$

BINOMIALFÖRDELNING

Låt $A_1, A_2, \ldots$ vara n oberoende händelser som alla har sannolikhet p. Summan $X = \sum_{k=1}^{n} I_{A_k}$ sägs vara binomialfördelad, med parametrar n och p. $X \sim \mathrm{Bin}(n, p)$

Tänk gärna på detta som att en slant som ger krona med sannolikhet p, singlas n gånger. Låt X vara antalet gånger man får krona totalt.
Man kan välja ut k kast på $\binom{n}{k}$ olika sätt. Sannolikheten att exakt dessa ger krona är $p^k(1-p)^{n-k}$. Alltså:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

OBS: att $k\binom{n}{k} = n\binom{n-1}{k-1}$

Detta ger

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

$$= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np$$

Den sista likheten följer av binomialsatsen för frekvensfunktionen för $X \sim Bin(n-1, p)$, och summerar därför till 1.

EXEMPEL: Lag A och B möts i bäst av sju matcher. Anta att sannolikheten att Lag A vinner en given match är 0.6, och att matcherna är oberoende. Vad är sannolikheten att A vinner matchserien?

Låt X vara antalet matcher som A vinner. Då är $X \sim Bin(7, 0.6)$

$$\mathbb{P}(X \geq 4) = \sum_{k=4}^{7} \binom{7}{k} 0.6^k \, 0.4^{7-k} \approx 0.7$$

$\square$

Två observationer:

- $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$ oberoende
  $\implies X_1 + X_2 \sim Bin(n_1 + n_2, p)$
- $X \sim Bin(n, p) \implies n - X \sim Bin(n, 1-p)$

GEOMETRISK FÖRDELNING

En slant klave ger, med sannolikhet p. Den singlas till första klaven. Låt X vara antal kast som behövdes. Vi har:

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots$$

Man säger att X är geometriskt fördelad med parameter p, kortform $X \sim Geo(p)$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}(X > k) = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{p}$$

Vi räknar senare ut att

$$\mathbb{V}ar(X) = \frac{1-p}{p^2}$$

EXEMPEL: X = antal tärningskast till och med första sexan

$X \sim Geo(\frac{1}{6})$

EXEMPEL: Spela 100 rader på lotto varje vecka. Låt X vara antal veckor man behöver spela för att få sju rätt första gången. Vi har $X \sim Geo(p)$ där p är sannolikhet att på sju rätt en given vecka.

Det finns $\binom{35}{7}$ olika rader, så:

$$p = \frac{100}{\binom{35}{7}}$$

Vi får då att $\mathbb{E}[X] = \frac{\binom{35}{7}}{100} \approx 67245$ veckor ≈ 1293 år.

POISSONFÖRDELNING

Om $\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 1,2,3,\ldots$
sägs vara Poissonfördelad med parameter λ. Kort form $X \sim Po(\lambda)$

OBS: $\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda}$ enligt Taylors formel.

- Antal vulkanutbrott i världen per år.
- Antal fordon som passerar på en enslig skogsväg under en dag.
- Antal lungcancerfall i Göteborg på en månad.
- Antal bankrutt i västra götaland under januari
- Antal mål i första halvlek på en fotbollsmatch

$$\mathbb{E}[X] = e^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

$$\mathbb{E}[X(X-1)] = e^{-\lambda}\sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2$$

Alltså blir $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda$

$$\mathbb{V}ar(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

EXEMPEL: På en enslig skogsväg passerar i genomsnitt 5,3 fordon per dag. Vad är sannolikheten att högst två fordon passerar under en given dag.
Det verkar rimligt att anta att X = antal fordon den given dagen är poisson fördelad $Po(5.3)$. (Parametern = väntevärde)
Vi får

$$\mathbb{P}(X \leq 2) = e^{-5.3}\left(\frac{5.3^0}{0!} + \frac{5.3^1}{1!} + \frac{5.3^2}{2!}\right) \approx 0.102$$

EXEMPEL: Sverige har 10 miljoner invånare. Det sker cirka 200 000 bankrutt per år. Vad är sannolikheten att en person som lever 100 år klarar sig utan bankrutt? Sannolikheten att få två bankrutt?
Det är rimligt att om X är antal bankrutt i livet X poisson fördelad. Väntevärdet är $\frac{200\,000 \cdot 100}{10\,000\,000} = 2$

$X \sim Po(2)$

$$\mathbb{P}(X=0) = e^{-2}, \quad \mathbb{P}(X=2) = e^{-2}\frac{2^2}{2!} = 2e^{-2}$$

$\square$

EXEMPEL: När n är stort och p är litet gäller det att $Bin(n,p) \approx Poi(np)$. Som exempel:

Låt c vara en konstant, låt $X \sim Bin(n, \frac{c}{n})$ och låt $n \to \infty$. För ett k som inte är för stort ges med

$$\mathbb{P}(X = k) = \binom{n}{k}\left(\frac{c}{n}\right)^k\left(1 - \frac{c}{n}\right)^{n-k}$$

Eftersom $\binom{n}{k}/n^k \to \frac{1}{k!}$ och $\left(1 - \frac{c}{n}\right)^{n-k} \to e^{-c}$ gäller att

$$\mathbb{P}(X = k) \to e^{-c}\frac{c^k}{k!}$$

Dvs, när n är stort är X ungefär $Poi(c)$-fördelad.

SANNOLIKHET OCH STATISTIK 15/4

IDAG: Exponentialfördelning
Poissonfördelning

EXPONENTIALFÖRDELNING.

En kontinuerlig stokastisk variabel som har täthet

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

sägs vara exponentialfördelad med parametern (intensitet) λ.
Kort form $X \sim \exp(\lambda)$

Uppstår som livslängd av saker som inte åldras t.ex:

- Livslängden av en LED-lampa.
- Väntetiden till första bil som passerar på en slingrig väg
- Tiden till första sönderfall hos ett radioaktivt material.

Vi ser att $F(x) = 1 - e^{-\lambda x}$, dvs $\mathbb{P}(X > x) = e^{-\lambda x}$

$$\mathbb{E}[X] = \int_0^\infty x \lambda e^{-\lambda x} dx = \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}$$

Det gäller också att

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}, \text{ så } \mathbb{V}\mathrm{ar}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

EXEMPEL: En LED-lampa har en livstid 10 år. Hur stor är sannolikheten att den håller minst 10 år? Anta att lampan inte åldras, och dessutom att dess livslängd X är $\exp(\lambda)$-fördelad, det verkar rimligt att välja ursprungsvärdet som $\mathbb{E}[X] = 10$, $\lambda = 0.1$

$$\mathbb{P}(X > 10) = e^{-0.1 \cdot 10} = e^{-1} \approx 0.37$$

Allmänt: $X \sim \exp(\lambda) \Rightarrow \mathbb{P}(X > \mathbb{E}[X]) = e^{-1}$

Exponentialfördelningen uppfyller <u>glömskeegenskapen</u> (minnet åldrandet)

$$\mathbb{P}(X > t+x \mid X > t) = \mathbb{P}(X > x)$$

ty $\mathbb{P}(X > t+x \mid X > t) = \dfrac{\mathbb{P}(X > t+x, X > t)}{\mathbb{P}(X > t)} = \dfrac{\mathbb{P}(X > t+x)}{\mathbb{P}(X > t)}$

$$= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} = \mathbb{P}(X > x)$$

Vi ser nu. Antag att $\mathbb{P}(X > t + x \mid X > t) = \mathbb{P}(X > x)$ för alla x och t. Om vi skriver $G(x) = \mathbb{P}(X > x)$. Enligt den andra

$$\mathbb{P}(X > t + x \mid X > t) = \frac{\mathbb{P}(X > t + x)}{\mathbb{P}(X > t)}$$

vilket antingen ger att $\mathbb{P}(X > x)$, vilket ger.

$$G(x+t) = G(t)\,G(x)$$

för alla x och t. Subtraktion $G(x)$ från vänster sidan, delta med t och let $t \to 0$ ger att

$$G'(x) = G'(0)\,G(x)$$

för alla x. Detta ger den unika lösningen

$$G(x) = e^{-\lambda x}$$

om $\lambda = -G'(0) > 0$ eftersom G är minskande. Alltså är X exponentialfördelad.

Vi har visat att X har oändlighetsegenskapen om och endast om X är exponentialfördelad.

EXEMPEL Exponentialfördelning och geometrisk fördelning.
Låt $X \sim \exp(\lambda)$ och $Y = \lceil X \rceil$. Då gäller

$$\mathbb{P}(Y = k) = \mathbb{P}(X \in (k-1, k]) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k)$$
$$= e^{-\lambda(k-1)} - e^{-\lambda k} = e^{-\lambda(k-1)}(1 - e^{-\lambda}) = p(1-p)^{k-1}$$

med $p = 1 - e^{-\lambda}$

Konkretet: $X \sim \exp(\lambda) \Rightarrow \lceil X \rceil \sim \mathrm{Geo}(1 - e^{-\lambda})$

$\square$

PROPOSITION: Om $X_1 \sim \exp(\lambda_1)$ och $X_2 \sim \exp(\lambda_2)$ och X_1, X_2 oberoende, så gäller $\min(X_1, X_2) \sim \exp(\lambda_1 + \lambda_2)$

BEVIS:

$$\mathbb{P}(\min(X_1, X_2) > x) = \mathbb{P}(X_1 > x, X_2 > x)$$
$$= \mathbb{P}(X_1 > x) \cdot \mathbb{P}(X_2 > x)$$
$$= e^{-\lambda_1 x} e^{-\lambda_2 x} = e^{-(\lambda_1 + \lambda_2)x}$$

$\square$

Låt $T_1, \dots, T_n$ vara oberoende och $\exp(\lambda)$-fördelade (samma λ för alla)

~~PROPOSITION: Om $X_1 \sim \exp(\lambda_1)$, $X_2 \sim \exp(\lambda_2)$ och X_1 och X_2 är oberoende att gäller $\min(X_1, X_2) \sim$~~

$$X = \sum_{k=1}^{n} T_k$$

sägs vara gammafördelad med parametrar n och λ. Kort form:

$$X \sim \Gamma(n, \lambda)$$

PROPOSITION: $\mathbb{P}(X > x) = e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$

POISSONPROCESSEN

Låt $0 < \tau_1 < \tau_2 < \tau_3 < \dots$ vara följd av tidspunkter då en viss typ av händelser (impulser) inträffar (t.ex jordbävningar, besökta bilar som passerar) Sätt

$$T_1 = \tau_1 \,, \quad T_2 = \tau_2 - \tau_1 \,, \quad T_3 = \tau_3 - \tau_2 \quad \dots$$

för tidsmellanrummen mellan impulser. Om $T_1, T_2, T_3 \dots$ är oberoende och $\exp(\lambda)$-fördelade, kallas följden $\{\tau_1, \tau_2, \dots\}$ för en poissonprocess med intensitet λ.

OBS: För en Poisson-process med int λ gäller $\tau_n \sim \Gamma(n, \lambda)$
Låt $X(t) = \max\{n : \tau_n \leq t\}$ = antal impulser som kommit vid tid t. Enligt proposition ovan:

$$\mathbb{P}(X(t) = n) = \mathbb{P}(\tau_n \leq t, \tau_{n+1} > t) = \mathbb{P}(\tau_{n+1} > t) - \mathbb{P}(\tau_n > t)$$
$$= e^{-\lambda t} \frac{\lambda t^n}{n!}$$

Alltså: $X(t) \sim Poi(\lambda t)$. Tack vare glömskegenskaperna gäller $X(s) - X(t) \sim Poi(\lambda(s-t))$ för alla $t < s$.

EXEMPEL: På en lågt trafikerad väg kommer i genomsnitt fem bilar per timme. Vad är sannolikheten att det passerar två bilar på en kvart?

Rimligt att bilar kommer enligt en Poissonprocess. Det måste då gälla att $\lambda = 5$ bilar / timme. Vi är intresserade av $X(\frac{1}{4})$ som är $Poi(\frac{5}{4})$-fördelad.

$$\mathbb{P}(X(\tfrac{1}{4}) \geq 2) = 1 - e^{-\frac{5}{4}}\left(\frac{(\frac{5}{4})^0}{0!} + \frac{(\frac{5}{4})^1}{1!}\right) \approx 0.36$$

$\square$

Betrakta två oberoende Poissonprocesser, kalla dem P1 och P2, med intensiteter λ_1 och λ_2. De gäller enligt glömskegenskaperna vid varje tidspunkt:

- Tiden T_{P1} till nästa impuls i P1 är $\exp(\lambda_1)$ fördelad
- Tiden T_{P2} till nästa impuls i P2 är $\exp(\lambda_2)$ fördelad
- Dessa två är oberoende

Tiden till nästa impuls i processen som räknar impulser i både P1 och P2 är då $\min(T_{P1}, T_{P2}) \sim \exp(\lambda_1 + \lambda_2)$
Alltså är summerade poissonprocessen en, med intensitet $\lambda_1 + \lambda_2$.
En konsekvens $X_1(1) + X_2(1) \sim Poi(\lambda_1 + \lambda_2)$

$\longrightarrow$ $Y_1 \sim Poi(\lambda_1)$ $Y_2 \sim Poi(\lambda_2)$, Y_1, Y_2 oberoende $\Longrightarrow Y_1 + Y_2 \sim Poi(\lambda_1 + \lambda_2)$

$$\mathbb{P}(Y_1 + Y_2 = k) = \sum_{j=0}^{k} \mathbb{P}(Y_1 = j)\mathbb{P}(Y_2 = k-j) = \sum_{j=0}^{k} e^{-\lambda_1} \frac{\lambda_1^j}{j!} \cdot e^{-\lambda_2} \frac{\lambda_2^{k-j}}{(k-j)!}$$

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

EXEMPEL: Pi an lägtrafikerad väg kommer i genomsnitt fem bilar per timme. Det kommer också i genomsnitt tre mc. Vad är sannolikheten att det passerar exakt tre fordon på halvtimmen?

Rimligt att anta två oberoende Poi-processer med intensitet 5 resp. 2, så bägge sorters fordon sammanvägda ger en Poi-process med int 7.
Låt $X(t)$ vara antal fordon på tid t. $X(\frac{1}{2}) \sim \text{Poi}(\frac{7}{2})$

$$\mathbb{P}(X(\tfrac{1}{2})=3) = e^{-\frac{7}{2}}\frac{(\frac{7}{2})^3}{3!} \approx 0.22$$

Om det passerar exakt n fordon på tid t, vad är den betingade sannolikheten att k av dessa är mc?
Skriv $X_b(t)$ resp $X_m(t)$

$$\mathbb{P}(X_m(t)=k \mid X(t)=n) = \frac{\mathbb{P}(X_m(t)=k,\ X(t)=n)}{\mathbb{P}(X(t)=n)}$$

$$= \frac{\mathbb{P}(X_m(t)=k)\,\mathbb{P}(X_b(t)=n-k)}{\mathbb{P}(X(t)=n)}$$

$$= \frac{e^{-2t}\frac{(2t)^k}{k!}\,e^{-5t}\frac{(5t)^{n-k}}{(n-k)!}}{e^{-7t}\frac{(7t)^n}{n!}} = \binom{n}{k}\left(\frac{2}{7}\right)^k\left(\frac{5}{7}\right)^{n-k}$$

dvs en $\text{Bin}(n, \frac{2}{7})$

$\square$

IDAG: Normalfördelning, Gaussisk fördelning.

