Exercise 4.1, Normal approximation

$$y_1, \dots, y_5 \overset{i.i.d.}{\sim} \text{Cauchy}(\theta, 1) \Rightarrow p(y_i|\theta) \propto \frac{1}{1+(y_i-\theta)^2}$$

$$\theta \sim U[0,1] \Longrightarrow p(\theta) = 1$$, observationer $(y_1, \dots, y_5) = (-2, -1, 0, 1.5, 2.5)$

a) Bestäm första och andra derivatan av log posterior funktionen

$$p(\theta|y) = c_N \cdot \prod_{j=1}^{5} \frac{1}{1+(y_j-\theta)^2}$$

$$\log p(\theta|y) = \log c_n + \sum_{j=1}^{5} \log\left[\frac{1}{1+(y_j-\theta)^2}\right]$$

$$= \log c_n - \sum_{j=1}^{5} \log\left(1+(y_j-\theta)^2\right)$$

$$\frac{d \log p(\theta|y)}{d\theta} = 2 \cdot \sum_{j=1}^{5} \frac{y_j-\theta}{1+(y_j-\theta)^2}$$

$$\frac{d^2 \log p(\theta|y)}{d\theta^2} = 2 \cdot \sum_{j=1}^{5} \frac{-1 \cdot \left(1+(y_j-\theta)^2\right) - (y_j-\theta) \cdot 2 \cdot (y_j-\theta) \cdot (-1)}{\left[1+(y_j-\theta)^2\right]^2}$$

$$= 2 \cdot \sum_{j=1}^{5} \frac{-1-(y_j-\theta)^2 + 2 \cdot (y_j-\theta)^2}{\left[1+(y_j-\theta)^2\right]^2} = 2 \cdot \sum_{j=1}^{5} \frac{(y_j-\theta+1) \cdot (y_j-\theta-1)}{\left[1+(y_j-\theta)^2\right]^2}$$

Exercise 4.1 forts.

b) Finn posteriorfunktionens typvärde genom att iterativt lösa ekvationen som bestäms genom att sätta derivatan av log-likelihooden till noll, i.e.

$$\frac{d \log p(\theta|y)}{d\theta} = \cancel{2} \cdot \sum_{j=1}^{5} \frac{y_j - \theta}{1+(y_j-\theta)^2} = 0$$

Vid utvärdering av ekvationen iterativt erhåller man $\hat{\theta} \approx -0{,}1376$

