MAD 2022-23, Assignment 4

Helga Rykov Ibsen <mcv462> Hold 6

1. september 2023

Opgave 1

Vi skal vise at MLE for paramenter θ er lige med det reciprokke af gennemsnittet $\overline{x}$:

$$\hat{\theta}_n = \frac{n}{\sum_{i=1}^{n} x_i} \tag{1}$$

Vi skal med andre ord vise at loss-funktionen har et maksimum i punktet θ. Vi starter med at opstille vores loss-funktion. Vi ved at $X_1, ..., X_n$ er uafhængige variabler som er fordelt med PMF $p_\theta(x) = (1-0)^{x-1}\theta$. Vi kan derfor opstille loss-funktionen som produktet af variablernes PMF:

$$L = \prod_{i=1}^{n} (1-\theta)^{x_i - 1} \cdot \theta$$

For at finde ud af om der ligger funktionens kritiske punkt i θ, skal vi først differentiere udtrykket mht. θ og derefter løse ligningen $\frac{\partial L}{\partial \theta} = 0$. Vi indfører den naturlige logaritme for at gøre det nemmere at differentiere produktet:

$$\begin{aligned} log(L) &= \sum_{i=1}^{n} log((1-\theta)^{x_1 - 1} \cdot \theta) \\ &= \sum_{i=1}^{n} ((x_i - 1) \cdot log((1-\theta) + log(\theta)) \end{aligned}$$

Nu differentierer vi funktionen mht. θ:

$$\frac{\partial log(L)}{\partial \theta} = \sum_{i=1}^{n} (-(x_i - 1)\frac{1}{1-\theta} + \frac{1}{\theta}) \tag{2}$$

Nu skal vi løse ligningen $\frac{\partial L}{\partial \theta} = 0$:

$$\sum_{i=1}^{n} \frac{1}{\theta} = \sum_{i=1}^{n} (\frac{x_i}{1-\theta} - \frac{1}{1-\theta})$$

Da θ-parameteren ikke er afhængig af x_i, kan vi smide $\sum$ ud:

$$n\frac{1}{\theta} = \frac{1}{1-\theta} \sum_{i=1}^{n} x_i - n\frac{1}{1-\theta}$$

Nu kan vi gange med $(1-\theta)$ på begge sider:

$$
\begin{aligned}
n(\frac{1}{\theta}-1) &= \sum_{i=1}^{n} x_i - n \\
\frac{n}{\theta} - n &= \sum_{i=1}^{n} x_i - n \\
\frac{n}{\theta} &= \sum_{i=1}^{n} x_i \\
\theta &= \frac{n}{\sum_{i=1}^{n} x_i}
\end{aligned}
$$

Vi har nu vist at der er en vandret tangent i punktet θ. Nu skal vi vise at dette punkt også er et maksimum. Vi skal derfor finde den anden afledede af loss mht. θ. Hvis den anden afledede er negativ, så har loss et maksimum i dette punkt.

$$
\begin{aligned}
\frac{\partial^2 log(L)}{\partial\theta^2} &= \sum_{i=1}^{n}(-(x_i-1)\frac{1}{(1-\theta)^2} - \frac{1}{\theta^2} \\
&= \frac{1}{(1-\theta)^2}\sum_{i=1}^{n}(1-x_i) - \frac{n}{\theta^2} \\
&= \frac{n}{(1-\theta)^2} - \frac{\sum_{i=1}^{n} x_i}{(1-\theta)^2} - \frac{n}{\theta^2}
\end{aligned}
$$

Vi indsætter θ-parameteren:

$$
\begin{aligned}
\frac{\partial^2 log(L)}{\partial\theta^2} &= \frac{n}{(1-\frac{n}{\sum x_i})^2} - \frac{\sum x_i}{(1-\frac{n}{\sum x_i})^2} - \frac{n}{(n/\sum x_i)^2} \\
&= n\left(\frac{1}{(1-\frac{1}{\overline{x}})^2} - \frac{\overline{x}}{(1-\frac{1}{\overline{x}})^2} - \frac{1}{(\frac{1}{\overline{x}})^2}\right)
\end{aligned}
$$

Da vi ved at $X_1, ..., X_n$ er positive heltal, så bliver den anden afledede indlysende negativ i θ. Og hvis $\overline{x}$-værdien > 1, så er det også indlysende at den anden afledede bliver negativ. Vi har dermed vist at der er et maksimum i θ.

