Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag – Sannsynlighet og statistikk I, MNFST101

Lørdag 24. mai 2003

Oppgave 1

a) La S være hendelsen at en tilfeldig mottatt melding er spam og la O være hendelsen at en melding inneholder ordet "sex". Opplysningene i oppgaven svarer da til at

$$\begin{aligned} P(O|S) &= 0.9 & P(O^c|S) &= 1 - 0.9 = 0.1 \\ P(O|S^c) &= 0.02 & P(O^c|S^c) &= 1 - 0.02 = 0.98. \end{aligned}$$

Dessuten er $P(S) = 0.7$ (og $P(S^c) = 0.3$). Lov om totalsannsynlighet gir at

$$P(O) = P(O|S)P(S) + P(O|S^c)P(S^c) = 0.9 \cdot 0.7 + 0.02 \cdot 0.3 = 0.636.$$

b) Vi ønsker å finne $P(S|O)$, d.v.s. en betinget sannsynlighet men med betinging motsatt vei av de sannsynlighetene som er gitt i oppgaven. Dette kan derfor løses ved hjelp av Bayes teorem som gir at

$$P(S|O) = \frac{P(O|S)P(S)}{P(O|S)P(S) + P(O|S^c)P(S^c)} = \frac{0.9 \cdot 0.7}{0.9 \cdot 0.7 + 0.02 \cdot 0.3} = 0.99.$$

En alternativ løsning er å bruke definisjonen av betinget sannsynlighet to ganger samt resultatet i a):

$$P(S|O) = \frac{P(S \cap O)}{P(O)} = \frac{P(O|S)P(S)}{P(O)}.$$

c) Bayes teorem gir her at betinget sannsynlighet for at en melding er spam gitt at meldingen ikke inneholder ordet "sex" blir

$$P(S|O^c) = \frac{P(O^c|S)P(S)}{P(O^c|S)P(S) + P(O^c|S^c)P(S^c)} = \frac{0.1 \cdot 0.7}{0.1 \cdot 0.7 + 0.98 \cdot 0.3} = 0.19.$$

d) Sannsynligheten for at hver enkelt av de 100 meldingene som har sluppet gjennom filteret er en spam-melding blir lik $P(S|O^c)$ i oppgave c). Antall meldinger X av de 100 som er spam blir dermed binomisk fordelt med parameter $p = 0.19$ og $n = 100$. Sannsynligheten for en eller flere spam-meldinger blant de 100 blir

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{100}{0} p^0 (1-p)^{100} = 1 - 0.81^{100} \approx 1.$$

Oppgave 2

a) Vi ser at W er den minste av Y_i'ene. Teorem for sannsynlighetstetthet til ordningsobservatorer (order statistics) gir da at

$$\begin{aligned} f_W(y) &= n f_Y(y)(1 - F_Y(y))^{n-1} \\ &= n\lambda e^{-\lambda y}[1 - (1 - e^{-\lambda y})]^{n-1} \\ &= n\lambda e^{-n\lambda y}. \end{aligned}$$

Vi gjenkjenner denne fordelingen som eksponentiell fordeling med parameter $n\lambda$. Dermed er forventningen gitt ved

$$E(W) = \frac{1}{n\lambda}.$$

b) Vi har i oppgave a) vist at ventetidene mellom de tidspunktene det trekkes opp fisk i løpet av intervallet er eksponentiellt fordelte med parameter $n\lambda$. I en generell Poissonprosess med intensitet (rate) λ er ventetidene mellom påfølgende hendelser eksponentiellt fordelte med parameter λ. Vi gjenkjenner derfor dette som en Poissonprosess med intensitet $n\lambda$. Derfor blir totalt antall hendelser i et tidsintervall av lengde t (antall fisk som trekkes opp) Poissonfordelt med parameter $n\lambda \cdot t$.

c) Sannsynligheten for at det henger fisk på krok i etter etter et tidsrom av lengde s blir

$$p = P(Y_i < s) = F_Y(s) = 1 - e^{-\lambda s}. \tag{1}$$

Uavhengighet medfører at det totale antall fisk X som henger i snøret etter tidsrom av lengde s blir binomisk fordelt med parameter n og p gitt ved (1). Videre gir setning for forventing under binomisk modell at

$$E(X) = np = n(1 - e^{-\lambda s}).$$

d) Når Z er lik summen av vekten til X fisker, hver med forventning μ og varians σ^2, blir betinget forventning til Z gitt at $X = x$ fisker henger i snøret