c) Posteriorfunktionen kan approximeras som

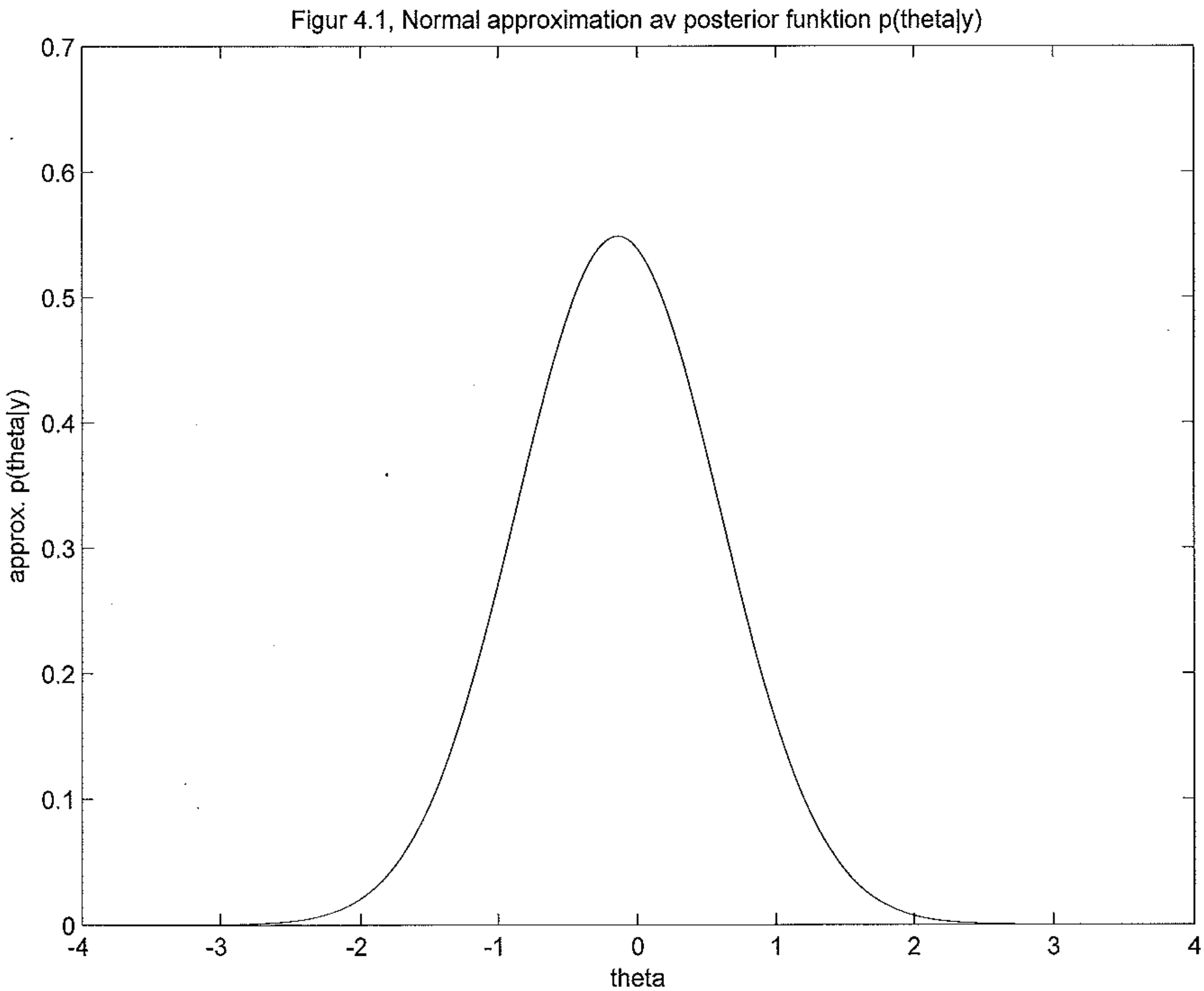
$$p(\theta|y) \approx N\left(\hat{\theta}, [I(\hat{\theta})]^{-1}\right)$$ där $\hat{\theta}$ gavs i b) och

$$I(\theta) = -\frac{d^2 \log p(\theta|y)}{d\theta^2} \implies$$

$$I(\hat{\theta}) = -2 \cdot \sum_{j=1}^{5} \frac{(y_j - (-0{,}1376) + 1) \cdot (y_j - (-0{,}1376) - 1)}{\left[1 + (y_j - (-0{,}1376))^2\right]^2}$$

$\approx 1{,}3749$, så att

$$p(\theta|y) \approx N(-0{,}1376, 0{,}7273)$$, se figur 4.1



Figur 4.1, Normal approximation av posterior funktion p(theta|y)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
approx. p(theta|y)
-4
-3
-2
-1
0
1
2
3
4
theta

Exercise 2.13, Discrete data:

I tabell 2.2 ges ant. livshotande olyckor och antalet döda för schemalagda flygningar per år mellan åren 1976–1985. Anpassning av diskreta datamodeller, år 1 är 1976.

a) Låt y_i = ant. livshotande olyckor för år i, $i = 1, 2, \ldots, 10$

$$y_i \overset{i.i.d.}{\sim} \text{Poisson}(\theta) \implies p(y_i \mid \theta) = \frac{1}{y_i!} \cdot \theta^{y_i} \cdot \exp(-\theta)$$

Ansätt en konjugerande prior för θ, i.e. en prior på den par. formen $p(\theta) \propto \theta^a \cdot \exp(-b\theta)$.