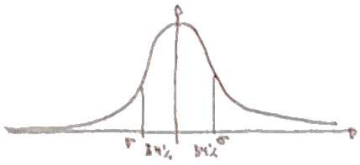
Normalfördelning

Uppstår (approximativt) av summor av alla många slumpmässiga bidrag.
Därför den vanligaste av alla fördelningar

DEFINITION: En kontinuerlig stokastisk variabel X sägs vara normalfördelad med parametrar μ och σ^2 om dess täthet ges av:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

Kort form $X \sim N(\mu, \sigma^2)$



För $X \sim N(0,1)$ skrivs tätheten som $\varphi(x)$ och fördelningsfunktionen som $\Phi(x)$. Med andra ord,

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

PROPOSITION: Låt $Z \sim N(0,1)$, och $X = \mu + \sigma Z$.
Då är $X \sim N(\mu, \sigma^2)$

OMVÄNT: $X \sim N(\mu, \sigma^2) \Longrightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$

BEVIS: $F_X(x) = \mathbb{P}(\mu + \sigma Z \leq x) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(t)\,dt$

Substituera $t = \frac{s-\mu}{\sigma}$, vilket ger $dt = \frac{1}{\sigma} ds$ och
$t = \frac{x-\mu}{\sigma} \Longleftrightarrow s = x$

$$\int_{-\infty}^{x} \frac{1}{\sigma} \varphi\left(\frac{s-\mu}{\sigma}\right) ds$$

Men $\frac{1}{\sigma}\varphi\left(\frac{s-\mu}{\sigma}\right)$ är tätheten för $N(\mu, \sigma^2)$, så X har ensamma fördelningsfunktion

$\square$

Låt $Z \sim N(0,1)$. Vi har $\mathbb{E}[Z] = \int_{-\infty}^{\infty} x\varphi(x)\,dx$. Integralen är 0 eftersom $\varphi(x)$ är en jämn funktion, så $\mathbb{E}[Z] = 0$

Vidare är $\mathbb{E}[Z^2] = \int_{-\infty}^{\infty} x^2 \varphi(x)\,dx$. Funktionen $x\varphi(x)$ har primitiv funktion $-\varphi(x)$. Partiell integration ger:

$$\mathbb{E}[Z^2] = [-x\varphi(x)]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \varphi(x)\,dx = 1$$

$\Longrightarrow \mathrm{Var}(Z) = 1$

För $X \sim N(\mu, \sigma^2)$, skriv $X = \mu + \sigma Z$ och ta

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

OBS att: $\mathbb{P}(X \le x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

så Φ är den enda fördelningsfunktion för normalfördelning man behöver veta. Tack vare symmetrin hos φ gäller:

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

så man behöver bara $\Phi(x)$ för $x > 0$

EXEMPEL: Längden X av en på måfå vald svensk man är approximativt normalfördelad med väntevärde 180 cm och standardavvikelse 5. Vi får t.ex:

$$\mathbb{P}(X \le 176) \approx \Phi\left(\frac{176-180}{5}\right) = \Phi(-0.8) = 1 - \Phi(0.8) \approx 0.21$$

$$\mathbb{P}(X \ge 210) \approx 1 - \Phi\left(\frac{210-180}{5}\right) = 1 - \Phi(6) \approx 9.9 \cdot 10^{-10}$$

Normalfördelning är i tillämpningar alltid approximativa och fungerar dåligt för extrema sannolikheter.

$\square$

EXEMPEL: Marmelbolaget säljer burkar med 400g apelsinmarmelad. På grund av slumpmekanismer utanför bolagets kontroll är den verkliga vikten av marmelad som fylls med i en given burk normalfördelad med standardavvikelse 4g.

Hur mycket marmelad ska bolaget sikta på att fylla i en given burk för att sannolikheten att den verkliga vikten blir minst 400g blir 0.95.

Vad man siktar på är alltså att tolka som väntevärdet av vikten på en burk. Skriv μ för detta väntevärde. Vi har alltså $X \sim N(\mu, 4^2)$ och ska välja μ så att $\mathbb{P}(X > 400) = 0.95$

$$\mathbb{P}(X > 400) = 1 - \Phi\left(\frac{400-\mu}{4}\right) = \Phi\left(\frac{\mu-400}{4}\right) = 0{,}95$$

Nu är $\Phi^{-1}(0{,}95) \approx 1.64 \implies \frac{\mu-400}{4} \approx 1.64$

$$\mu \approx 400 + 4 \cdot 1{,}64 \approx 406.6$$

Lite standardvärden på Φ^{-1}:

x	$\Phi^{-1}(x)$
0.9	1.28
0.95	1.64
0.975	1.96
0.99	2.33
0.995	2.55

Sannolikhet och statistik 23/4

FLERDIMENSIONELLA FÖRDELNINGAR.

Om X och Y är två stokastiska variabler definierade på samma utfallsrum, kallar vi paret (X,Y) för en två dimensionell stokastisk variabel. Den bivariata fördelningsfunktionen för (X,Y) ges av:

$$F(x,y) = \mathbb{P}(X \le x,\ Y \le y) \quad , \quad (x,y) \in \mathbb{R}^2$$

Vi har $F(x,\infty) = \mathbb{P}(X \le x,\ Y < \infty) = \mathbb{P}(X \le x) = F_X(x)$
och på samma sätt $F(\infty, y) = F_Y(y)$

DISKRETA TVÅDIMENSIONELLA STOKASTISKA VARIABLER.

Om X och Y är diskreta, säger man att (X,Y) är diskreta.
OBS: $V_X \times V_Y = \{(x_j, y_k) \mid x_j \in V_X,\ y_k \in V_Y\} \supseteq V_{X,Y}$

Vi antar att $V_{X,Y} = V_X \times V_Y$. (det förekommer ovanligt att par av värden som (X,Y) antar med sannolikhet 0)

Den bivariata frekvensfunktionen för (X,Y) ges av.

$$p(x_j, y_k) = \mathbb{P}(X = x_j,\ Y = y_k) \quad , \quad (x_j, y_k) \in V_X \times V_Y$$

Direkt observation

$$p_X(x_j) = \mathbb{P}(X = x_j,\ Y \in V_Y) = \sum_{y_k \in V_Y} p(x_j, y_k)$$

och $$p_Y(y_k) = \sum_{x_j \in V_X} p(x_j, y_k)$$

EXEMPEL: Antag att (X,Y) antar vart och ett av alla heltalspar (x,y) så att $1 \le x \le y \le 4$ med samma sannolikhet.
Det finns $4+3+2+1 = 10$ värden så

$$p(x,y) = \frac{1}{10} \ , \ 1 \le x \le y \le 10$$

$$p_X(x) = \sum_{y=x}^{4} \frac{1}{10} = \frac{5-x}{10} \quad , \quad x = 1,2,3,4$$

$$p_Y(y) = \sum_{x=1}^{y} \frac{1}{10} = \frac{y}{10} \quad , \quad y = 1,2,3,4$$

EXEMPEL: Slå två tärningar. Låt X vara första tärningens utfall och Y vara summan. Bestäm $p(x,y)$
Paret (x,y) är möjligt värde om $x \in \{1,2,3,4,5,6\}$ och $y \in \{x+1, \dots, x+6\}$ vilka kan inträffa på exakt ett sätt

$$p(x,y) = \frac{1}{36} \ , \ 1 \le x \le 6, \quad x+1 \le y \le x+6$$

EXEMPEL: Låt $X \sim Po(\lambda)$, och sedan givet $X = x$ låt $Y \sim Bin(x, p)$

V. fås $p(x,y) = e^{-\lambda} \frac{\lambda^x}{x!} \binom{x}{y} p^y (1-p)^{x-y}$, $0 \le y \le x < \infty$

Marginalfördelning för Y blir

$$p_Y(y) = \sum_{x=y}^{\infty} p(x,y) = e^{-\lambda} \sum_{x=y}^{\infty} \frac{\lambda^x}{x!} \binom{x}{y} p^y (1-p)^{x-y}$$

$$= \frac{e^{-\lambda} \lambda^y p^y}{y!} \sum_{x=y}^{\infty} \frac{(\lambda(1-p))^{x-y}}{(x-y)!}$$

$$= \frac{e^{-\lambda} (\lambda p)^y}{y!} e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^y}{y!}$$

Med mera ord: $Y \sim Po(\lambda p)$

Konsekvens: Om man i en Poi-process med in λ räknar med impulserna med sannolikhet p, får man en Poi-process med intensitet λp. (Uttunnad Poisson-process)

SATS: Om X, Y diskreta och oberoende $\iff$ $p(x_j, y_k) = p_X(x_j) p_Y(y_k)$ för alla $x_j \in V_X$, $y_k \in V_Y$

BEVIS: Om X, Y oberoende gäller

$$p(x_j, y_k) = \mathbb{P}(X = x_j, Y = y_k) = \mathbb{P}(X = x_j)\mathbb{P}(Y = y_k) = p_X(x_j) p_Y(y_k)$$

Å andra sidan om p kan faktoriseras för vi

$$\mathbb{P}(X \in A, Y \in B) = \sum_{x_j \in A} \sum_{y_k \in B} p(x_j, y_k) = \sum_{x_j \in A} p_X(x_j) \sum_{y_k \in B} p_Y(y_k)$$

$$= \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

$\square$

KONTINUERLIG TVÅDIMENSIONELL STOKASTISK VARIABEL

Paret (X,Y) sägs vara en kontinuerlig tvådimensionell stokastisk variabel om det finns en täthetsfunktion $f : \mathbb{R}^2 \to \mathbb{R}$ sådan att:

$$\mathbb{P}((X,Y) \in B) = \iint_B f(x,y)\, dx\, dy$$

för alla $B \subseteq \mathbb{R}^2$

Vägen observationer

- $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\, dx\, dy$
- $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$
- $\mathbb{P}(X \in A) = \mathbb{P}((X,Y) \in A \times \mathbb{R}) = \int_A \int_{-\infty}^{\infty} f(x,y)\, dy\, dx$ så
  - $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$
  - $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$

EXEMPEL: Låt (X,Y) ha täthet $f(x,y) = c(x+3y)$, $0 \le x \le 1$, $0 \le y \le 1$
Vad är c, samt vad är f_X och f_Y

Det måste gälla att $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$, så

$$1 = c\int_0^1\int_0^1 (x+3y)\,dx\,dy = c\left(\int_0^1 x\,dx + 3\int_0^1 y\,dy\right) = 2c$$

$\Longrightarrow c = \frac{1}{2}$

$$f_X(x) = \frac{1}{2}\int_0^1 (x+3y)\,dy = \frac{1}{2}\left(x + \frac{3}{2}\right), \quad 0 \le x \le 1$$

$$\text{och } f_Y(y) = \frac{1}{2}\int_0^1 (x+3y)\,dx = \frac{1}{2}\left(3y + \frac{1}{2}\right), \quad 0 \le y \le 1.$$

EXEMPEL: Låt $D = \{(x,y) \mid x^2+y^2 \le 1\}$. Låt $(X,Y) \sim \text{likf}(D)$
Dvs: $f(x,y) = \begin{cases} \frac{1}{\pi} & x \in D \\ 0 & x \notin D \end{cases}$

Låt $Z = |(X,Y)| = \sqrt{X^2+Y^2}$. Vad är fördelningen för Z

$$F_Z(z) = \mathbb{P}((X,Y) \in B_z) = \frac{1}{\pi}\,\text{Area}(B_z)$$

där $B_z = \{(x,y) \mid x^2+y^2 \le z^2\}$. Alltså är $\text{Area}(B_z) = \pi z^2$.

$$F_Z(z) = z^2, \quad 0 \le z \le 1$$

$$f_Z(z) = 2z \quad 0 \le z \le 1$$

Vad är fördelning för X?

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi}\sqrt{1-x^2}$$

Om $f(x,y) = f_X(x)\,f_Y(y)$ för alla (x,y) gäller

$$\mathbb{P}(X \in A, Y \in B) = \int_A\int_B f(x,y)\,dx\,dy = \int_A f_X(x)\,dx \int_B f_Y(y)\,dy$$
$$= \mathbb{P}(X \in A)\,\mathbb{P}(X \in B)$$

dvs X, Y oberoende
Omvänt X och Y oberoende $\Longrightarrow f(x,y) = f_X(x)\,f_Y(y)$ för nästan alla (x,y).

FUNKTIONER AV TVÅ STOKASTISKA VARIABLER.
Om (X,Y) är en tvådimensionell stokastisk variabel och $g: \mathbb{R}^2 \longrightarrow \mathbb{R}$, vad är fördelningen av $g(X,Y)$. Ändra från fall till fall.

EXEMPEL: $(X,Y) \sim \text{likf}([0,1]^2)$. Låt $A = XY$. För $0 < a < 1$

$$F_A(a) = \mathbb{P}(XY \le a) = \int_0^1\int_0^{\min(\frac{a}{x},1)} dy\,dx = \left(a + a\int_a^1 \frac{1}{x}\,dx\right) = a - a\ln a$$

Derivering ger: $f_A(a) = -\ln a \quad 0 < a < 1$

$\square$

OSL I TVÅ DIMENSIONER

- Diskret: $\mathbb{E}[g(X,Y)] = \sum_{(x_j,y_k)\in V_{X,Y}} g(x_j,y_k)\, p(x_j,y_k)$

- Kontinuerligt: $\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\, dy$

BEVIS: Låt $V_g = g(V_{X,Y})$

$$\mathbb{E}[g(X,Y)] = \sum_{z\in V_g} z\,\mathbb{P}(g(X,Y)=z) = \sum_{z\in V_g} z \sum_{(x_j,y_k)\,|\,g(x_j,y_k)=z} p(x_j,y_k)$$

$$= \sum_{z\in V_g}\sum_{(x_j,y_k)\,|\,g(x_j,y_k)=z} g(x_j,y_k)\,p(x_j,y_k) = \sum_{(x_j,y_k)\in V_{X,Y}} g(x_j,y_k)\,p(x_j,y_k)$$

$\square$

FÖLJDSATS: För alla stokastiska variabler X och Y gäller $\mathbb{E}[X+Y] = \mathbb{E}[X]+\mathbb{E}[Y]$

BEVIS: Enligt OSL:

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

$$\cancel{\int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} f(x,y)\,dy\,dx} = \int_{-\infty}^{\infty} x\, f_X(x)\,dx + \int_{-\infty}^{\infty} y\, f_Y(y)\,dy = \mathbb{E}[X]+\mathbb{E}[Y]$$

EXEMPEL: Låt $I_1, I_2, \dots, I_n$ vara sådana att

$\mathbb{P}(I_k=1) = 1-\mathbb{P}(I_k=0) = p$ Låt $X = \sum_{k=1}^{n} I_k$

D: ger följdsatsen att

$$\mathbb{E}[X] = \sum_{k=1}^{n}\mathbb{E}[I_k] = np$$

Exempelvis visar detta att $X \sim \mathrm{Bin}(n,p) \Longrightarrow \mathbb{E}[X] = np$

SANNOLIKHET OCH STATISTIK 26/4

BETINGADE FÖRDELNINGAR

Antag att X och Y är diskreta. Man skriver:

$$p_{Y|X}(y|x) = \mathbb{P}(Y=y \mid X=x) = \frac{p(x,y)}{p_X(x)}, \quad (x,y) \in V_{X,Y}$$

EXEMPEL: Låt $p(x,y)$ ges ut sidan:

x/y	0	1
0	$\frac{1}{10}$	$\frac{1}{4}$
1	$\frac{1}{5}$	$\frac{9}{20}$

$$p_{Y|X}(0|1) = \frac{p(1,0)}{p_X(1)} = \frac{\frac{1}{5}}{\frac{1}{5}+\frac{9}{20}} = \frac{4}{13}, \quad p_{X|Y}(1|1) = \frac{p(1,1)}{p_Y(1)} = \frac{\frac{9}{20}}{\frac{1}{4}+\frac{9}{20}} = \frac{9}{14}$$