Opgave 2

(a)

Da vi ved at punkterne (X_1, Y_1)...,(X_4, Y_4) er ligefordelt uafhængige variable og vi ved at pdf er karakteriseret ved at den aldrig er større end 1, kan vi definere pdf som 1 over arealet af vinduet:

$$f_\theta(x,y) = \begin{cases} \frac{1}{x_{max}-x_{min}} \cdot \frac{1}{y_{max}-y_{min}} & \text{for } x_{min} \leq x \leq x_{max}, y_{min} \leq y \leq y_{max} \\ 0 & \text{ellers} \end{cases}$$

(b)

Vi starter med at finde sandsynlighed for θ=$(-1, 4, -1, 3)$:

$$f_\theta(x,y) = \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{20}$$

Da der er tale om ligefordelte sandsynligheder, er de fire punkter (0,0), (0,1), (1,1), (2,2) lige sandsynlige:

$$f_\theta(0,0) = \frac{1}{20}$$
$$f_\theta(0,1) = \frac{1}{20}$$
$$f_\theta(1,1) = \frac{1}{20}$$
$$f_\theta(2,2) = \frac{1}{20}$$

Og da der er tale om de fire punkter er uafhængige variabler, finder vi den samlede sandsynlighed — *likelihood* — som produkt af dem:

$$L = \prod_{i=1}^{4} f_\theta(x_i, y_i)$$
$$L = (\frac{1}{20})^4 = \frac{1}{160000}$$

Vi gentager processen og gør det samme for θ=$(-2, 5, -3, 6)$:

$$f_\theta(x,y) = \frac{1}{7} \cdot \frac{1}{9} = \frac{1}{63}$$
$$L = (\frac{1}{63})^4 = \frac{1}{15752961}$$

(c)

Givet sættet af stjerner (0,0), (0,1), (1,1) og (2,2), det bedste bud på hvor grænserne for vinduet går må bestemmes som *likelihood*, $L \geq 0$, fordi tæthedsfunktion (pdf) ikke kan være mindre end 0.

$$\begin{aligned} L > 0 \qquad & \text{hvis} \quad x_{min} \geq 0 \quad \text{og} \quad x_{max} \leq 2 \\ & \text{og} \quad\; y_{min} \geq 0 \quad \text{og} \quad y_{max} \leq 2 \end{aligned}$$

Ellers vil L være mindre end 0.

Opgave 3

(a)

Da vi skal finde posterior fordeling $p(r|Y_N)$, bruger vi Bayes' teorem, som siger at posterior fordeling er proportionelt med produkt af prior fordeling og sandsynligheden for et bestemt udfald:

$$p(r|Y_N) \propto p(r)^{\delta-1} \cdot p(Y_N|r)^{\gamma-1}$$

Vi ved at $p(r)$ er en beta fordeling og $p(Y_N|r)$ er binomialfordelt og kan tage de to fordelinger og gange dem sammen, så det giver:

$$p(r|Y_N) \propto \left[\binom{N}{Y_N} r^{Y_N} (1-r)^{N-Y_N} \right] \cdot \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot r^{\alpha-1} \cdot (1-r)^{\beta-1} \right]$$

Vi kan samle konstante led sammen i det første udtryk og få et pænere udtryk:

$$p(r|Y_N) = \left[\binom{N}{Y_N} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \cdot \left[r^{Y_N} (1-r)^{N-Y_N} \cdot r^{\alpha-1} \cdot (1-r)^{\beta-1} \right]$$

Da det første led er en konstant kan vi tage det ud og udtrykke posterior fordeling vha. proportionalitet som produkt af de to fordelinger:

$$p(r|Y_N) \propto r^{Y_N} (1-r)^{N-Y_N} \cdot r^{\alpha-1} \cdot (1-r)^{\beta-1} \tag{3}$$

Vi korkorter udtrykket ved at anvende potensregler og indsætter $\alpha = \beta = 1$:

$$\begin{aligned} p(r|Y_N) &\propto r^{Y_N+\alpha-1} (1-r)^{N-Y_N+\beta-1} \\ &\propto r^{Y_N+1-1} (1-r)^{N-Y_N+1-1} \end{aligned}$$