Välj $a = \alpha - 1$, $b = \beta$, så att

$$\theta \sim \text{Gamma}(\alpha, \beta) \implies p(\theta) \propto \theta^{\alpha-1} \cdot e^{-\beta\theta}, \quad \theta > 0.$$

Alltså blir

$$p(\theta \mid y_i) \propto p(\theta) \cdot \prod_{i=1}^{10} p(y_i \mid \theta) \propto \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot \theta^{\sum_{i=1}^{10} y_i} \cdot e^{-10\theta}$$

$$= \theta^{\alpha + 10\bar{y} - 1} \cdot e^{-(\beta+10)\theta}$$

$$\theta \mid y_i \sim \text{Gamma}(\alpha + 10\bar{y}, \beta + 10)$$

$\bar{y} = 23{,}8$ från tabell och ingen vetskap om θ medför att vi kan tagga en icke-informativ prior fördelning, d.v.s. $\alpha = \beta = 0$, så att $\left(p(\theta) \propto \frac{1}{\theta}\right)$

$$\theta \mid y_i \sim \text{Gamma}(238, 10)$$

Exercise 2.13 forts.

a forts.

Ett 95%-igt prediktionsintervall för en ny observation $\tilde{y}$ kan bestämmas genom simulering el. normalappr.

Fall 1: Simulering

$$\tilde{y} \mid \theta, y \sim Po(\theta) \quad , \text{ty} \quad y_i \overset{i.i.d.}{\sim} Po(\theta)$$

Drag först $\theta_{(k)}$ från posteriorn $p(\theta \mid y_i)$, $k = 1, 2, \ldots, 1000$, därefter drag $\tilde{y}_{(k)}$ från $p(\tilde{y} \mid \theta)$. Ordna observationerna $\tilde{y}_{(k)}$ i storleksordning och skatta den undre och övre intervallsgränsen som

$\hat{\pi}_{0.025} = \tilde{y}_{(k),(25)}$ och $\hat{\pi}_{0.975} = \tilde{y}_{(k),(975)}$, där

$\hat{\pi}_p = \tilde{y}_{(k),(1000 \cdot p)}$ är den skattade $p\%$-percentilen för den $(1000 \cdot p)$:e lägsta observationen

Erhåller intervallet till $[14, 35]$

Fall 2: Normalapproximation

Antag $\tilde{y} \mid y \approx N\left[E[\tilde{y} \mid y], Var[\tilde{y} \mid y]\right]$

$$E[\tilde{y} \mid y] = E\left[E(\tilde{y} \mid \theta, y) \mid y\right] = E[\theta \mid y] = \frac{238}{10} = 23{,}8$$

$$Var[\tilde{y} \mid y] = E\left[Var(\tilde{y} \mid \theta, y) \mid y\right] + Var\left[E[\tilde{y} \mid \theta, y] \mid y\right] =$$

Exercise 2.13 forts.

$$= E[\theta|y] + Var(\theta|y) = 23{,}8 + \frac{238}{10^2} = 26{,}18$$

Alltså, $\tilde{y}|y \sim N\left[E[\tilde{y}|y], Var[\tilde{y}|y]\right] = N(23.8, 26.18)$

och ett 95%-igt prediktionsintervall ges som

$$23{,}8 - 1{,}96 \cdot \sqrt{26{,}18} < \tilde{y} < 23{,}8 + 1{,}96 \cdot \sqrt{26{,}18}$$

$$\Longleftrightarrow \quad 13{,}77 < \tilde{y} < 33{,}83 \; .$$

Med åtminstone 95% sannolikhet kan vi hävda att det sanna värdet $\tilde{y}$ ligger i intervallet $[13, 34]$.