EXEMPEL: Slå tre tärningar. Låt X vara första tärningens utslag, och Y summan. Sannolikhet att $p_{Y|X}(y|x) = \frac{1}{6}$, $x+1 \le y \le x+6$.
Då får vi enligt TSL:

$$p_Y(a) = \sum_{x \in V_X} p_{Y|X}(y|x)\, p_X(x) = \frac{1}{6}\left(0+0+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\right) = \frac{1}{9}$$

$$\left(p_Y(y) = \frac{6-|y-7|}{36} \right.$$

$\square$

Motsvarighet för kontinuerlig stokastisk variabel? Vill betrakta ej $\{X=x\}$ ifrån om X kontinuerlig och ändrade $\mathbb{P}(X=x)=0$.
Antag först att (X,Y) är kontinuerlig. Vi borde ha

$$\mathbb{P}(Y \in B \mid X=x) = \mathbb{P}(Y \in B \mid X \in x \pm \Delta x) = \frac{\mathbb{P}(Y \in B,\ X \in x \pm \Delta x)}{\mathbb{P}(X \in x \pm \Delta x)}$$

$$= \frac{\int_B \int_{x-\Delta x}^{x+\Delta x} f(x,y)\,dx\,dy}{\int_{x-\Delta x}^{x+\Delta x} f_X(x)\,dx} \approx \frac{2\Delta x \int_B f(x,y)\,dy}{2\Delta x\, f_X(x)}$$

$$= \int_B \frac{f(x,y)}{f_X(x)}\,dy$$

Därför definierar vi täthetsfunktionen för Y givet $X=x$ som

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Då får vi per definition att

$$\mathbb{P}(Y \in B \mid X=x) = \int_B f_{Y|X}(y|x)\,dy$$

Notera: $\mathbb{P}(Y \in B) = \int_{-\infty}^{\infty}\int_B f(x,y)\,dy\,dx = \int_{-\infty}^{\infty} f_X(x) \int_B f_{Y|X}(y|x)\,dy\,dx$

$$= \int_{-\infty}^{\infty} \mathbb{P}(Y \in B \mid X=x)\, f_X(x)\,dx$$

$\Rightarrow$ Kontinuerlig version av totala sannolikhetslagen.

En annan variant:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, f_X(x)\,dx$$

Lite inom parentes: Y diskret och X kontinuerlig

Antag $\exists f : V_Y \times \mathbb{R} : \forall y \in V_Y \times \forall B \subseteq \mathbb{R}$

$$\mathbb{P}(X \in B,\ Y = y_k) = \int_B f(x, y_k)\,dx$$

DEFINITION: $\mathbb{P}(Y = y_k \mid X = x) = \dfrac{f(x, y_k)}{f_X(x)}$

TSL: $\mathbb{P}(Y = y_k) = \int_{-\infty}^{\infty} \mathbb{P}(Y = y_k \mid X = x)\, f_X(x)\,dx$

Vi betecknar också $f_{X|Y}(x \mid y_k) = \dfrac{f(x, y_k)}{p_Y(y_k)}$

med TSL: $f_X(x) = \sum_{y_k \in V_Y} f_{X|Y}(x \mid y_k)\, p_Y(y_k)$ $\square$

I de flesta tillämpningar är det intuitivt mer klart vad den betingade fördelningen är än den bivariata.

Om (X, Y) kontinuerlig och X, Y oberoende så vi:

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

EXEMPEL: $(X, Y) \sim \text{likf}(D)$, $D = \{(x,y) \mid x^2 + y^2 < 1\}$. Vi har sett tidigare att

$$f(x,y) = \frac{1}{\pi}, \quad (x,y) \in D$$

och $f_X(x) = \frac{2}{\pi}\sqrt{1 - x^2} \quad 0 < x < 1$

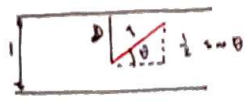
Alltså:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{2\sqrt{1-x^2}}, \quad -\sqrt{1-x^2} < y < \sqrt{1-x^2}$$

Med mera allt: Betingat fördelas för Y givet $X = x$ är $\text{likf}\left(-\sqrt{1-x^2}, \sqrt{1-x^2}\right)$

EXEMPEL: Välj $X \sim \text{likf}(0,1)$, och sedan, givet $X = x$ $Y \sim \text{likf}(0, x)$. Vi har $f_X(x) = 1$, $0 < x < 1$ och $f_{Y|X}(y|x) = \frac{1}{x}$ $\quad 0 < y < x$. Vad är $\mathbb{P}(Y < X^2)$?

$$\mathbb{P}(Y < X^2) = \int_0^1 \mathbb{P}(Y < x^2 \mid X = x)\, f_X(x)\,dx = \int_0^1\int_0^{x^2} \frac{1}{x}\,dy\,dx = \int_0^1 x\,dx = \frac{1}{2}$$



Buffons nålproblem

Rimlig modell $D \sim \text{likf}(0, \frac{1}{2})$, $\Theta \sim \text{likf}(0, \frac{\pi}{2})$, D, Θ oberoende

Vad är sannolikhet att nålen skär linje? Vi söker $\mathbb{P}(\frac{1}{2}\sin\Theta > D)$

$$\mathbb{P}(\sin\Theta > 2D) = \int_0^{\frac{\pi}{2}} \mathbb{P}(2D < \sin\theta \mid \Theta = \theta)\, f_\Theta(\theta)\,d\theta$$

$$= \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \mathbb{P}(2D < \sin\theta)\,d\theta = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \sin\theta\,d\theta = \frac{2}{\pi}$$

EXEMPEL: Om $T_1, T_2, \ldots$ är oberoende och $\exp(\lambda)$-fördelade så är $X_n = \sum_{k=1}^{n} T_k \sim F(n, \lambda)$. Vi har visat att

$$\mathbb{P}(X_n > x) = e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$$

Sant för $n=1$, så antag att sant för $n = m$. Då är $X_{m+1} = X_m + T_{m+1}$

$$\mathbb{P}(X_{m+1} > x) = \int_0^\infty \mathbb{P}(X_{m+1} > x \mid T_{m+1} = t) f_T(t)\, dt$$

$$= \int_0^\infty \mathbb{P}(X_m > x - t \mid T_{m+1} = t) f_T(t)\, dt$$

$$= \int_0^x \lambda e^{-\lambda t} e^{-\lambda(x-t)} \sum_{k=0}^{m-1} \frac{(\lambda(x-t))^k}{k!}\, dt + e^{-\lambda x}$$

Men $\int_0^x \frac{\lambda^{k+1}(x-t)^k}{k!} = \frac{(\lambda x)^{k+1}}{(k+1)!}$, så vår uttryck blir

$$e^{-(\lambda x)} \sum_{k=0}^{m} \frac{(\lambda x)^k}{k!}$$

PROPOSITION: Om X och Y är oberoende gäller

a) $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

b) $\mathbb{V}ar(X+Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y)$

BEVIS: a) Observerat att

$$\mathbb{E}[XY] = \iint xy\, f(x,y)\, dx\, dy = \iint xy\, f_X(x) f_Y(y)\, dx\, dy$$

$$= \int x f_X(x)\, dx \int y f_Y(y)\, dy = \mathbb{E}[X]\mathbb{E}[Y]$$

b) $\mathbb{V}ar(X+Y) = \mathbb{E}[(X+Y-\mu_X-\mu_Y)^2] = \mathbb{E}[(X-\mu_X)^2] + \mathbb{E}[(Y-\mu_Y)^2] + 2\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$

$= \mathbb{V}ar(X) + \mathbb{V}ar(Y)$

ty extra termen är enligt observerat 0 $\square$

Mer generellt: $\mathbb{E}[\prod X_k] = \prod \mathbb{E}[X_k]$, $\mathbb{V}ar(\sum X_k) = \sum \mathbb{V}ar(X_k)$

EXEMPEL: $X \sim Bin(n,p)$. Skriv $X = \sum_{k=1}^{n} I_k$, där I_k är oberoende

$\mathbb{V}ar(X) = \sum_{k=1}^{n} \mathbb{V}ar(I_k) = np(1-p)$

Låt $X_1, X_2, \ldots$ vara oberoende och likafördelade med $\mathbb{E}[X_i] = \mu$ och $\mathbb{V}ar(X_i) = \sigma^2 < \infty$

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad \mathbb{V}ar(\bar{X}_n) = \frac{1}{n^2}\sum_{k=1}^{n} \mathbb{V}ar(X_k) = \frac{\sigma^2}{n}$$

STORA TALENS LAG: För varje $\varepsilon > 0$ gäller

$$\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

BEVIS: Chebyshevs olikhet ger

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathbb{V}ar(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0 \text{ då } n \to \infty$$

SANNOLIKHET OCH STATISTIK. 27/4

BETINGADE VÄNTEVÄRDEN

Vi ska se på att $\mathbb{E}[Y|X=x]$ är väntevärdet av den stokastiska variabel som har samma fördelning som Y givet $X=x$

DEFINITION:

Diskret Y: $\mathbb{E}[Y|X=x] = \sum_{y_k \in V_Y} y_k p_{Y|X}(y_k|x)$

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

Totala sannolikhetslagen för väntevärden:

X diskret: $\mathbb{E}[Y] = \sum_{x_j \in V_X} \mathbb{E}[Y|X=x_j] p_X(x_j)$

X kontinuerlig: $\mathbb{E}[Y] = \int_{-\infty}^{\infty} \mathbb{E}[Y|X=x] f_X(x)\, dx$

BEVIS: Fallet (X,Y) kontinuerligt

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

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

där sista likheten är OSL.

EXEMPEL: $X \sim \text{likf}(0,1)$ och sedan $Y \sim \text{likf}(0,X)$. Vi vill visa att $f(x,y) = \frac{1}{x}$, $0 \le y < x < 1$. Enligt OSL:

$$\mathbb{E}[Y] = \int_0^1 \int_0^x y f(x,y)\, dy\, dx = \int_0^1 \int_0^x y \frac{1}{x}\, dy\, dx = \frac{1}{2}\int_0^1 x\, dx = \frac{1}{4}$$

Alt:

$$\mathbb{E}[Y] = \int_0^1 \mathbb{E}[Y|X=x] f_X(x)\, dx = \int_0^1 \frac{x}{2}\, dx$$

DEFINITION: $\mathbb{E}[Y|X]$ är den stokastiska variabel som antar värdet $\mathbb{E}[Y|X=x]$ då $X=x$

Med andra ord om funktionen θ ges av $g(x) = \mathbb{E}[Y|X=x]$ så är $\mathbb{E}[Y|X] = g(X)$

Kort form för TSL för betingat väntevärde. Antag (X,Y) kontinuerlig (funkar lika bra diskret). Enligt OSL, TSL gäller

$$\mathbb{E}[Y] = \int \mathbb{E}[Y|X=x] f_X(x)\, dx = \int g(x) f_X(x)\, dx = \mathbb{E}[g(X)] = \mathbb{E}[\mathbb{E}[Y|X]]$$

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

EXEMPEL: Låt $X \sim \text{Geo}(p)$ och sedan $Y \sim \text{Bin}(X,r)$. Vad är det $\mathbb{E}[Y]$

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Xr] = \frac{r}{p}$$

EXEMPEL: Låt $X_1, X_2, \ldots$ vara oberoende och $\mathbb{E}[X_k] = \mu$ för alla k.
Låt N vara oberoende av X_k:na och positivt heltalsvärd:

$$\mathbb{E}\left[\sum_{k=1}^{N} X_k\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right] = \mathbb{E}[N\mu] = \mu\mathbb{E}[N]$$

OBS: $\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A|X)]$, ty $\mathbb{H} = \int \mathbb{P}(A|X=x) f_X(x)\,dx = \mathbb{P}(A)$

KOVARIANS OCH KORRELATION.

Låt X och Y vara två stokastiska variabler.

DEFINITION: $\mathbb{C}ov(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$.

Ett mått på hur X och Y samvarierar.

$\mathbb{C}ov(X,Y) > 0$ betyder positivt beroende

$\mathbb{C}ov(X,Y) < 0$ betyder negativt beroende

Formel: $\mathbb{C}ov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Följer av: $\mathbb{C}ov(X,Y) = \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] = \mathbb{E}[XY] - \mu_X\mu_Y$

Om X, Y oberoende $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, så följer det att

X och Y oberoende $\Longrightarrow \mathbb{C}ov(X,Y) = 0$

Varning: $\mathbb{C}ov(X,Y) = 0$ medför i allmänhet inte att X, Y oberoende.

Några observationer:

- $\mathbb{C}ov(X,a) = 0$
- $\mathbb{C}ov(X,X) = \mathbb{V}ar(X)$
- $\mathbb{C}ov(X,Y) = \mathbb{C}ov(Y,X)$

Formeln ger också att kovariansen bilinjär:

$$\mathbb{C}ov(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) = a_1b_1\mathbb{C}ov(X_1,Y_1) + a_1b_2\mathbb{C}ov(X_1,Y_2) + a_2b_1\mathbb{C}ov(X_2,Y_1) + a_2b_2\mathbb{C}ov(X_2,Y_2)$$

Exempelvis $\mathbb{C}ov(X+a, Y+b) = \mathbb{C}ov(X,Y) + \mathbb{C}ov(X,b) + \mathbb{C}ov(a,Y) + \mathbb{C}ov(a,b)$
$= \mathbb{C}ov(X,Y)$

Varianserna för en summa:

$$\mathbb{V}ar(X+Y) = \mathbb{C}ov(X+Y, X+Y) = \mathbb{V}ar(X) + 2\mathbb{C}ov(X,Y) + \mathbb{V}ar(Y)$$

EXEMPEL: Låt $\Theta \sim \mathrm{Likt}(0,2\pi)$ och $X = \cos\Theta$, $Y = \sin\Theta$. Vi har

$$\mathbb{E}[X] = \int_0^{2\pi} \cos\theta \frac{1}{2\pi}\,d\theta = 0, \quad \mathbb{E}[Y] = \int_0^{2\pi} \sin\theta \frac{1}{2\pi}\,d\theta = 0$$

$$\mathbb{E}[XY] = \int_0^{2\pi} \cos\theta \sin\theta \frac{1}{2\pi}\,d\theta = 0$$

Alltså $\mathbb{C}ov(X,Y) = 0$, men X, Y är inte oberoende.

EXEMPEL: Låt X och Y vara oberoende och likf$(0,1)$. Låt $Z = XY$, $W = X + Y$. Bestäm $\mathbb{C}ov(Z,W)$

$$\mathbb{E}[Z] = \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = \frac{1}{4}$$

$$\mathbb{E}[W] = \mathbb{E}[X] + \mathbb{E}[Y] = 1$$

$$\mathbb{E}[ZW] = \mathbb{E}[XY(X+Y)] = \mathbb{E}[X^2]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y^2] = \frac{1}{3}$$

Alltså:

$$\mathbb{C}ov(Z,W) = \frac{1}{3} - \frac{1}{4}\cdot 1 = \frac{1}{12} \quad \Rightarrow \text{ Positivt beroende.}$$

$\square$

I detta exempel var kovariansen $\frac{1}{12}$, men vad säger det om hur <u>starkt</u> beroendet (positivt) är? Måste sättas i relation till varianserna. Dags att prata om korrelation.

