Institut für Angewandte Mathematik
Sommersemester 2022
Andreas Eberle, Luis Aniello La Rocca

13. Übungsblatt „Einführung in die Statistik"

Für dieses Übungsblatt ist keine Abgabe erforderlich.

Hinweise zur Klausur:

- **Diese findet am Freitag 29.7. von 9.00 bis 11.00 Uhr im GHS statt.**
- **Unterlagen, Taschenrechner, Handys etc. sind nicht zugelassen.**
- **Mit der Klausur erhalten Sie auch die Verteilungstabellen unten.**

---

1. (Bayes I) Seien $X_1, X_2, \dots, X_n$ unabhängig mit $P[X_i = 1] = p$ und $P[X_i = 0] = 1-p$. Wir betrachten ein Bayessches Modell, in dem die a priori Verteilung für den unbekannten Parameter p eine Beta-Verteilung mit Parametern $a, b > 0$ ist.

a) Bestimmen Sie die a posteriori Verteilung gegeben die Beobachtungswerte $X_1 = x_1$, $X_2 = x_2$, ... ,$X_n = x_n$.

b) Berechnen Sie den Bayes-Schätzer.

c) Sei $x = (0, 1, 0, 1, 0, 0, 0, 0, 0, 0)^T$. Skizzieren Sie die Graphen der posteriori Dichten für die a priori Verteilungen $\beta(1/2, 1/2)$, $\beta(1, 1)$, $\beta(10, 10)$ und $\beta(100, 100)$.

d) Interpretieren Sie die Ergebnisse.

Hinweis: Der Erwartungswert der Beta-Verteilung ist $b/(a + b)$.

Lösung:

a) Die Likelihood Funktion der Beobachtungswert gegeben dem Parameter p ist auf Grund der Unabhängigkeit gegeben durch

$$f_{X|p}(x_1, \dots, x_n, p) = p^{s_n}(1-p)^{n-s_n}$$

wobei $s_n = x_1 + \ldots + x_n$. Die a priori Verteilung des Parameters p ist eine Beta-Verteilung mit Parametern $a, b > 0$. Diese hat eine Dichte, die gegeben ist durch

$$f_p(p) = \frac{1}{\mathrm{B}(a, b)} p^{a-1}(1-p)^{b-1}.$$

wobei $\mathrm{B}(a, b)$ die Beta-Funktion ist. Dann ist die a posteriori Verteilung des Parameters p gegeben der Beobachtungen gegeben durch

$$
\begin{aligned}
f_{p|X}(p, x_1, \dots, x_n) &= \frac{p^{s_n}(1-p)^{n-s_n}\frac{1}{\mathrm{B}}(a,b)p^{a-1}(1-p)^{b-1}}{\int\limits_0^1 q^{s_n}(1-q)^{n-s_n}\frac{1}{\mathrm{B}}(a,b)q^{a-1}(1-q)^{b-1}\mathrm{d}q} \\
&= \frac{p^{s_n+a-1}(1-p)^{n-s_n+b-1}}{\int\limits_0^1 q^{s_n+a-1}(1-q)^{n-s_n+b-1}\mathrm{d}q} \\
&= \frac{1}{\mathrm{B}(s_n+a, n-s_n+b)}p^{s_n+a-1}(1-p)^{n-s_n+b-1}
\end{aligned}
$$

Also ist die a posteriori Verteilung des Parameters p gegeben der Beobachtungen wieder eine Beta-Verteilung mit Parametern $S_n + a$ und $n - S_n + b$.

b) Der Bayes-Schätzer ist der Erwartungswert der a posteriori Verteilung, also gegeben durch

$$
p_{\text{post}} = \frac{n - S_n + b}{S_n + a + n - S_n + b} = \frac{n - S_n + b}{n + a + b}.
$$

