

Aufgabe 1

Fortsetzung Blatt 2, Aufgabe 1: Exponentialverteilungsmodell für `babyboom.dat`-Daten.

- (a) Berechnen Sie die erwartete Fisher-Information und leiten Sie daraus Jeffreys' Priori her.

Fisher Information and Jeffreys' Prior for the Exponential Distribution

Let $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exponential}(\lambda)$, with density

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0, \lambda > 0.$$

Log-Likelihood and Fisher Information

The log-likelihood function for the sample is

$$\ell(\lambda) = \sum_{i=1}^n \log f(X_i; \lambda) = \sum_{i=1}^n [\log \lambda - \lambda X_i] = n \log \lambda - \lambda \sum_{i=1}^n X_i.$$

The score function is

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n X_i.$$

The observed information is

$$-\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} = \frac{n}{\lambda^2}.$$

Since this does not depend on the data, the Fisher information is

$$\mathcal{I}(\lambda) = \mathbb{E} \left[-\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} \right] = \frac{n}{\lambda^2}.$$

Jeffreys' Prior

Jeffreys' prior is defined as

$$\pi_J(\lambda) \propto \sqrt{\mathcal{I}(\lambda)}.$$

Using the Fisher information for a single observation ($n = 1$):

$$\mathcal{I}_1(\lambda) = \frac{1}{\lambda^2},$$

so

$$\pi_J(\lambda) \propto \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}.$$

Thus, Jeffreys' prior for the exponential distribution is

$$\boxed{\pi_J(\lambda) \propto \frac{1}{\lambda}, \quad \lambda > 0.}$$

This is an improper prior, as it does not integrate to 1 over $(0, \infty)$, but it is often used in Bayesian inference due to its invariance properties.

(b) Wie ändern sich die Schätzer, wenn wir statt dessen Jeffreys' Prior benutzen?

Aufgabe 2

Aim:

- Posteriori berechnen (bis auf multiplikative Konstante) für stetige Verteilung
- quadratische Ergänzung bei Normalverteilung
- Einfluss von Priori und Likelihood auf Posteriori: Posteriori-Erwartungswert als gewichtetes Mittel von Priori und Daten; graphisch; Einfluss der Stichprobengröße

Sei x_1, \dots, x_n eine iid-Stichprobe von $X \sim N(\mu, \sigma^2)$ mit unbekanntem Mittelwert $\mu \in \mathbb{R}$ und bekannter Varianz $\sigma^2 > 0$. Für μ wird a priori eine $N(m, s)$ -Verteilung angenommen.

(a) Zeigen Sie, dass die Posteriori-Verteilung $\mu|x$ gegeben ist durch $N(\tilde{m}, \tilde{s})$, wobei

$$\tilde{m} = \frac{s\bar{x} + m\frac{\sigma^2}{n}}{s + \frac{\sigma^2}{n}} \quad \text{und} \quad \tilde{s} = \frac{s\frac{\sigma^2}{n}}{s + \frac{\sigma^2}{n}}$$

mit $x = (x_1, \dots, x_n)^\top$ und $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Hinweis: Quadratische Ergänzung.

Lösung:

Likelihood:

$$\begin{aligned} p(x|\mu) &= p(x_1, \dots, x_n|\mu) \stackrel{\text{iid}}{=} \prod_{i=1}^n p(x_i|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \end{aligned}$$

Priori:

$$p(\mu) = \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{1}{2s}(\mu - m)^2\right) \propto \exp\left(-\frac{1}{2s}(\mu - m)^2\right)$$

Posteriori:

$$\begin{aligned}
 p(\mu|x) &\propto p(x|\mu)p(\mu) \\
 &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right)\exp\left(-\frac{1}{2s}(\mu-m)^2\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^n(x_i^2-2\mu x_i+\mu^2)+\frac{1}{s}(\mu^2-2\mu m+m^2)\right]\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}\left(\sum_{i=1}^nx_i^2-2\mu\sum_{i=1}^nx_i+n\mu^2\right)+\frac{1}{s}(\mu^2-2\mu m+m^2)\right]\right) \\
 &\propto \exp\left(\frac{1}{2}\left[\underbrace{\left(\frac{1}{\sigma^2}n+\frac{1}{s}\right)\mu^2}_{=:a}-2\underbrace{\left(\frac{1}{\sigma^2}\sum_{i=1}^nx_i+\frac{1}{s}m\right)\mu}_{=:b}+\underbrace{\left(\frac{1}{\sigma^2}\sum_{i=1}^nx_i^2+\frac{1}{s}m^2\right)}_{=:c}\right]\right) \\
 &\propto \exp\left(-\frac{1}{2}a\left[\mu^2-2\frac{b}{a}\mu+\frac{c}{a}\right]\right) \\
 &\propto \exp\left(-\frac{1}{2a^{-1}}\left[\mu^2-2\frac{b}{a}\mu+\left(\frac{b}{a}\right)^2-\left(\frac{b}{a}\right)^2+\frac{c}{a}\right]\right) \\
 &\propto \exp\left(-\frac{1}{2a^{-1}}\left[\left(\mu-\frac{b}{a}\right)^2-\left(\frac{b}{a}\right)^2+\frac{c}{a}\right]\right) \\
 &\propto \exp\left(-\frac{1}{2a^{-1}}\left(\mu-\frac{b}{a}\right)^2\right)\exp\left(-\frac{1}{2a^{-1}}\left[-\left(\frac{b}{a}\right)^2+\frac{c}{a}\right]\right) \\
 &\propto \exp\left(-\frac{1}{2a^{-1}}\left(\mu-\frac{b}{a}\right)^2\right)
 \end{aligned}$$

Das ist der Kern einer $N(\tilde{m}, \tilde{s})$ -Verteilung (d.h. Dichte ohne multiplikative Konstanten) mit

$$\tilde{m} = \frac{b}{a} = \frac{\frac{1}{\sigma^2}\sum_{i=1}^nx_i+\frac{1}{s}m}{\frac{1}{\sigma^2}n+\frac{1}{s}} = \frac{\frac{s}{n}\sum_{i=1}^nx_i+\frac{\sigma^2}{n}m}{s+\frac{\sigma^2}{n}} = \frac{s\bar{x}+\frac{\sigma^2}{n}m}{s+\frac{\sigma^2}{n}}$$

und

$$\tilde{s} = \frac{1}{a} = \frac{1}{\frac{1}{\sigma^2}n+\frac{1}{s}} = \frac{s\frac{\sigma^2}{n}}{s+\frac{\sigma^2}{n}}.$$

Da es genügt, die Posteriori nur bis auf eine multiplikative Konstante zu berechnen, gilt

$$\mu|x \sim N(\tilde{m}, \tilde{s}).$$

Bem:

Es gilt $\tilde{m} = \frac{s}{s+\frac{\sigma^2}{n}}\bar{x} + \frac{\frac{\sigma^2}{n}}{s+\frac{\sigma^2}{n}}m$, d.h. gewichtetes Mittel aus Daten-Mittelwert (\bar{x} , ML-Schätzer für μ) und Priori Erwartungswert für μ (m).

- (b) Erzeugen Sie in R $n = 10$ Realisierungen einer $N(4, 2)$ -verteilten Zufallsgröße X . Stellen Sie die Likelihood sowie die Dichten der Priori und Posteriori für den Erwartungswert μ graphisch dar, wenn als Priori $\mu \sim N(0, 10)$ angenommen wird.

Lösung:

R-Code *Likelihood als Funktion in μ*

$$p(x|\mu) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \propto \dots \propto \exp\left(-\frac{1}{2\sigma^2/n}(\mu - \bar{x})^2\right)$$

Kern von $N(\bar{x}, \sigma^2/n)$

- Priori: $(N(0, 10))$ mit hoher Varianz, d.h. flach.
- Likelihood: Dichte der Realisierungen ($\mu = 4$)
- Posteriori: Maximum bei $\mu \approx 4$, deutlich geringere Varianz als Priori.

- (c) Wie ändert sich die Darstellung mit wachsender Stichprobengröße n ?

