

Løsningsforslag eksamen TALM1005

november 2019

Oppgave 1

a) Et 8-bits digitalt signal er altså på formen

$$b \ b \ b \ b \ b \ b \ b \ b,$$

der hvert siffer b enten er 0 eller 1, dvs. hvert siffer har 2 muligheter. Det totale antall mulige 8-bits signaler blir da gitt ved multiplikasjonssetningen:

$$2^8 = \underline{\underline{256}}$$

b) La X angi antall 1-ere i det digitale signalet. Når sifrene genereres tilfeldig og uavhengig av hverandre, med konstant sannsynlighet lik $p = \frac{1}{2}$ for at et gitt siffer blir 1, får X binomisk fordeling.

c) Sannsynligheten for å få et signal bestående av 8 1-ere blir altså

$$\begin{aligned} P(X=8) &= \binom{8}{8} \cdot p^8 \cdot (1-p)^{8-8} \\ &= 1 \cdot \left(\frac{1}{2}\right)^8 \cdot \left(\frac{1}{2}\right)^0 \\ &= \frac{1}{\underline{\underline{256}}} \end{aligned}$$

d) Vi får oppgitt at automaten utbetaler premie i form av 10 kr dersom signalet inneholder tre 1-ere, og 200 kr dersom det inneholder sju 1-ere. Vi skal finne forventningsverdien for utbetalt premie.

La Y angi utbetalt premie. Hvis X som tidligere angir antall 1-ere i signalet, blir forventningsverdien for Y gitt ved

$$E(Y) = 10 \cdot P(X=3) + 200 \cdot P(X=7)$$

Ut i fra binomiske punktsannsynligheter får vi følgende forventningsverdi i antall kroner for forventet premie:

$$\begin{aligned} E(Y) &= 10 \cdot \binom{8}{3} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(1 - \frac{1}{2}\right)^{8-3} + 200 \cdot \binom{8}{7} \cdot \left(\frac{1}{2}\right)^7 \cdot \left(1 - \frac{1}{2}\right)^{8-7} \\ &= 100 \cdot \binom{8}{3} \cdot \left(\frac{1}{2}\right)^8 + 200 \cdot \binom{8}{7} \cdot \left(\frac{1}{2}\right)^8 \\ &= \frac{135}{16} \\ &\approx \underline{\underline{8,44}} \end{aligned}$$

e) Definerer følgende hendelser:

M_1 : Mottar 1

M_0 : Mottar 0

S_1 : Sender 1

S_0 : Sender 0

Vi får oppgitt at $P(M_1|S_0) = 0,05$ og $P(M_0|S_1) = 0,02$. Ettersom et mottatt siffer er enten 1 eller 0, betyr de oppgitte sannsynlighetene at $P(M_1|S_1) = 1 - 0,02 = 0,98$.

Vi skal finne sannsynligheten for at en mottatt 1-er ble sendt som en 1-er, dvs. vi skal finne $P(S_1|M_1)$. Fra Bayes' teorem:

$$P(S_1|M_1) = \frac{P(M_1|S_1) \cdot P(S_1)}{P(M_1)}$$

Her er $P(M_1)$, altså sannsynligheten for å motta en 1-er, gitt ved loven om total sannsynlighet:

$$\begin{aligned} P(M_1) &= P(M_1|S_1) \cdot P(S_1) + P(M_1|S_0) \cdot P(S_0) \\ &= 0,98 \cdot \frac{1}{2} + 0,05 \cdot \frac{1}{2} \\ &= \underline{\underline{0,515}} \end{aligned}$$

Innsatt gir dette at

$$\begin{aligned} P(S_1|M_1) &= \frac{P(M_1|S_1) \cdot P(S_1)}{P(M_1)} \\ &= \frac{0,98 \cdot \frac{1}{2}}{0,515} \\ &= 0,951 \\ &\approx \underline{\underline{0,95}} \end{aligned}$$

