Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

Rettet oktober 2018

ST0103 2016S – løsningsskisse

Oppgave 1

a) La A være hendelsen at prøven er alunskifer og S hendelsen at prøven gir svart strek. Opplysningen i oppgaven er da at $P(\bar{A}) = 0{,}8$, $P(S \mid A) = 1$ og $P(S \mid \bar{A}) = 0{,}3$. Sannsynligheten for at en prøve som gir svart strek, er alunskifer, er da ved Bayes' setning

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S \mid A)P(A) + P(S \mid \bar{A})P(\bar{A})}$$
$$= \frac{1 \cdot (1 - 0{,}8)}{1 \cdot (1 - 0{,}8) + 0{,}3 \cdot 0{,}8} = \frac{0{,}2}{0{,}2 + 0{,}24} = \frac{0{,}2}{0{,}44} = 0{,}45.$$

b) Hvis $\mu = 90$, er testobservatoren

$$T = \frac{\bar{X} - 90}{S/\sqrt{n}}$$

t-fordelt med $n - 1$ frihetsgrader, der n er utvalgets størrelse, $\bar{X}$ gjennomsnittlig uraninnhold og S utvalgsstandardavviket. En stor verdi av T tyder på at det er den alternative hypotesen som er riktig, slik at nullhypotesen skal forkastes. Med signifikansnivå på 0,05 blir kritisk verdi 0,95-kvantilen til t-fordelingen med $n - 1 = 10 - 1 = 9$ frihetsgrader, som er 1,833 (fra tabell). Forkastningsområdet er altså fra 1,833 og oppover. Med de oppgitte tallene får vi verdien

$$\frac{90{,}8 - 90}{1{,}4/\sqrt{10}} = 1{,}81.$$

Vi forkaster dermed ikke nullhypotesen, og har ikke noe grunnlag for å påstå at $\mu > 90$. (Hvis $\mu < 90$, som også dekkes av nullhypotesen, er $P(T \geq 1{,}833)$ enda mindre enn 0,05, så vi har sikret oss at sannsynligheten for feilaktig forkastning av nullhypotesen i alle tilfeller høyst er 0,05.)

Oppgave 2

a) For at X skal være binomisk fordelt, må alle karbonatomer i molekylet være ^{13}C eller ikke, og sannsynligheten for at et karbonatom er ^{13}C, skal være den samme for alle karbonatomene i molekylet. Dette er allerede spesifisert i oppgaven. Det som ikke er nevnt, og som vi må anta, er at hvorvidt et karbonatom er ^{13}C eller ikke, skal være uavhengig av hvorvidt hvert av de andre karbonatomene i molekylet er det.

La X være antall ^{13}C-atomer i biopolymermolekylet. X er binomisk fordelt med parametre $n = 100$ og $p = 0{,}011$. $P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{x=0}^{2} \binom{n}{x} p^x (1-p)^{n-x} = 1 - \binom{100}{0} \cdot 0{,}01^0 \cdot 0{,}989^{100} - \binom{100}{1} \cdot 0{,}01^1 \cdot 0{,}989^{99} - \binom{100}{2} \cdot 0{,}011^2 \cdot 0{,}989^{98} = 1 - 0{,}989^{100} - 100 \cdot 0{,}011 \cdot 0{,}989^{99} - \binom{100}{2} \cdot 0{,}011^2 \cdot 0{,}989^{98} = 0{,}0986.$

b) La Y være poissonfordelt med forventningsverdi 1,1.

$$P(Y \le 2) = \frac{1{,}1^0}{0!}e^{-1{,}1} + \frac{1{,}1^1}{1!}e^{-1{,}1} + \frac{1{,}1^2}{2!}e^{-1{,}1} = \Big(1 + 1{,}1 + \frac{1{,}1^2}{2!}\Big)e^{-1{,}1} = 0{,}9004.$$

Siden $n > 10$ og $p < 0{,}1$ i den binomiske fordelingen fra (a), kan vi bruke poissonfordeling med parameter $np = 100 \cdot 0{,}011 = 1{,}1$ som tilnærming. Det gir $P(X \ge 3) \approx P(Y \ge 3) = 1 - P(Y \le 2) = 1 - 0{,}9004 = 0{,}0996$, der den nest siste likheten følger fra utregningen over.

