KAPITEL 5

# Methoden zur Konstruktion von Schätzern

## 5.1. Parametrisches Modell

Sei $(x_1, \ldots, x_n)$ eine Stichprobe. In der parametrischen Statistik nimmt man an, dass die Stichprobe $(x_1, \ldots, x_n)$ eine Realisierung von unabhängigen und identisch verteilten Zufallsvariablen $(X_1, \ldots, X_n)$ mit Verteilungsfunktion $F_\theta(x)$ ist. Dabei hängt die Verteilungsfunktion $F_\theta$ von einem unbekannten Wert (*Parameter*) $\theta$ ab. In den meisten Fällen nimmt man außerdem an, dass entweder die Zufallsvariablen $X_i$ für alle Werte des Parameters $\theta$ absolut stetig sind und eine Dichte $h_\theta$ besitzen, oder dass sie für alle Werte von $\theta$ diskret sind und eine Zähldichte besitzen, die ebenfalls mit $h_\theta$ bezeichnet wird. Die Aufgabe der parametrischen Statistik besteht darin, den unbekannten Parameter $\theta$ anhand der bekannten Stichprobe $(x_1, \ldots, x_n)$ zu schätzen.

Die Menge aller möglichen Werte des Parameters $\theta$ wird der *Parameterraum* genannt und mit $\Theta$ bezeichnet. In den meisten Fällen ist $\theta = (\theta_1, \ldots, \theta_p)$ ein Vektor mit Komponenten $\theta_1, \ldots, \theta_p$. In diesem Fall muss der Parameterraum $\Theta$ eine Teilmenge von $\mathbb{R}^p$ sein.

Um den Parameter $\theta$ anhand der Stichprobe $(x_1, \ldots, x_n)$ zu schätzen, konstruiert man einen Schätzer.

Definition 5.1.1. Ein *Schätzer* ist eine Abbildung

$$\hat{\theta} : \mathbb{R}^n \to \Theta, \quad (x_1, \ldots, x_n) \mapsto \hat{\theta}(x_1, \ldots, x_n).$$

Man muss versuchen, den Schätzer so zu konstruieren, dass $\hat{\theta}(x_1, \ldots, x_n)$ den wahren Wert des Parameters $\theta$ möglichst gut approximiert. Wie das geht, werden wir im Weiteren sehen.

Beispiel 5.1.2. Wir betrachten ein physikalisches Experiment, bei dem eine physikalische Konstante (z.B. die Lichtgeschwindigkeit) bestimmt werden soll. Bei $n$ unabhängigen Messungen der Konstanten ergaben sich die Werte $(x_1, \ldots, x_n)$. Normalerweise nimmt man an, dass diese Stichprobe eine Realisierung von $n$ unabhängigen und identisch verteilten Zufallsvariablen $(X_1, \ldots, X_n)$ mit einer Normalverteilung ist:

$$X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2).$$

Dabei ist $\mu$ der wahre Wert der zu bestimmenden Konstanten und $\sigma^2$ die quadratische Streuung des Experiments. Beide Parameter sind unbekannt. Somit besteht das Problem, den Parameter $\theta = (\mu, \sigma^2)$ aus den gegebenen Daten $(x_1, \ldots, x_n)$ zu schätzen. In diesem Beispiel ist der Parameterraum gegeben durch

$$\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\} = \mathbb{R} \times (0, \infty).$$

Die Dichte von $X_i$ ist gegeben durch (siehe auch Abbildung 1)