Schwarz olikhet: För alla stokastiska variabler X och Y gäller

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

BEVIS: Vi använder att för alla konstanter t:

$$0 \leq \mathbb{E}[(X - tY)^2] = \mathbb{E}[X^2] - 2t\mathbb{E}[XY] + t^2\mathbb{E}[Y^2]$$

HL blir som minst för $t = \dfrac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$. Med detta t i HL och erhåll:

$$0 \leq \mathbb{E}[X^2] - 2\frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} + \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} = \mathbb{E}[X^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]}$$

Flytta med $\mathbb{E}[Y^2]$ och sätt är klart.

$\square$

DEFINITION: Korrelationskoefficienten för (X,Y) ges av

$$\rho(X,Y) = \frac{\mathbb{C}ov(X,Y)}{\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}}.$$

Vi har $-1 \leq \rho(X,Y) \leq 1$ enligt Schwarz använd på $X - m_X$ och $Y - m_Y$. Vi har också

- $\rho(X,Y) = \rho(aX, bY)$ skalningsinvarians.
- $Y = aX + b$ för konstant $a > 0 \iff \rho(X,Y) = 1$
- $Y = aX + b$ för konstant $a < 0 \iff \rho(X,Y) = -1$

DEFINITION: Om $\rho(X,Y) = 0$ kallas X och Y okorrelerade

EXEMPEL: Vi har X, Y oberoende och likf $(0,1)$ och $Z = XY$, $W = X + Y$. Vi finner $\mathbb{C}ov(Z, W) = \frac{1}{12}$. Vad är $\rho(X,Y)$? Vi behöver varianserna.

$$\mathbb{V}ar(W) = \mathbb{V}ar(X+Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y) = \frac{1}{6}$$

Eftersom varianser av likf $(0,1)$ är $\frac{1}{12}$

$$\begin{aligned}\mathbb{V}ar(Z) &= \mathbb{V}ar(XY) = \mathbb{E}[X^2Y^2] - \mathbb{E}[XY]^2 \\ &= \mathbb{E}[X^2]\mathbb{E}[Y^2] - (\mathbb{E}[X]\mathbb{E}[Y])^2 \\ &= \left(\tfrac{1}{3}\right)^2 - \left(\tfrac{1}{4}\right)^2 = \frac{7}{144}\end{aligned}$$

Eftersom $\mathbb{E}[X^2] = \frac{1}{3}$

Alltså: $\rho(Z,W) = \dfrac{\frac{1}{12}}{\sqrt{\frac{1}{6} \cdot \frac{7}{144}}} = \sqrt{\frac{6}{7}}$

EXEMPEL: Bästa linjära prediktor: Vilka a och b ger att minimera $\mathbb{E}[(Y - (aX+b))^2]$

Antag att $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ och $\mathbb{V}ar(X) = \mathbb{V}ar(Y) = 1$

Då är $\rho = \mathbb{C}ov(X,Y) = \mathbb{E}[XY]$ och

$$\begin{aligned}\mathbb{E}[(Y-(aX+b))^2] &= \mathbb{E}[Y^2] - 2\mathbb{E}[Y(aX+b)] + \mathbb{E}[(aX+b)^2] \\ &= 1 - 2a\rho + a^2 + b^2\end{aligned}$$

Detta min med $b = 0$ och $a = \rho$. Man säger alltså att i detta fall är bästa linjära prediktor av Y given X lika med ρX. I det allmänna fallet ges den bästa linjära prediktorn av

$$\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

$\square$

CENTRALA GRÄNSVÄRDESSATSEN

Låt $X_1, X_2, \ldots$ vara oberoende och lika fördelade med $\mathbb{E}[X_i] = \mu$ och $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Låt

$$S_n = \sum_{k=1}^{n} X_k$$

Då gäller för alla $x \in \mathbb{R}$ att

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) \longrightarrow \Phi(x)$$

då $n \longrightarrow \infty$. Med andra ord: S_n är approximativt $N(n\mu, n\sigma^2)$-fördelad när n är stort. (Oavsett vilken fördelning X_k:na har, men ju större n som krävs för att approximationen ska vara god beror förstås på X_k:nas fördelning)

EXEMPEL: I världen inträffar i genomsnitt 100 stora jordbävningar per år. Vad är sannolikheten att det ett givet år inträffar minst 110 stora jordbävningar?

Rimligt att tro att jb kommer som en Poi-process med int 100. Antal jb $\sim \mathrm{Poi}(100)$, lika sannolikt att räkna på:
Alt: Tiderna $T_1, T_2, \ldots$ mellan jb $\exp(100)$-fördelade. Låt:

$$S_n = \sum_{k=1}^{n} T_k$$

Vi söker $\mathbb{P}(S_{110} \le 1)$. Vi har $\mu = \mathbb{E}[T_k] = \frac{1}{100} = 0.01$
$\sigma^2 = \mathrm{Var}(T_k) = \frac{1}{100^2} = 0{,}0001$. Alltså, enligt CGS.

$$\mathbb{P}(S_{110} \le 1) = \mathbb{P}\left(\frac{S_{110} - 110 \cdot 0.01}{0.01 \cdot \sqrt{110}} \le \frac{1 - 100 \cdot 0.01}{0.01 \cdot \sqrt{110}}\right)$$
$$\approx \Phi\left(-\frac{0{,}1}{\sqrt{0.011}}\right) \approx 0.17$$

EXEMPEL: Låt $X \sim \mathrm{Bin}(n, p)$ med n stort och p inte extremt ($\frac{1}{n} \ll p \ll 1 - \frac{1}{n}$; vi tänker att $n \longrightarrow \infty$ och p fixt. Skriv

$$X = \sum_{k=1}^{n} I_k$$

Vi har $\mu = \mathbb{E}[I_k] = p$ och $\sigma^2 = \mathrm{Var}(I_k) = p(1-p)$. Enligt CGS

$$\mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \le x\right) \longrightarrow \Phi(x)$$

Exempelvis slår en tärning 700 gånger och låt X vara antal 6.
Vad är $\mathbb{P}(X \ge 100)$?
X diskret, så $\mathbb{P}(X \ge 100) = \mathbb{P}(X > 99) = \mathbb{P}(X > 99.5)$

$$\mathbb{P}(X > 99.5) = 1 - \mathbb{P}\left(\frac{X - 700 \cdot \frac{1}{6}}{\sqrt{700 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \le \frac{99.5 - 700 \cdot \frac{1}{6}}{\sqrt{700 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right)$$
$$\approx 1 - \Phi(-1.74) = \Phi(1.74) \approx 0.959$$

CGS fungerar också utan lika fördelning för enkla frekvensfunktioner. Även utan oberoende under lite tuffare krav. I båda fallen måste variansen förstås justeras. Se upp med extrema sannolikheter.

STATISTIK

Sannolikhetsparametrar okända. Skattar parametrar utifrån data.
Skolusitt: iid = independent and identically distributed
= oberoende och likafördelade.

DEFINITION: Låt $X_1, \dots, X_n$ vara iid och fördelade som X.
De kallas $X_1, \dots, X_n$ för stickprov (sample) på X
(eller på F_X)

Antag att F_X beror på en parameter, θ. Exempelvis $X \sim Po(\theta)$,
$X \sim \exp(\theta)$, $X \sim N(\theta, 1)$ eller $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = \theta$.
Parametern kan vara flerdimensionell, t.ex. $\theta = (\mu, \sigma^2)$ och $X \sim N(\mu, \sigma^2)$

En funktion av $X_1, \dots, X_n$ som används till att skatta θ kallas för en punktskattning (estimator) av θ. Standardbeteckningar:
$\hat{\theta}$, $\hat{\theta}_n$, $\hat{\theta}(X_1, \dots, X_n)$

OBSERVERA att $\hat{\theta}$ är en stokastisk variabel. Efter data $X_1 = x_1, \dots, X_n = x_n$ observerats får man ett givet värde $\hat{\theta}(x_1, \dots, x_n)$, ett estimat av θ.

DEFINITION: Om det, oavsett det korrekta värdet på θ, gäller att
$$\mathbb{E}[\hat{\theta}] = \theta$$
kallas $\hat{\theta}$ för väntevärdesriktig skattning av θ
(unbiased, VVR)

DEFINITION: Om $\lim_{n \to \infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) = 0$ kallas $\hat{\theta}$ för konsistent skattning av θ

PROPOSITION: Om $\hat{\theta}_n$ är VVR och $\mathbb{V}ar(\hat{\theta}_n) \longrightarrow 0$, så är $\hat{\theta}_n$ konsistent.
BEVIS: Enligt Chebyshevs's olikhet:
$$\mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{\mathbb{V}ar(\hat{\theta}_n)}{\varepsilon^2} \longrightarrow 0$$
Låt $X_1, \dots, X_n$ vara ett stickprov på en stokastisk variabel X med $\mathbb{E}[X] = \mu$ och $\mathbb{V}ar(X) = \sigma^2 < \infty$. Kom ihåg att:
$$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k.$$
Vi vet att $\mathbb{E}[\bar{X}] = \mu$ och $\mathbb{V}ar(\bar{X}) = \frac{\sigma^2}{n}$, så enligt propositionen är $\bar{X}$ en konsistent skattning av μ. $\square$

Man kan se konsistens som ett mycket basalt krav och som i sig inte räcker för att säga att en skattning är bra.

DEFINITION: Om $\hat{\theta}$ och $\tilde{\theta}$ är två VVR skattningar av θ och
$$\mathbb{V}ar(\hat{\theta}) < \mathbb{V}ar(\tilde{\theta})$$
för alla θ, säger man att $\hat{\theta}$ är mer effektiv än $\tilde{\theta}$

SKATTNING AV VARIANS: Låt $X_1, \dots, X_n$ vara stickprov från en X med $\mathbb{E}[X] = \mu$ och $\mathbb{V}ar(X) = \sigma^2 < \infty$. Man brukar skatta σ^2 med:

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2$$

PROPOSITION: S^2 är vvr. Om $\mathbb{E}[X^4] < \infty$ gäller också att S^2 är konsistent.

DELBEVIS: Det är lätt att se att $S^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}^2\right)$. Det gäller att:

$$\mathbb{E}[\bar{X}^2] = \mathbb{V}ar(\bar{X}) + \mathbb{E}[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$$

Vi har också att $\mathbb{E}[X_k^2] = \mathbb{V}ar(X_k) + \mathbb{E}[X_k^2] = \sigma^2 + \mu^2$.

Detta ger:

$$\mathbb{E}[S^2] = \frac{1}{n-1}\left[n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right] = \sigma^2$$

$\square$

KONFIDENSINTERVALL

Låt $X_1, \dots, X_n$ vara ett stickprov på en stokastisk variabel X, vars fördelning beror på en okänd parameter θ. Om T_1 och T_2 är två funktioner av $X_1, \dots, X_n$ och

$$\mathbb{P}(T_1 \le \theta \le T_2) = q$$

så kallas $[T_1, T_2]$ för ett konfidensintervall för θ av konfidensnivå q och skriver

$$T_1 \le \theta \le T_2 \ (q)$$

OBS! Sannolikheten gäller innan T_1 och T_2 observerats. När vi observerar $T_1 = t_1$ och $T_2 = t_2$ och skriver $t_1 \le \theta \le t_2 (q)$ är detta ett påstående som antingen är sant eller falskt, men som innan data observerats, hade sannolikhet q att bli utfallen t_1 och t_2 som skulle göra det sant.

Låt $X \sim \text{likf}(0, \theta)$, θ okänd och $X_1, \dots, X_n$ ett stickprov på X.
Vi vet att $\mathbb{E}[\bar{X}] = \frac{\theta}{2}$, så

$$\hat{\theta} = 2\bar{X}$$

är ett väntevärdesriktig skattning av θ.

Alt: Skriv $M = \max\{X_1, \dots, X_n\}$. Skulle vara dumt att ett tecken θ som mindre än M. Låt oss göra en skattning baserad på M.

$$F_M(x) = \mathbb{P}(M \le x) = \mathbb{P}(X \le x)^n = \left(\frac{x}{\theta}\right)^n$$

$$f_M(x) = \frac{1}{\theta^n} n x^{n-1}, \quad x \in (0, \theta)$$

$$\mathbb{E}[M] = \int_0^\theta x \frac{1}{\theta^n} n x^{n-1} \, dx = \frac{n}{\theta^n} \int_0^\theta x^n \, dx = \frac{n}{n+1}\theta$$

Alltså $\tilde{\theta} = \frac{n+1}{n} M$ är en annan väntevärdesriktig skattning av θ.
Vilken är mest effektiv?

$$\mathbb{V}ar(\hat{\theta}) = \frac{4}{n}\mathbb{V}ar(X) = \frac{\theta^2}{3n}, \quad \mathbb{E}[M^2] = \frac{n}{\theta^n}\int_0^\theta x^{n+1} \, dx = \frac{n}{n+2}\theta^2$$

$$\mathbb{V}ar(M) = \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)\theta^2, \quad \mathbb{V}ar(\tilde{\theta}) = \left[\left(\frac{n+1}{n}\right)^2 \frac{n}{n+2} - 1\right]\theta^2 = \frac{\theta^2}{n(n+2)}$$

Vi har alltså att $\tilde{\theta}$ är mycket mer effektiv än $\hat{\theta}$.

KONFIDENSINTERVALL: Verkar bra att låta T_1 och T_2 vara funktioner av M. V: så ett symmetriskt konfidensintervall med konfidensintervall med konfidensgrad 95 %, dvs tar $\mathbb{P}(T_1 \geq \theta) = \mathbb{P}(T_2 \leq \theta) = 0{,}025$

$$\mathbb{P}(M \leq x) = \frac{x^n}{\theta^n} = 0{,}025$$

så: $x = 0{,}025^{\frac{1}{n}}\theta$, Alltså

$$0.025 = \mathbb{P}(M \leq 0{,}025^{\frac{1}{n}}\theta) = \mathbb{P}\left(\theta \geq \frac{M}{0.025^{\frac{1}{n}}}\right)$$

På samma sätt:

$$\mathbb{P}\left(\theta \leq \frac{M}{0{,}975^{\frac{1}{n}}}\right) = 0{,}025 \quad , \text{ Alltså: } \frac{M}{0{,}975^{\frac{1}{n}}} \leq \theta \leq \frac{M}{0.025^{\frac{1}{n}}}$$

I boken finns ett exempel med $n=10$ och $M = 8.69$. Vi får

$$8{,}71 \leq \theta \leq 12.57$$

med konfidensgrad 95 %.