Lösung:

R-Code

Für wachsendes n wird der Einfluss der Priori geringer, die Posteriori hat geringere Varianz und wird maximal rund um den wahren Wert $\mu = 4$, d.h.

$$\tilde{m} \xrightarrow{n \rightarrow \infty} \mu, \quad \tilde{s} \xrightarrow{n \rightarrow \infty} 0$$

(vgl. auch Formeln aus Teilaufgabe a))

Genauer:

$$\tilde{m} = \underbrace{\frac{s}{s + \frac{\sigma^2}{n}}}_{\xrightarrow{n \rightarrow \infty} 1 \Rightarrow \xrightarrow{P} 1} \cdot \underbrace{\bar{x}}_{\xrightarrow{P} \mu (\text{SchwGGZ})} + \underbrace{\frac{\sigma^2}{s + \frac{\sigma^2}{n}} \cdot \underbrace{m}_{\xrightarrow{n \rightarrow \infty} 0}}_{\text{const}} \xrightarrow{P} 1 \cdot \mu + 0 = \mu$$

Aufgabe 3

Aim:

- Entscheidungstheoretische Begriffe: Verlustfunktion, optimale Entscheidung, a posteriori erwarteter Verlust
- Technik: a posteriori erwarteten Verlust als Funktion in d (Entscheidung) betrachten und diese optimieren
- Spezialbeispiel: absolute Verlustfunktion

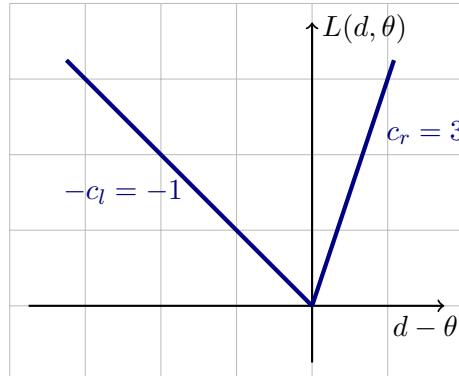
Betrachten Sie die asymmetrische Verlustfunktion

$$L(d, \theta) = \begin{cases} -c_l(d - \theta) & \text{falls } d - \theta \leq 0, \\ c_r(d - \theta) & \text{falls } d - \theta > 0, \end{cases}$$

mit $c_l, c_r > 0$ und $\theta \in \mathbb{R}$.

- (a) Skizzieren Sie $L(d, \theta)$ als Funktion von $(d - \theta)$ für $c_l = 1$ und $c_r = 3$.

Lösung:



- (b) Bestimmen Sie die Bayes-optimale Entscheidung d^* bezüglich der Verlustfunktion $L(d, \theta)$.

Lösung:

Es gilt: $d^* = \arg \min_d r(d, p|x)$.

Bestimme zunächst a posteriori erwarteten Verlust $r(d, p|x)$ für feste Entscheidung d :

$$\begin{aligned}
 r(d, p|x) &= \mathbb{E}[L((d(X), \theta)|x)] \\
 &= \int_{\Theta} L(\underbrace{d(x)}_{=:d}, \theta) p(\theta|x) d\theta \\
 &= \int_{-\infty}^{\infty} L(d, \theta) p(\theta|x) d\theta \\
 &= \int_d^{\infty} -c_l(d - \theta) p(\theta|x) d\theta + \int_{-\infty}^d c_r(d - \theta) p(\theta|x) d\theta \\
 &= \lim_{z \rightarrow \infty} \int_d^z -c_l(d - \theta) p(\theta|x) d\theta + \lim_{z \rightarrow \infty} \int_{-z}^d c_r(d - \theta) p(\theta|x) d\theta \\
 &=: \lim_{z \rightarrow \infty} r_1(d, z) + \lim_{z \rightarrow \infty} r_2(d, z)
 \end{aligned}$$

Die Bayes-optimale Entscheidung d^* minimiert $r(d, p|x)$. Betrachte also

$$\begin{aligned}
 \frac{\partial}{\partial d} r(d, p|x) \stackrel{!}{=} 0 &\Leftrightarrow \frac{\partial}{\partial d} \lim_{z \rightarrow \infty} r_1(d, z) + \lim_{z \rightarrow \infty} r_2(d, z) \stackrel{!}{=} 0 \\
 &\stackrel{\text{Hinweis}}{\Leftrightarrow} \lim_{z \rightarrow \infty} \frac{\partial}{\partial d} r_1(d, z) + \lim_{z \rightarrow \infty} \frac{\partial}{\partial d} r_2(d, z) \stackrel{!}{=} 0
 \end{aligned}$$

Betrachte die Ableitungen einzeln:

•

$$\begin{aligned}
 \frac{\partial}{\partial d} r_1(d) &= \frac{\partial}{\partial d} \int_d^z \underbrace{-c_l(d-\theta)p(\theta|x)}_{=:g_1(d,\theta)} d\theta \\
 &\stackrel{\text{Hinweis}}{=} \int_d^z \frac{\partial}{\partial d} g_1(d,\theta) d\theta + g_1(d,z) \frac{\partial}{\partial d} z - g_1(d,d) \frac{\partial}{\partial d} d \\
 &= \int_d^z -c_l p(\theta|x) d\theta + g_1(d,z) \cdot 0 - [-c_l(d-d)p(\theta|x)] \cdot 1 \\
 &= -c_l \int_d^z p(\theta|x) d\theta \\
 &= -c_l [F_{\theta|x}(z) - F_{\theta|x}(d)]
 \end{aligned}$$

•

$$\begin{aligned}
 \frac{\partial}{\partial d} r_2(d) &= \frac{\partial}{\partial d} \int_{-z}^d \underbrace{c_r(d-\theta)p(\theta|x)}_{=:g_2(d,\theta)} d\theta \\
 &\stackrel{\text{Hinweis}}{=} \int_{-z}^d \frac{\partial}{\partial d} g_2(d,\theta) d\theta + g_2(d,d) \frac{\partial}{\partial d} d - g_2(d,-z) \frac{\partial}{\partial d} -z \\
 &= \int_{-z}^d c_r p(\theta|x) d\theta + c_r(d-d)p(\theta|x) \cdot 1 - g_2(d,-z) \cdot 0 \\
 &= c_r \int_{-z}^d p(\theta|x) d\theta \\
 &= c_r [F_{\theta|x}(d) - F_{\theta|x}(-z)]
 \end{aligned}$$

Insgesamt erhält man also:

$$\begin{aligned}
 \frac{\partial}{\partial d} r(d, p|x) &\stackrel{!}{=} 0 \Leftrightarrow \lim_{z \rightarrow \infty} -c_l \left[\underbrace{F_{\theta|x}(z)}_{\rightarrow 1 (z \rightarrow \infty)} - F_{\theta|x}(d) \right] + \lim_{z \rightarrow \infty} c_r \left[F_{\theta|x}(d) - \underbrace{F_{\theta|x}(-z)}_{\rightarrow 0 (z \rightarrow \infty)} \right] = 0 \\
 &\Leftrightarrow -c_l + c_l F_{\theta|x}(d) + c_r F_{\theta|x}(d) = 0 \\
 &\Leftrightarrow F_{\theta|x}(d) = \frac{c_l}{c_l + c_r}
 \end{aligned}$$

Kandidat für das Optimum ist also das $\frac{c_l}{c_l + c_r}$ -Quantil der Posteriori-Verteilung.

Überprüfe, ob tatsächlich ein Minimum vorliegt:

$$\frac{\partial^2}{\partial d^2} r(d, p|x) = \frac{\partial}{\partial d} -c_l + c_l F_{\theta|x}(d) + c_r F_{\theta|x}(d) = (c_l + c_r)p(d|x) > 0 \quad \checkmark$$

Also ist die Bayes-optimale Entscheidung d^* gleich dem $\frac{c_l}{c_l + c_r}$ -Quantil der Posteriori-Verteilung.

Hinweise:

- Sie dürfen annehmen, dass $\frac{\partial}{\partial d} \lim_{z \rightarrow \infty} r_i(d, z) = \lim_{z \rightarrow \infty} \frac{\partial}{\partial d} r_i(d, z)$ für $i = 1, 2$ und

$$r_1(d, z) = \int_d^z L(d, \theta) p(\theta|x) d\theta, \quad r_2(d, z) = \int_{-z}^d L(d, \theta) p(\theta|x) d\theta.$$

- Für das Differenzieren einer Funktion in einem Integral gilt

$$\frac{\partial}{\partial x} \int_{a(x)}^{b(x)} g(x, t) dt = \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} g(x, t) dt + g(x, b(x)) \frac{\partial}{\partial x} b(x) - g(x, a(x)) \frac{\partial}{\partial x} a(x).$$

- (c) Benutzen Sie dieses allgemeine Resultat, um zu zeigen, dass für die absolute Verlustfunktion $L(d, \theta) = |d - \theta|$ der Posteriori-Median Bayes-optimal ist.

