

Løsningsforslag statistikkeksamen desember 2014

Oppgave 1

- a) i. To hendelser er disjunkte hvis det er intet overlapp mellom hendelsene, altså hvis $A \cap B = \emptyset$. Siden vi har en sannsynlighet for å finne $A \cap B$ som er ulik null, ser vi at det må være mulig at hendelsene A og B inntreffer samtidig, og hendelsene er ikke disjunkte.
- ii. To hendelser er uavhengige dersom $P(A \cap B) = P(A) \cdot P(B)$. Fra opplysningene i teksten, får vi:

$$P(A) \cdot P(B) = 0.40 \cdot 0.70 = 0.28 = P(A \cap B)$$

A og B er med andre ord uavhengige.

- b) i. Vi skal finne sannsynligheten for at enten A , B eller begge inntreffer, dvs. $A \cup B$. Fra formelarket ser vi at:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.40 + 0.70 - 0.28 = \underline{\underline{0.82}}$$

- ii. Vi skal finne sannsynligheten for at A inntreffer, samtidig som at B ikke inntreffer, altså $P(A \cap \overline{B})$. Fra formelarket ser vi da at:

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

Siden A og B er uavhengige, har vi $P(A|B) = P(A)$, og vi får:

$$P(A \cap \overline{B}) = P(\overline{B})P(A) = (1 - P(B))P(A) = (1 - 0.70) \cdot 0.40 = \underline{\underline{0.12}}$$

Oppgave 2

- a) i. Vi bruker definisjonen på forventning fra formelarket og får:

$$E(X) = \mu = \sum_x x \cdot P(X = x) = 1 \cdot 0.35 + 2 \cdot 0.28 + 3 \cdot 0.21 + 4 \cdot 0.16 = \underline{\underline{2.18}}$$

Vi gjør så tilsvarende for variansen:

$$\begin{aligned} \text{Var}(X) = \sigma^2 &= \sum_x (x - \mu)^2 \cdot P(X = x) = \sum_x x^2 \cdot P(X = x) - \mu^2 \\ &= 1^2 \cdot 0.35 + 2^2 \cdot 0.28 + 3^2 \cdot 0.21 + 4^2 \cdot 0.16 - 2.18^2 = \underline{\underline{1.1676}} \end{aligned}$$

- ii. Siden vi har mange målinger ($n > 30$), og hver X er uavhengige og identisk fordelte, har vi fra sentralgrensesetningen at $X \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}}^2) = N(\mu, \frac{\sigma^2}{n}) = N(2.18, \frac{1.1676}{100})$. Vi har dermed:

$$\begin{aligned} P(\bar{X} \leq 2.1) &\approx G\left(\frac{2.1 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) = G\left(\frac{2.1 - 2.18}{\sqrt{\frac{1.1676}{100}}}\right) \\ &\approx G(-0.74) = 1 - G(0.74) \stackrel{\text{tabell}}{=} 1 - 0.7704 \approx \underline{\underline{0.23}} \end{aligned}$$

- b) For hver måling får vi enten $X = 1$ eller ikke. Sannsynligheten for at en $X = 1$, er alltid 0.35. I tillegg er hver måling uavhengige av hverandre. Vi har dermed at antall ganger $X = 1$ på $n = 100$ målinger, Y , er binomisk fordelt, $Y \sim \text{bin}(100, 0.35)$. Vi får da:

$$P(Y = 30) = \binom{100}{30} 0.35^{30} (1 - 0.35)^{100-30} \approx \underline{\underline{0.0494}}$$

- c) Siden $np(1-p) = 100 \cdot 0.35 \cdot (1 - 0.35) = 22.75 > 10$, vil Y være tilnærmet normalfordelt, $Y \sim N(\mu_Y, \sigma_Y^2) = N(np, np(1-p)) = N(35, 22.75)$. Dette trengte vi ikke utnytte i forrige oppgave siden det var lett å regne ut sannsynligheten direkte, men det blir ganske mye arbeid å regne ut $P(Y \leq 30)$ direkte med binomisk fordeling. Når vi vet at Y er tilnærmet normalfordelt, blir regnestykket enklere. Siden Y er en diskret variabel, bruker vi i tillegg heltallskorreksjon for å gjøre tilnærmelsen bedre:

$$\begin{aligned} P(Y \leq 30) &\approx G\left(\frac{30 + 0.5 - \mu_Y}{\sigma_Y}\right) = G\left(\frac{30 + 0.5 - 35}{\sqrt{22.75}}\right) = G(-0.9435) \\ &\approx G(-0.94) = 1 - G(0.94) \stackrel{\text{tabell}}{=} 1 - 0.8264 \approx \underline{\underline{0.17}} \end{aligned}$$