FORELÄSNING 3/8 TMI

TRANSFORMERING AV STOKASTISK VARIABEL

Upprepat: Om vi har en stokastisk variabel med fördelning F_1. Finns en funktion g så att $g(X)$ har fördelning F_2

PROPOSITION: Om F_1 och F_2 är inverterbara och $X \sim F_1$ gäller att

$$Y = F_2^{-1}(F_1(X)) \sim F_2$$

BEVIS: $\mathbb{P}(Y \le y) = \mathbb{P}(F_1(X) \le F_2(y)) = \mathbb{P}(X \le F_1^{-1}(F_2(y))) = F_2(y)$

Speciellt fall: Om $U \sim \text{likt}(0,1)$ gäller $F^{-1}(U) \sim F$

$\square$

EXEMPEL: Använd en likt(0,1)-fördelad stokastisk variabel U till att simulera en $\exp(\lambda)$-fördelning. Det gäller att $F(x) = 1 - e^{-\lambda x}$, Alltså:

$$F^{-1}(u) = -\frac{\ln(1-u)}{\lambda}$$

Alltså är $-\frac{\ln(1-U)}{\lambda} \sim \exp(\lambda)$

$\square$

Minns att $X_1, \dots, X_n$ oberoende och $\exp(\lambda)$ medelvärde $\sum_{i=1}^{n} X_i \sim \Gamma(n, \lambda)$

Vi vet:

$$F(x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$$

Derivera ger

$$f(x) = \frac{1}{(n+1)!} \lambda^n x^{n-1} e^{-\lambda x}$$

Mer allmänt säger man för parametrarna $a, \lambda > 0$ att $X \sim \Gamma(a, \lambda)$ om

$$f(x) = \frac{1}{C} \lambda^a x^{a-1} e^{-\lambda x}$$

<u>χ^2-fördelning.</u>

Om $Z_1, \dots, Z_n$ oberoende och $N(0,1)$-fördelade säger vi att

$$X = \sum_{i=1}^{n} Z_i^2$$

är χ^2-fördelad med n frihetsgrader. Vi har

$$\mathbb{E}[X] = n, \quad \mathbb{V}\text{ar}(X) = 2n$$

Intressant fakrum

$$\mathbb{P}(Z_1^2 + Z_2^2 \le x) = \mathbb{P}((Z_1, Z_2) \in B(0, \sqrt{x})) = \frac{1}{2\pi} \iint_{B(0,\sqrt{x})} e^{-\frac{1}{2}(z_1^2+z_2^2)} dz_1 dz_2$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\sqrt{x}} r e^{-\frac{r^2}{2}} dr d\theta = \left[-e^{-\frac{1}{2}r^2}\right]_0^{\sqrt{x}} = 1 - e^{-\frac{1}{2}x}$$

Alltså: $Z_1^2 + Z_2^2 \sim \exp(\frac{1}{2})$, ty $X_i^2 = \exp(\frac{1}{2})$. Det visar att om n är jämnt gäller

$$\sum_{i=1}^{n} Z_i^2 \sim \Gamma(\tfrac{n}{2}, \tfrac{1}{2})$$

Detta gäller även om n är udda $\Rightarrow$

$$X_n^2 = F(\tfrac{n}{2}, \tfrac{1}{2})$$

Låt $\mathbb{U} = (U_1, \dots, U_n)$ vara oberoende om $N(0,1)$. Låt M vara en $n \times n$ ortogonalmatris. Låt $\mathbb{X} = M\mathbb{U}$. Då är $X_1, \dots, X_n$ oberoende om $N(0,1)$

Låt $M = [c_1, \dots c_n]$, där $c_1 = [\frac{1}{\sqrt{n}} \dots \frac{1}{\sqrt{n}}]^T = \frac{1}{\sqrt{n}} \mathbb{1}$

Då är $\mathbb{X} = \sum_{k=1}^{n} U_k c_k$

U_1 för $\bar{X} = \frac{1}{n} \mathbb{1}^T \mathbb{X} = \frac{1}{\sqrt{n}} U_1$ eftersom $c_2 \dots c_n$ ortogonala mot $\mathbb{1}$

Alltså

$$\sum_{k=1}^{n} (X_k - \bar{X})^2 = (\mathbb{X} - \bar{X}\mathbb{1})^T (\mathbb{X} - \bar{X}\mathbb{1}) = \left(\mathbb{X} - \frac{1}{\sqrt{n}} U_1 \mathbb{1}\right)^T \left(\mathbb{X} - \frac{1}{\sqrt{n}} U_1 \mathbb{1}\right)$$

$$= \left(\sum_{k=2}^{n} U_k c_k\right)^T \left(\sum_{k=2}^{n} U_k c_k\right) = \sum_{k=2}^{n} U_k^2$$

Det följer att $\sum_{k=1}^{n} (X_k - \bar{X})^2 \sim \chi^2_{n-1}$ och är oberoende av $\bar{X}$, dvs $(n-1)s^2 \sim \chi^2_{n-1}$ och oberoende av $\bar{X}$. Med X_k omvandlade $N(\mu, \sigma^2)$, standardisera och få

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

KONFIDENSINTERVALL FÖR σ^2

Låt $X_1, \dots, X_n$ vara sp p: $N(\mu, \sigma^2)$. Det följer att

$$F^{-1}_{\chi^2_{n-1}}\left(\frac{\alpha}{2}\right) \le \frac{(n-1)s^2}{\sigma^2} \le F^{-1}_{\chi^2_{n-1}}\left(1 - \frac{\alpha}{2}\right)$$

med sannolikhet $1-\alpha$. Vi får

$$\frac{(n-1)s^2}{F^{-1}_{\chi^2_{n-1}}\left(1 - \frac{\alpha}{2}\right)} \le \sigma^2 \le \frac{(n-1)s^2}{F^{-1}_{\chi^2_{n-1}}\left(\frac{\alpha}{2}\right)}$$

med konfidensgrad $1-\alpha$. Ett vänstra begränsat konfidensintervall ges av.

$$\sigma^2 \le \frac{(n-1)s^2}{F^{-1}_{\chi^2_{n-1}}(\alpha)}$$

FELINTENSITET

Låt T vara en positiv stokastisk variabel. Se som en livstid. Skriv $G(t) = \mathbb{P}(T > t)$. Givet överlevnad till tid t, vad är risken för livet allaster fram vid?

$$\mathbb{P}(T \in (t, t+\Delta t) \mid T > t) = \frac{\mathbb{P}(T \in (t, t+\Delta t))}{\mathbb{P}(T > t)} \approx \frac{f(t)\Delta t}{G(t)}$$

DEFINITION: Felintensitetsfunktionen för T ges av

$$r(t) = \frac{f(t)}{G(t)}$$

Oftare läsbara att specificera om kravet på f eller G

OBS att $r(t) = -\frac{d}{dt} \ln(G(t))$, alltså $\ln G(t) = -\int_0^t r(s)\,ds$

$\Rightarrow G(t) = e^{-\int_0^t r(s)\,ds}$

EXEMPEL: $T \sim \exp(\lambda) \iff r(t) = \lambda$

EXEMPEL: $T \sim \text{likt}(0,1) \Longrightarrow r(t) = \dfrac{f(t)}{G(t)} = \dfrac{1}{1-t}$, $t \in (0,1)$

kraftigt åldrande.

EXEMPEL: Låt $G(t) = \dfrac{1}{(1+t)^2}$, $t \geq 0$. En enkelsidig tungsumsad fördelning.

$$r(t) = \frac{f(t)}{G(t)} = \frac{\frac{2}{1+t^3}}{\frac{1}{1+t^2}} = \frac{2}{1+t}$$

Föryngring.

EXEMPEL: Låt $r(t) = t$, $t \geq 0$. Då blir

$$G(t) = e^{-\int_0^t s\,ds} = e^{-\frac{1}{2}t^2}$$

Gränsen likt normalfördelat begränsad till att vara större än 0.

FÖRELÄSNING 13

KONFIDENSINTERVALL FÖR μ I NORMALFÖRDELNING.

Vi har $X_1, \dots, X_n$ sp på $N(\mu, \sigma^2)$. Börjar med antan att σ^2 är känd. Låt oss koncentrera oss på $\bar{X}$, (man kan visa att detta är den mest effektiva av alla skattningar av μ). Det gäller att

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \text{ dvs}$$

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Skriv $z_\alpha = \Phi^{-1}(1-\alpha)$. Det gäller att

$$\mathbb{P}\left(-z_{\alpha/2} \le \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le z_{\alpha/2}\right) = 1 - \alpha$$

Lös ut μ och få det symmetriska konfidensintervallet

$$\mu \in \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad (1-\alpha)$$

Uppåt begränsat konfidensintervall: Utnyttja att

$$\mathbb{P}\left(-z_\alpha \le \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

Lös ut μ och få

$$\mu \le \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Om σ^2 är okänd använd S^2 istället för σ^2. Det gäller att

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Härifrån nu rakt av

$$\mu \in \bar{X} \pm F^{-1}_{t_{n-1}}\left(\frac{1-\alpha}{2}\right) \frac{S}{\sqrt{n}}$$

EXEMPEL: I ett elverk mäts spänningen en dag om dagen i en vecka. Resultat:

230.7, 226.9, 228.8, 232.2, 227.3, 227.0, 229.1

Rimligt anta att data är $N(\mu, \sigma^2)$. Vi har $\bar{X} = 228.9$ och $s^2 = 2.014^2$, så $\frac{s}{\sqrt{n}} = \frac{2.014}{\sqrt{7}} = 0.761$. För t_6-fördelning gäller att $F^{-1}_{t_6}(0.975) = 2.46$, $F^{-1}_{t_6}(0.995) = 3.71$, $F^{-1}_{t_6}(0.9) = 1.44$.

Alltså:

$$\mu = 228.9 \pm 2.46 \cdot 0.761 = 228.9 \pm 1.9 \qquad (95\%)$$

$$\mu = 228.9 \pm 3.71 \cdot 0.761 = 229.9 \pm 2.8 \qquad (99\%)$$

$$\mu \le 228.9 + 1.44 \cdot 0.761 = 230.0 \qquad (90\%)$$

Om vi nu skulle veta att $\sigma^2 = 2^2$ och vill ha ett 95% symmetriskt konfidensintervall utnyttjar vi att $z_{0.975} = 1.96$ och får

$$\mu = 228.9 \pm 1.96 \frac{2}{\sqrt{7}} = 228.9 \pm 1.5 \qquad (95\%)$$

Det gäller att $F^{-1}_{t_n}(\alpha) \longrightarrow \Phi^{-1}(\alpha)$ då $n \longrightarrow \infty$. Tumregel: $n \ge 100 \longrightarrow t_n \approx N(0,1)$. Vi har t.ex. att $F^{-1}_{t_{100}}(0.975) = 1.98$ och $\Phi^{-1}(0.975) = 1.96$.

Konfidensintervall för σ^2: Vi hoppar över. Men behövs:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

PREDIKTION (σ^2 känd)

Vad kommer spänningen i utfallet vara vista dag? Vi har $X_1, \dots, X_n$ sp p: $N(\mu, \sigma^2)$ och n mätvärden och vi kan säga om en ny observation Y. Vi har:

$$Y - \bar{X} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right)$$

ty $\mathbb{V}ar(Y - \bar{X}) = \mathbb{V}ar\, Y + \mathbb{V}ar\, \bar{X} = \sigma^2 + \frac{\sigma^2}{n}$, alltså

$$\mathbb{P}\left(-z_{\alpha/2} \le \frac{Y - \bar{X}}{\sigma\sqrt{1 + \frac{1}{n}}} \le z_{\alpha/2}\right) = 1 - \alpha$$

Vi får ett prediktionsintervall:

$$Y = \bar{X} \pm z_{\alpha/2}\, \sigma \sqrt{1 + \frac{1}{n}} \qquad (1-\alpha)$$

I vårt exempel med $\sigma^2 = 2^2$ och $n = 7$

$$Y = 228.9 \pm 1.96 \cdot 2 \sqrt{\frac{8}{7}} = 228.9 \pm 4.2 \qquad (95\%)$$

Reflektioner:

- Normalfördelningsantagandet alltid lite fel, ffa i svansarna
- Ju större n, desto mer normal blir $\bar{X}$, och därmed kan vi klara mer extrema konfidensgrader.
- Om beroende mellan observationer kan $\bar{X}$ fortfarande vara normal, men konfidensgraderna blir väldigt fel, eftersom variansen hos $\bar{X}$ inte blir den samma.
- Det finns tester (goodness of fit) och normalfördelningsplot för att kolla normalfördelning.

KONFIDENSINTERVALL FÖR p

I en opinionsstudie med 10 000 deltagare säger sig 6.4 procent stödja parti A. Ge ett konfidensintervall för stödet för A i hela populationen.

Låt p vara andelen väljare som stödjer A. $\sum_n X_n \sim Bin(n, p)$

Enligt CGS är

$$\bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right), \text{ dvs}$$

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Så om $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, gäller:

$$\mathbb{P}\left(-z_{\alpha/2} \le \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right) \approx 1 - \alpha$$

Man kan lösa ut p ur olikheten och bilda konfidensintervall för p. Lite besvärligt. Man kan utnyttja att $\frac{\bar{X}}{p}$ är approximativt 1 med stor sannolikhet och få att.

$$\frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \approx N(0,1)$$

Detta är De-Moivre-Laplace gränsvärdessats.

Det är lätt att räkna ut p:

$$p = \bar{X} \pm z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \qquad (1-\alpha)$$

I vårt exempel under vi $\bar{X} = 0.084$. För 95% observerar vi att $z_{0.025} = 1.96$. Vi får

$$p = 0.084 \pm 1.96\sqrt{\frac{0.084 \cdot 0.916}{10\,000}} = 0.084 \pm 0.0054$$

dvs 8.4% ± 0.54%

<u>Reflektioner:</u>

- Ingen problem i svarsrutan!
- Mycket bra skattning av p bland
  - de som svarar
  - de som talar sanning

Svarsbortfall, tegner ofta falsvar för stora problem i opinions undersökningar.

MAXIMUM LIKELIHOOD (ML)

Punktskatta θ genom att maximera sannolikheten att få de data man fick.

Mer precist: Om $X_1, \dots, X_n$ är ett <u>sp</u> på en <u>sv</u> med en frekvensfunktion $p(x) = p_\theta(x)$ som beror av θ, om vi observerar $X_1 = x_1, \dots, X_n = x_n$ finns det θ som max

$$\prod_{k=1}^{n} p_\theta(x_k) =: L(\theta; x_1, \dots, x_n)$$

Detta maximum, betecknat $\hat{\theta}$, kallas för ML-skattningen av θ.

EXEMPEL: Låt $X \sim Bin(n, \theta)$. Finn ML-skattningen av θ.

$$L(\theta; x) = p_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Om $x = 0$ är L avtagande, så max fås då $\theta = 0$.

Om $x = n$ är L växande, så max fås då $\theta = 1$

Annars, låt:

$$L'(\theta; x) \propto x\theta^{x-1}(1-\theta)^{n-x} - (n-x)\theta^x(1-\theta)^{n-1-x} = 0$$

Detta ger $x(1-\theta) - (n-x)\theta = 0$

Lösningen är $\theta = \frac{x}{n}$. Sammanfattat: $\hat{\theta} = \frac{x}{n}$

$\square$

OBS: Som underlättar: $L(\theta; x_1, \dots, x_n)$ maximeras precis då $\ell(\theta; x_1, \dots, x_n) := \ln L(\theta; x_1, \dots, x_n)$ maximeras

EXEMPEL: Låt $X \sim Geo(\theta)$. ML-skatta θ. Vi har $L(\theta; x) = \theta(1-\theta)^{x-1}$, så

$$\ell(\theta; x) = \ln\theta + (x-1)\ln(1-\theta)$$

Vi får

$$\ell'(\theta; x) = \frac{1}{\theta} - \frac{x-1}{1-\theta} = 0$$

Vilket ger $\theta = \frac{1}{x}$. ML-skattningen av θ är alltså: $\hat{\theta} = \frac{1}{\bar{x}}$.

EXEMPEL: Låt $X_1, \dots, X_n$ vara ett sp på $Po(\theta)$. Vi har

$$L(\theta) = \prod_k e^{-\theta} \frac{\theta^{x_k}}{x_k!}. \text{ Alltså:}$$

$$\ell(\theta) = -n\theta + \left[\sum_k x_k\right] \ln\theta - \sum_k \ln(x_k!)$$

Alltså: $\ell'(\theta) = -n + \frac{\sum_k x_k}{\theta} = 0$

Detta ger $\hat{\theta} = \frac{1}{n}\sum_k x_k = \bar{x}$

För kontinuerliga fördelningar maximerar vi istället täthaten av stickprovet som funktion av θ.

