Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

TMA4245 Statistikk
Vår 2013

Øving nummer 12, blokk II
Løsningsskisse

Oppgave 1

a) μ = populasjonsgjennomsnitt, dvs. eit gjennomsnitt for alle bilane som køyrer på vegstrekningen i ein gitt periode.

$$\widehat{\mu} = \bar{X} = \frac{\sum_{i=1}^{12} X_i}{12}\ , \quad S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{12}(X_i - \bar{X})^2}$$

$$\bar{X} = \frac{\sum_{i=1}^{12} X_i}{12} = \frac{880}{12} = \underline{\underline{73.33}}\ , \quad S = \sqrt{\frac{1}{11}\sum_{i=1}^{12}(X_i - \bar{X})^2} = \sqrt{\frac{1034.7}{11}} = \underline{\underline{9.7}}$$

b) Type 1 feil er å forkaste H_0 når H_0 er rett.

$$H_0 :\ \mu \geq 77 \qquad H_1 :\ \mu < 77$$

$\alpha = 0.05$, forkast om:

$$\frac{\bar{X} - 77}{\frac{S}{\sqrt{n}}} < -t_{0.05,11} = -1.8$$

$$\frac{73.33 - 77}{\frac{9.7}{\sqrt{12}}} = -1.31 > -1.8$$

dvs. ikkje grunnlag for å påstå at farten er blitt lågare på 5 % nivå.

c) Type 2 feil er å ikkje forkaste når H_0 er gal. La $\beta = P(\text{type 2 feil})$. Då er styrken $1-\beta$.

$$\begin{aligned} P\left(\frac{\bar{X} - 77}{\frac{10}{\sqrt{12}}} < -1.645 \mid \mu = 74\right) &= P\left(\frac{\bar{X} - 74}{\frac{10}{\sqrt{12}}} < -1.645 + \frac{3}{\frac{10}{\sqrt{12}}} \mid \mu = 74\right) \\ &= \Phi\left(-1.645 + \frac{3\sqrt{12}}{10}\right) = \Phi(-0.61) \\ &= 1 - 0.729 = \underline{\underline{0.271}} \end{aligned}$$

$$
\begin{aligned}
& P\left(\frac{\bar{X}-77}{\frac{10}{\sqrt{n}}} < -1.645 \mid \mu = 74\right) = 0.9 \\
\Leftrightarrow\ & P\left(\frac{\bar{X}-74}{\frac{10}{\sqrt{n}}} < -1.645 + \frac{3}{\frac{10}{\sqrt{n}}} \mid \mu = 74\right) = 0.9 \\
\Leftrightarrow\ & \Phi\left(-1.645 + \frac{3\sqrt{n}}{10}\right) = 0.9 \\
\Leftrightarrow\ & -1.645 + \frac{3\sqrt{n}}{10} = 1.28 \\
\Leftrightarrow\ & \frac{3\sqrt{n}}{10} = 1.28 + 1.645 = 2.925 \\
\Leftrightarrow\ & n = \frac{(2.925)^2 \cdot 10^2}{3^2} = 95.06
\end{aligned}
$$

Dvs. vi må måle farten på 96 bilar eller fleir.

Oppgave 2

1. Figur 1 viser eksempel på histogram som plottes når man kjører koden *hypoth.m*. Vi ser at med $n = 10000$ data ligner histogrammet ganske godt normalfordelingen den kommer fra, sentrert rundt forventningsverdien $\mu = 1$.

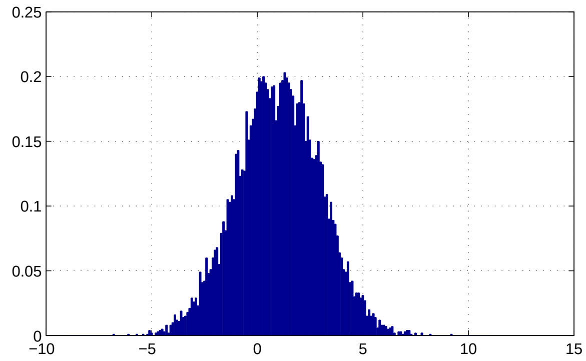
2. Tolker vi koden ser vi at parameteren *mu_hypoth* er forventningsverdien i nullhypotesen, altså med $mu\_hypoth = \mu = 1$ tester vi

