

Inlämningsuppgift 2

Erik Ödmann, David Carlsson

2022-03-18

Deluppgift 1

Vi har ett stickprov X_1, \dots, X_n från $X \sim Poi(\theta)$, där θ är okänd ($\theta > 0$).

Fall 1

I första fallet använder vi apriorifördelningen $p(\theta) \propto 1/\sqrt{\theta}$ som tillhör klassen av Jeffreys fördelningar, där $p(\theta) \propto \sqrt{I(\theta)}$. För att kontrollera detta behöver vi beräkna fisherinformationen för en poissonfördelad variabel. Fisherinformationen ges utav formeln

$$I(\theta) = -E[l''(\theta)]$$

Där likelihood funktionen för en observation är

$$L(\theta) = \frac{\theta^x e^{-\theta}}{x!}$$

Vilket ger oss log-likelihood funktionen för en observation

$$\begin{aligned} l(\theta) &= \ln\left(\frac{\theta^x e^{-\theta}}{x!}\right) \\ &= \ln(\theta^x e^{-\theta}) - \ln(x!) \\ &= \ln(\theta^x) + \ln(e^{-\theta}) - \ln(x!) \\ &= x \ln(\theta) - \theta - \ln(x!) \end{aligned}$$

Förstaderivatan utav log-likelihood funktionen för en observation blir

$$\begin{aligned} l'(\theta) &= \frac{d}{d\theta} \left(x \ln(\theta) - \theta - \ln(x!) \right) \\ &= \frac{x}{\theta} - 1 \end{aligned}$$

och andraderivatan kan då beräknas till

$$\begin{aligned} l''(\theta) &= \frac{d}{d\theta} \left(\frac{x}{\theta} - 1 \right) \\ &= -\frac{x}{\theta^2} \end{aligned}$$

Eftersom att $E[x] = \theta$ kan vi då beräkna Jeffreys fördelning till

$$\begin{aligned} p(\theta) &\propto \sqrt{I(\theta)} = \sqrt{-E[l''(\theta)]} \\ &= \sqrt{-E\left[-\frac{x}{\theta^2}\right]} \\ &= \frac{1}{\sqrt{\theta}} \end{aligned}$$

Vilket är lika med apriorifördelningen som vi skulle utgå ifrån. Nu kan vi fortsätta med att härleda aposteriorifördelningen som ges utav uttrycket

$$p(x|\theta) = \frac{p(\theta|x)p(\theta)}{p(x)} \propto p(\theta|x)p(\theta)$$

Där $p(\theta|x)$ är likelihood funktionen för hela stickprovet och $p(\theta)$ är apriorifördelningen som vi blev tilldelade. Likelihooden för hela stickprovet ges utav

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\ &= \frac{\theta^{\sum_{i=1}^n x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!} \\ &= \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^n x_i!} \\ &\propto \theta^{n\bar{x}} e^{-n\theta} \end{aligned}$$

Aposteriorifördelningen blir då

$$\begin{aligned} p(x|\theta) &\propto \theta^{n\bar{x}} e^{-n\theta} \frac{1}{\sqrt{\theta}} \\ &= \theta^{n\bar{x}-1/2} e^{-n\theta} \end{aligned}$$

och eftersom $\Gamma(\alpha, \beta) \propto \theta^{\alpha-1} e^{-\beta\theta}$ kan vi säga att aposteriorifördelningen är proportionell mot $\Gamma(n\bar{x} + 1/2, n)$ vilket är en gammafördelning.

Fall 2

Nu utgår vi istället från apriorifördelningen $p(\theta) \sim \Gamma(\alpha, \beta) \propto \theta^{\alpha-1} e^{-\beta\theta}$ samt likelihood funktionen från fall 1. Detta ger oss aposteriorifördelningen

$$\begin{aligned} p(x|\theta) &\propto \theta^{n\bar{x}} e^{-n\theta} \theta^{\alpha-1} e^{-\beta\theta} \\ &= \theta^{\alpha+n\bar{x}-1} e^{-(\beta+n)\theta} \end{aligned}$$

Vilket är proportionell mot $\Gamma(\alpha + n\bar{x}, \beta + n)$. Vi kan därmed dra slutsatsen att aposteriorifördelningen i fall 2 är proportionell mot samma gammafördelning som i fall 1 då $\alpha = 1/2$ och $\beta = 0$.

