

Løsningsforslag, eksamen statistikk, juni 2015

Oppgave 1

Siden det spørres om tall fra et intervall, som oppgaven viser kan være et reelle, er det tydelig at tallene er tatt fra en kontinuerlig fordeling. Siden det ikke sies noe mer om denne fordelingen enn at tallene velges tilfeldig, kan vi anta at det er snakk om en uniform fordeling, eller rektangulærfordeling om man vil. Sannsynlighetstettheten har da formen:

$$f(x) = \begin{cases} a, & \text{hvis } x \in [0, 1] \\ 0, & \text{ellers.} \end{cases}$$

der a er en konstant som vi finner ved å kreve at

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^1 a \, dx = 1$$

Utfører vi integralet, finner vi at $a = 1$. Sannsynligheten for at et tilfeldig valgt tall ligger mellom 0.5 og 1, blir da

$$P(0.5 \leq X \leq 1) = \int_{0.5}^1 f(x) dx = \int_{0.5}^1 dx = [x]_{0.5}^1 = 1 - 0.5 = 0.5$$

Altså 50%. Dette kunne vi lett argumentert for uten å sette opp et så formelt regnestykke som vi har gjort over. Siden intervallet $[0.5, 1]$ står for halvparten av tallene fra rektangulærfordelingen, er det opplagt at sannsynligheten må være 50%.

La Y være antall tall som ligger i dette intervallet blant de 7 tallene som blir generert. Siden sannsynligheten er det samme for hvert tall, har vi at $Y \sim \text{bin}(7, 0.5)$. Sannsynligheten for at nøyaktig 3 tall er i det oppgitte intervallet, blir da:

$$P(Y = 3) = \binom{7}{3} 0.5^3 (1 - 0.5)^{7-3} = 35 \cdot 0.125 \cdot 0.0625 \approx \underline{\underline{0.27}}$$

Oppgave 2

Siden alle lineærkombinasjoner av normalfordelte, stokastiske variable også er normalfordelte, vil Z være normalfordelt med forventning:

$$E(Z) = E(X - 2Y) = E(X) - 3E(Y) = 10 - 3 \cdot 3 = 1$$

Siden X og Y er uavhengige, blir variansen:

$$\text{Var}(Z) = \text{Var}(X - 3Y) = \text{Var}(X) + (-3)^2 \text{Var}(Y) = 2^2 + 9 \cdot 0.1^2 = 4.09$$

Vi får da:

$$\begin{aligned} P(Z > 2) &= 1 - P(Z \leq 2) = 1 - G\left(\frac{2-1}{\sqrt{4.09}}\right) \approx 1 - G(0.49) \\ &= 1 - 0.6874 = 0.3126 \approx \underline{\underline{0.31}} \end{aligned}$$

Oppgave 3

a) Siden utfallsrommet er $V_x = \{0, 1, 2, 3\}$ må vi ha at:

$$\sum_{x=0}^3 a \left(\frac{1}{2}\right)^x = 1.$$

Vi får da:

$$\begin{aligned} 0.5^0 a + 0.5^1 a + 0.5^2 a + 0.5^3 a &= 1 \\ a(0.5^0 + 0.5^1 + 0.5^2 + 0.5^3) &= 1 \end{aligned}$$

som gir oss

$$a = \frac{1}{0.5^0 + 0.5^1 + 0.5^2 + 0.5^3} = \frac{1}{1.875} = \frac{8}{\underline{\underline{15}}}$$

b) Forventinga blir:

$$\begin{aligned} \mu = E(S) &= \sum_{x=0}^3 x \cdot P(S = x) \\ &= \sum_{x=0}^3 x \cdot \frac{8}{15} \left(\frac{1}{2}\right)^x \\ &= \frac{8}{15} (0 \cdot 0.5^0 + 1 \cdot 0.5^1 + 2 \cdot 0.5^2 + 3 \cdot 0.5^3) = \frac{11}{\underline{\underline{15}}} \approx 0.73 \end{aligned}$$

Variansen blir:

$$\begin{aligned}\sigma^2 = \text{Var}(X) &= \sum_{x=0}^3 x^2 \cdot P(X=x) - \mu^2 \\ &= \sum_{x=0}^3 x^2 \cdot \frac{8}{15} \left(\frac{1}{2}\right)^x - \left(\frac{11}{15}\right)^2 \\ &= \frac{8}{15} (0^2 \cdot 0.5^0 + 1^2 \cdot 0.5^1 + 2^2 \cdot 0.5^2 + 3^2 \cdot 0.5^3) - \left(\frac{11}{15}\right)^2 \\ &= \frac{7}{5} - \left(\frac{11}{15}\right)^2 = \frac{194}{225} \approx \underline{\underline{0.86}}\end{aligned}$$

c) Vi definerer to utfall:

A: Studenten dupper av minst 1 gang.