$$E(Z|X = x) = \mu x$$

og betinget varians

$$\mathrm{Var}(Z|X = x) = \sigma^2 x.$$

Anvendes setningen som er gitt i oppgaven på problemet får vi at den ubetingede variansen til Z blir

$$\begin{aligned}
\mathrm{Var}\, Z &= E_X \,\mathrm{Var}(Z|X) + \mathrm{Var}_X\, E(Z|X) \\
&= E_X(\sigma^2 X) + \mathrm{Var}_X(\mu X) \\
&= \sigma^2 EX + \mu^2 \,\mathrm{Var}\, X \\
&= \sigma^2 np + \mu^2 np(1-p) \\
&= \sigma^2 n(1 - e^{-\lambda s}) + \mu^2 n(1 - e^{-\lambda s})(e^{-\lambda s}) \\
&= n(1 - e^{-\lambda s})(\sigma^2 + \mu^2 e^{-\lambda s}).
\end{aligned}$$

Oppgave 3

a) Levetidene er eksponentiellt fordelte med parameter $1/\theta$. Dermed blir forventet levetid $E(Y) = 1/(1/\theta) = \theta$. Likelihood funksjonen blir

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{y_i}{\theta}} \\
&= \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} y_i}
\end{aligned}$$

og

$$\ln L(\theta) = -n \ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} y_i.$$

Setter vi den deriverte lik null får vi

$$\begin{aligned}
\frac{d}{d\theta} \ln L(\theta) &= 0 \\
-\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} y_i &= 0
\end{aligned}$$

som løst med hensyn på θ gir at sannsynlighetsmaksimeringsestimatoren av θ blir

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Forventningen blir

$$E\hat{\theta} = E(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n} nE(Y_i) = \theta$$

slik at estimatoren altså er forventningsrett.

b) Vi vet at ventetiden til den n'te hendelsen i en Poissonprosess (lik summen av n eksponentiellt fordelte variable med parameter λ) er gammafordelt med form parameter n og skalaparameter λ. Derfor er $\sum_{i=1}^{n} Y_i$ gammafordelt med med parameter n og $\frac{1}{\theta}$. Sannsynlighetstettheten til $\sum_{i=1}^{n} Y_i$ er derfor

$$f_{\sum Y_i}(y) = \frac{1}{\theta^n \Gamma(n)} y^{n-1} e^{-\frac{1}{\theta} y}.$$

Transformasjonen $a\sum_{i=1}^{n} Y_i$ har dermed sannsynlighetstetthet

$$\begin{aligned} f_{a\sum Y_i}(y) &= \frac{1}{a}\frac{1}{\theta^n \Gamma(n)} \left(\frac{y}{a}\right)^{n-1} e^{-\frac{1}{\theta}\frac{y}{a}} \\ &= \frac{1}{(a\theta)^n \Gamma(n)} y^{n-1} e^{-\frac{1}{a\theta} y} \end{aligned}.$$

For at dette skal bli en kji-kvadrat fordeling skal skalaparameteren i fordelingen,

$$\frac{1}{a\theta} = \frac{1}{2},$$

som betyr at vi må ha $a = \frac{2}{\theta}$. Lar vi n' betegne antall frihetsgrader og setter eksponentene $\frac{n'}{2} - 1$ til y i uttrykket for sannsynlighetstettheten til kji-kvadrat fordelingen lik eksponenten $n-1$ i uttrykket over får vi at antall frihetsgrader $n' = 2n$.

c) I b) har vi vist at $U = \frac{2}{\theta}\sum_{i=1}^{n} Y_i$ er kji-kvadrat fordelt med $2n$ frihetsgrader. Dermed er $V = \frac{2}{\theta'}\sum_{i=1}^{m} X_i$ kji-kvadrat fordelt med $2m$ frihetsgrader. Vi vet da at

$$\frac{U/(2n)}{V/(2m)} = \frac{(\frac{2}{\theta}\sum_{i=1}^{n} Y_i)/(2n)}{(\frac{2}{\theta'}\sum_{i=1}^{m} X_i)/(2m)}$$

er F-fordelt med $2n$ og $2m$ frihetsgrader. Gitt at H_0 er sann er $\theta = \theta'$ og vi ser da at uttrykket over blir lik $\bar{Y}/\bar{X}$.

Velger vi signifikansnivå lik α blir beslutningsregelen for en test å forkaste H_0 når

$$\frac{\bar{Y}}{\bar{X}} < F_{\alpha,2n,2m},$$

i og med at små verdier av $\bar{Y}/\bar{X}$ tyder på at $EX > EY$ ($\theta' > \theta$, d.v.s. H_1), den kritiske verdien er m.a.o. nedre α-kvantil i F-fordelingen.