EXEMPEL: Låt $X_1, \dots, X_n$ vara SP på $N(\mu, \sigma^2)$. ML-skatta (μ, σ^2)

$$L(\mu, \sigma) = \prod_k \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x_k-\mu)^2}{\sigma^2}} = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_k \frac{(x_k-\mu)^2}{\sigma^2}}$$

Vi har att

$$l(\mu, \sigma) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2}\frac{\sum_k (x_k-\mu)^2}{\sigma^2}$$

$$\Rightarrow \quad \frac{\partial l}{\partial \mu} = \frac{\sum_k (x_k - \mu)}{\sigma^2}, \quad \text{som är } 0 \text{ då } \mu = \bar{x}$$

$$\Rightarrow \hat{\mu} = \bar{x}$$

Vidare är $\frac{\partial l}{\partial \sigma}(\hat{\mu}, \sigma) = -\frac{n}{\sigma} + \frac{\sum_k (x_k - \hat{\mu})^2}{\sigma^3}$

Detta är 0 då $\sigma^2 = \frac{1}{n}\sum_k (x_k - \hat{\mu})^2$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2$$

HYPOTESPRÖVNING / TESTER

Är slanten rättvis? Den neutrala hypotesen är att $p = \frac{1}{2}$, men vi misstänker att $p \neq \frac{1}{2}$. Vi vill samla data för att avgöra.

Filosofi: Utgå från att den neutrala hypotesen är sann. Om data utifrån detta antagande får utfall som är osannolika och samtidigt skulle varit mer sannolikt om den alternativa hypotesen är sann, drar vi slutsatsen att den senare är den sanna. (All analys bör göras innan data samlas in)
(E.g. crooked casinos, men oftast naturligt val)

I vårt slantsinglingsexempel: Vi vill testa nollhypotesen $H_0: p = \frac{1}{2}$ mot alternativhypotesen $H_A: p \neq \frac{1}{2}$

Fixera ett litet tal $\alpha > 0$ (ofta $\alpha = 0.05$ eller $\alpha = 0.01$). Betecknar data med X och bestäm en mängd B_0 s.a.

$$\mathbb{P}_{H_0}(X \notin B_0) = \alpha$$

och $X \notin B_0$ talar till förmån för H_A över H_0. Om det sådan visar sig att $X \notin B_0$ förkastar vi H_0 till förmån för H_A på signifikansnivån α. Om $X \in B_0$ accepterar vi H_0.

I exemplet: Dutt n slantsinglingar. Låt X vara antal klara. Om X avviker mycket från $\frac{n}{2}$ talar detta till förmån för H_A över H_0. Enligt CGS gäller under H_0 (dvs om H_0 är sann)

$$\frac{X - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \approx N(0,1)$$

Med $z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$, gäller alltså:

$$\mathbb{P}_{H_0}\left(\frac{X - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \in \pm z_{\alpha/2}\right) \approx \alpha$$

Det vill säga

$$\mathbb{P}_{H_0}\left(X \in \frac{n}{2} \pm z_{\alpha/2}\sqrt{\frac{n}{4}} \right) \approx \alpha$$

och vi förkastar H_0 till förmån för H_a om $X \notin \frac{n}{2} \pm \sqrt{\frac{n}{4}}$

Exempelvis med $n = 100$ och $\alpha = 0.05$, blir $z_{0.025}\sqrt{\frac{n}{4}} = 1.96\sqrt{25} \approx 10$ och vi förkastar om X avviker med minst 10 från 50

Med $n = 10000$ och $\alpha = 0.01$ blir $z_{0.005}\sqrt{\frac{n}{4}} = 2.58\sqrt{2500} \approx 129$ och vi förkastar om X avviker med minst 129 från 5000

EXEMPEL: Låt $X_1, \dots, X_n$ vara ett sp på $N(\mu, \sigma^2)$. Testa $H_0: \mu = \mu_0$ mot $H_a: \mu \neq \mu_0$

Om H_0 är sann var $\bar{X}$ ligga nära μ_0, och stora avvikelser från det talar för H_a.

Under H_0: $\dfrac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$

Alltså: $\mathbb{P}_{H_0}\left(\bar{X} \in \mu_0 \pm F^{-1}_{t_{n-1}}\left(1 - \frac{\alpha}{2}\right)\frac{s}{\sqrt{n}} \right) = 1 - \alpha$

och vi förkastar H_0 till förmån av H_a på signifikansnivå α om $\bar{X} \notin \mu_0 \pm F^{-1}_{t_{n-1}}\left(1 - \frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}$

KONKRET FALL: Marmeladburkar som innehåller 400 g. Vi misstänker att det inte stämmer. Vi väljer 10 burkar på måfå. Rimligt att anta att burkars vikt är $N(\mu, \sigma^2)$-fördelade.

Vi vill testa $H_0: \mu = 400$ mot $H_a: \mu \neq 400$. Vi förkastar H_0 till förmån för H_a på signifikansnivå 5% om:

$$\bar{X} \notin 400 \pm F^{-1}_{t_9}(0{,}975)\frac{s}{\sqrt{10}}$$

Vi slår upp att $F^{-1}_{t_9}(0{,}975) \approx 2.26$ vilket ger att vi förkastar om

$$|\bar{X} - 400| \geq 0{,}716$$

OK, men vi kunder är mer intresserade av $H_0: \mu \geq 400$ mot $H_a: \mu < 400$. Bara verken blivit mindre än 400 som talar för H_a. Ger ett rimligt test. Börjar med H_0; testa $H_0: \mu = \mu_0$ mot $H_a: \mu < \mu_0$

Utnyttja att

$$\mathbb{P}_{H_0}\left(\bar{X} \leq \mu_0 - F^{-1}_{t_{n-1}}(1-\alpha)\frac{s}{\sqrt{n}} \right) = \alpha$$

Förkasta H_0 på signifikansnivå α om $\bar{X} < \mu_0 - F^{-1}_{t_{n-1}}(1-\alpha)\frac{s}{\sqrt{n}}$

För marmeladburkarna förkastar vi om $H_0: \mu = 400$ mot $H_a: \mu < 400$ på signifikansnivå 5% om

$$\bar{X} < 400 - F^{-1}_{t_9}(0{,}95)\frac{s}{\sqrt{10}} \approx 400 - 0.586.$$

FÖRELÄSNING 15

KORRESPONDENSEN MELLAN TEST OCH KONFIDENSINTERVALL

Antag θ är endimensionell. Skapa ett konfidensintervall

$$T_1 \le \theta \le T_2$$

med konfidens~~intervall~~grad $1-\alpha$. Detta betyder att

$$\mathbb{P}(T_1 \le \theta \le T_2) = 1-\alpha$$

Vill testa $H_0: \theta = \theta_0$ mot någon alternativ hypotes H_A.
Förkasta H_0 om $\theta_0 \notin [T_1, T_2]$. Detta ger test på signifikansnivå α ty.

$$\mathbb{P}_{H_0}(\theta_0 \notin [T_1, T_2]) = \alpha$$

Konfidensintervallet väljs så att $\theta_0 \notin [T_1, T_2]$ talar till förmån för H_A framför H_0.

EXEMPEL: $X_1, \dots, X_n$ op på $N(\mu, \sigma^2)$. Skapa symmetriskt konfidensintervall för μ.

$$\mu < \bar{X} \pm F^{-1}_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}$$

med konfidens~~intervall~~grad $1-\alpha$. Enligt recept ovan skall vi förkasta $H_0: \mu = \mu_0$ till förmån för $H_A: \mu \neq \mu_0$ om

$$\mu_0 \notin \bar{X} \pm F^{-1}_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}$$

Exakt det vi kom ~~fram~~ till tidut

$\square$

ATT JÄMFÖRA TVÅ STICKPROV

Testa om en medicin sänker blodsockret. Gör en dubbelblind studie: samla ett stickprov $X_1, \dots, X_n$ av blodsockervärden från försökspersonen som tagit medicinen och ett stickprov $Y_1, \dots, Y_m$ från personer som tagit placebo.

Antag att $X_k \sim N(\mu_X, \sigma_X^2)$ och $Y_k \sim N(\mu_Y, \sigma_Y^2)$
Vill testa $H_0: \mu_X = \mu_Y$ mot $H_A: \mu_X \neq \mu_Y$ (eller $H_0: \mu_X < \mu_Y$)
Gör konfidensintervall för $\mu_X - \mu_Y$. Baserat på $\bar{X}, \bar{Y}$.

FALLET σ_X^2 OCH σ_Y^2 KÄNDA

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{1}{n}\sigma_X^2 + \frac{1}{m}\sigma_Y^2\right)$$

ger konfidensintervall

$$\mu_X - \mu_Y < \bar{X} - \bar{Y} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n}\sigma_X^2 + \frac{1}{m}\sigma_Y^2}$$

FALLET σ_X^2 OCH σ_Y^2 OKÄNDA

Antag att σ_X^2 och σ_Y^2 är lika: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$

Skatta σ^2 med den poolade stickprovsvariansen:

$$s_p^2 = \frac{\sum_{k=1}^{n}(X_k - \bar{X})^2 + \sum_{k=1}^{m}(Y_k - \bar{Y})^2}{n+m-2} = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$$

Det visar sig att

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

ger symmetriskt konfidensintervall.

$$\mu_X - \mu_Y \in \bar{X} - \bar{Y} \pm F^{-1}_{t_{n+m-2}}\left(1 - \frac{\alpha}{2}\right) s_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

med konfidensgrad $1-\alpha$. Nedåt begränsat konfidensintervall:

$$\mu_X - \mu_Y \geq \bar{X} - \bar{Y} - F^{-1}_{t_{n+m-2}}(1-\alpha)\, s_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

EXEMPEL: Böjhållfasthet hos tegel vid olika bränntemperatur:

700 C°: 147 140 121 138 120 139

800 C°: 193 227 201 212 207

Lät μ_X och μ_Y förväntad hållfasthet vid 700 resp 800 grader.

Gör 99.9 procent symmetriskt konfidensintervall för $\mu_Y - \mu_X$

Vi antar att data är normalfördelning med samma varians

Vi har: $\bar{X} = 132.8$, $s_X^2 = 10.83^2$, $\bar{Y} = 209.0$, $s_Y^2 = 12.77^2$

$$s_p^2 = \frac{5s_X^2 + 4s_Y^2}{9} = 11.73^2$$

Konfidensintervallet blir nu:

$$\mu_Y - \mu_X \in \bar{Y} - \bar{X} \pm F^{-1}_{t_9}(0.9995)\, s_p\sqrt{\frac{1}{6} + \frac{1}{5}}$$

$$= 75.2 \pm 4.78 \cdot 11.73\sqrt{\frac{1}{6} + \frac{1}{5}}$$

$$= 75.2 \pm 34.0$$

TJUVKIKANDE

I alla datamängder finns slumpmässigt uppkomna mönster.

Sådana mönster kan aldrig bevisas av den datamängd som fann den ! De kan dock ge hypoteser som kan prövas med ny data.

Sensmoral: Bestäm alltid vilka tester som skall göras innan du tittar på data. Helst redan innan du samlar in data.

MULTIPELTESTNIVÅ

Antag att vi testar 100 sanna nollhypoteser på 5% signifikansnivå. I genomsnitt kommer vi att fem gånger felaktigt förkasta nollhypotesen. Det kan bli många fel i att statistikerliv.

Sensmoral: testa inte om du inte har goda skäl att tro att nollhypotesen är falsk.

FÖRELÄSNING TM2

SUMMOR AV OBEROENDE STOKASTISKA VARIABLER.

Om X och Y oberoende, vad är fördelning för $Z = X+Y$?
Antag att (X,Y) kontinuerlig.

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}(X+Y \le z) = \int_{-\infty}^{\infty} \mathbb{P}(x+Y \le z \mid X = x) f_X(x)\,dx$$
$$= \int_{-\infty}^{\infty} F_Y(z-x) f_X(x)\,dx$$

Under enkla villkor (som vi inte går in på) kan man derivera under integraltecknet och få:

$$f_Z(z) = \frac{d}{dz}\int_{-\infty}^{\infty} F_Y(z-x) f_X(x)\,dx = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x)\,dx$$

Om $X, Y > 0$ fås man

$$f_Z(z) = \int_0^z f_Y(z-x) f_X(x)\,dx$$

Om (X,Y) diskret får man

$$p_Z(z) = \sum_{x \in U_X} p_Y(z-x) p_X(x)$$

Vi har redan utnyttjat detta till att visa att X_1, X_2 oberoende och

$$X_1 \sim Po(\lambda_1),\ X_2 \sim Po(\lambda_2) \implies X_1 + X_2 \sim Po(\lambda_1+\lambda_2)$$

EXEMPEL: Enligt CGS måste det gälla att om $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ och X_1, X_2 oberoende medför att $X_1 + X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. Låt oss kolla i fallet $\mu_1 = \mu_2 = 0$ och $\sigma_1 = \sigma_2 = 1$.

$$f_{X_1+X_2}(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\,dt$$
$$= \frac{1}{2\pi}\int e^{-(t-\frac{1}{2}z)^2 - \frac{1}{4}z^2}\,dt$$
$$= \frac{1}{\sqrt{4\pi}} e^{-\frac{1}{4}z^2} \int \frac{1}{\pi} e^{-(t-\frac{1}{2}z)^2}\,dt$$

Under integraltecknet står tätheten för $N(\frac{z}{2}, \frac{1}{2})$, vi.

$$f_{X_1+X_2}(z) = \frac{1}{\sqrt{4\pi}} e^{-\frac{1}{4}z^2}$$

dvs täthet för $N(0,2)$.

$\square$

EXEMPEL: Låt X, Y vara oberoende och $\exp(\lambda)$-fördelade

$$f_{X+Y}(z) = \int_0^z \lambda e^{-\lambda(z-t)} \lambda e^{-\lambda t}\,dt = \lambda^2 e^{-\lambda z}\int_0^z dt = \lambda^2 z e^{-\lambda z}$$

Med andra ord: $X+Y \sim \Gamma(2,\lambda)$

MOMENTGENERERANDE FUNKTION.

För en sv ges den momentgenererande funktionen M_X av

$$M_X(t) = \mathbb{E}[e^{tX}], \quad t \in \mathbb{R}$$

Om X är kontinuerlig och positiv med täthet f, har man:

$$M_X(t) = \int_0^\infty e^{tx} f(x)\, dx = \mathcal{L}_f(-t)$$

Därför kan vi om (men inte visa) att om man vet M_X för alla t i en omgivning av 0, så vet man även f. Detta är sant och gäller för alla sv X.

PROPOSITION: Om X och Y är oberoende gäller

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

BEVIS: $M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}e^{tY}] \underset{\text{oberoende}}{=} \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}] = M_X(t)M_Y(t)$

Om vi får derivera innan väntevärdet (och det får vi) får vi:

$$M_X'(t) = \frac{d}{dt}\mathbb{E}[e^{tX}] = \mathbb{E}[Xe^{tX}]$$

Detta ger $M'_X(0) = \mathbb{E}[X]$. Vi får också

$$M_X''(t) = \frac{d^2}{dt^2}\mathbb{E}[e^{tX}] = \mathbb{E}[X^2e^{tX}]$$

Alltså $M_X''(0) = \mathbb{E}[X^2]$

Generellt: $M_X^{(k)}(0) = \mathbb{E}[X^k]$, $k = 0, 1, 2, \ldots$

EXEMPEL: Låt $X \sim Po(\lambda)$

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^\infty e^{tk}e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^\infty \frac{(\lambda e^t)^k}{k!} = e^{\lambda(e^t-1)}$$

Om X_1, X_2 oberoende och $Po(\lambda_1)$ respektive $Po(\lambda_2)$, har vi

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

Vi ser att $X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$

EXEMPEL: Låt $X \sim N(\mu, \sigma^2)$

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

Kvadratkompletterar exponenten och så ser HL ut:

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

Funktionen i integralen är täthet för $N(\mu+\sigma^2t, \sigma^2)$, så:

$$M_X(t) = e^{\mu t + \frac{\sigma^2t^2}{2}}$$

Om nu X_1, X_2 oberoende och $N(\mu_1, \sigma_1^2)$ respektive $N(\mu_2, \sigma_2^2)$, har vi:

$$M_{X_1+X_2}(t) = e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

dvs $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

$\square$

En generalisering av inverterbarheten: Antag att X om $X_1, X_2, \dots$ är kontinuerliga. Då har vi att om $M_{X_n}(t) \longrightarrow M_X(t)$ för alla t, så gäller att $\mathbb{P}(X_n \le x) \longrightarrow \mathbb{P}(X \le x)$ för alla x.

