Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

Rettet 30. desember 2015 og seinere.
Sist oppdatert 11. oktober 2018.

ST0103 2015H – løsningsskisse

Oppgave 1

a) Hver av de n forsøkene (boringene) må resultere i suksess (mineralet blir funnet) eller fiasko (mineralet blir ikke funnet), og sannsynligheten for suksess må være den samme (p) hver gang. Disse forutsetningene er oppfylt ifølge oppgaveteksten. I tillegg må utfallene (suksess eller fiasko) være uavhengig mellom forsøkene.

$P(3 \leq X \leq 4) = P(X = 3) + P(X = 4) = \binom{20}{3} \cdot 0{,}3^3 \cdot (1-0{,}3)^{20-3} + \binom{20}{4} \cdot 0{,}3^4 \cdot (1-0{,}3)^{20-4} = 0{,}072 + 0{,}130 = 0{,}202$. Eller ved å bruke tabell: $P(3 \leq X \leq 4) = P(X \leq 4) - P(X \leq 2) = 0{,}238 - 0{,}035 = 0{,}203$ (liten avrundingsfeil).

b) Forventningsverdien til X er $EX = np = 200 \cdot 0{,}3 = 60$ og variansen $\operatorname{Var} X = np(1-p) = 200 \cdot 0{,}3 \cdot 0{,}7 = 42$, som gir standardavvik $\sqrt{\operatorname{Var} X} = \sqrt{42} = 6{,}48$.

Siden $\operatorname{Var} X \geq 5$, kan vi bruke normaltilnærming, som med heltallskorreksjon gir

$$\begin{aligned} P(50 \leq X \leq 75) = P(49{,}5 \leq X \leq 75{,}5) &= P\left(\frac{49{,}5 - 60}{6{,}48} \leq \frac{X - 60}{6{,}48} \leq \frac{75{,}5 - 60}{6{,}48}\right) \\ &\approx P(-1{,}62 \leq Z \leq 2{,}39) = P(Z \leq 2{,}39) - P(Z < -1{,}62) \\ &= 0{,}9916 - 0{,}0526 = 0{,}939, \end{aligned}$$

der Z er standardnormalfordelt og den nest siste likheten følger fra oppslag i tabell. Hvis heltallskorreksjon ikke brukes, får vi i stedet

$$\begin{aligned} P(50 \leq X \leq 75) &= P\left(\frac{50 - 60}{6{,}48} \leq \frac{X - 60}{6{,}48} \leq \frac{75 - 60}{6{,}48}\right) \approx P(-1{,}54 \leq Z \leq 2{,}31) \\ &= P(Z \leq 2{,}31) - P(Z < -1{,}54) = 0{,}9896 - 0{,}0618 = 0{,}928, \end{aligned}$$

mens det eksakte svaret er $\sum_{x=50}^{75} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=50}^{75} \binom{200}{x} 0{,}3^x \cdot 0{,}7^{200-x} = 0{,}940$.

c)

$$
\begin{aligned}
E\hat{p} &= E\Big(\frac{1}{n+m}(X+Y)\Big) = \frac{1}{n+m}E(X+Y) = \frac{1}{n+m}(EX+EY) \\
&= \frac{1}{n+m}(np+mp) = \frac{1}{n+m}(n+m)p = p
\end{aligned}
$$

og

$$
\begin{aligned}
E\tilde{p} &= E\Big(\frac{1}{2}\Big(\frac{1}{n}X + \frac{1}{m}Y\Big)\Big) = \frac{1}{2}E\Big(\frac{1}{n}X + \frac{1}{m}Y\Big) = \frac{1}{2}\Big(\frac{1}{n}EX + \frac{1}{m}EY\Big) \\
&= \frac{1}{2}\Big(\frac{1}{n}np + \frac{1}{m}mp\Big) = \frac{1}{2}(p+p) = p,
\end{aligned}
$$

så både $\hat{p}$ og $\tilde{p}$ er forventningsrette.

$$
\begin{aligned}
\operatorname{Var}\hat{p} &= \operatorname{Var}\left(\frac{1}{n+m}(X+Y)\right) = \left(\frac{1}{n+m}\right)^2 \operatorname{Var}(X+Y) = \frac{1}{(n+m)^2}(\operatorname{Var}X + \operatorname{Var}Y) \\
&= \frac{1}{(n+m)^2}(np(1-p) + mp(1-p)) = \frac{1}{(n+m)^2}(n+m)p(1-p) \\
&= \frac{1}{n+m}p(1-p)
\end{aligned}
$$

og

$$
\begin{aligned}
\operatorname{Var}\tilde{p} &= \operatorname{Var}\left(\frac{1}{2}\Big(\frac{1}{n}X + \frac{1}{m}Y\Big)\right) = \left(\frac{1}{2}\right)^2 \operatorname{Var}\left(\frac{1}{n}X + \frac{1}{m}Y\right) \\
&= \frac{1}{4}\left(\left(\frac{1}{n}\right)^2 \operatorname{Var}X + \left(\frac{1}{m}\right)^2 \operatorname{Var}Y\right) = \frac{1}{4}\Big(\frac{1}{n^2}np(1-p) + \frac{1}{m^2}mp(1-p)\Big) \\
&= \frac{1}{4}\Big(\frac{1}{n}p(1-p) + \frac{1}{m}p(1-p)\Big) = \frac{1}{4}\Big(\frac{1}{n} + \frac{1}{m}\Big)p(1-p).
\end{aligned}
$$