$$\mathrm{H}_0: \quad \mu = 1 \quad vs \quad \mathrm{H}_1: \quad \mu \neq 1$$

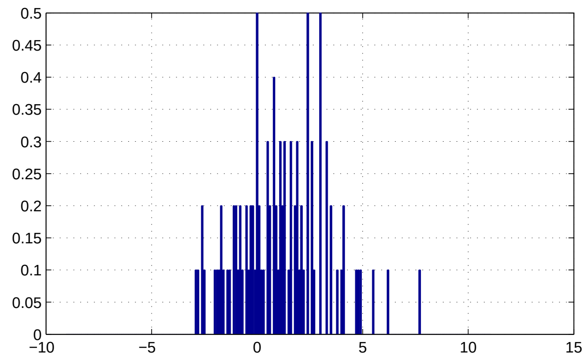
I koden beregner vi da T-observatoren $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ som parameteren t. Legg merke til at vi har antar at variansen er ukjent, slik at dette blir en t-test. Outputparametrene er

- $p1$: Sannsynligheten i en t-fordeling for større verdier av testobservatoren enn vår observerte verdi av t, altså $1 - P(-t \leq T \leq t)$
- h: Resultat av Matlabs innebygde t-test, $h = 0$ dersom vi ikke kan forkaste nullhypotesen med 5% signifikansnivå, $h = 1$ dersom vi kan.
- $p2$: Resultat av Matlabs innebygde t-test, hvor denne parameteren er den beregnede T-observatoren. Når vi kjører koden observerer vi at $p1 = p2$.
- ci: Resultat av Matlabs innebygde t-test, dette er det beregnede 95% konfidensintervalet for μ.

For dataene i figur 1 får vi $p1 = p2 = 0.9400,\ h = 0,\ ci = [0.9597, 1.0374]$. Altså forkaster vi ikke nullhypotesen (om at forventningsverdien til dataene er 1) med 95% signifikansnivå. Dette ser vi fra parameteren $h = 0$ i tillegg til at verdien 1 ligger godt innenfor 95% konfidensintervalet ci.
Figur 2 viser et tilsvarende forsøk hvor vi har kjørt koden med $n = 100$ data (histogrammet ser såpass spredt ut fodi oppløsningen er høy). Output er her $p1 = p2 = 0.6190,\ h = 0,\ ci = [0.6912, 1.5162]$. Igjen beholder vi altså nullhypotesen, men vi observerer at konfidensintervalet er bredere siden vi har simulert færre data.



Figur 1: Histogram av $n = 10000$ simulerte data fra normalfordeling med $\mu = 1$ og $\sigma^2 = 2^2$



Figur 2: Histogram av $n = 100$ simulerte data fra normalfordeling med $\mu = 1$ og $\sigma^2 = 2^2$.

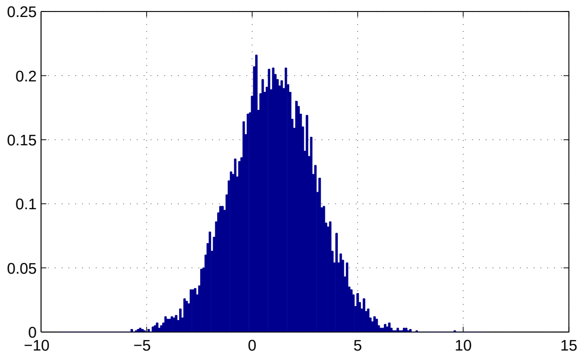
3. Dersom vi endrer hypotesetesten vår til f.eks.

$$\mathrm{H}_0 : \quad \mu = 1.2 \quad vs \quad \mathrm{H}_1 : \quad \mu \neq 1.2$$

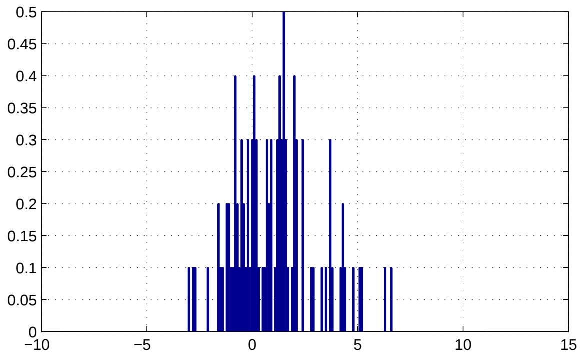
(setter $mu\_mypoth = 1.2$ i koden) tester vi for om forventningsverdien er 1.2. Siden datene kommer fra normalfordelingen $N(1, 2^2)$ burde vi altså optimalt forkaste nullhypotesen.

