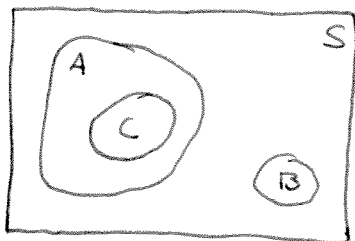


Lösningar: Tenta i Dataanalys & statistik TMS/36
2010-01-11.

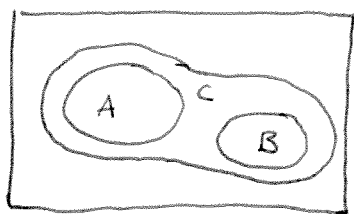
1a)



$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(C)}{P(C)} = 1$$

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{P(\emptyset)}{P(C)} = 0$$

b)

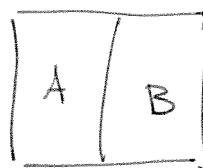


$$P(C|A) = \frac{P(A \cap C)}{P(A)} = 1 \text{ om } A \subseteq C$$

och likadant för $P(C|B)$

Dvs. $P(C|A) = P(C|B) = 1$ för alla
C s.a. $A \cup B \subseteq C$.

c) Ja, självklart. T.ex.



Om $P(A) \neq 0 \neq P(B)$

Så

$$P(A \cap B) = P(\emptyset) = 0 \text{ men } P(A)P(B) > 0.$$

2) a) Tex. X = antal "krona" vid 10 kast med ett
mynt.

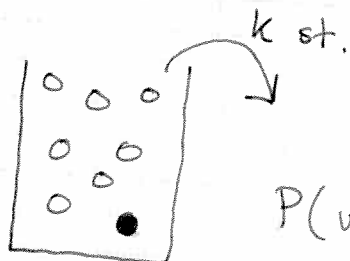
$$X \sim \text{Bin}(10, \frac{1}{2})$$

$$b) P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) =$$

$$= \left(\frac{1}{2}\right)^{10} + \binom{10}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^9 + \binom{10}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^8 =$$

$$= \left(\frac{1}{2}\right)^{10} [1 + 10 + 45] = 0,0547$$

3)



$$P(\text{vi drar den svarta}) = \frac{\binom{1}{1} \binom{n-1}{k-1}}{\binom{n}{k}} =$$

$$= \frac{(n-1)! \cdot \cancel{(n-k)!} \cdot k!}{(n-1-k+1)! \cdot \cancel{(k-1)!} \cdot n!} = \frac{k}{n}$$

4.)

T = tiden mellan två samtal (minuter), $T \sim \text{Exp}(\lambda = \dots)$
 Då vet vi att:

$N(t)$ = antal samtal på t minuter $\sim \text{Poi}(\lambda t) = \text{Poi}(\dots)$

Alltså: ~~$P(N(2))$~~

$$P(\text{exakt 3 samtal på 2 minuter}) =$$

$$= P(N(2) = 3) = \frac{e^{-\lambda t} (\lambda t)^3}{3!} = \frac{e^{-1} (1)^3}{3!} = \frac{e^{-1}}{6}$$

5) a)

$$L(\theta) = \prod_{i=1}^n f(x_i | \theta) = \left\{ \begin{array}{l} \text{stickprov} \\ \text{oberoende} \end{array} \right\} = f(x_1, \dots, x_n | \theta) =$$

$$\stackrel{\text{definition}}{=} P(\bar{X}_1 = x_1, \bar{X}_2 = x_2, \dots, \bar{X}_n = x_n | \theta)$$

d.v.s. $L(\theta)$ är sannolikheten att observera
 det stickprov vi har observerat, givet
 ett värde på θ .

b) $X \sim \text{Geo}(p)$, $f(x) = (1-p)^{x-1} p$.

$$L(p) = \prod_{i=1}^n (1-p)^{x_i-1} p = (1-p)^{\sum x_i - n} p^n \quad \leftarrow \text{Maximera m.a.p } p.$$

logaritmera: $\ln L(p) = (\sum x_i - n) \ln(1-p) + n \ln p$

derivera: $\frac{d}{dp} \ln L(p) = \frac{n - \sum x_i}{1-p} + \frac{n}{p} = 0 \Leftrightarrow 1 - \bar{X} + \frac{1-p}{p} = 0 \Leftrightarrow$

$$\Leftrightarrow \bar{X} = 1 + \frac{1-p}{p} \Leftrightarrow p = \frac{1}{\bar{X} + 1}$$

$$5c) \quad \bar{X} = 6,3$$

$$\text{MLE}(p) = \frac{1}{6,3}$$

Tex: X = antal kast med en tärning tills jag fick en sexa.