(Her ville vi ha fått et litt dårlig resultat ved normaltilnærming, som i (c), siden $np(1-p) \approx 1 < 5$. Hvis vi hadde gjort det, ville vi ha fått $1 - P\Big(Z \le (2{,}5 - 100 \cdot 0{,}011)/\sqrt{100 \cdot 0{,}011 \cdot 0{,}989}\Big) = 1 - P(Z \le 1{,}34) = 1 - 0{,}9099 = 0{,}090$.)

c) La X være antall ^{13}C-atomer. X er binomisk fordelt med parametre $n = 332\,652$ og $p = 0{,}011$. Forventningsverdien til X er $np = 332\,652 \cdot 0{,}011 = 3659{,}2$ og variansen $np(1-p) = 332\,652 \cdot 0{,}011 \cdot 0{,}989 = 3618{,}9$, som gir standardavvik $\sqrt{3618{,}9} = 60{,}2$.

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

$$\begin{aligned} P(X \ge 3700) = P(X \ge 3699{,}5) &= P\Big(\frac{X - 3659{,}2}{60{,}2} \ge \frac{3699{,}5 - 3659{,}2}{60{,}2}\Big) \\ &\approx P(Z \ge 0{,}67) = 1 - P(Z \le 0{,}67) = 1 - 0{,}7486 = 0{,}251, \end{aligned}$$

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

$$\begin{aligned} P(X \ge 3700) = P(X \ge 3700) &= P\Big(\frac{X - 3659{,}2}{60{,}2} \ge \frac{3700 - 3659{,}2}{60{,}2}\Big) \\ &\approx P(Z \ge 0{,}68) = 1 - P(Z \le 0{,}68) = 1 - 0{,}7517 = 0{,}248, \end{aligned}$$

mens det eksakte svaret er $1 - \sum_{x=0}^{3699} \binom{332\,652}{x} 0{,}011^x \cdot 0{,}989^{332\,652 - x} = 0{,}2508$.

Oppgave 3

a)

$$P(T \le t) = \int_0^t f(x)\,dx = \int_0^t \frac{1}{\mu}e^{-x/\mu}\,dx = [-e^{-x/\mu}]_0^t = 1 - e^{-t/\mu}$$

for $t > 0$. Vi kan også argumentere med at $F'(t) = f(t)$ når $F(t) = 1 - e^{-t/\mu}$, men må da også sjekke at $F(0) = 0$, som den skal (eller at $\lim_{t\to\infty} F(t) = 1$), for det fins uendelig mange funksjoner som har f som derivert, men bare én som er kumulativ sannsynlighetsfordeling.

b) $P(T \geq 20) = 1 - P(T < 20) = 1 - (1 - e^{-20/20}) = e^{-1} = 0{,}368$.

Ved hukommelsesløsheten til den eksponentielle fordelingen er $P(T \geq 30 \mid T \geq 10) = P(T \geq 20) = e^{-1} = 0{,}368$. Eller vi kan bruke definisjonen av betinget sannsynlighet:

$$P(T \geq 30 \mid T \geq 10) = \frac{P(T \geq 30 \cap T \geq 10)}{P(T \geq 10)} = \frac{P(T \geq 30)}{P(T \geq 10)} = \frac{e^{-30/20}}{e^{-10/20}} = e^{-3/2-(-1/2)} = e^{-1} = 0{,}368$$

c) Simultantettheten er gitt ved $L = (\frac{1}{\mu}e^{-t_1/\mu})(\frac{1}{\mu}e^{-t_2/\mu})\cdots(\frac{1}{\mu}e^{-t_n/\mu}) = e^{-\sum t_i/\mu}/\mu^n$, som også er likelihoodfunksjonen sett som en funksjon av μ. Da er $\ln L = -\sum t_i/\mu - n\ln\mu$, som har maksimum når $0 = d\ln L/d\mu = \sum t_i/\mu^2 - n/\mu$, altså $\sum t_i = n\mu$, eller $\mu = \sum t_i/n = \bar{t}$. Sannsynlighetsmaksimeringsestimatoren er altså $\bar{T}$.