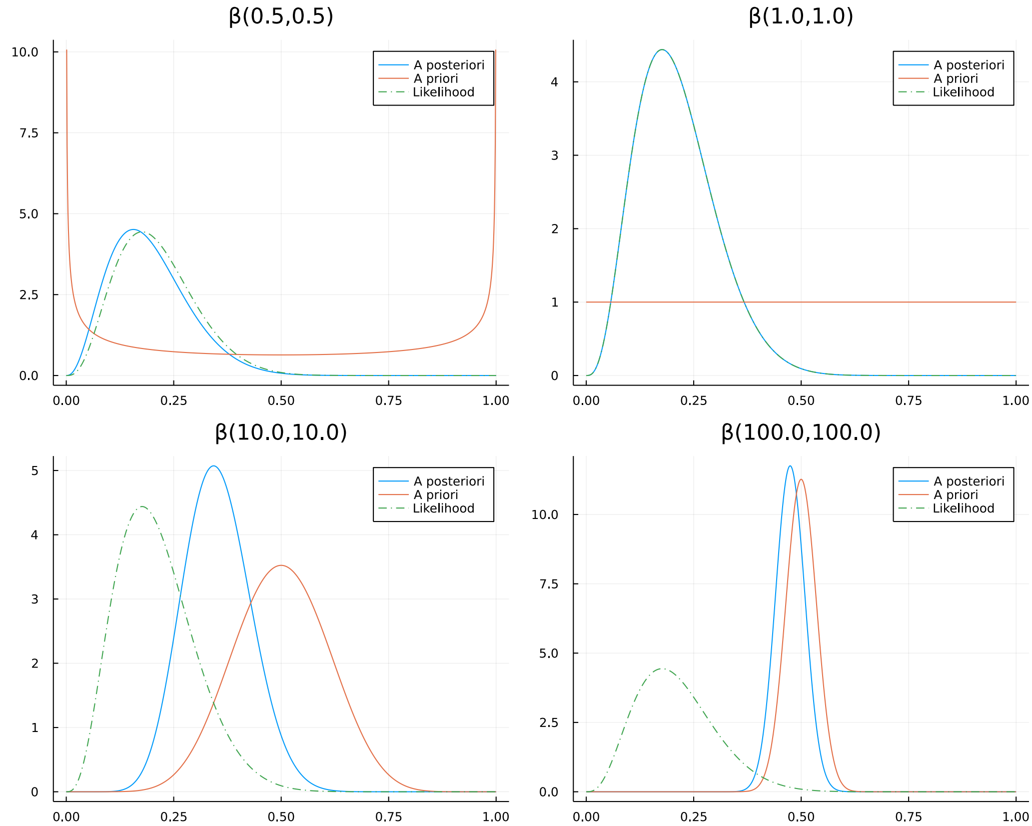

Abbildung 1: Skizzen aus Aufgabenteil (c)

d) In den Graphen aus Aufgabenteil (c) sieht man zu den a priori und a posteriori Dichten für den Parameter p die Dichte einer Beta-Verteilung mit Parametern $S_n + 1$ und

$n - S_n + 1$. Daran erkennt man, dass die schwachen a priori Verteilungen die posteriori Verteilung kaum verändern, wohingegen die starken a posteriori Verteilungen kaum von den Informationen aus den Daten bewegt werden. Lediglich im für die a priori Verteilung $\beta(10, 10)$ sieht man, dass die a posteriori Verteilung ein Kompromiss aus der Likelihood und der a priori Verteilung ist. Darüber hinaus sieht man, dass die Gleichverteilung (hier als $\beta(1, 1)$) keine Informationen als a priori Verteilung liefert.

2. (Bayes II: Das lineare Gauß-Modell) Sei

$$Y \;=\; Aw + \varepsilon$$

mit unbekanntem Parameter $w \in \mathbb{R}^d$, Design-Matrix $A \in \mathbb{R}^{n\times d}$ und Störterm $\varepsilon \sim N(0, C)$, wobei die Kovarianzmatrix $C \in \mathbb{R}^{n\times n}$ symmetrisch und positiv definit ist. Wir betrachten das Bayes-Modell mit a priori Verteilung $w \sim N(0, C_{\text{prior}})$, $C_{\text{prior}} \in \mathbb{R}^{d\times d}$ symmetrisch und positiv definit.

a) Bestimmen Sie die a posteriori Verteilung im Gaußschen Produktmodell, d.h. im Fall $d = 1$, $A = (1, 1, \ldots, 1)^T$ und $C = v \cdot I_n$. Vergleichen Sie die den Bayes-Schätzer mit dem Maximum-Likelihood-Schätzer.

b) Bestimmen Sie die a posteriori Verteilung im allgemeinen Fall.

c) Interpretieren Sie das Ergebnis aus b) im Fall $C = v \cdot I_n$. Welchen Zusammenhang gibt es mit dem kleinste Quadrate Verfahren?

