Norges teknisk–naturvitenskapelige universitet *Rettet 12. desember og 18. desember*
Institutt for matematiske fag

ST0103 2014H – løsningsskisse

Oppgave 1

a) La X være radius til blomsten. Sannsynligheten for at radius er mindre enn 17 mm, er

$$P(X < 17) = P\left(\frac{X-15}{2} < \frac{17-15}{2}\right) = P(Z < 1) = 0{,}8413,$$

(fra tabell), der Z er standard normalfordelt. Antall blomster med radius mindre enn 17 mm er binomisk fordelt med parametre $n = 10$ og $p = 0{,}8413$. Sannsynligheten for at 8 eller flere har radius mindre enn 17 mm, er $\binom{10}{8}0{,}8413^8(1-0{,}8413)^2 + \binom{10}{9}0{,}8413^9(1-0{,}8413)^1 + \binom{10}{10}0{,}8413^{10}(1-0{,}8413)^0 = 0{,}797$.

b) Siden omkretsen er radien multiplisert med en konstant, er også omkretsen normalfordelt. Forventningsverdien er $E(2\pi X) = 2\pi EX = 2\pi \cdot 15 = 30\pi$. Variansen er $\operatorname{Var}(2\pi X) = (2\pi)^2 \operatorname{Var} X = (2\pi)^2 \cdot 2^2$, slik at standardavviket er 4π.

c) Et 99 %-konfidensintervall har grenser $\bar{x} \pm t_{\alpha/2} s/\sqrt{n}$. Her er $\bar{x} = 19{,}0$, $\alpha = 0{,}01$, $t_{0{,}005} = 2{,}861$ (fra tabell, 19 frihetsgrader), $s = 3{,}0$ og $n = 20$, og vi får grenser $19{,}0 \pm 2{,}861 \cdot 3{,}0/\sqrt{20} = 19{,}0 \pm 1{,}9$, og konfidensintervallet blir $[17{,}1,\ 20{,}9]$.

d) Vi bruker testobservatoren

$$\frac{\bar{X}-17}{S/\sqrt{n}},$$

som er t-fordelt med 19 frihetsgrader hvis $\mu = 17$. Vi forkaster nullhypotesen for *store* verdier av testobservatoren – nærmere bestemt hvis den er større enn eller lik $t_{0{,}05} = 1{,}729$ (fra tabell, 19 frihetsgrader) – da er nemlig sannsynligheten for feilaktig forkastning av nullhypotesen lik 0,05 hvis $\mu = 17$. Vi får verdien $(19{,}0 - 17{,}0)/(3{,}0/\sqrt{20}) = 3{,}0$, og vi forkaster nullhypotesen.

e) La A være hendelsen at biologen plukker en blomst av art A, og la C være hendelsen at radius er mindre enn eller lik 17 mm. Da sier opplysningene i oppgaven at $P(A) = 0{,}4$, $P(\bar{A}) = 0{,}6$, Videre har vi fra (a) at $P(C \mid A) = 0{,}841$, mens vi på samme måte finner $P(C \mid \bar{A}) = P(Z < (17-19)/3) = P(Z < -0{,}67) = 0{,}251$. Loven om total sannsynlighet gir at sannsynligheten for at blomsten har radius mindre enn eller lik 17 mm, er $P(C) = P(C \mid A)P(A) + P(C \mid \bar{A})P(\bar{A}) = 0{,}841 \cdot 0{,}4 + 0{,}251 \cdot 0{,}6 = 0{,}487$.

Fra Bayes' setning har vi at sannsynligheten for at blomsten er av art A hvis den har radius mindre enn eller lik 17 mm, er $P(A \mid C) = P(C \mid A)P(A)/P(C) = 0{,}841 \cdot 0{,}4/0{,}487 = 0{,}691$.

f) Vi lar nå C være hendelsen at radius er mindre enn eller lik c mm. Feilklassifisering inntreffer nøyaktig når en av de to disjunkte hendelsene $\bar{C} \cap A$ eller $C \cap \bar{A}$ inntreffer. Når $c = 17$, er sannsynligheten for dette

$$\begin{aligned} P((\bar{C} \cap A) \cup (C \cap \bar{A})) &= P(\bar{C} \cap A) + P(C \cap \bar{A}) = P(\bar{C} \mid A)P(A) + P(C \mid \bar{A})P(\bar{A}) \\ &= (1 - P(C \mid A))P(A) + P(C \mid \bar{A})P(\bar{A}) \\ &= (1 - 0{,}841) \cdot 0{,}4 + 0{,}251 \cdot 0{,}6 = 0{,}214. \end{aligned}$$