$E\bar{T} = E(\frac{1}{n}(T_1 + T_2 + \cdots + T_n)) = \frac{1}{n}E(T_1 + T_2 + \cdots + T_n) = \frac{1}{n}(ET_1 + ET_2 + \cdots + ET_n) = \frac{1}{n}(\mu + \mu + \cdots + \mu) = \frac{1}{n} \cdot n\mu = \mu$. $\operatorname{Var}\bar{T} = \operatorname{Var}(\frac{1}{n}(T_1 + T_2 + \cdots + T_n)) = \frac{1}{n^2}\operatorname{Var}(T_1 + T_2 + \cdots + T_n) = \frac{1}{n^2}(\operatorname{Var}T_1 + \operatorname{Var}T_2 + \cdots + \operatorname{Var}T_n) = \frac{1}{n^2}(\mu^2 + \mu^2 + \cdots + \mu^2) = \frac{1}{n^2} \cdot n\mu^2 = \frac{1}{n}\mu^2$. I den tredje likheten for variansen er det brukt at $T_1, T_2, \ldots, T_n$ er uavhengige. I den fjerde likheten er det brukt at $\operatorname{Var}T_i = \mu^2$ når T_i er eksponentielt fordelt med forventningsverdi μ.

d) Vi finner kumulativ fordelingsfunksjon for $\frac{2}{\mu}T$ uttrykt ved kumulativ fordelingsfunksjon for T: $P(\frac{2}{\mu}T \leq x) = P(T \leq \frac{\mu}{2}x)$. Da kan vi finne sannsynlighetstettheten f_X for $\frac{2}{\mu}T$ ved å derivere,

$$f_X(x) = \frac{d}{dx}P\Big(\frac{2}{\mu}T \leq x\Big) = \frac{d}{dx}P\Big(T \leq \frac{\mu}{2}x\Big) = f\Big(\frac{\mu}{2}x\Big) \cdot \frac{d}{dx}\Big(\frac{\mu}{2}x\Big) = \frac{1}{\mu}e^{-\frac{\mu x/2}{\mu}} \cdot \frac{\mu}{2} = \frac{1}{2}e^{-x/2}$$

for $x > 0$. I den tredje overgangen er kjerneregelen brukt. I tabellsamlingen ser vi at sannsynlighetstettheten for en khikvadatfordelt variabel med to frihetsgrader nettopp er gitt ved $\frac{1}{2}e^{-x/2}$.

e) Hendelsen $\chi_{1-\alpha/2} \leq \frac{2}{\mu}\sum_{i=1}^n T_i \leq \chi_{\alpha/2}$ har sannsynlighet $1-\alpha$. Den første ulikheten er ekvivalent med $\mu \leq \frac{2}{\chi_{1-\alpha/2}}\sum_{i=1}^n T_i$. Den andre er ekvivalent med $\frac{2}{\chi_{\alpha/2}}\sum_{i=1}^n T_i \leq \mu$. Vi kan sette sammen disse to ulikhetene så vi får hendelsen $\frac{2}{\chi_{\alpha/2}}\sum_{i=1}^n T_i \leq \mu \leq \frac{2}{\chi_{1-\alpha/2}}\sum_{i=1}^n T_i$. Denne hendelsen er den samme som vi startet med, og har altså sannsynlighet $1-\alpha$. Men dette er nettopp definisjonen på at $\left[\frac{2}{\chi_{\alpha/2}}\sum_{i=1}^n T_i, \frac{2}{\chi_{1-\alpha/2}}\sum_{i=1}^n T_i\right]$ er et $100(1-\alpha)$ %-konfidensintervall for μ.

Hvis $n = 40$ og $\bar{T} = 28{,}3$, er $\sum_{i=1}^{40} T_i = 40 \cdot 28{,}3 = 1132$. Fra tabell finner vi $\chi_{1-\alpha/2} = 57{,}153$ og $\chi_{\alpha/2} = 106{,}629$ når $\alpha = 0{,}05$ og antall frihetsgrader er $2n = 80$, og konfidensintervallet blir $[2 \cdot 1132/106{,}629, 2 \cdot 1132/57{,}153] = [21{,}2, 39{,}6]$.