BEVISSKISS av CGS: Kom ihåg CGS: Om $X_1, X_2, \dots$ är iid och fördelade som en sv X med $\mathbb{E}[X] = \mu$ och $\mathbb{V}ar(X) = \sigma^2 < \infty$ gäller

$$\cancel{\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)} \quad \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) \longrightarrow \Phi(x)$$

Räcker att visa för $\mu = 0$ och $\sigma^2 = 1$. Taylorutveckla M_X kring 0 till ordning 2:

$$M_X(s) \approx M_X(0) + s M'_X(0) + \frac{1}{2} s^2 M''_X(0) = 1 + s\mu + \frac{1}{2} s^2 (\mu^2 + \sigma^2) = 1 + \frac{1}{2} s^2$$

Vill visa att $M_{S/\sqrt{n}}(t) \longrightarrow M_{N(0,1)}(t) = e^{\frac{t^2}{2}}$

Men $M_{S_n/\sqrt{n}}(t) = \mathbb{E}[e^{t S_n/\sqrt{n}}] = M_{S_n}\left(\frac{t}{\sqrt{n}}\right) = M_X\left(\frac{t}{\sqrt{n}}\right)^n \approx \left(1 + \frac{t^2}{2n}\right)^n \longrightarrow e^{t^2/2}$

$\square$

P-värde

Test av H_0 mot H_A. Testets p-värde, P är det minsta tal α sådant att H_0 kan förkastas till förmån för H_A på signifikansnivå α. Dvs P är det α där H_0 är precis på gränsen att förkastas.

Mer formellt och generellt: Låt $\mathcal{X}$ beteckna data. För varje α, finn lämpligt A_α så att $\mathbb{P}_{H_0}(\mathcal{X} \notin A_\alpha) = \alpha$. Testet förkastar H_0 på signifikansnivå α om $\mathcal{X} \notin A_\alpha$. Antag att A_α:na är valda så att $\alpha_1 \le \alpha_2 \Longleftrightarrow A_{\alpha_1} \supseteq A_{\alpha_2}$. Då är

$$P = \min\{\alpha : \mathcal{X} \in A_\alpha\}$$

Om θ är en endimensionell parameter, gör konfidensintervall

$$T_1(\mathcal{X}, \alpha) \le \theta \le T_2(\mathcal{X}, \alpha) \qquad (1-\alpha)$$

Man förkastar $H_0: \theta = \theta_0$ på signifikansnivå α om $\theta_0 \notin [T_1(\mathcal{X}, \alpha), T_2(\mathcal{X}, \alpha)]$. Då är

$$P = \min\{\alpha : \theta_0 \notin [T_1(\mathcal{X}, \alpha), T_2(\mathcal{X}, \alpha)]\}$$

Typiskt är intervallgränserna kontinuerliga i α och så blir P lika med det α som ger $\theta_0 = T_1(\mathcal{X}, \alpha)$ eller $\theta_0 = T_2(\mathcal{X}, \alpha)$

EXEMPEL: Låt $X_1, \dots, X_n$ vara ett sp på $N(\mu, \sigma^2)$. Testa $\mu = 0$ mot $\mu \neq 0$. Antag att $n = 10$, $\bar{X} = 1.7$, $s^2 = 1.5$.
Vad är testets p-värde? Konfidensintervall

$$\mu \in \bar{X} \pm F^{-1}_{t_9}\left(1 - \frac{\alpha}{2}\right) \frac{s}{\sqrt{n}}$$

Punkten 0 hamnar på gränsen om $\bar{X} = F^{-1}_{t_9}\left(1 - \frac{\alpha}{2}\right) \frac{s}{\sqrt{n}}$, dvs

$$F^{-1}_{t_9}\left(1 - \frac{\alpha}{2}\right) = \frac{\bar{X}\sqrt{n}}{s} = \frac{1.7\sqrt{10}}{\sqrt{1.5}} = 4.39.$$

Alltså $1 - \frac{\alpha}{2} = F_{t_9}(4.39) = 1 - 0.0004$, så $\alpha = 0.0018$

Vi har $P \approx 0.0018$

FÖRELÄSNING TM3

STYRKA

Signifikansnivån av ett test sannolikheten att felaktigt förkasta H_0. Man vill heller inte felaktigt acceptera H_0. Styrkan är sannolikheten att förkasta nollhypotesen som funktion av vad sanningen verkligen är. Konkret:

Testa $H_0: \theta = \theta_0$ mot $H_A: \theta \neq \theta_0$. Styrkan är en funktion av vad θ verkligen är

$$g(\theta_1) = \mathbb{P}_{\theta_1}(H_0 \text{ förkastas})$$

Det gäller testets att $g(\theta_0) = \alpha$.

EXEMPEL: Slantsingling. Test av $H_0: p = \frac{1}{2}$ mot $H_A: p \neq \frac{1}{2}$
$n = 100$. Låt X vara kvast. Vi kom fram till att förkasta på 5% signifikansnivå om $|X - 50| \geq 10$. Vad är $g(0.7)$? Om $p = 0.7$ gäller

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{X - 70}{\sqrt{21}} \approx N(0,1)$$

Alltså $\mathbb{P}_{0.7}(|X - 10| \geq 50) = \mathbb{P}_{0.7}(X \geq 59.5) + \mathbb{P}_{0.7}(X \leq 40.5)$

Den andra termen är försvinnande och

$$\mathbb{P}_{0.7}(X \geq 59.5) = \mathbb{P}_{0.7}\left(\frac{X-70}{\sqrt{21}} \geq \frac{59.5 - 70}{\sqrt{21}}\right) \approx 1 - \Phi(-2.291)$$

$$\approx 0.989$$

Alltså: $g(0.7) \approx 0.989$

Vad är $g(0.6)$? Med samma beräkningar blir

$$g(0.6) \approx 1 - \Phi\left(\frac{59.5 - 60}{\sqrt{24}}\right) \approx 1 - \Phi(-0.102) \approx 0.541$$

Stickprov $X_1, \dots, X_n$ på $N(\mu, \sigma^2)$ med σ^2 känd. Testa $H_0: \mu = 0$ mot $H_A: \mu > 0$ på 1% signifikansnivå. Hur stort behöver n vara för att få $g(1) = 0.9$ (Dvs sannolikheten att förkasta om vår 90% om det verkliga värdet på μ är 1). Testet förkastas då om

$$\frac{\bar{X}}{\frac{\sigma}{\sqrt{n}}} \geq \Phi^{-1}(0.99) \approx 2.33$$

Vi skall alltså beräkna $\mathbb{P}_1\left(\frac{\bar{X}}{\frac{\sigma}{\sqrt{n}}} \geq 2.33\right)$

Om $\mu = 1$ är $\frac{\bar{X} - 1}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ och

$$\mathbb{P}_1\left(\frac{\bar{X}}{\frac{\sigma}{\sqrt{n}}} \geq 2.33\right) = \mathbb{P}_1\left(\frac{\bar{X} - 1}{\frac{\sigma}{\sqrt{n}}} \geq 2.33 - \frac{1}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi\left(\frac{\sqrt{n}}{\sigma} - 2.33\right)$$

Eftersom $\Phi^{-1}(0.9) \approx 1.28$ är villkoret ≥ 0.9 så

$$\frac{\sqrt{n}}{\sigma} - 2.33 \geq 1.28$$

dvs så $n \geq (3.61\sigma)^2$

BAYESIANSK STATISTIK

Istället för att se okända parametrar som fixa tal, se dem som stokastiska variabler. Inte naturligt, men vad sägs vi tro att en parameter har för fördelning? Med mycket data spelar det inte alltid stor roll. Ibland naturligt, ibland inte.

EXEMPEL: Ett mynt singlas 10 ggr. Vi vet att myntet antingen ger klave m.s. $\frac{2}{3}$ eller med m.s. $\frac{1}{3}$. Om vi får klave 6 ggr, vad säger vi tro om sannolikhet att myntet ger klave?

Data $X \sim Bin(n, \theta)$. Vi fick $X = 6$. Är θ lika med $\frac{1}{3}$ eller $\frac{2}{3}$? Vi kan tänka oss att θ är en av dem $\mathbb{P}(\theta = \frac{1}{3}) = \mathbb{P}(\theta = \frac{2}{3}) = \frac{1}{2}$

Istället:

$$\mathbb{P}(\theta = \tfrac{2}{3} \mid X = 6) = \frac{\mathbb{P}(X = 6 \mid \theta = \frac{2}{3})\mathbb{P}(\theta = \frac{2}{3})}{\mathbb{P}(X = 6 \mid \theta = \frac{2}{3})\mathbb{P}(\theta = \frac{2}{3}) + \mathbb{P}(X = 6 \mid \theta = \frac{1}{3})\mathbb{P}(\theta = \frac{1}{3})}$$

$$= \frac{\binom{10}{6}\left(\frac{2}{3}\right)^6\left(\frac{1}{3}\right)^4 \cdot \left(\frac{1}{2}\right)}{\binom{10}{6}\left(\frac{2}{3}\right)^6\left(\frac{1}{3}\right)^4 \cdot \frac{1}{2} + \binom{10}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^4 \frac{1}{2}} = \frac{2^2}{2^2 + 1^2} = \frac{4}{5}$$

Den fördelning av parametern vi tror på innan experimentet (I exemplet $\mathbb{P}(\theta = \frac{2}{3}) = \frac{1}{2}$) kallas för θ:s prior och den betingade fördelningen givet data (i ex. $\mathbb{P}(\theta = \frac{2}{3} \mid X = 6) = \frac{4}{5}$) kallas för θ:s posterior.

(På svenska: a-priori-fördelning och a-posteriori-fördelning)

BAYES FORMEL FÖR TÄTHETER.

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y \mid x)\, f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y \mid t)\, f_X(t)\, dt} = C f(x,y) = C f_{Y|X}(y \mid x) f_X(x)$$

De sista likheterna visar att nämnaren inte har med x att göra.

I Bayesiansk statistik har vi alltid en parameter θ vars prior $f_\theta(t)$ är känd, liksom den betingade fördelningen för data givet θ: $f_{X|\theta}(x \mid t)$. Vi vill veta posterior $f_{\theta|X}(t \mid x)$

Att ge en objektiv prior är oftast omöjligt, så därför är Bayesiansk statistik ibland utesluten (tänk läkemedelsbolag som rapporterar till regulatoriska myndigheter). Däremot mycket vanligt inom AI.

En AI-algoritm har en uppfattning om världen i termer av sannolikheter. Den får data och uppdaterar sin uppfattning.

EXEMPEL: Om du vet att $X \sim \exp(\theta)$, att $0 < \theta < 1$ och att $X = x$, vad tror du om θ?

Man kan tänka sig en prior som är likformig:

$f_\theta(t) = 1, \quad 0 < t < 1$. Då får vi:

$$f_{\theta|X}(t \mid x) = C f_{X|\theta}(x \mid t) f_\theta(t) = C t e^{-tx}, \quad 0 < t < 1$$

OBS att detta sista ses som en funktion av t.

Vi känner igen den som täthet för $\Gamma(2, x)$-fördelning, men bara för $0 < t < 1$, så posterior är en trunkerad gammafördelning

EXEMPEL: Error Correction

En datasträm av 0/1:or kommer så att den startar på noll och sedan följs en etta av en etta m.s 0.8 och en nolla av en nolla m.s 0.8. Varje siffra avlöses fel m.s 0.1. Om man ser 0100, vad är den mest sannolika korrekta dataströmmen, <u>0100</u> eller <u>0000</u>

kalla obs för Y och det korrekta för X.

$$\mathbb{P}(X = 0100 \mid Y = 0100) \propto \mathbb{P}(Y = 0100 \mid X = 0100)\,\mathbb{P}(X = 0100)$$

$$= 0.9^4 \cdot 0.5 \cdot 0.2 \cdot 0.2 \cdot 0.8$$

$$\approx 0.0105$$

$$\mathbb{P}(X = 0000 \mid Y = 0100) \propto \mathbb{P}(Y = 0100 \mid X = 0000)\,\mathbb{P}(X = 0000)$$

$$= 0.9^3 \cdot 0.1 \cdot 0.5 \cdot 0.8^3$$

$$\approx 0.0187$$

FÖRELÄSNING (TM4)

EXEMPEL: En slant singlas 3 ggr. Det blev kronor alla tre gångerna. Vad säger vi tro om sannolikheten att f: klave θ? ML-skattningen blir 0. Verker orimligt. Bayesianskt: Antag att prior är likf, dvs $f_\theta(t) = 1$, $0 < t < 1$

$$f_{\theta|X}(t|0) = C f_{X|\theta}(0|t) f_\theta(t) = C(1-t)^3$$

Man ser snabbt att $C = 4$. Man kan skatta θ med väntevärdet i ~~prior~~ posterior, dvs

$$\mathbb{E}[\theta | X = 0] = 4\int_0^1 t(1-t)^3\,dt = \frac{1}{5}$$

Man kan känna igen $f_{\theta|X}(t|0)$ som en β-fördelning.

DEFINITION: En stokastisk variabel Y sägs vara β-fördelad med parametrarna $a > 0$ och $b > 0$ om

$$f_Y(y) = C y^{a-1}(1-y)^{b-1}, \quad 0 < y < 1.$$

Kort form $Y \sim \beta(a,b)$

OBS: $\beta(1,1)$ är likf $(0,1)$

Exemplet ovan är ett specialfall av följande observation:
Låt $\theta \sim \beta(a,b)$ och sedan $X_1, \dots, X_n$ oberoende och

$$\mathbb{P}(X_k = x) = \begin{cases} \theta, & x = 1 \\ 1-\theta, & x = 2 \end{cases}$$

Låt n_1 vara antal k sådana att $X_k = 1$ och $n_2 = n - n_1$.
Då gäller att

$$f_{\theta|X_1,\dots,X_n}(t|x_1,\dots,x_n) = C t^{n_1}(1-t)^{n_2} t^{a-1}(1-t)^{b-1}$$
$$= C t^{a+n_1-1}(1-t)^{b+n_2-1}$$

$\theta | X_1 = x_1, \dots, X_n = x_n$ är $\theta \sim \beta(a+n_1, b+n_2)$. Troligt lätträknat.

Om istället $X \sim Bin(n,\theta)$, dyker det bara upp en faktor $\binom{n}{n_1}$ i räkningarna ovan. Den bevarar inte på t, så vi får att
$\theta | X = n_1 \sim \beta(a+n_1, b+n_2)$ som nu.