Setter vi inn $n = 20$ og $m = 10$, får vi $\operatorname{Var}\hat{p} = p(1-p)/(20+10) = 0{,}0333p(1-p)$, mens $\operatorname{Var}\tilde{p} = (1/20 + 1/10)p(1-p)/4 = 0{,}0375p(1-p)$. Siden $\hat{p}$ har minst varians, foretrekker vi den.[1]

[1] Vi kan også vise at $\operatorname{Var}\hat{p} \leq \operatorname{Var}\tilde{p}$ for alle n og m: $\operatorname{Var}\hat{p} \leq \operatorname{Var}\tilde{p}$ er ekvivalent med at $1/(n+m) \leq (1/n + 1/m)/4$. Multipliserer vi ulikheten med $4nm(n+m)$, får vi den ekvivalente ulikheten $4nm \leq m(n+m) + n(n+m) = (n+m)^2$. Subtraherer vi $4nm$ fra hver side, blir ulikheten $0 \leq (n+m)^2 - 4nm = (n-m)^2$ (verifiser den siste identiteten). Men denne ulikheten er *alltid* sann (kvadratet av et tall er aldri negativt), så $\operatorname{Var}\hat{p} \leq \operatorname{Var}\tilde{p}$ uansett hva n og m er. Vi har likhet bare når $n = m$, men det er lett å verifisere at estimatorene i dette tilfellet er like.

Oppgave 2

a) Antall ganger mønsteret forekommer, X, er poissonfordelt med parameter $0{,}03 \cdot 100 = 3$. $P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 1 - 3^0 e^{-3}/0! - 3^1 e^{-3}/1! = 1 - (1+3)e^{-3} = 0{,}8009$. Eller fra tabell: $P(X \geq 2) = 1 - P(X \leq 1) = 0{,}1991 = 0{,}8009$.

$P(X = 2 \mid X \geq 2) = P(X = 2 \ \cap \ X \geq 2)/P(X \geq 2) = P(X = 2)/P(X \geq 2) = 3^2 e^{-3}/2!/0{,}8009 = 0{,}280$.

b) Avstand til neste sted mønsteret forekommer i poissonprossesen, som er i lengde, tilsvarer det vi kaller ventetid i en poissonprosess i tid. Ventetida er eksponentielt fordelt med intensiteten i prosessen som parameter.

Hvis Y er avstanden, er dermed $P(Y > 50) = 1 - P(Y \leq 50) = 1 - (1 - e^{-\lambda \cdot 50}) = e^{-0{,}03 \cdot 50} = 0{,}22$.

c) La Y_1, Y_2 og Y_3 være de tre avstandene. At den minste av dem er større enn $1/\lambda$, er det samme som at alle tre er større enn $1/\lambda$, og $P(Y_1 > 1/\lambda \ \cap \ Y_2 > 1/\lambda \ \cap \ Y_3 > 1/\lambda) = (P(Y > 1/\lambda))^3 = (1 - P(Y \leq 1/\lambda))^3 = (1 - (1 - e^{-\lambda \cdot 1/\lambda}))^3 = (e^{-\lambda \cdot 1/\lambda})^3 = (e^{-1})^3 = e^{-3} = 0{,}050$, der Y er en variabel med samme fordeling som Y_1, Y_2 og Y_3.

Oppgave 3

a)

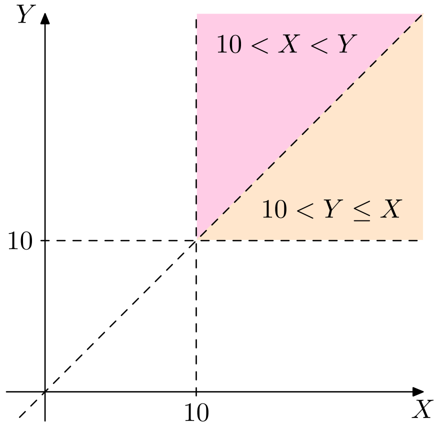
$$P(X > 10{,}0) = P\left(\frac{X - 9{,}0}{2{,}0} > \frac{10{,}0 - 9{,}0}{2{,}0}\right) = P(Z > 0{,}50) = 1 - P(Z \leq 0{,}50)$$
$$= 1 - 0{,}6915 = 0{,}3085,$$

der Z er standardnormalfordelt.