Deluppgift 2

I deluppgift 2 ska vi konstruera konfidensintervall med hjälp av bootstrap.

Percentilmetoden

Ett $100(1 - \alpha)\%$ -igt intervall för θ har gränserna

$$\left[\theta_{(m)}^*, \theta_{(B+1-m)}^* \right]$$

Där

$$\begin{aligned} \theta_i^* &= \text{Medelvärde för de inte genererade stickprovet} \\ B &= \text{Antalet stickprov genererade} \\ \alpha &= \text{Signifikansnivå} \\ m &= \alpha/2 * B \end{aligned}$$

“Basic” metoden

Ett $100(1 - \alpha)\%$ -igt intervall för θ har gränserna

$$\left[2\hat{\theta} - \theta_{(B+1-m)}^*, 2\hat{\theta} - \theta_{(m)}^* \right]$$

Där

$$\begin{aligned} \hat{\theta} &= \text{Medelvärde för det ursprungliga stickprovet} \\ \theta_i^* &= \text{Medelvärde för de inte genererade stickprovet} \\ B &= \text{Antalet stickprov genererade} \\ \alpha &= \text{Signifikansnivå} \\ m &= \alpha/2 * B \end{aligned}$$

Icke-parametrisk bootstrap

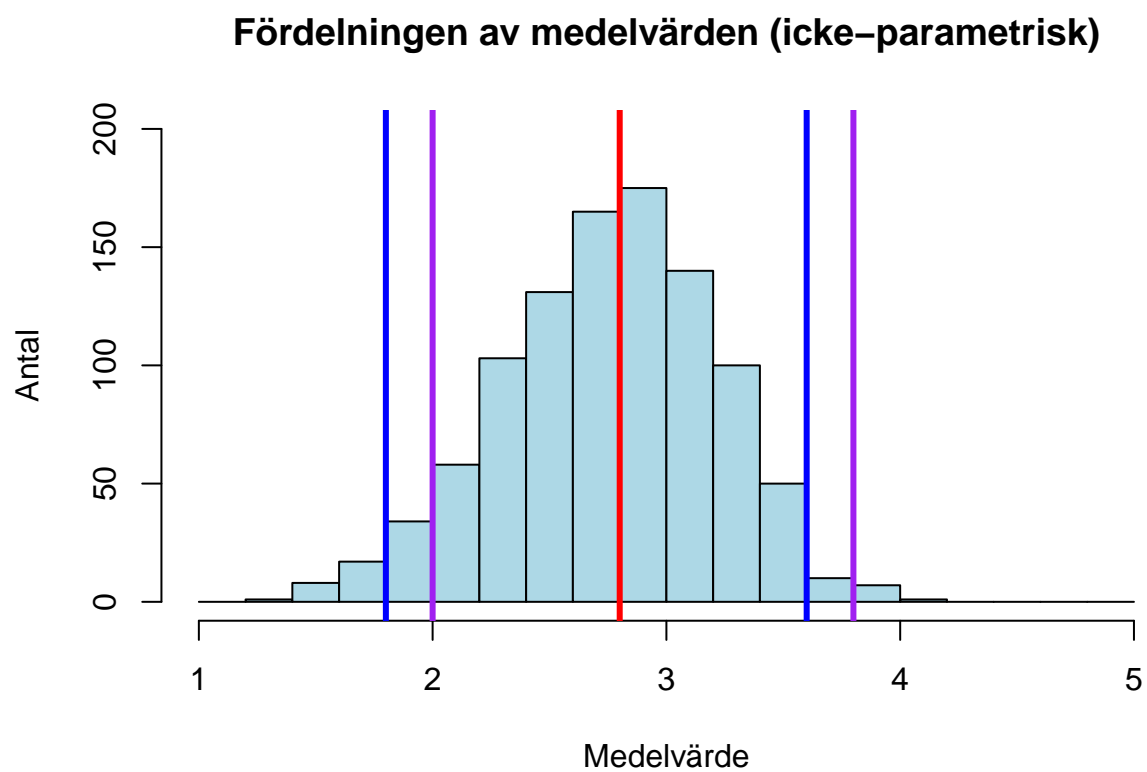
Applicerar vi percentilintervallet på medelvärdena som genererades från en icke-parametrisk bootstrap får vi följande gränser.

$$[1.8, 3.6]$$

I figur 1 är intervallets gränser markerade som två blåa linjer. Applicerar vi istället “basic” metoden får vi följande gränser

$$[2, 3.8]$$

och dessa är markerade i figur 1 som två lila linjer. Vi kan notera att “basic” intervallet ligger till höger om percentil intervallet. Detta är en indikation på att fördelningen inte är helt symmetrisk eftersom då hade intervallgränserna varit samma för båda metoder.



