TMA4240 Statistikk
Høst 2020

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalte oppgaver 8
Løsningsskisse

Oppgave 1

Vi vet at alle X_i er binomisk fordelte med parametre n_i og p, for $i = 1, \ldots, 6$. Dette betyr at vi har $E[X_i] = n_i p$ og $\text{Var}(X_i) = n_i p(1-p)$. Vi ønsker å finne konstanten k slik at estimatoren $\hat{p}$ er forventningsrett, d.v.s. $E[\hat{p}] = p$. Dette gir

$$\begin{aligned} E[\hat{p}] &= E\left[k \sum_{i=1}^{6} X_i\right] = k \sum_{i=1}^{6} E[X_i] \\ &= k \sum_{i=1}^{6} n_i p = kp \sum_{i=1}^{6} n_i \\ &= kp \cdot 730. \end{aligned}$$

For at $\hat{p}$ skal være forventningsrett, må vi derfor ha $E[\hat{p}] = 730kp = p$ som betyr at $730k = 1$ og dermed $\underline{\underline{k = 1/730}}$.

Estimatorens varians kan nå finnes som

$$\begin{aligned} \text{Var}(\hat{p}) &= \text{Var}\left(k \sum_{i=1}^{6} X_i\right) = k^2 \sum_{i=1}^{6} \text{Var}(X_i) \\ &= k^2 \sum_{i=1}^{6} n_i p(1-p) = k^2 p(1-p) \sum_{i=1}^{6} n_i \\ &= p(1-p) \cdot \frac{730}{730^2} = \underline{\underline{\frac{p(1-p)}{730}}}. \end{aligned}$$

Estimatet for p kan finnes fra tabellen. Vi får

$$\hat{p} = k \sum_{i=1}^{6} X_i = \frac{1}{730} \cdot 694 \approx \underline{\underline{0.9507}}.$$

Oppgave 2

Poisson-fordeling med forventningsverdi $\mu = \lambda t$, der λ skadefrekvens pr skidag og t er eksponeringstid i antall skidager.

a) I dette punktet er det kjent at $\lambda = 1/1000$ for "Alpinfjellet".

Sannsynligheten for at det skjer akkurat én ulykke i løpet av $t = 2000$ skidager:

$$P(X = 1) = \frac{(0.001 \cdot 2000)^1}{1!} \exp(-0.001 \cdot 2000) = 2 \cdot \exp(-2) = \underline{\underline{0.27}}$$

Sannsynligheten for at du utsettes for en eller flere ulykker ved opphold 10 skidager i "Alpinfjellet":

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{(0.001 \cdot 10)^0}{0!} \exp(-0.001 \cdot 10) = 1 - \exp(-0.01) = \underline{\underline{0.01}}. \end{aligned}$$

Sannsynligheten for minst en ulykke ved opphold i t skidager er:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \exp(-0.001 \cdot t)$$

Vi skal finne t slik at denne sannsynligheten blir større enn 0.1.

$$\begin{aligned} P(X \geq 1) &> 0.1 \\ 1 - \exp(-0.001 \cdot t) &> 0.1 \\ \exp(-0.001 \cdot t) &< 0.9 \\ -0.001 \cdot t &< \ln 0.9 \\ t &> -1000 \cdot \ln 0.9 = 105.4 \end{aligned}$$

Du må tilbringe minst $\underline{\underline{106}}$ skidager i "Alpinfjellet" for at din sannsynlighet for minst en ulykke skal bli større enn 0.1.

b) Vi har observasjoner av samhørende verdier av antall ulykker, X_i, og eksponering i antall skidager, t_i $(i = 1, ..., n)$, for n tilfeldig valgte dager anlegget var åpent.