Gjennomsnittet av X og Y, $\bar{X} = \frac{1}{2}(X+Y) = \frac{1}{2}X + \frac{1}{2}Y$, er en lineærkombinasjon av uavhengige normalfordelte variabler, og dermed normalfordelt. Forventningsverdien er $E\bar{X} = E(\frac{1}{2}(X+Y)) = \frac{1}{2}E(X+Y) = \frac{1}{2}(EX + EY) = \frac{1}{2}(9{,}0 + 9{,}0) = 9{,}0$ og variansen $\operatorname{Var}\bar{X} = \operatorname{Var}(\frac{1}{2}(X+Y)) = (\frac{1}{2})^2 \operatorname{Var}(X+Y) = \frac{1}{4}(\operatorname{Var}X + \operatorname{Var}Y) = \frac{1}{4}(2{,}0^2 + 2{,}0^2) = 2{,}0$, slik at

$$P(\bar{X} > 10{,}0) = P\left(\frac{\bar{X} - 9{,}0}{\sqrt{2{,}0}} > \frac{10{,}0 - 9{,}0}{\sqrt{2{,}0}}\right) = P(Z > 0{,}71) = 1 - P(Z \leq 0{,}71)$$
$$= 1 - 0{,}7611 = 0{,}24.$$

b) Hendelsen $X > 10 \ \cap \ Y > 10$ (farget område) kan skrives som unionen av to disjunkte hendelser, $10 < X < Y$ (rosa område) og $10 < Y \leq X$ (oransje område). Fordi X og Y er kontinuerlige og uavhengige og har samme fordeling, er sannsynligheten for disse to disjunkte hendelsene like, slik at $P(10 < X < Y) = \frac{1}{2}P(X > 10 \ \cap \ Y > 10) = \frac{1}{2}P(X > 10)\,P(Y > 10) = \frac{1}{2} \cdot 0{,}3085^2 = 0{,}048$.



Oppgave 4

a) $\left[s\sqrt{(n-1)/\chi_{0,025}}, s\sqrt{(n-1)/\chi_{0,975}}\right]$ er et 95 %-konfidensintervall for standardavviket, σ, der $n = 8$ er antall observasjoner, $s = 0{,}00138$ er utvalgsstandardavviket, og $\chi_{0,025} = 16{,}013$ og $\chi_{0,975} = 1{,}690$ henholdsvis er 0,975- og 0,025-kvantil (eller *øvre* 0,025- og 0,975-kvantil) i en khikvadratfordeling med $n-1 = 7$ frihetsgrader (fra tabell). Vi får intervallet $\left[0{,}00138\sqrt{7/16{,}013}, 0{,}00138\sqrt{7/1{,}690}\right] = [0{,}00091, 0{,}00281]$.

b) Hvis nullhypotesen er sann, er testobservatoren

$$T = \frac{\bar{X} - \bar{Y}}{S_\mathrm{p}\sqrt{1/n + 1/m}}$$

t-fordelt med $n - m - 2$ frihetsgrader, der n og m er utvalgsstørrelsene, $\bar{X}$ og $\bar{Y}$ gjennomsnittene i de to utvalgene og S_p^2 er interpolert utvalgsvarians. En liten eller en stor verdi av T tyder på at det er den alternative hypotesen som er riktig, slik at nullhypotesen skal forkastes. Med signifikansnivå på 0,001 blir kritiske verdier 0,0005- og 0,9995-kvantilene til t-fordelingen med 13 frihetsgrader. De har samme absoluttverdi, men motsatt fortegn, og er lik $\pm 4{,}221$ (fra tabell). Vi forkaster altså nullhypotesen hvis $T \leq -4{,}221$ eller $T \geq 4{,}221$.

Interpolert utvalgsvarians er definert ved $S_\mathrm{p}^2 = ((n-1)S_x^2 + (m-1)S_y^2)/(n+m-2)$, der S_x^2 og S_y^2 er utvalgsvariansene. Med våre tall får S_p verdien

$$s_\mathrm{p} = \sqrt{\frac{7 \cdot 0{,}00138^2 + 6 \cdot 0{,}00014^2}{8 + 7 - 2}} = 0{,}00102,$$

og T får verdi

$$t = \frac{2{,}29947 - 2{,}31011}{0{,}00102\sqrt{1/8 + 1/7}} = -20{,}2.$$

Vi forkaster altså nullhypotesen, og påstår at forventningsverdien for målt masse ved de to metodene er ulike.

Testmetoden forutsetter at de to utvalgene, som skal være uavhengige, består av uavhengige observasjoner fra hver sin normalfordeling, som begge skal ha samme varians. Om den siste forutsetningen er oppfylt her, er tvilsomt, siden utvalgsstandardavvikene er så forskjellige (det ene er ti ganger større enn det andre).

(Lord Rayleighs forsøk bidrog til oppdagelsen av argon, som utgjør omtrent 0,93 % av lufta. Når nitrogen ble utvunnet fra luft, skjedde det ved at oksygen ble fjernet. Det ble antatt at nesten all gassen som var igjen var nitrogen. Men argon, som er tyngre enn nitrogen, var også igjen, og blandingen fikk større tetthet enn ren nitrogen.)