b) Låt θ_{pm} = förväntat antal olyckor per år och passagerarmile

x_i = ant. passagerarmiles för år i, så att

$$y_i \mid \theta_{pm}, x_i \sim Po(\theta_{pm} \cdot x_i)$$

x_i kan bestämmas från tabell 2.2 som

$$x_i = \frac{\text{ant. döda passagerare år } i}{\text{ant. döda passagerare per mile år } i}, \quad \text{t.ex. } x_1 = \frac{734}{0{,}19/100 \cdot 10^6} \approx 3{,}863 \cdot 10^{11}$$

vilket ger en tabell för x_i som

år	x_i
1	$3{,}863 \cdot 10^{11}$
⋮	⋮

Posteriorfunktionen för θ_{pm} ges som

$$p(\theta_{pm} | y_i, x_i) \propto p(\theta_{pm} | x_i) \cdot \prod_{i=1}^{10} p(y_i | x_i, \theta_{pm}),$$

där vi kan välja $p(\theta_{pm} | x_i) = \text{Gamma}(0,0) \propto \frac{1}{\theta_{pm}}$

p.s.s. som i a)

$$\prod_{i=1}^{10} p(y_i | x_i, \theta_{pm}) \overset{\text{m.a.p. } \theta}{\propto} \prod_{i=1}^{10} (\theta_{pm} \cdot x_i)^{y_i} \cdot \exp(-\theta_{pm} \cdot x_i)$$

$$\propto \theta_{pm}^{\sum_{i=1}^{10} y_i} \cdot \exp\left(-\sum_{i=1}^{10} x_i \cdot \theta\right) = \theta_{pm}^{10 \cdot \bar{y}} \cdot \exp(-10 \cdot \bar{x} \cdot \theta_{pm}),$$

så att

$$\theta_{pm} | y_i, x_i \sim \text{Gamma}(10 \cdot \bar{y}, 10 \cdot \bar{x}) = \text{Gamma}(238, 5.716 \cdot 10^{12})$$

$$\tilde{y} | \tilde{x}, \theta_{pm} \sim \text{Po}(\tilde{x} \cdot \theta_{pm}) \overset{\text{i uppg.}}{=} \text{Po}(8 \cdot 10^{11} \cdot \theta_{pm}).$$

P.s.s. som i a) kan ett 95%-igt prediktionsintervall för $\tilde{y}$ bestämmas med simulering el. normalapproximation.

Fall 1, Simulering

Drag $\theta_{(k),pm}$ från $p(\theta_{pm} | y)$ och givet $\theta_{(k),pm}$ drag $\tilde{y}$ från $p(\tilde{y} | \tilde{x}, \theta_{pm})$; $k = 1, 2, \ldots, 1000$

Fall 2, normal approximation

$$p(\tilde{y} | y, \tilde{x}) \approx N\left[E[\tilde{y} | y, \tilde{x}], \text{Var}[\tilde{y} | y, \tilde{x}]\right]$$

$$E[\tilde{y}|y,\tilde{x}] = E\left[E[\tilde{y}|\theta_{pm},y,\tilde{x}]\,|y,x\right] =$$

$$= E\left[8\cdot 10^{11}\cdot \theta_{pm}|y,x\right] = 8\cdot 10^{11}\cdot \frac{238}{5{,}716\cdot 10^{12}} \approx 33{,}3$$

$$Var[\tilde{y}|y,\tilde{x}] = E\left[Var(\tilde{y}|\theta_{pm},y,\tilde{x})|y,x\right]$$

$$+ Var\left[E[\tilde{y}|\theta_{pm},y,\tilde{x}]\,|y,x\right] = E\left[8\cdot 10^{11}\cdot \theta_{pm}|y,x\right]$$

$$+ Var\left[8\cdot 10^{11}\cdot \theta_{pm}|y,x\right] = 8\cdot 10^{11}\cdot \frac{238}{5{,}716\cdot 10^{12}} + (8\cdot 10^{11})^2\cdot \frac{238}{(5{,}716\cdot 10^{12})^2}$$

$\approx 38{,}0$, vilket ger det 95%-iga prediktionsintervallet för $\tilde{y}$ till $[21{,}1, 45.5]$ så vi kan med åtminstone 95%:s sannolikhet hävda att $\tilde{y}$ ligger i intervallet $[21, 46]$.

Prediktionsintervallet blir längre i b, än i a, och det beror på att man i b, även tar hänsyn till variabiliteten i ant. passagerarmiles per år.