I figur 3 ser vi et resultat for $n = 10000$ simulerte data, som resulterte i $p1 = p2 = 0$, $h = 1$, $ci = [0.9569, 1.0341]$. Her forkaster vi altså nullhypotesen siden $h = 1$ og 1.2 ligger godt utenfor 95% konfidensintervalet.

I figur 4 ser vi et resultat for $n = 100$ simulerte data, som resulterte i $p1 = p2 = 0.3407$, $h = 0$, $ci = [0.6285, 1.3995]$. Her forkaster vi altså ikke nullhypotesen siden $h = 0$ og 1.2 ligger innenfor 95% konfidensintervalet. Vi observerer altså at i dette tilfellet er $n = 100$ for lite data til å konkludere med at forventningsverdien ikke er 1.2 når den i virkeligheten er 1.



Figur 3: Histogram av $n = 10000$ simulerte data fra normalfordeling med $\mu = 1$ og $\sigma^2 = 2^2$.



Figur 4: Histogram av $n = 100$ simulerte data fra normalfordeling med $\mu = 1$ og $\sigma^2 = 2^2$.

Oppgave 3

$E(X) = \lambda\nu$

a) λ : forventet (gjennonsnitlig) antall bakterier pr. liter vann.

$P(X = 0) = \frac{(\lambda\nu)^0}{0!} e^{-\lambda\nu} = e^{-3(0.5)} = \underline{0.223}$

$P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{x=0}^{3} \frac{(1.5)^x}{x!} e^{-1.5} = 1 - 0.934 = \underline{0.066}$

b) $P(X > 3 \mid x > 0) = \frac{P(X>3 \cap x>0)}{P(X>0)} = \frac{P(X>3)}{1-P(X=0)} = \frac{0.066}{1-0.223} = \underline{0.0849}$

$Y = X_1 + X_2 \sim Po(\lambda\nu_1 + \lambda\nu_2)$; med $\lambda\nu_1 + \lambda\nu_2 = 3(3) = 9$

$P(X_1 + X_2 > 3) = 1 - P(X_1 + X_2 \leq 3) = 1 - \sum_{x=0}^{3} \frac{9^x}{x!} e^{-9} = \underline{0.97877}$

c) $P(X_1 + X_2 > 3 \mid X_1 > 0 \cap X_2 > 0) = 1 - P(X_1 + X_2 \leq 3 \mid X_1 > 0 \cap X_2 > 0)$
$= 1 - \frac{P(X_1+X_2 \leq 3 \cap X_1>0 \cap X_2>0)}{P(X_1>0 \cap X_2>0)}$

$= 1 - \frac{P(X_1=1 \cap X_2=1)+P(X_1=1 \cap X_2=2)+P(X_1=2 \cap X_2=1)}{P(X_1>0).P(X_2>0)}$
$= 1 - \frac{P(X_1=1)P(X_2=1)+P(X_1=1)P(X_2=2)+P(X_1=2)P(X_2=1)}{(1-P(X_1=0))(1-P(X_2=0))}$
$= 1 - \frac{(3.6+3.6^2/2+3^2.6/2).e^{-3}.e^{-6}}{(1-e^{-3})(1-e^{-6})}$
$= \underline{0.987}$

Og

$L(\lambda) = \pi_{i=1}^{n} \frac{(\lambda\nu_i)^{x_i}}{x_i!} e^{-\lambda\nu_i}$

$l(\lambda) = \sum_{i=1}^{n} \left(x_i ln(\lambda\nu_i) + ln(x_i!) - \lambda\nu_i\right)$