For generell c er $P(C \mid A) = P(Z < (c-15)/2) = G((c-15)/2)$ og $P(C \mid \bar{A}) = P(Z < (c-19)/3) = G((c-19)/3)$ istedenfor henholdsvis 0,841 og 0,251, der G er kumulativ fordelingsfunksjon for standard normalfordeling. Sannsynligheten for feilklassifisering blir nå, som ovenfor, $h(c) = 0{,}4\,(1 - G((c-15)/2)) + 0{,}6\,G((c-19)/3)$. Funksjonen h har minimum når $h'(c) = 0$, det vil si når $-0{,}4G'((c-15)/2)) \cdot \frac{1}{2} + 0{,}6G'((c-19)/3) \cdot \frac{1}{3} = 0$. Nå er G' sannsynlighetstettheten for standard normalfordeling, altså $g(z) = e^{-z^2/2}/\sqrt{2\pi}$, slik at vi har minimum når

$$0{,}4 \cdot \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{c-15}{2}\right)^2} = 0{,}6 \cdot \frac{1}{3}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{c-19}{3}\right)^2},$$

altså når

$$e^{-\frac{1}{2}\left(\frac{c-15}{2}\right)^2} = e^{-\frac{1}{2}\left(\frac{c-19}{3}\right)^2}, \qquad \left(\frac{c-15}{2}\right)^2 = \left(\frac{c-19}{3}\right)^2,$$

som gir

$$\frac{c-15}{2} = \frac{c-19}{3} \qquad \text{eller} \qquad \frac{c-15}{2} = -\frac{c-19}{3}.$$

Dette er to lineære likninger. Bare den siste gir løsning mellom 15 og 19, nemlig $c = 16{,}6$. Den foreløpige artsbestemmelsen bør derfor gjøres ved å klassifisere blomster med mindre radius enn 16,6 mm som art A og resten som B. (Sannsynligheten for feilklassifisering blir da $h(16{,}6) = 0{,}212$, altså ikke store forbedringen fra å sette skillet ved 17 mm.)

Oppgave 2

a) Antall utbrudd X i løpet av de neste 10 årene er poissonfordelt med parameter $0{,}2 \cdot 10 = 2$. Sannsynligheten for to eller flere utbrudd er $P(X \geq 2) = 1 - P(X \leq 1) = 1 - 2^0/0! \cdot e^{-2} - 2^1/1! \cdot e^{-2} = 0{,}59$. Den betingede sannsynligheten for to eller flere utbrudd gitt minst ett utbrudd er $P(X \geq 2 \mid X \geq 1) = P(X \geq 2 \cap X \geq 1)/P(X \geq 1) = P(X \geq 2)/(1 - P(X = 0)) = 0{,}59/(1 - 2^0/0! \cdot e^{-2}) = 0{,}69$.

b) Tida mellom to hendelser i en poissonprossess med intensitet λ er eksponentielt fordelt med parameter λ (det vil si med sannsynlighetstetthet gitt ved $\lambda e^{-\lambda t}$ for $t > 0$). Forventningsverdien er $1/\lambda$, slik at gjennomsnittet $\bar{t}$ av de oppgitte tidsintervallet er et estimat av $1/\lambda$. Et estimat av $\lambda = 1/\frac{1}{\lambda}$ er dermed $1/\bar{t}$. Her er $\bar{t} = (7{,}9 + 2{,}5 + 5{,}4 + 10{,}8 + 10{,}8 + 0{,}3 + 2{,}6 + 8{,}4 + 13{,}7 + 6{,}8)/10 = 6{,}92$, og vi får $1/6{,}92 = 0{,}14$ som et estimat av λ.

Istedenfor å estimere λ som parameter i eksponentiell fordeling, kan vi også estimere λ som parameter i poissonfordeling. Her hadde vi $x = 10$ hendelser i et intervall av lengde $t = 69{,}2$, og vi får estimatet $x/t = 0{,}14$. (Vi finner de nevnte estimatene hvis vi bruker sannsynlighetsmaksimeringsmetoden.)

Oppgave 3

a) Minste kvadraters metode er å velge som estimater av α og β de $\hat{\alpha}$ og $\hat{\beta}$ som gjør $\sum_{i=1}^{10}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$ minst mulig – det vil si som gjør summen av kvadratene av de vertikale avstandene mellom punktene (x_i, y_i), $1 \leq i \leq 10$, og linja $y = \hat{\alpha} + \hat{\beta}x$ minst mulig.

b) Vi skal teste $H_0\colon \beta = 0$ mot $H_1\colon \beta \neq 0$ (også $H_0\colon \beta \leq 0$ mot $H_1\colon \beta > 0$ blir godtatt, da en katalysator skal øke reaksjonsfarten, og vi antakeligvis vil være interessert i om den virker). Vi bruker testobservatoren

$$T = \frac{\hat{\beta}}{\sqrt{s^2/\sum_{i=1}^{10}(x_i - \bar{x})^2}},$$

som er t-fordelt med 8 frihetsgrader hvis $\beta = 0$. Vi forkaster nullhypotesen for *små og store* verdier av testobservatoren – nærmere bestemt hvis $T \geq t_{0{,}025} = 2{,}306$ (fra tabell, 8 frihetsgrader) eller $T \leq -t_{0{,}025} = -2{,}306$ – da er nemlig sannsynligheten for feilaktig forkastning av nullhypotesen lik 0,05 hvis $\beta = 0$. Vi får verdien $t = 1{,}12/\sqrt{2{,}3/4{,}1} = 1{,}50$, og vi forkaster dermed ikke nullhypotesen. Det er grunn til å tvile på om katalysatoren har noen virkning.