Exercise 2.13 c

Som i a, med olycksstatistik ersatt med statistik över antalet döda pass. Simulering gav prediktionsintervallet till [638, 750]

Exercise 2.13 d

Som i b, med olycksstatistik ersatt med statistik över antalet döda pass. Simulering gav prediktionsintervallet till [900, 1035]

Exercise 2.13 e

(10)
ō2

I vilket av fallen a) – d) verkar Poissonmodellen mer eller mindre rimlig?

- Modell b) och d) verkar vara mer rimliga, eftersom antalet olyckor bör öka med antalet passagerarmiles
- Inte säkert dock, eftersom fler antal flygtimmar kan medföra ökade resurser på flygsäkerheten.
- Poisson modell för c) och d) kan vara missvisande. Om ett plan kraschar kan det medföra flera dödsoffer på samma gång.

Exercise 2.19, Posteriorintervall

Det högsta posteriorintervallet är inte invariant till transformation, som det centrala posteriorintervallet är.

T.ex. antag att givet σ^2, $\frac{nv}{\sigma^2} \sim \chi^2_n$, och att σ har den icke-informativa prior funktionen

$$p(\sigma) \propto \frac{1}{\sigma}, \quad \sigma > 0$$

a) Låt $X = \sigma$ och $Y = u(X) = X^2$,

$$f_X(x) \propto \frac{1}{x}$$

Funktionen $Y = u(X)$:s invers existerar, ty $x > 0$, så att

$$x = v(y) = u^{-1}(y) = \sqrt{y}, \quad v'(y) = \frac{1}{2\cdot\sqrt{y}}.$$

Med variabelsubstitutionsteknik får vi täth.fkn för Y till

$$g_Y(y) = f_X(v(y)) \cdot v'(y) \propto \frac{1}{\sqrt{y}} \cdot \frac{1}{2\cdot\sqrt{y}} \propto \frac{1}{y}$$

$$\overset{y = x^2 = \sigma^2}{\Leftrightarrow} \quad p(\sigma^2) \propto \frac{1}{\sigma^2} = \sigma^{-2}$$

Exercise 2.19, Posteriorintervall forts.

b) Bestämmer först de respektive posteriorfunktionerna för σ och σ^2. Vi har att

$$p(\sigma) \propto \sigma^{-1} \overset{a}{\Longrightarrow} p(\sigma^2) \propto \sigma^{-2}$$

$$\frac{nv}{\sigma^2} \Big| \sigma^2 \sim \chi^2_n \Longrightarrow p\left(\frac{nv}{\sigma^2} \Big| \sigma^2\right) \propto \left[\frac{nv}{\sigma^2}\right]^{\frac{n}{2}-1} \cdot e^{-\frac{nv}{2} \cdot \frac{1}{\sigma^2}}$$

$$\overset{c=\frac{nv}{2}}{\propto} \sigma^{2-n} \cdot e^{-c/\sigma^2}$$

. Sätt $Q = \frac{nv}{\sigma^2}$, så att

$$p(\sigma|Q) \propto \sigma^{-1} \cdot \sigma^{2-n} \cdot e^{-c/\sigma^2} = (\sigma^2)^{\frac{1}{2}-\frac{n}{2}} \cdot e^{-c/\sigma^2}$$

$$p(\sigma^2|Q) \propto \sigma^{-2} \cdot \sigma^{2-n} \cdot e^{-c/\sigma^2} = (\sigma^2)^{-\frac{n}{2}} \cdot e^{-c/\sigma^2}$$

$$\frac{d\, p(\sigma|Q)}{d\sigma} = 0 \iff (1-n) \cdot \sigma^{-n} \cdot e^{-c/\sigma^2} + \sigma^{1-n} \cdot e^{-c/\sigma^2} \cdot \frac{2c}{\sigma^3}$$

$$= e^{-c/\sigma^2} \cdot \sigma^{-n} \left[1-n+2c \cdot \sigma^{-2}\right] = 0 \iff$$

$$(1-n) \cdot \sigma^2 + 2c = 0 \iff \sigma^2 = \frac{2c}{n-1} = \frac{2 \cdot n \cdot v}{n-1}$$

$$\overset{\sigma>0}{\iff} \sigma = \sqrt{\frac{2nv}{n-1}}$$

, alltså är $p(\sigma|Q)$ unimodal ty endast en maximipunkt. Detta innebär att det 95%-iga intervallet med högst täthet är endast ett intervall.