Lösung:

Absolute Verlustfunktion:

$$L(d, \theta) = |d - \theta| = \begin{cases} -(d - \theta) & d - \theta \leq 0 \\ (d - \theta) & d - \theta > 0 \end{cases}$$

Die absolute Verlustfunktion ist also ein Spezialfall der Verlustfunktion aus (b) mit $c_l = c_r = 1$. Die Bayes-optimale Entscheidung ist das $\frac{1}{1+1} = \frac{1}{2}$ -Quantil (also der Median) der Posteriori.

Aufgabe 4

Aim:

- Entscheidungstheoretische Begriffe: Verlustfunktion, optimale Entscheidung, a posteriori erwarteter Verlust, Bayes-Risiko
- Technik: a posteriori erwarteten Verlust als Funktion in d (Entscheidung) betrachten und diese optimieren
- Spezialbeispiel: quadratische Verlustfunktion

- (a) Zeigen Sie, dass für die quadratische Verlustfunktion $L(d, \theta) = (d - \theta)^2$ der Posteriori-Erwartungswert $\mathbb{E}[\theta|x]$ Bayes-optimal ist.

Lösung:

Es gilt: $d^* = \arg \min_d r(d, p|x)$.

Bestimme zunächst a posteriori erwarteten Verlust $r(d, p|x)$ für feste Entscheidung d :

$$\begin{aligned} r(d, p|x) &= \mathbb{E} \left[L \left(\underbrace{d(X)}_{=:d}, \theta \right) \middle| x \right] \\ &= \mathbb{E} \left[(d - \theta)^2 \middle| x \right] \\ &= \mathbb{E} [d^2 - 2d\theta + \theta^2 \middle| x] \\ &= d^2 - 2d \cdot \mathbb{E} [\theta|x] + \mathbb{E} [\theta^2|x] \end{aligned}$$

Die Bayes-optimale Entscheidung d^* minimiert $r(d, p|x)$. Betrachte also

$$\frac{\partial}{\partial d} r(d, p|x) \stackrel{!}{=} 0 \Leftrightarrow 2d - 2 \cdot \mathbb{E}[\theta|x] \stackrel{!}{=} 0 \Leftrightarrow d = \mathbb{E}[\theta|x].$$

Überprüfe, ob tatsächlich ein Minimum vorliegt:

$$\frac{\partial^2}{\partial d^2} r(d, p|x) = 2 > 0 \quad \forall d$$

Die Bayes-optimale Entscheidung für die quadratische Verlustfunktion ist $d^* = \mathbb{E}[\theta|x]$ (Posteriori-Erwartungswert).

Sei nun x eine Realisierung von einer Zufallsgröße $X \sim N(\mu, 1)$. Für den unbekannten Erwartungswert werde a priori $\mu \sim N(0, s)$, $s > 0$ angenommen.

(b) Bestimmen Sie den Posteriori-Erwartungswert $d^* = \mathbb{E}(\mu|x)$.

Lösung:

Konjugierte Priori, daher Posteriori

$\mu|x \sim N(\tilde{m}, \tilde{s})$ mit

$$\tilde{m} = \frac{sx + 0 \cdot 1}{s + 1} = \frac{sx}{s + 1} \quad \tilde{s} = \frac{s \cdot 1}{s + 1} = \frac{s}{s + 1}.$$

$(x = x_1 \Rightarrow \bar{x} = x, n = 1; \sigma^2 = 1, m = 0)$. Der Posteriori-Erwartungswert ist also $\mathbb{E}(\mu|x) = \frac{xs}{s+1}$.

Bem: $\mathbb{E}(\mu|x)$ ist nach Aufgabe (a) Bayes-optimal für die quadratische Verlustfunktion.

(c) Bestimmen Sie das Bayes-Risiko $r^*(p) = r(d^*, p|x)$.

Lösung:

$$r^*(p) = r(d^*, p|x) = \mathbb{E}[L(d^*, \mu)|x] = \mathbb{E}\left[(\mathbb{E}(\mu|x) - \mu)^2 \middle| x\right] = \text{Var}[\mu|x] = \frac{s}{s+1}.$$