Oppgave 3

- a) La X være antall mål Rosenborg skårer i en kamp. Vi får oppgitt at X er poissonfordelt, dvs:

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda},$$

der størrelsen på λ avhenger av om Rosenborg spiller på hjemmebane eller bortebane. Hvis Rosenborg spiller på hjemmebane, har vi at $\lambda = 1.5$ og vi får:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \frac{1.5^0}{0!} e^{-1.5} - \frac{1.5^1}{1!} e^{-1.5} = 1 - 0.2231 - 0.3347 = \underline{\underline{0.4422}} \end{aligned}$$

- b) i. Vi vet ikke om kampen blir spilt på hjemmebane eller bortebane, så vi må ta hensyn til dette i vår utregning. La oss kalle hendelsen “en kamp blir spilt på hjemmebane” for H og “en kamp blir spilt på bortebane” for B . Hvis vi lar X være antall skårte mål i en tilfeldig kamp og bruker loven om total sannsynlighet, har vi:

$$P(X=3) = P(H) \cdot P(X = 3|H) + P(B) \cdot P(X = 3|B)$$

Siden det spilles like mange kamper hjemmebane som bortebane, er det 50/50 sjanse for at en tilfeldig kamp spilles på bortebane eller hjemmebane. Dette gir oss $P(H) = P(B) = 0.50$. Vi får da:

$$\begin{aligned} P(X = 3) &= 0.50 \cdot P(X = 3|H) + 0.50 \cdot P(X = 3|B) \\ &= 0.50 \cdot \frac{1.5^3}{3!} e^{-1.5} + 0.50 \cdot \frac{1.0^3}{3!} e^{-1.0} \approx \underline{\underline{0.0934}} \end{aligned}$$

- ii. Vi skal finne sannsynligheten for at en kamp spilles på hjemmebane, gitt at det ble skåret 3 mål, altså $P(H|X = 3)$. Hvis vi bruker Bayes lov, får vi:

$$P(H|X = 3) = \frac{P(H)P(X = 3|H)}{P(X = 3)} = \frac{0.50 \cdot \frac{1.5^3}{3!} e^{-1.5}}{0.0934} \approx \underline{\underline{0.672}}$$

- c) Hvis en kamp varer i 90 minutter og forventet antall mål i en kamp er 1.0, har vi en intensitet på $\alpha = \frac{1}{90}$ mål per minutt. La T være tiden det tar før første mål. Vi har da at T er eksponensiellfordelt med parameter α , altså $T \sim \exp(\alpha) = \exp(\frac{1}{90})$. Vi får da:

$$P(T \leq 5) = F(5) = 1 - e^{-\frac{1}{90} \cdot 5} = 1 - 0.94596 \approx \underline{\underline{0.054}}$$

Oppgave 4

- a) La X være resultatet av en måling av arsenkonsentrasjon (målt i μg per liter vann). Vi får oppgitt at $X \sim N(5, 0.10^2)$. Vi får da:

$$\begin{aligned} P(4.9 < X < 5.1) &= P(4.9 < X \leq 5.1) = P(X \leq 5.1) - P(X \leq 4.9) \\ &= G\left(\frac{5.1 - 5.0}{0.10}\right) - G\left(\frac{4.9 - 5.0}{0.10}\right) = G(1) - G(-1) \\ &= G(1) - (1 - G(1)) = 2G(1) - 1 \stackrel{\text{tabell}}{=} 2 \cdot 0.8413 - 1 = \underline{\underline{0.6826}} \end{aligned}$$