Vi finner sannsynlighetsmaksimeringsestimatoren (SME) for λ basert på de n uavhengige observasjonsparene $(X_1, t_1), (X_2, t_2), \ldots, (X_n, t_n)$. Vi ser først på rimelighetsfunksjonen, og siden observasjonsparene er uavhengige finner vi den ved å multiplisere sammen marginalsannsynlighetene. Vi innfører $f_i(x_i; \lambda)$:

$$f_i(x_i; \lambda) = \frac{(\lambda t_i)^{x_i}}{x_i!} \exp(-\lambda t_i)$$

Rimelighetsfunksjonen er gitt ved:

$$\begin{aligned} L(\lambda) &= L(\lambda; x_1, ..., x_n) = f(x_1, ..., x_n; \lambda) \\ &= f_1(x_1; \lambda) \cdots f_n(x_n; \lambda) \\ &= \prod_{i=1}^{n} \frac{(\lambda t_i)^{x_i}}{x_i!} \exp(-\lambda t_i) \end{aligned}$$

Tar logaritmen:

$$\begin{aligned} l(\lambda; x_1, ..., x_n) &= \ln [L(\lambda)] \\ &= \ln \left(\prod_{i=1}^{n} \frac{1}{x_i!} \right) + \sum_{i=1}^{n} x_i \ln(\lambda t_i) - \lambda \sum_{i=1}^{n} t_i \end{aligned}$$

Finn maksimumspunkt ved å derivere ln-rimelighetsfunksjonen og sette lik 0.

$$\frac{\partial l}{\partial \lambda} = 0 + \sum_{i=1}^{n}(x_i \frac{1}{\lambda t_i} \cdot t_i) - \sum_{i=1}^{n} t_i = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} t_i$$

Settes dette uttrykket lik 0, får vi løsningen

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} t_i}$$

SME for λ blir da $\underline{\underline{\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} t_i}}}$.

Er $\hat{\lambda}$ forventningsrett:

$$\begin{aligned} \mathrm{E}[\hat{\lambda}] &= \mathrm{E}\left[\frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} t_i}\right] = \frac{1}{\sum_{i=1}^{n} t_i}\mathrm{E}(\sum_{i=1}^{n} X_i) \\ &= \frac{1}{\sum_{i=1}^{n} t_i}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{\sum_{i=1}^{n} t_i}\sum_{i=1}^{n}\lambda t_i \\ &= \underline{\underline{\lambda}}. \end{aligned}$$

Ja, estimatoren er forventningsrett.

Oppgave 3

a) The probability is $\int_{0.5}^{0.9} 6x(1-x)\,dx = \int_{0.5}^{0.9}(6x - 6x^2)\,dx = [3x^2 - 2x^3]_{0.5}^{0.9} = 0.472$.

b) The likelihood function is given by

$$L(\beta) = \prod_{i=1}^{n} \beta(\beta+1)x_i(1-x_i)^{\beta-1} = \beta^n(\beta+1)^n\left(\prod_{i=1}^{n} x_i\right)\prod_{i=1}^{n}(1-x_i)^{\beta-1},$$

and the log likelihood

$$\ln L(\beta) = n\ln\beta + n\ln(\beta+1) + \sum_{i=1}^{n}\ln x_i + (\beta-1)\sum_{i=1}^{n}\ln(1-x_i),$$

which has derivative

$$(\ln L)'(\beta) = \frac{n}{\beta} + \frac{n}{\beta+1} + \sum_{i=1}^{n}\ln(1-x_i).$$

$(\ln L)'$ is decreasing on $(0,\infty)$ and the sum of two first terms tends to ∞ when $\beta \to 0^+$ and to 0 when $\beta \to \infty$, so that $(\ln L)'$ will have a single zero (the third term is negative) for $\beta > 0$ and be positive left of the zero and negative right of the zero. This means that L has its maximum at this zero. Solving for the zero,

$$\beta^2\sum_{i=1}^{n}\ln(1-x_i) + \Big(2n + \sum_{i=1}^{n}\ln(1-x_i)\Big)\beta + n = 0,$$

we get

$$\beta = \frac{-2n - \sum_{i=1}^n \ln(1-x_i) \pm \sqrt{4n^2 + (\sum_{i=1}^n \ln(1-x_i))^2}}{2\sum_{i=1}^n \ln(1-x_i)}$$
$$= -\frac{n}{\sum_{i=1}^n \ln(1-x_i)} - \frac{1}{2} \pm \sqrt{\Big(\frac{n}{\sum_{i=1}^n \ln(1-x_i)}\Big)^2 + \frac{1}{4}}.$$