Lösung:

a) Im Gaußschen Produktmodell gilt $A = (1, 1, \ldots, 1)^T$, also $A^T A = n$. Damit erhalten wir als Spezialfall des allgemeinen Resultats in b) (oder mit separater Rechnung):

$$\begin{aligned}
\underbrace{C_{\text{post}}^{-1}}_{\text{posterior Präzision}} &= \underbrace{C_{\text{prior}}^{-1}}_{\text{prior Präzision}} + \underbrace{\frac{1}{v/n}}_{\text{Präzision des MLE } \bar{Y}} \\
m_{\text{post}} &= \left(vC_{\text{prior}}^{-1} + n\right)^{-1} \sum_{i=1}^{n} y_i = \frac{1}{1 + \frac{v/n}{C_{\text{prior}}}} \bar{y} \\
&= \frac{1}{1 + \frac{v/n}{C_{\text{prior}}}} \underbrace{\bar{y}}_{\text{MLE}} + \frac{\frac{v/n}{C_{\text{prior}}}}{1 + \frac{v/n}{C_{\text{prior}}}} \underbrace{0}_{\text{Mittelwert des priors}}
\end{aligned}$$

b)

$$Y = Aw + \varepsilon \quad , \quad w \sim \mathcal{N}\left(0, C_{\text{prior}}\right) \quad , \quad \varepsilon \sim \mathcal{N}\left(0, C\right) \;\text{ unabh.}$$

Nach der Bayesschen Formel gilt für die posteriori Verteilung

$$
\begin{aligned}
f(w|y) &= C(y) \cdot f(w) \cdot f(y|w) \quad \text{bzw.} \\
\log f(w|y) &= \log C(y) + \log f(w) + \log f(y|w) \\
&= \mathrm{const}(y) + \frac{1}{2} w^T C_{\mathrm{prior}}^{-1} w + \frac{1}{2}(y - Aw)^T C^{-1} (y - Aw) \\
&= \widetilde{\mathrm{const}}(y) + \frac{1}{2} w^T \underbrace{\left(C_{\mathrm{prior}}^{-1} + A^T C^{-1} A\right)}_{=:C_{\mathrm{post}}^{-1}} w - w^T A^T C^{-1} y \\
&= \widetilde{\widetilde{\mathrm{const}}}(y) + \frac{1}{2} \left\| C_{\mathrm{post}}^{-1/2} \left(w - \underbrace{C_{\mathrm{post}} A^T C^{-1} y}_{=:m_{\mathrm{post}}}\right) \right\|^2
\end{aligned}
$$

Also

$$
f(w|y) \propto e^{-\frac{1}{2}(w - m_{\mathrm{post}})^T C_{\mathrm{post}}^{-1} (w - m_{\mathrm{post}})}
$$

d.h. die posteriori Verteiung ist $\mathcal{N}(m_{\mathrm{post}}, C_{\mathrm{post}})$.

c) Für $C = vI_n$ erhalten wir

$$
\begin{aligned}
C_{\mathrm{post}}^{-1} &= C_{\mathrm{prior}}^{-1} + \frac{1}{v} A^T A m_{\mathrm{post}} & &= \left(C_{\mathrm{prior}}^{-1} + \frac{1}{v} A^T A\right)^{-1} A^T \frac{1}{v} y \\
&= \left(v C_{\mathrm{prior}}^{-1} + A^T A\right)^{-1} A^T y
\end{aligned}
$$

Der kleinste Quadrate Schätzer ergibt sich also, fall die Präzision C_{prior}^{-1} der a-priori Verteilung verschwindet ("flache a priori Verteilung").