- b) Vi setter opp følgende hypotese:

$$H_0 : \mu = 5.0 \quad \text{mot} \quad H_1 : \mu > 5.0$$

Vi får oppgitt $\alpha = 0.05$. Fra tabellen finner vi da $U_{0.05} = 1.645$. Fra formelarket har vi testobservatoren:

$$U_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}},$$

der n er antall målinger, altså $n = 5$. Vi må regne ut gjennomsnittet:

$$\bar{X} = \frac{1}{5}(5.12 + 5.15 + 4.99 + 5.04 + 5.10) = 5.08$$

Dette gir oss følgende verdi for testobservatoren:

$$U_0 = \frac{5.08 - 5.0}{\frac{0.10}{\sqrt{5}}} = 1.789$$

Siden $U_0 > U_\alpha$, kan vi forkaste H_0 (eller påstå H_1), og vi konkluderer med at vi har grunnlag til å påstå at konsentrasjonen av arsen er høyere enn $5.0 \mu g$ per liter vann.

- c) Signifikanssannsynligheten, p , er definert som sannsynligheten for å få et resultat lik eller mer ekstremt enn det vi målte, gitt at H_0 er sann. I denne settingen er det gjennomsnittet av målingene som er resultatet vi er interessert i. Dersom H_0 stemmer, så betyr det at $\mu = 5.0$. Fra våre målinger fikk vi et gjennomsnitt på 5.08. Med andre ord:

$$\begin{aligned} p &= P(\bar{X} \geq 5.08 | \mu = 5.0) = 1 - P(\bar{X} < 5.08 | \mu = 5.0) = 1 - P(\bar{X} \leq 5.08 | \mu = 5.0) \\ &= 1 - G\left(\frac{5.08 - 5.0}{\frac{0.10}{\sqrt{5}}}\right) \approx 1 - G(1.79) \stackrel{\text{tabell}}{=} 1 - 0.9633 = 0.0367 \approx \underline{\underline{0.037}} \end{aligned}$$

Vi kunne brukt denne til å avgjøre hypotesetesten i forrige oppgave: siden $p < \alpha$, kan vi forkaste H_0 . Vi ser derimot at dersom vi hadde brukt $\alpha = 0.01$, kunne vi ikke forkastet H_0 (siden $p > \alpha$), dvs. da hadde vi ikke hatt grunnlag til å påstå at konsentrasjonen var høyere enn $5.0\mu g$ per liter vann.

- d) Styrkefunksjonen, $\beta(\theta)$, er definert som sannsynligheten for å forkaste H_0 som funksjon av θ . I vår oppgave er θ forventningen til X , dvs. μ , som vi får oppgitt til å være $5.1\mu g$. Vi forkaster H_0 dersom testobservatoren vår, U_0 , er høyere enn kvantilen, $U_{0.05}$. Vi får da:

$$P(U_0 > U_{0.05} | \mu = 5.1) = P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > U_{0.05} | \mu = 5.1\right)$$

Dersom vi skriver om ulikheten $\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > U_{0.05}$, får vi:

$$\begin{aligned} P(\bar{X} > \mu_0 + U_{0.05} \frac{\sigma}{\sqrt{n}} | \mu = 5.1) &= P(\bar{X} > 5.0 + 1.645 \frac{0.10}{\sqrt{5}} | \mu = 5.1) \\ &= P(\bar{X} > 5.0736 | \mu = 5.1) \\ &= 1 - P(\bar{X} \leq 5.0736 | \mu = 5.1) \\ &= 1 - G\left(\frac{5.0736 - 5.1}{\frac{0.10}{\sqrt{5}}}\right) \\ &\approx 1 - G(-0.59) = 1 - (1 - G(0.59)) \\ &= G(0.59) \stackrel{\text{tabell}}{=} 0.7224 \approx \underline{\underline{0.72}} \end{aligned}$$

Kommentar: her kunne vi regnet ut $P(U_0 > U_{0.05} | \mu = 5.1)$ uten å skrive om ulikheten siden vi har oppgitt fordelingen til U_0 i formelheftet (det er heller ikke så vanskelig å finne ut hva fordelingen til U_0 er). Vi kunne rett og slett standardisert direkte der vi bruker forventningen og variansen til U_0 , som står oppgitt i formelheftet.