B: Studenten dupper av nøyaktig 3 ganger.

Oppgaven ber oss finne sannsynligheten for at studenten dupper av nøyaktig 3 ganger, gitt at vi vet at han/ho har duppa av minst 1 gang. Med andre ord, vi skal finne $P(B|A)$. Vi bruker loven om betinget sannsynlighet og får:

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

$P(B \cap A)$ er sannsynligheten for at en student både dupper av nøyaktig 3 ganger og minst 1 gang. Dersom student dupper av nøyaktig 3 ganger, har han/ho automatisk duppet av minst 1 gang. Vi får da at $P(B \cap A) = P(B)$ slik at:

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

Sannsynligheten for at studenten dupper av nøyaktig 3 ganger er:

$$P(B) = P(S=3) = \frac{8}{15} \left(\frac{1}{2}\right)^3 = \frac{1}{15}.$$

Sannsynligheten for at studenten dupper av minst 1 gang er:

$$P(A) = P(S \geq 1) = \frac{8}{15} \left(\frac{1}{2}\right)^1 + \frac{8}{15} \left(\frac{1}{2}\right)^2 + \frac{8}{15} \left(\frac{1}{2}\right)^3 = \frac{7}{15}$$

Vi får da:

$$P(B|A) = \frac{P(B)}{P(A)} = \frac{1/15}{7/15} = \underline{\underline{\frac{1}{7}}}$$

- d) La antall avduppinger for student i være angitt ved en stokastisk variabel X_i , der $i = 1, 2, 3, \dots$ og der hver X_i har fordelingen gitt i starten av oppgaven. La S angi det totale antall avduppinger iløpet av en time. Vi har da:

$$S = \sum_{i=1}^{300} X_i$$

Siden hver X_i har samme fordeling, og antallet av X_i 'er vi legger sammen er stor (større enn 30), sier sentralgrensesetningen at:

$$S \stackrel{\text{tilnærm.}}{\sim} N(n\mu, n\sigma^2) = N(300 \cdot 11/15, 300 \cdot 194/225) \approx N(220, \frac{776}{3})$$

Vi skal regne ut:

$$P(200 \leq S \leq 250) = P(S \leq 250) - P(S < 200)$$

Dette er tilfelle uansett om S er tilnærmet normalfordelt eller ikke. Siden S egentlig er diskret, har vi $P(S < 200) = P(S \leq 199)$. Vi bruker så normaltilnærmelsen (med heltallskorreksjon):

$$\begin{aligned} P(S \leq 250) - P(S \leq 199) &\approx G\left(\frac{250 + 0.5 - 220}{\sqrt{\frac{776}{3}}}\right) - G\left(\frac{199 + 0.5 - 220}{\sqrt{\frac{776}{3}}}\right) \\ &\approx G(1.90) - G(-1.21) = G(1.90) - (1 - G(1.27)) \\ &= G(1.90) + G(1.27) - 1 = 0.9713 + 0.8980 - 1 = 0.8693 \\ &\approx \underline{\underline{0.869}} \end{aligned}$$

Kommentar 1: Det er også mulig å si at S er tilnærmet normalfordelt allerede før man setter opp regnestykket. Forskjellen i regnestykket blir da at vi får $P(S < 200) = P(S \leq 200)$ dersom vi behandler S som kontinuerlig. I dette tilfelle ville faktisk heltallskorrekksjonen $+0.5$ gitt en mindre korrekt tilnærming siden det egentlig skal være $P(S < 200) = P(S \leq 199)$. Riktig bruk av heltallskorreksjon i dette tilfelle, ville vært å legge til -0.5 istedenfor $+0.5$. Vi hadde da endt opp med samme verdi som når vi regnet ut $P(S \leq 199)$ med $+0.5$ heltallskorreksjon.

Kommentar 2: Det er også mulig å tolke oppgaven slik at “mellom 200 og 250” betyr at 200 og 250 ikke skal være med, altså at man skal regne ut $P(200 < S < 250)$. Dette ansees som en akseptabel måte å tolke oppgaveteksten og vil ikke gi trekk.

Oppgave 4

Vi definerer to utfall:

A: Artikkel nr. 1 blir akseptert.

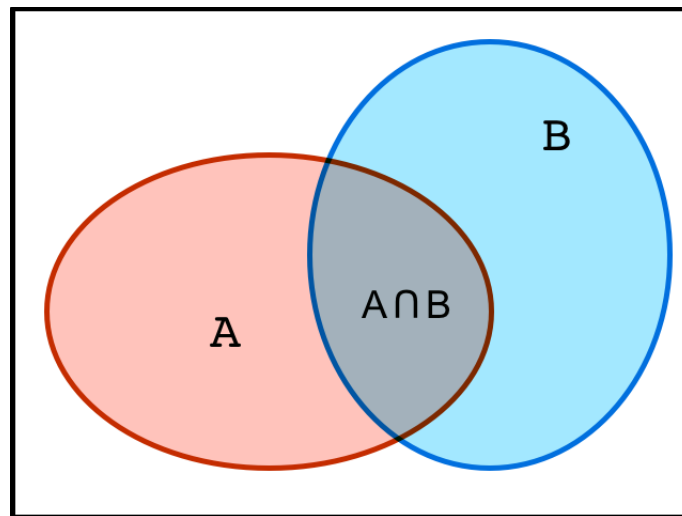
B: Artikkel nr. 2 blir akseptert.