Generalisering: Vektorn $Y = (Y_1, \dots, Y_{k-1})$ sägs vara Dirichlet fördelad med parametrar $b_1 > 0, b_2 > 0, \dots, b_k > 0$ om det för alla $y = (y_1, \dots, y_{k-1})$, sådana att $y_i \geq 0$ för alla i och $y_k = 1 - \sum_{i=1}^{k-1} y_i \geq 0$ gäller att

$$f(y) = C y_1^{b_1-1} y_2^{b_2-1} \cdots y_k^{b_k-1}$$

Betafördelningen är specialfallet $k = 2$. Vi skriver $Y \sim Dir(b_1, \dots, b_k)$.

Antag att $\theta = (\theta_1, \dots, \theta_{k-1}) \sim Dir(b_1, \dots, b_k)$ och låtom $X = (X_1, \dots, X_k)$ oberoende och sådana att $\mathbb{P}(X_i = j) = \theta_j$, $j = 1, \dots, k$.

~~Antag att $\theta = (\theta_1, \dots, \theta$~~

$$f_{\theta|X}(t_1, \dots, t_{k-1} \mid x) = C t_1^{n_1} \dots t_k^{n_k} \cdot t_1^{b_1 - 1} \dots t_k^{b_k - 1}$$
$$= t_1^{b_1 + n_1 - 1} \dots t_k^{b_k + n_k - 1}$$

Dvs $\theta | X = x \sim Dir(b_1 + n_1, \dots, b_k + n_k)$

Detta är ett exempel på ett <u>konjugerande prior</u>: Om data X har en viss fördelning, $f_{X|\theta}$, och det då visar sig att posterior $f_{\theta|X}$ är av samma typ som prior f_θ, säger man att f_θ är en konjugerande prior till $f_{X|\theta}$

Vi har alltså sett att betafördelningen är en konjugerande prior till binomialfördelningen och mer generellt att Dirichletfördelningen är konjugerande prior till multinomialfördelningen.

EXEMPEL: Låt $\theta \sim N(0,1)$ och sedan $X \sim N(\theta, 1)$

$$f_{\theta|X}(t|x) = C e^{-\frac{1}{2}(x-t)^2} \cdot e^{-\frac{1}{2}t^2} = C e^{-\left(\frac{t-x}{2}\right)^2}$$

Dvs $\theta | X = x \sim N\left(\frac{x}{2}, \frac{1}{2}\right)$

Vi ser att normalfördelningen är en konjugerande prior till normalfördelningen med given varians. $\square$

PROBLEM MED NÄMNAREN.

Latent Dirichlet allocation: Låt $\theta_i \sim Dir(a_1, \dots, a_k)$, $i = 1, \dots, D$.
Låt $\psi_i \sim Dir(b_1, \dots, b_k)$, $i = 1, \dots, K$. Alla oberoende. För alla $i \neq j$, $i = 1, \dots, D$, $j = 1, \dots, L$, låt $Z_{ij} \sim \theta_i$ och sedan $W_{ij} \sim \psi_{Z_j}$

Vad är $f_{Z,\psi,\theta|W}(\psi, \theta, z \mid w)$? Tillåtom: Bayes formel hjälper enkel men nämnaren blir

$$f_W(w) = \sum_z \int \dots \int f_{W|\psi,\theta,z}(w|\psi,\theta,z) f(\psi,\theta,z) \, d\psi \, d\theta$$

Detta är K^{DL} termer som alla är integraler!

GIBBS SAMPLING

Både $f_{Z|\psi,\theta,w}$, $f_{\psi|\theta,z,w}$ och $f_{\theta|\psi,z,w}$ är lätta. Då till sistnämnda är Dirichlet och given allt annat är alla Z_{ij} oberoende med lätt hittade fördelningar

Gör så här:

1. Starta med vilken värden som helst på ψ, θ, z.
2. Uppdatera Z enligt $f_{Z|\psi,\theta,w}$
3. Uppdatera ψ enligt $f_{\psi|\theta,z,w}$
4. Uppdatera θ enligt $f_{\theta|\psi,z,w}$
5. Upprepa 2-4

Detta konvergerar mot den vätta fördelningen

FÖRELÄSNING 16

LINJÄR REGRESSION

Mycket vanligt med linjära samband, t.ex

- Ström som funktion av spänning
- CIY som funktion av ålder.
- Rödförskjutning som funktion av avstånd till galax
- Längd hos dotter som funktion av mammas längd.

Sällan perfekta samband: data kommer inte riktigt på en linje. Modell:

$$Y_k = a + bx_k + \varepsilon_k, \quad k = 1, 2, 3, \dots, n$$

där $\varepsilon_1, \dots, \varepsilon_n$ är oberoende och $N(0, \sigma^2)$. a, b, σ okända parametrar. Talen $x_1, \dots, x_n$ kan ses som givna fixa tal

ML-skattning av a och b: Det gäller att $Y_k \sim N(a + bx_k, \sigma^2)$

$$f_{Y_k}(y_k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_k - a - bx_k)^2}$$

Därför blir
$$L(a, b, \sigma; y_1, \dots, y_n) = f_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$$
$$= \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{k=1}^n (y_k - a - bx_k)^2}$$

Att maximera m.a.p a och b är att minimera
$$\sum_{k=1}^n (y_k - a - bx_k)^2 =: \mathcal{L}(a, b),$$ alltså minsta kvadratmetoden

Vi skall alltså lösa
$$\frac{\partial \mathcal{L}}{\partial a} = -2\sum_{k=1}^n (y_k - a - bx_k) = 0$$
$$\frac{\partial \mathcal{L}}{\partial b} = -2\sum_{k=1}^n (y_k - a - bx_k)x_k = 0$$

Detta ger
$$\hat{b} = \frac{\sum_{k=1}^n (\bar{y}_k - \bar{y})(x_k - \bar{x})}{\sum_{k=1}^n (x_k - \bar{x})^2}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

Man brukar skriva
$$S_{xx} = \sum_{k=1}^n (x_k - \bar{x})^2, \quad S_{xy} = \sum_{k=1}^n (y_k - \bar{y})(x_k - \bar{x}), \quad S_{yy} = \sum_{k=1}^n (y_k - \bar{y})^2$$

Då blir
$$\hat{b} = \frac{S_{xy}}{S_{xx}}$$

Man kan beräkna att $\hat{b} \sim N\left(b, \frac{\sigma^2}{S_{xx}}\right)$, $\hat{a} \sim N\left(a, \frac{\sigma^2 \sum_{k=1}^n x_k^2}{n S_{xx}}\right)$

Om σ^2 känd, ger konfidensintervall/test för a resp. b med dessa.

T.ex $\dfrac{\hat{b} - b}{\frac{\sigma}{\sqrt{S_{xx}}}} \sim N(0,1)$, s.: $b \in \hat{b} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{S_{xx}}}$ $(1-\alpha)$

Om σ^2 är okänd, ersätt med s^2.

$$\frac{\hat{b} - b}{\frac{s}{\sqrt{S_{xx}}}} \sim t_{n-2} \quad \text{och} \quad b \in \hat{b} \pm F^{-1}_{t_{n-2}}\left(1 - \frac{\alpha}{2}\right) \frac{s}{\sqrt{S_{xx}}} \qquad (1-\alpha)$$

Analogt för a (men inte lika intressant)

Ok, s: vad är s^2?

$$s^2 = \frac{1}{n-2} \sum_{k=1}^{n} (y_k - \hat{a} - \hat{b} x_k)^2$$

Den är väntevärdesriktig för σ^2. Formel

$$s^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$

Formler som gör det lättare att räkna

$$S_{xy} = \sum (x_k - \bar{x})(y_k - \bar{y}) = \sum x_k y_k - \frac{1}{n}\left(\sum x_k\right)\left(\sum y_k\right)$$

D: för v. också

$$S_{xx} = \sum (x_k - \bar{x})^2 = \sum x_k^2 - \frac{1}{n}\left(\sum x_k\right)^2 \text{ , och analogt för } y.$$

EXEMPEL: För att observera hur en kvinnas längd bror av hennes mors längd mätte Sex mödra-dotter par. Vi tror på ett linjärt samband. Data

(170.1, 173.4), (161.3, 162.6), (177.1, 170.9)
(167.0, 162.8), (168.6, 171.2), (162.2, 165.5)

Vi får:

$$S_{xy} = \sum x_k y_k - \frac{1}{n}\left(\sum x_k\right)\left(\sum y_k\right) = 91.823$$

$$S_{xx} = \sum x_k^2 - \frac{1}{n}\left(\sum x_k\right)^2 = 166.628$$

Detta ger $\hat{b} = \dfrac{S_{xy}}{S_{xx}} = 0.548$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 75.74$$

Den skattade regressionslinjen blir alltså:

$$y = 75.74 + 0.548x$$

Vi får också:

$$s^2 = \frac{1}{4} \sum_{k=1}^{6} (y_k - \hat{a} - \hat{b} x_k)^2 = 14.34$$

Symmetriskt 95% konfidensintervall för b:

$$b \in \hat{b} \pm F^{-1}_{t_4}(0.975) \frac{s}{\sqrt{S_{xx}}} = 0.548 \pm 2.776 \frac{\sqrt{14.34}}{\sqrt{166.6}} = 0.55 \pm 0.81$$

Om vi gör ett test av $H_0: b = 0$ mot $H_a: b \neq 0$, kan vi alltså inte förkasta på 5% signifikansnivå.

EXEMPEL: Vid Universitet Polytechnique mätte man mängden syre y som krävs för nedbrytning av partiklar som funktion av mängden utsläppta partiklar x vid gasning. Linjärt samband. Sifferdata:

$$n = 33, \quad \sum x_k y_k = 41355, \quad \sum x_k^2 = 41086, \quad \sum y_k^2 = 43117$$
$$\sum x_k = 1104, \quad \sum y_k = 1124$$

Skatta regressionslinjen $y = a + bx$ och ge 95% konfidensintervall för b.

$$S_{xy} = \sum x_k y_k - \frac{1}{n}\left(\sum x_k\right)\left(\sum y_k\right) = 41355 - \frac{1}{33}\cdot 1104 \cdot 1124 = 3572$$
$$S_{xx} = \sum x_k^2 - \frac{1}{n}\left(\sum x_k\right)^2 = 41086 - \frac{1}{33}\cdot 1104^2 = 4152$$

Detta ger:

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{3572}{4152} = 0.86$$
$$\hat{a} = \bar{y} - \hat{b}\bar{x} = \frac{1}{33}(1124 - 0.86 \cdot 1104) = 5.29$$

Regressionslinjen:

$$y = 5.29 + 0.86x$$

Konfidensintervall:

Vi har $F^{-1}_{t_{31}}(0.975) = 2.04$ Vi behöver s.

$$S_{yy} = \sum y_k^2 - \frac{1}{n}\left(\sum y_k\right)^2 = 43117 - \frac{1}{33}\,1124^2 = 4833$$
$$s^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{1}{31}\left(4833 - \frac{3572^2}{4152}\right) = 56.74$$

Alltså:

$$b = 0.86 \pm 2.04 \cdot \frac{\sqrt{56.74}}{\sqrt{4152}} = \cancel{0.86 \pm 2.04} \; 0.86 \pm 0.29 \qquad (95\%) \qquad \square$$

PREDIKTION I LINJÄR REGRESSION

Antag att vi vill ge ett prediktionsintervall för en ny observation

$$Y = a + bx + \varepsilon$$

för ett givet värde x. Rimlig punktskattning: $\hat{a} + \hat{b}x$. Skriv

$$D = Y - (\hat{a} + \hat{b}x)$$

Vi har $\mathbb{E}[D] = 0$. Vad är $\mathrm{Var}(D)$? Eftersom $\hat{a} = \bar{Y} - \hat{b}\bar{x}$ har vi

$$D = Y - \bar{Y} - \hat{b}(x - \bar{x}).$$

OBS att Y och $\bar{Y}$ är oberoende, ty Y är en ny observation. Därför gäller det att Y och $\hat{b}$ är oberoende. Det gäller också att

$$\mathrm{Cov}(\bar{Y}, S_{xy}) = \mathrm{Cov}\left(\bar{Y}, \sum x_k Y_k - \frac{1}{n}\left(\sum x_k\right)\left(\sum Y_k\right)\right)$$
$$= \sum x_k \mathrm{Cov}(\bar{Y}, Y_k) - \bar{x}\sum \mathrm{Cov}(\bar{Y}, Y_k)$$
$$= \frac{1}{n}\sigma^2 \sum x_k - \bar{x}\sigma^2 = 0$$

Vi har

$$\mathrm{Var}(D) = \sigma^2 + \frac{1}{n}\sigma^2 + (x-\bar{x})^2 \mathrm{Var}(\hat{b}) = \left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)\sigma^2$$

Alltså

$$\frac{D}{\sigma\sqrt{1 + \frac{1}{n}\frac{(x-\bar{x})^2}{S_{xx}}}} \sim N(0,1)$$

Man kan visa att

$$\frac{D}{s\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t_{n-2}$$

Detta ger ett prediktionsintervall

$$Y \in \hat{a} + \hat{b}x \pm F^{-1}_{t_{n-2}}\left(1 - \frac{\alpha}{2}\right) s\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$ med prediktionsgrad $(1-\alpha)$

EXEMPEL: Ge ett 90% prediktionsintervall för en ny observation Y vid $x = 40$: Virginia Polytechnic-exemplet. Vi har $F^{-1}_{t_{31}}(0.95) = 1.696$, $\bar{x} = \frac{1108}{33} = 33.45$, $s = x - \bar{x} = 6.55$, så

$$Y \in 5.29 + 0.86 \cdot 40 \pm 1.696\sqrt{56.74}\sqrt{\frac{34}{33} + \frac{6.55^2}{4152}} = 86.7 \pm 13$$

□

EXEMPEL: Ge en linjär regression med data $(1,1)$, $(2,2)$, $(3,5)$, ett 90% konfidensintervall för b, ett 90% prediktionsintervall för $x = 4$ och ett 90% prediktionsintervall för $x = 2$.

Vi har $n = 3$, $\sum x_i = 6$, $\sum y_i = 8$, $\sum x_i y_i = 20$
$\sum x_i^2 = 14$, $\sum y_i^2 = 30$

Vi får $S_{xy} = 20 - \frac{1}{3} 8 \cdot 6 = 4$
$S_{xx} = 14 - \frac{1}{3} 6^2 = 2$
$S_{yy} = 30 - \frac{1}{3} 8^2$

$$\hat{b} = \frac{4}{2} = 2$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = \frac{1}{3}(8 - 2 \cdot 6) = -\frac{4}{3}$$

Regressionslinjen blir $y = -\frac{4}{3} + 2x$
Vidare är $s^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{26}{3} - \frac{4^2}{2} = \frac{2}{3}$

Konfidensintervallet är

$$b \in \hat{b} \pm F^{-1}_{t_1}(0.95)\frac{s}{\sqrt{S_{xx}}} = 2 \pm 6.31\frac{\sqrt{\frac{2}{3}}}{\sqrt{2}} = 2 \pm 3.64 \quad (90\%)$$

Prediktionsintervall för $x = 4$: Det gäller att $x - \bar{x} = 2$, så det 90% prediktionsintervallet blir

$$Y \in \hat{a} + \hat{b}x \pm F^{-1}_{t_{n-2}}(0.95)\, s\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} = 6.7 \pm 9.4$$

Prediktionsintervall för $x = 2$: Här är $x - \bar{x} = 0$, så svårigt samma siffror.

$$Y \in \hat{a} + \hat{b}x \pm F^{-1}_{t_{n-2}}(0.95)\, s\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} = 2.7 \pm 6.0$$

Svårare med extrapolation än interpolation.