$l'(\lambda) = \sum_{i=1}^{n} \left(x_i \frac{\nu_i}{\lambda\nu_i} + 0 - \nu_i\right) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \nu_i$

$l'(\lambda) = 0 \Rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} \nu_i}$
dvs. $\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} \nu_i}$

$E(\hat{\lambda}) = \frac{E\left(\sum_{i=1}^{n} X_i\right)}{\sum_{i=1}^{n} \nu_i} = \frac{\sum_{i=1}^{n} E(X_i)}{\sum_{i=1}^{n} \nu_i} = \frac{\sum_{i=1}^{n} \lambda\nu_i}{\sum_{i=1}^{n} \nu_i} = \lambda \frac{\sum_{i=1}^{n} \nu_i}{\sum_{i=1}^{n} \nu_i} = \lambda$

$Var(\hat{\lambda}) = \frac{Var\left(\sum_{i=1}^{n} X_i\right)}{\left(\sum_{i=1}^{n} \nu_i\right)^2} = \frac{\sum_{i=1}^{n} \lambda\nu_i}{\left(\sum_{i=1}^{n} \nu_i\right)^2} = \lambda \frac{\sum_{i=1}^{n} \nu_i}{\left(\sum_{i=1}^{n} \nu_i\right)^2} = \frac{\lambda}{\sum_{i=1}^{n} \nu_i}$

d) $H_0 : \lambda = \lambda_0 = 3$ mot $H_1 : \lambda > 3$

Test obs.

$U = \frac{\hat{\lambda}-\lambda_0}{\sqrt{\frac{\lambda_0}{\sum_{i=1}^{n} \nu_i}}} \approx N(0,1)$ under H_0.

Forkaster dersom $U > k$ der k bestemmer fra $P(U > k$ når H_0 er riktig $) = \alpha$

dvs. $k = z_\alpha$

dvs. Forkaster dersom $U > z_\alpha$.

Innsatt tall: $z_{0.025} = 1.96$, og $\hat{\lambda} = 78/(20) = 3.9$

$u = \frac{3.9-3}{\sqrt{3/20}} = 2.32 > 1.96$

dvs. Forkaster H_0.

e) Under $H_0 : Z \sim \text{bin}(n = 10, p_0)$
der $p_0 = P(X > 6 \mid \lambda = \lambda_0) = 1 - P(X \leq 5 \mid \lambda = \lambda_0) = 1 - 0.4457 = \underline{0.5543}$

Forkaster H_0 hvis $Z \geq k$ der k bestemmes fra kravet:
$P(Z \geq k$ hvis H_0 er riktig$) \leq 0.025$.

For ulike verdier for k,

Z	$P(Z = z$ hvis H_0 er riktig$)$	$P(Z \geq k$ når H_0 er riktig$)$
10	0.002738	0.02738
9	0.022016	0.024754
8	0.07966	0.104414

Ser at en må ha $\underline{k = 9}$.

Innsatt data: $z = 6 < k = 9$

dvs. Forkaster ikke H_0.

f) La $\lambda = 3.5$:

$$P(U > 1.96) = P\left(\frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda_0}{\sum \nu_i}}} > 1.96\right) = P(\hat{\lambda} > 1.96\sqrt{\frac{\lambda_0}{\sum \nu_i}} + \lambda_0)$$
$$= P\left(\frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda_0}{\sum \nu_i}}} > \frac{1.96\sqrt{\frac{\lambda_0}{\sum \nu_i}}+\lambda_0-\lambda}{\sqrt{\frac{\lambda}{\sum \nu_i}}}\right)$$
$$= P\left(\frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda_0}{\sum \nu_i}}} > 0.619\right) = 1 - \Phi(0.619) = 1 - 0.7324 = \underline{0.2676}.$$

$$P(Z \geq 9) = \textstyle\sum_{z=9}^{10} \frac{10!}{z!(10-z)!} p^z (1-p)^{10-z}$$

der $p = P(X > 6 \mid \lambda = 3.5) = 1 - P(X \leq 5 \mid \lambda = 3.5) = 1 - 0.3007 = 0.6993$

$\Rightarrow P(Z \geq 9) = 0.02796 + 0.12025 = \underline{0.1482}$