6

a) En skattare är en funktion av stickprovet.

b) $\hat{\mu}_{\text{egen}} = \frac{X_1 + X_2}{2} \quad (\text{t.ex.})$

c) $E(\bar{X}) = E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} \sum E(X_i) = \frac{1}{n} \sum \mu = \mu.$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} \sum \text{Var}(X_i) = \frac{1}{n^2} \sum \sigma^2 = \frac{\sigma^2}{n}$$

$$E(\hat{\mu}_{\text{egen}}) = \frac{1}{2} [E(X_1) + E(X_2)] = \frac{2\mu}{2} = \mu \quad \therefore \text{väntevärdesriktig.}$$

$$\text{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{\text{Var}(X_1) + \text{Var}(X_2)}{4} = \frac{\sigma^2}{2}$$

Båda är väntevärdesriktiga, men $\bar{X}$ har mindre varians om $n > 2$.

$\therefore \bar{X}$ bäst.

d) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{2}\right)$ ty ~~det~~ den är en lin.komb av normalfördelade variabler.

$\hat{\mu}_{\text{egen}}$ är också normalfördelade, $\hat{\mu}_{\text{egen}} \sim N\left(\mu, \frac{\sigma^2}{2}\right)$

6) e) X_i = alkoholhalten i den i-te flaskan.

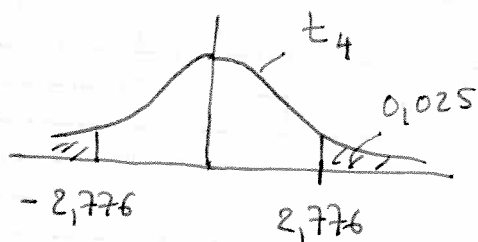
Alkoholhalten i ~~den~~ karaffen = $\frac{1}{5} \sum_{i=1}^5 X_i = \bar{X}$

$$P\left(\underset{\substack{\uparrow \\ N(21; \frac{2}{5})}}{\bar{X}} \geq 22\right) = P\left(\underset{\substack{\uparrow \\ N(0,1)}}{Z} \geq \frac{22-21}{\sqrt{\frac{2}{5}}}\right) = P\left(Z \geq \sqrt{\frac{5}{2}}\right) =$$

$$= P(Z \geq 1,58) = \{\text{tabell}\} = 0,057$$

f) μ = genomsnittliga alkoholhalten.

$$P(A \leq \mu \leq B) = P\left(\frac{\bar{X}-B}{\frac{S}{\sqrt{n}}} \leq \underbrace{\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}}_{\sim t_{n-1}} \leq \frac{\bar{X}-A}{\frac{S}{\sqrt{n}}}\right) = \begin{cases} \bar{X} = 19,99 \\ S^2 = 3,8135 \end{cases} = 0,95$$



Jag väljer att göra ett 2-sidigt CI med kont.grad 0,95

$$\Rightarrow B = \bar{X} + 2,776 \frac{S}{\sqrt{5}} = 19,99 + 2,776 \sqrt{\frac{3,8135}{5}} = 22,41$$

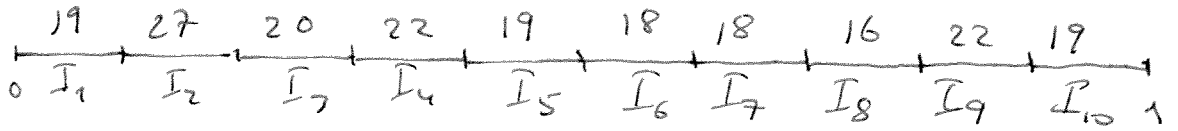
$$A = \bar{X} - 2,776 \frac{S}{\sqrt{5}} = 19,99 - 2,776 \sqrt{\frac{3,8135}{5}} = 17,57$$

95% CI ~~22,41~~ $[17,57; 22,41]$

7.

8

a)

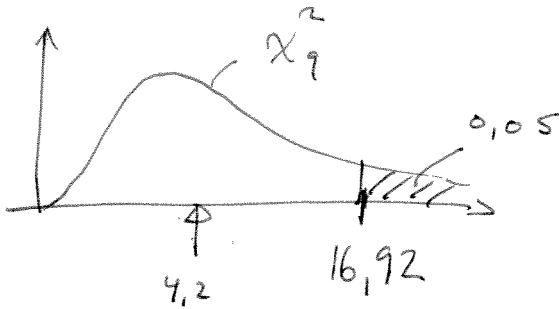


$$H_0: X \sim \text{Uniform}$$

$$E_i = 20$$

$$H_1: X \text{ not uniform.}$$

$$\chi^2_{n-p-1} = \chi^2_9 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i} = 4,2$$



Vi kan alltså inte förkasta att generatoren är likfördelat.

b) Så får man inte göra!!!

Att ~~titta~~ titta på data och sedan bestämma sin hypotes och göra testet på samma data är förbjudet.

Man kommer alltid kunna hitta någon egenskap hos stickprovet som inte är representativ för populationen.

"If you torture data long enough, it will confess."
- R. Coase