Oppgave 2

a) Hvis X angir antall fartsovertredelser innen et tidsintervall, er X poissonfordelt dersom følgende kriterier er oppfylte:

- i) Det er konstant sannsynlighet for at en fartsovertredelse skal skje innen tidsintervallet (dvs. uavhengig av tidspunkt på dagen)
- ii) Sannsynlighetene for én fartsovertredelse er uavhengig av eventuelle andre (dvs. bilførerne advarer ikke hverandre e.l.)
- iii) Sannsynligheten for at to overtredelser skal skje samtidig er neglisjerbar.

b) Gitt at intensiteten for fartsovertredelser er $\alpha = 4/\text{time}$, blir forventet antall overtredelser i løpet av 3 timer lik

$$\begin{aligned} \lambda &= \alpha t \\ &= \frac{4}{\text{time}} \cdot 3 \text{ timer} \\ &= \underline{\underline{12}} \end{aligned}$$

Vi skal finne $P(X > 14)$ dersom vi antar at X er poissonfordelt. Får at

$$\begin{aligned} P(X > 14) &= 1 - P(X \leq 14) && (\text{Husk at } X \text{ er diskret}) \\ &= 1 - 0,772 && (\text{Kumulativ poissonford. på kalkulator el. tabell}) \\ &= 0,228 \\ &\approx \underline{\underline{0,23}} \end{aligned}$$

c)

i) La T betegne tiden fra kontrollen starter til første overtredelse registreres. Ettersom dette angir tiden til første hendelse i en poissonprosess, vet vi at T blir eksponentialfordelt med $\alpha = \frac{4}{\text{time}} = \frac{4}{60 \text{ min}} = \frac{1}{15 \text{ min}}$. Hvi vi angir T i minutter, skal vi finne

$$\begin{aligned} P(T > 15) &= 1 - P(T \leq 15) \\ &= 1 - (1 - e^{-\alpha \cdot 15}) \\ &= e^{-\frac{1}{15} \cdot 15} \\ &= e^{-1} \\ &\approx \underline{\underline{0,37}} \end{aligned}$$

ii) Finner forventningsverdi og varians for T fra formelarket (under eksponentialfordelingen):

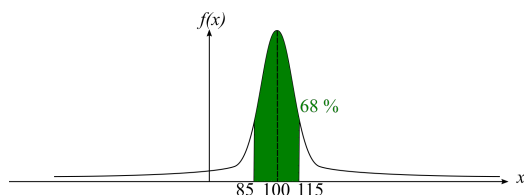
$$\mu = E(T) = \frac{1}{\alpha} = \frac{1}{\frac{1}{15 \text{ min}}} = \underline{\underline{15 \text{ min}}}$$

$$\sigma = \sqrt{\text{Var}(T)} = \sqrt{\frac{1}{\alpha^2}} = \frac{1}{\alpha} = \underline{\underline{15 \text{ min}}}$$

Oppgave 3

a) Gitt at IQ-score X i en befolkning er normalfordelt med $\mu = 100$ og $\sigma = 15$.

i) Figuren under viser en skisse av sannsynlighetstettheten for IQ i befolkningen. Her ligger toppunktet i $x = \mu = 100$, og standardavviket er slik at ca. 68 % av grafen ligger i intervallet $[x - \sigma, x + \sigma] = [85, 115]$.



ii) Vi skal bestemme sanns. for at et tilfeldig person har en IQ mellom 95 og 105. Dette kan vi enten gjøre via kumulativ normalfordeling på kalkulator, eller ved å omforme til standard normalfordeling og så bruke tabell:

$$\begin{aligned} P(95 < X < 105) &= P\left(\frac{95 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{105 - \mu}{\sigma}\right) && \text{(Standardiserer)} \\ &= P\left(\frac{95 - 100}{15} < Z < \frac{105 - 100}{15}\right) \\ &= P\left(-\frac{1}{3} < Z < \frac{1}{3}\right) \\ &= P\left(Z \leq \frac{1}{3}\right) - P\left(Z \leq -\frac{1}{3}\right) \\ &= P\left(Z \leq \frac{1}{3}\right) - \left(1 - P\left(Z \leq \frac{1}{3}\right)\right) && \text{(Grep for negative verdier)} \\ &= 2P\left(Z \leq \frac{1}{3}\right) - 1 \\ &\approx 2 \cdot (Z \leq 0,33) - 1 \\ &= 2 \cdot 0,6293 - 1 \\ &= 0,2586 \\ &\approx \underline{\underline{0,26}} \end{aligned}$$

iii) Sannsynligheten for at en tilfeldig person har en IQ større enn 150 er

$$\begin{aligned} P(X \geq 150) &= 1 - P(X \leq 150) \\ &= 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{150 - \mu}{\sigma}\right) \\ &= 1 - P\left(Z \leq \frac{150 - 100}{15}\right) \\ &= 1 - P\left(Z \leq \frac{10}{3}\right) \\ &\approx 1 - P(Z \leq 3,33) \\ &= 1 - 0,9996 \\ &= \underline{\underline{0,0004}} \end{aligned}$$

b) Vi skal bestemme 2 %-kvantilen for IQ i en befolkning, dvs. den IQ-verdien som er slik at kun 2 % av befolkningen har høyere IQ enn denne verdien. Vi skal altså bestemme en verdi $x_{0,02}$ som er slik at

$$\begin{aligned}P(X \geq x_{0,02}) &= 0,02 \\1 - P(X \leq x_{0,02}) &= 0,02 \\P(X \leq x_{0,02}) &= 0,98\end{aligned}$$

Her kan vi så klart prøve oss fram på kalkulator, men det er enklere - og mer presist - å omforme til standard normalfordeling:

$$\begin{aligned}P\left(\frac{X - \mu}{\sigma} \leq \frac{x_{0,02} - \mu}{\sigma}\right) &= 0,98 \\P\left(Z \leq \frac{x_{0,02} - 100}{15}\right) &= 0,98\end{aligned}$$

Ut i fra definisjonen av en kvantil, vil Z -verdien som tilsvarer dette, være lik 2 %-kvantilen $u_{0,02}$ for en standard normalfordeling, altså

$$\frac{x_{0,02} - 100}{15} = u_{0,02} = 2,054 \Rightarrow x_{0,02} = 100 + 15 \cdot 2,054 = 130,8 \approx \underline{\underline{131}}$$

c) Gitt en klasse på 100 personer fra befolkningen, skal vi finne sannsynligheten for at gjennomsnittlig IQ $\bar{X}$ for de 100 studentene er mellom 95 og 105. Vi skal altså finne

$$P(95 < \bar{X} < 105),$$

og da trenger vi forventningsverdi og standardavvik for et gjennomsnitt av normalfordelte variable; disse er hhv. $\mu_{\bar{X}} = \mu = 100$ og $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{100}} = \frac{15}{10} = 1,5$. Finner sannsynligheten på kalkulator (eller standardisering og tabell, som tidligere vist):

```
Normal C.D
Lower : 95
Upper : 105
σ      : 1.5
μ      : 100
Save Res: None
Execute
None List
```

Får svaret at

$$P(95 < \bar{X} < 105) = \underline{\underline{0,9991}}$$

ii) Skal finne sannsynligheten for at minst halvparten av studentene i en gruppe på 100 har IQ større enn 100. Sannsynligheten for at én tilfeldig student har en IQ større enn 100 er

$$\begin{aligned}P(X > 100) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{100 - \mu}{\sigma}\right) \\&= P\left(Z \leq \frac{100 - 100}{15}\right) \\&= P(Z \leq 0) \\&= \frac{1}{2}\end{aligned}\quad (\text{Symmetrisk om } x = 0)$$