Figur 1: Simulerad fördelning av medelvärdet för icke-parametriskt bootstrap. Medelvärdet av fördelningen är markerat med den röda linjen. Konfidensintervallsgränserna är markerade i blått (percentil) och lila (“basic”).

Parametrisk bootstrap

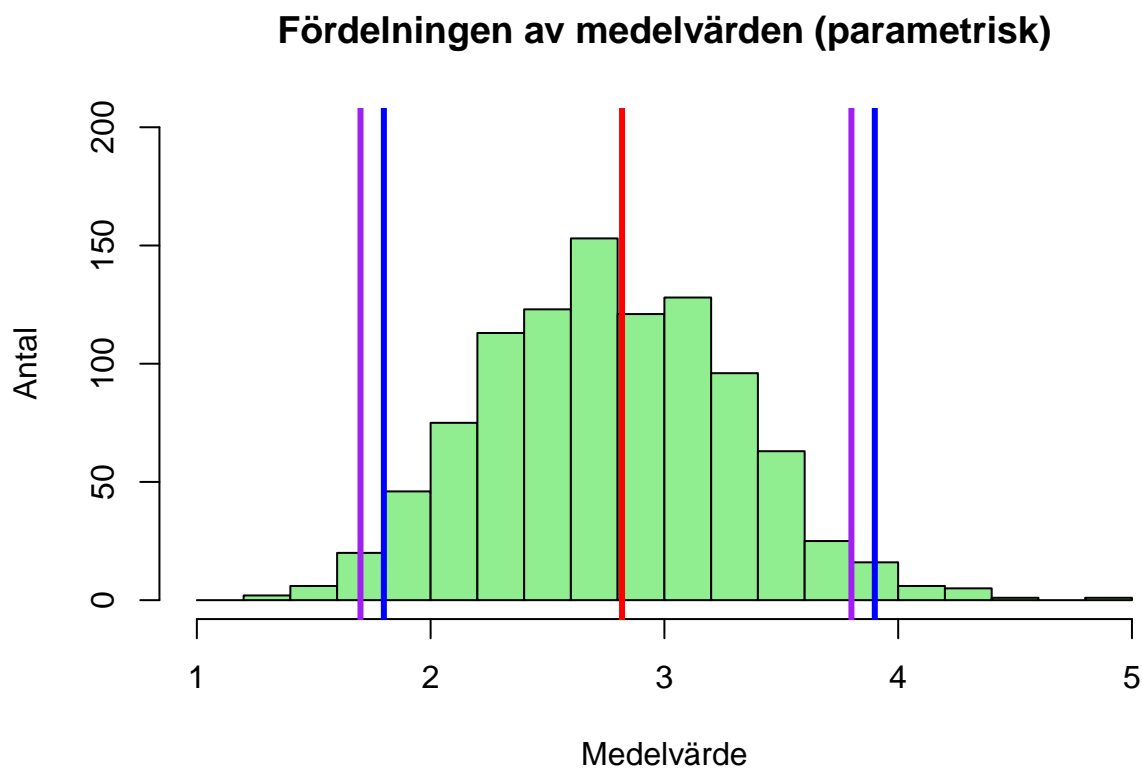
Applicerar vi percentilintervallet på medelvärdena som genererades från en parametriska bootstrap får vi följande gränser.

$$[1.8, 3.9]$$

I figur 2 är intervallets gränser markerade som två blåa linjer. Applicerar vi istället "basic" metoden får vi följande gränser

$$[1.7, 3.8]$$

Om vi jämför med det icke-parametriska datasetet så är intervallen i detta fall något bredare. Det verkar alltså som att fördelningen för det parametriska datasetet är bredare (eller plattare) än i det icke-parametriska fallet. Vi kan dock notera att gränserna ligger närmare varandra vilket kan vara en indikation på att detta datasetet är relativt symmetriskt.



Figur 2: Simulerad fördelning av medelvärdet för parametrisk bootstrap. Medelvärdet av fördelningen är markerat med den röda linjen. Konfidensintervallsgränserna är markerade i blått (percentil) och lila ("basic").