I oppgaveteksten får vi oppgitt at $P(A) = 0.7$ og $P(B) = 0.4$. Sannsynligheten for at minst 1 artikkel blir avvist, blir da unionen av $\bar{A}$ (artikkel 1 blir avvist) og $\bar{B}$ (artikkel 2 blir avvist). Med andre ord: $P(\bar{A} \cup \bar{B}) = 0.75$.

Vi skal finne sannsynligheten for at minst 1 artikkel blir akseptert, altså $P(A \cup B)$. Vi buker den generelle addisjonssetningen og får:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

For å se sammenhengen mellom $P(A \cap B)$ og $P(\bar{A} \cup \bar{B})$, er det en fordel å tegne et venn-diagram:



$P(\bar{A} \cup \bar{B})$ er det området hvor minst 1 artikkel blir avvist. Dette blir alt som ligger enten utenfor A , utenfor B eller begge deler. Dette er alt utenom det området hvor begge blir akseptert, altså $P(A \cap B)$. Vi har med andre ord:

$$P(A \cap B) = 1 - P(\bar{A} \cup \bar{B}) = 1 - 0.75 = 0.25.$$

Vi får da:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7 + 0.4 - 0.25 = \underline{\underline{0.85}}$$

Oppgave 5

Siden både forventinga og variansen er ukjent, må vi bruke t-fordelingen. Konfidenintervallet er da gitt ved:

$$\bar{y} \pm t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}$$

Vi finner den estimerte variansen, S^2 :

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right)$$

Vi har

$$\bar{y} = \frac{1}{19} \sum_{i=1}^{19} y_i = \frac{1}{19} \cdot 5261 = 276.8947368$$

Den estimerte variansen blir da:

$$\begin{aligned} S^2 &= \frac{1}{19-1} \left(\sum_{i=1}^{19} y_i^2 - n\bar{y}^2 \right) = \frac{1}{18} (1469945 - 19 \cdot (276.8947368)^2) \\ &= 733.4327485 \end{aligned}$$

Dette gir oss $S = \sqrt{733.4327485} = 27.08196353$. Siden vi skal finne et 95% konfidensintervall, har vi $\alpha = 0.05$. Fra tabellen finner vi $t_{\alpha/2, n-1} = t_{0.025, 18} = 2.101$. Konfidensintervallet blir dermed:

$$\begin{aligned} &276.8947368 \pm 2.101 \cdot \frac{27.08196353}{\sqrt{19}} \\ &276.8947368 \pm 13.05257318 \end{aligned}$$

Dersom vi runder av, får vi dermed [264, 290]

Oppgave 6

a) Vi skal finne $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$. Fra formelheftet, har vi:

$$\hat{\beta}_1 = \frac{1}{M} \left(\sum_{i=1}^n x_i Y_i - \bar{x}\bar{Y} \right)$$

hvor

$$M = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

Vi har:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{9} \cdot 234 = 26$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{9} \cdot 811.3 = 90.1444\dots$$

Vi får da:

$$M = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = 10144 - 9 \cdot (26)^2 = 4060$$

$$\hat{\beta}_1 = \frac{1}{M} \left(\sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y} \right) = \frac{1}{4060} (24628.6 - 9 \cdot 26 \cdot 90.1444\dots) = 0.8706403941 \approx 0.87$$

Videre har vi

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} = 90.1444\dots - 0.8706403941 \cdot 26 \approx 67.5$$

Den estimerte regresjonslinjen blir dermed:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x = \underline{\underline{67.5 + 0.87x}}$$

b) Vi setter opp følgende hypoteser:

$$H_0 : \beta_1 = 0.85 \quad \text{mot} \quad H_1 : \beta_1 > 0.85$$

For å utføre testen, bruker vi testobservatoren:

$$U_0 = \frac{\hat{\beta}_1 - \beta_1^0}{\frac{\sigma}{\sqrt{M}}} = \frac{0.8706403941 - 0.85}{\sqrt{\frac{0.01}{4060}}} \approx 13.15$$

Vi sammenlikner denne med $U_{0.01} = 2.326$ som vi finner i tabellen. Siden $U_0 > U_{0.01}$, kan vi forkaste nullhypotesen og vi har dermed grunnlag til å påste at tallet 0.85 var for lite.