La Y angi antall studenter i gruppa med IQ større enn 100. Y blir da binomisk fordelt med $n = 100$ og $p = \frac{1}{2}$. Vi skal altså finne

$$\begin{aligned}P(Y \geq 50) &= 1 - P(X \leq 49) \\&= 1 - 0,46 \\&= \underline{\underline{0,54}}\end{aligned}\quad (\text{Binomisk kum. ford. på kalkulator})$$

d) Skal bestemme antall studenter n som måtte ha vært i gruppen for at det skulle ha vært 99,9 % sannsynlig at minst én student hadde en IQ større enn 130. Sannsynligheten for at én tilfeldig student

har IQ større enn 130 er

$$\begin{aligned}P(X \geq 130) &= 1 - P(X \leq 130) \\&= 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{130 - \mu}{\sigma}\right) \\&= 1 - P\left(Z \leq \frac{130 - 100}{15}\right) \\&= 1 - P(Z \leq 2) \\&= 1 - 0,9772 \\&= \underline{0,0228}\end{aligned}$$

Hvis W angir antall studenter med IQ større enn 130, blir W binomisk fordelt med n som skal bestemmes, og $p = 0,0228$. Vi skal altså bestemme n slik at

$$\begin{aligned}P(W \geq 1) &= 0,999 \\1 - P(W = 0) &= 0,999 \\P(W = 0) &= 0,001\end{aligned}$$

Binomisk punktsannsynlighet er gitt ved

$$\begin{aligned}P(W = 0) &= \binom{n}{0} p^0 (1 - p)^{n-0} \\&= (1 - p)^n \quad (\text{Her er } \binom{n}{0} = 1)\end{aligned}$$

Likninga for n blir altså

$$\begin{aligned}P(W = 0) &= 0,001 \\(1 - p)^n &= 0,001 \\n \ln(1 - p) &= \ln 0,001 \\n &= \frac{\ln 0,001}{\ln(1 - p)} \\&= \frac{\ln 0,001}{\ln(1 - 0,0228)} \\&= 299,5 \\&\approx \underline{\underline{300}}\end{aligned}$$

Det trengs ca. 300 studenter i gruppa.

Oppgave 4

a)

i) En hypotesetest tar utgangspunkt i en **stikkprøve** for å undersøke en hypotese om en befolkningsparameter (typisk gjennomsnittet μ). **Nullhypotesen** H_0 representerer den “forsiktige”, “konservative”, “status quo”-situasjonen, mens **mothypotesen** H_1 er at den relle verdien for parameteren er enten mindre enn, større enn, eller forskjellig fra (for en tosidig test) den verdien som H_0 påstår. Vi antar alltid at H_0 er riktig, inntil vi har sterke bevis for det motstatte - spesifikt at stikkprøven er så “unormal” eller “usannsynlig” at H_0 antakelig er feil. **Signifikansnivået** α er sannsynligheten/risikoen for at vi forkaster H_0 selv om den er riktig, og velges typisk som en “liten” verdi, ofte $\alpha = 5\%$. **Kritisk verdi** k tilsvarer den verdien for gjennomsnittet av stikkprøven der vi forkaster H_0 - når vi når kritisk verdi, er stikkprøven så “unormal” hvis H_0 er riktig, at H_0 antakelig er feil.

ii) For å undersøke om responstiden er **høyere** enn den garanterte verdien, blir hypotesetesten

$$\begin{aligned}H_0 : \mu &= 10 \\H_1 : \mu &> 10\end{aligned}$$

La X angi responstiden i minutter. Vi kan avgjøre hypotesten med valgfri metode - her vises samtlige metoder (trenger kun én for full uttelling):

Kritisk verdi

Vi skal finne den kritiske verdien for hva dnt målte gjennomsnittlige responstiden $\bar{X}$ må være for at vi skal konkludere med at responstiden er høyere enn dette, ved et signifikansnivå på 5 %.