Vi indfører to nye parametre for at gøre potensudtrykkene mere læsebare:

$$\delta = Y_N + 1 \quad \text{og} \quad \gamma = N - Y_N + 1 \tag{4}$$

Nu kan vi indsætte parametrene δ og γ ind i posterior fordelingsfunktionen:

$$p(r|Y_N) \propto r^{\delta-1} \cdot (1-r)^{\gamma-1}$$

Til sidst indsæter vi normaliseringskonstanten, som vi tog ud af ligningen før og får udtrykket for posterior fordelingsfunktionen:

$$\begin{aligned} p(r|Y_N) &= \frac{\Gamma(\delta+\gamma)}{\Gamma(\delta)\cdot\Gamma(\gamma)} r^{\delta-1} \cdot (1-r)^{\gamma-1} \\ &= \frac{\Gamma(N+2)}{\Gamma(Y_N+1)\cdot\Gamma(N-Y_N+1)} r^{Y_N} \cdot (1-r)^{N-Y_N} \end{aligned}$$

(b)

Vi anvender den samme tankegang som i 3(a). Denne gang er vores beta-prior givet som $p(r) = 2r$. Vi kan derfor skrive dens forskrift som:

$$\begin{aligned} p(r) &= \binom{N}{Y_N} r^{Y_N} (1-r)^{N-Y_N} \\ &= 2 \cdot r^{Y_N} (1-r)^{N-Y_N} \qquad \text{fordi K(2,1) = 2} \end{aligned}$$

Dvs. $N = 2$ og $Y_N = 1$.

Og da vi ved at posterior fordeling kan udtrykkes som produkt af beta-prior og binomialfordelingsfuntionen, kan vi springe direkte til trinnet (cf. ligning (3) i 3(a)), hvor $N = 2$ og $Y_N = 1$:

$$\begin{aligned} p(r|Y_N) &\propto r^{Y_N} (1-r)^{N-Y_N} \cdot r^{\alpha-1} \cdot (1-r)^{\beta-1} \\ &\propto r^1 (1-r)^{2-1} \cdot r^{\alpha-1} \cdot (1-r)^{\beta-1} \\ &\propto r^{\alpha} \cdot (1-r)^{\beta} \end{aligned}$$

Parametrene $\delta = \alpha$ og $\gamma = \beta$:

$$p(r|Y_N) \propto r^{\delta} \cdot (1-r)^{\gamma}$$

Til sidst tilføjer vi normaliseringskonstanten og får det endelige udtryk for posterior beta-fordeling:

$$p(r|Y_N) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} r^{\alpha} \cdot (1-r)^{\beta}$$

(c)

Vi følger samme logik som i 3(a) og 3(b) og starter med at se på den givne beta-prior $p(r) = 3r^2$. Vi ved at den konstante led af beta-prior (N, Y_N) svarer i vores tilfald til enten $(3,1)$ fordi $K(3,1) = 3$ eller $(3,2)$ fordi $K(3,2) = 3$. Idet vi har r^2, kan vi se bort fra den sidste og bruge $N = 3$ og $Y_N = 2$. Beta-prior kan derfor skrives som:

$$\begin{aligned} p(r) &= \binom{N}{Y_N} r^{Y_N}(1-r)^{N-Y_N} \\ &= \binom{3}{2} \cdot r^2(1-r)^{3-2} \end{aligned}$$

Vi kan nu udtrykke posterior fordeling (NB: kontanterne er udeladt) vha. proportionalitet:

$$\begin{aligned} p(r|Y_N) &\propto r^2(1-r)^{3-2} \cdot r^{\alpha-1} \cdot (1-r)^{\beta-1} \\ &\propto r^2(1-r)^1 \cdot r^{\alpha-1} \cdot (1-r)^{\beta-1} \\ &\propto r^{2+\alpha-1} \cdot (1-r)^{1+\beta-1} \end{aligned}$$

Parametrene $\delta = \alpha + 1$ og $\gamma = \beta$:

$$p(r|Y_N) \propto r^{\delta} \cdot (1-r)^{\gamma}$$

Til sidst tilføjer vi normaliseringskonstanten og får det endelige udtryk for posterior beta-fordeling:

$$\begin{aligned} p(r|Y_N) &= \frac{\Gamma(\delta+\gamma)}{\Gamma(\delta)\cdot\Gamma(\gamma)} r^{\delta} \cdot (1-r)^{\gamma} \\ &= \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\cdot\Gamma(\beta)} r^{\alpha+1} \cdot (1-r)^{\beta} \end{aligned}$$