13

Ö2

Låt $(\sqrt{a'}, \sqrt{b'})$ vara det 95%-iga intervallet av högst täthet för $p(\sigma \mid data)$. Då gäller att

$$a^{\frac{1}{2}-\frac{n}{2}} \cdot \exp(-c/a) = b^{\frac{1}{2}-\frac{n}{2}} \cdot \exp(-c/b)$$

$$\Longleftrightarrow \left(\frac{1}{2}-\frac{n}{2}\right) \cdot \log a - \frac{c}{a} = \left(\frac{1}{2}-\frac{n}{2}\right) \cdot \log b - \frac{c}{b} \qquad (1)$$

Om vi kvadrerar posteriorintervallet $(\sqrt{a'}, \sqrt{b'})$ till det 95%-iga intervallet av högst täthet för $p(\sigma^2 \mid data)$, gäller

$$a^{-\frac{n}{2}} \cdot \exp(-c/a) = b^{-\frac{n}{2}} \cdot \exp(-c/b)$$

$$\Longleftrightarrow -\frac{n}{2} \cdot \log a - \frac{c}{a} = -\frac{n}{2} \cdot \log b - \frac{c}{b} \qquad (2)$$

$$(1)-(2) \Longleftrightarrow \frac{1}{2} \log a = \frac{1}{2} \log b \Longrightarrow a = b \quad \lightning$$

motsägelse! $[a,b] = [a,a]$ kan inte vara något posterior intervall, ty intervallet är endast en punkt.

Exercise 2.22, Censored and uncensored data in the exponential model

14

02

a) $y|\theta \sim \text{Exp}(\theta) \implies p(y|\theta) \overset{\text{m.a.p. }\theta}{\propto} \theta \cdot e^{-\theta y}$

$\theta \sim \text{Gamma}(\alpha, \beta) \implies p(\theta) \propto \theta^{\alpha-1} \cdot e^{-\beta\theta}$

Observerar $y \geq 100$

Posteriorfunktionen, $p(\theta | y \geq 100)$, som en funktion av α, β

är $p(\theta|y \geq 100) \propto \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot p(y \geq 100 | \theta)$

$$= \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot \int_{100}^{\infty} \theta \cdot e^{-\theta \cdot y}\, dy = \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot \theta \cdot \left(-\frac{1}{\theta}\right) \cdot \left[e^{-\theta y}\right]_{100}^{\infty}$$

$$= \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot e^{-100\cdot\theta} = \theta^{\alpha-1} \cdot e^{-(\beta+100)\cdot\theta} = \text{Gamma}(\alpha, \beta+100)$$

$$E[\theta | y \geq 100] = \frac{\alpha}{\beta+100}, \quad Var[\theta|y \geq 100] = \frac{\alpha}{(\beta+100)^2}$$

b) $p(y = 100|\theta) \propto \theta \cdot e^{-\theta \cdot 100}$, så att

$$p(\theta|y=100) \propto \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot \theta \cdot e^{-\theta\cdot 100} = \theta^{\alpha} \cdot e^{-(\beta+100)\cdot\theta}$$

$$= \text{Gamma}(\alpha+1, \beta+100)$$

$$E[\theta|y=100] = \frac{\alpha+1}{\beta+100}, \quad Var[\theta|y=100] = \frac{\alpha+1}{(\beta+100)^2}$$

c) $$Var(\theta|y \geq k) = \frac{\alpha}{(\beta+k)^2}, \quad Var(\theta|y=k) = \frac{\alpha+1}{(\beta+k)^2}$$

Jämför!