Kritisk verdi k er bestemt ved at

$$P(\bar{X} \geq k | H_0) = \alpha = 0,05$$

Får at

$$\begin{aligned} P(\bar{X} \geq k | H_0) &= P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq \frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) && \text{(Normaliserer)} \\ &= P\left(Z \geq \frac{k - 10}{\frac{1}{\sqrt{20}}}\right) \\ &= 0,05 \end{aligned}$$

Vi skal altså bestemme den z -verdien som er slik at 5 % av verdiene ligger over denne verdien. Dette er selve definisjonen av en kvantil, og ut i fra kvantiltabellen, der $u_{0,05} = 1,645$, får vi at

$$\begin{aligned} \frac{k - 10}{\frac{1}{\sqrt{20}}} = 1,645 &\Rightarrow k = 10 + 1,645 \cdot \frac{1}{\sqrt{20}} \\ k &= \underline{10,4} \end{aligned}$$

Det observerte gjennomsnittet var $\bar{x} = 10,5$ min, dvs. **større** en kritisk verdi for et signifikansnivå ved 5 %. Ved dette signifikansnivået vil vi **forkaste nullhypotesen**, og konkluderer med at responstiden er høyere enn garantert.

Testobservator

Testobservatoren her er

$$\begin{aligned} U_0 &= \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \\ &= \frac{10,5 - 10}{\frac{1}{\sqrt{20}}} \\ &= \underline{2,24} \end{aligned}$$

Sammenlikner dette med 5 %-kvantilen $u_{0,02} = 1,645$, og ser at $|U_0| > u_{0,05}$, dvs. vi forkaster H_0 .

Signifikanssannsynlighet (p-verdi)

Vi finner sannsynligheten for å få et minst like “ekstremt” gjennomsnitt som det observerte, gitt at H_0 er riktig:

$$\begin{aligned} p &= P(\bar{X} \geq 10,5 | H_0) \\ &= P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq \frac{10,5 - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) \\ &= P\left(Z \geq \frac{10,5 - 10}{\frac{1}{\sqrt{20}}}\right) \\ &= P(Z \geq 2,24) \\ &= 1 - P(Z \leq 2,24) \\ &= 1 - 0,9875 \\ &= 0,0125 \\ &= \underline{0,013} \end{aligned}$$

Etter $p < \alpha$, er stikkprøven såpass “usannsynlig” at vi forkaster H_0 .

b) Vi skal beregne sannsynligheten for at hypotesetesten klarer å “fange opp” at responstiden er for høy dersom den “sanne” verdien for responstiden er $\mu = 11$ min. Dette tilsvarer å finne styrkefunksjonen $\beta(\mu = 11)$, dvs. vi skal beregne

$$\beta = P(\text{akseptere } H_1 | \mu = 11)$$

Ettersom vi tidligere har beregnet kritisk verdi til $k = 10,4$, tilsvarer dette å beregne sannsynligheten for å observere et snitt i stikkprøven som er større enn den kritiske verdien. Vi får:

$$\begin{aligned}\beta &= P(\text{akseptere } H_1 | \mu = 11) \\ &= P(\bar{X} \geq k | \mu = 11) \\ &= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{k - \mu}{\frac{\sigma}{\sqrt{n}}} | \mu = 11\right) \\ &= P\left(Z \geq \frac{10,4 - 11}{\frac{1}{\sqrt{20}}}\right) \\ &= P(Z \geq -2,68) \\ &= 1 - P(Z \leq -2,68) \\ &= 1 - (1 - P(Z \leq 2,68)) && \text{(Triks for neg. verdier)} \\ &= P(Z \leq 2,68) \\ &\approx \underline{0,9963}\end{aligned}$$

Dersom vi altså gjennomfører “mange” stikkprøver av 20 samtaler fra en “pool” der gjennomsnittlig responstid μ faktisk er 11 min, vil vi i 99,6 % av tilfellene få en verdi som gjør at vi konkluderer (riktig) med **for høy** responstid.