We choose the larger zero since $(\ln L)'$ has only one zero for positive arguments (the other we found must be negative), and get the maximum likelihood estimator

$$\sqrt{\Big(\frac{n}{\sum_{i=1}^n \ln(1-X_i)}\Big)^2 + \frac{1}{4}} - \frac{n}{\sum_{i=1}^n \ln(1-X_i)} - \frac{1}{2} = \sqrt{\frac{1}{\big(\overline{\ln(1-X)}\big)^2} + \frac{1}{4}} - \frac{1}{\overline{\ln(1-X)}} - \frac{1}{2}.$$

For $n = 100$ and $\sum_{i=1}^n \ln(1-x_i) = -104.0$ the estimate is $\sqrt{1/1.04^2 + 1/4} + 1/1.04 - 1/2 = 1.545$.

(The discussion of actual attainment of maximum at the zero and of which zero to be chosen, is not required.)

Oppgave 4

a)

$$P(X > 1000) = P(\frac{X - 800}{100} > 2) = P(Z > 2) = 0.023$$

$$P(500 < X < 1000) = P(X < 1000) - P(X < 500) = (1 - 0.023) - P(Z < -3) = 0.976$$

b) From the probability of 0 in the Poisson:

$$P(\text{Ingen fisk}) = e^{-3} = 0.05$$

By conditional probability and the Poisson distribution:

$$P(X > 3|X > 0) = \frac{1 - P(X \leq 3)}{1 - 0.05} = \frac{1 - (0.05 + 0.15 + 0.22 + 0.22)}{0.95} = \frac{1 - 0.647}{0.95} = 0.37$$

c)

$$P(X > 0) = 1 - P(X = 0) = 1 - (\theta + (1-\theta)\cdot e^{-\mu}) = 1 - (0.5 + 0.5e^{-4}) = 0.49$$

$$E(X) = \sum_{x=0}^{\infty} xP(X = x) = \sum_{x=1}^{\infty} xP(X = x) = (1-\theta)\sum_{x=1}^{\infty} x\frac{\mu^x}{x!}e^{-\mu} = (1-\theta)\mu$$

where the last sum is the expected value in the usual Poisson distribution (Here μ).

This gives $E(X) = 0.5 \cdot 4 = 2$.

d) The likelihood function is the product of the independent variables, regarded as a function of the unknown parameters:

$$L(\theta,\mu)=\prod_{i=1}^{n}P(X=x_i)=(\theta+(1-\theta)e^{-\mu})^r\prod_{x_i>0}(1-\theta)\frac{\mu^{x_i}}{x_i!}e^{-\mu}$$

Log likelihood becomes:

$$l(\theta,\mu)=\ln L(\theta,\mu)=r\ln(\theta+(1-\theta)e^{-\mu})+(n-r)\ln(1-\theta)+\ln\mu\sum_{x_i>0}x_i-(n-r)\mu-\sum_{x_i>0}\ln x_i!$$

The maximum likelihood estimator for θ is found by differentiation with respect to θ:

$$\frac{dl}{d\theta}=r\frac{1-e^{-\mu}}{\theta+(1-\theta)e^{-\mu}}-\frac{(n-r)}{1-\theta}$$

Solving for $\frac{dl}{d\theta}=0$ gives:

$$r(1-\theta)(1-e^{-\mu})=(n-r)(\theta+(1-\theta)e^{-\mu})$$

And separating for θ gives the desired solution:

$$\hat{\theta}=\frac{r-ne^{-\mu}}{n(1-e^{-\mu})}$$

The plot peaks at about $\hat{\mu}=3$, which is the maximum likelihood estimate for μ.

Inserting this we get

$$\hat{\theta}=\frac{8-20e^{-3}}{20(1-e^{-3})}=0.37$$