$$h_{\mu,\sigma^2}(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

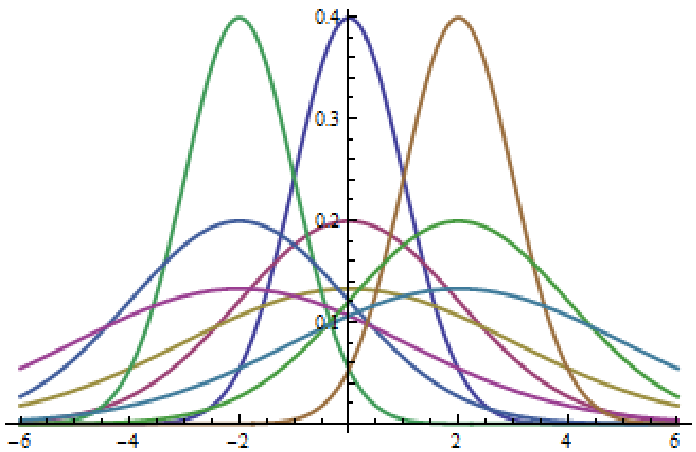


Abbildung 1. Das Bild zeigt die Dichten der Normalverteilungen, die zu verschiedenen Werten der Parameter $\mu$ und $\sigma^2$ gehören. Die Aufgabe der parametrischen Statistik ist es, zu entscheiden, zu welchen Parameterwerten eine gegebene Stichprobe gehört.

Als Schätzer für $\mu$ und $\sigma^2$ können wir z.B. den empirischen Mittelwert und die empirische Varianz verwenden:

$$\hat{\mu}(x_1, \ldots, x_n) = \frac{x_1 + \ldots + x_n}{n} = \bar{x}_n, \quad \hat{\sigma}^2(x_1, \ldots, x_n) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = s_n^2.$$

In den nächsten drei Abschnitten werden wir die drei wichtigsten Methoden zur Konstruktion von Schätzern betrachten: die *Momentenmethode*, die *Maximum-Likelihood-Methode* und die *Bayes-Methode*.

An dieser Stelle müssen wir noch eine Notation einführen. Um im parametrischen Modell die Verteilung der Zufallsvariablen $X_1, \ldots, X_n$ eindeutig festzugelegen, muss man den Wert des Parameters $\theta$ angeben. Bevor man von der Wahrscheinlichkeit eines mit $X_1, \ldots, X_n$ verbundenen Ereignisses spricht, muss man also sagen, welchen Wert der Parameter $\theta$ annehmen soll. Wir werden deshalb sehr oft die folgende Notation verwenden. Mit $\mathbb{P}_\theta[A]$ bezeichnen wir die Wahrscheinlichkeit eines Ereignisses $A$ unter der Annahme, dass die Zufallsvariablen $X_i$ unabhängig und identisch verteilt mit Verteilungsfunktion $F_\theta$ (bzw. mit Dichte/ Zähldichte $h_\theta$) sind. Dabei können sich $\mathbb{P}_{\theta_1}[A]$ und $\mathbb{P}_{\theta_2}[A]$ durchaus unterscheiden. Analog bezeichnen wir mit $\mathbb{E}_\theta Z$ und $\operatorname{Var}_\theta Z$ den Erwartungswert bzw. die Varianz einer Zufallsvariable $Z$ unter der Annahme, dass die Zufallsvariablen $X_i$ unabhängig und identisch verteilt mit Verteilungsfunktion $F_\theta$ (bzw. mit Dichte/ Zähldichte $h_\theta$) sind.
Die Zufallsvariablen $X_1, \ldots, X_n$ kann man sich als messbare Funktionen auf einem Messraum $(\Omega, \mathcal{A})$ denken. In der Wahrscheinlichkeitstheorie musste man außerdem ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf diesem Raum angeben. Im parametrischen Modell brauchen wir nicht *ein*

Wahrscheinlichkeitsmaß, sondern eine durch $\theta$ parametrisierte *Familie* von Wahrscheinlichkeitsmaßen $\{\mathbb{P}_\theta : \theta \in \Theta\}$ auf $(\Omega, \mathcal{A})$. Je nachdem welchen Wert der Parameter $\theta$ annimmt, können wir eines dieser Wahrscheinlichkeitsmaße verwenden.

## 5.2. Momentenmethode

Wie in der parametrischen Statistik üblich, nehmen wir an, dass die Stichprobe $(x_1, \ldots, x_n)$ eine Realisierung der unabhängigen und identisch verteilten Zufallsvariablen $(X_1, \ldots, X_n)$ mit Verteilungsfunktion $F_\theta$ ist. Dabei ist $\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$ der unbekannte Parameter. Für die Momentenmethode brauchen wir die folgenden Begriffe.

Definition 5.2.1. Das $k$-te *theoretische Moment* (mit $k \in \mathbb{N}$) der Zufallsvariable $X_i$ ist definiert durch

$$m_k(\theta) = \mathbb{E}_\theta[X_i^k].$$

Zum Beispiel ist $m_1(\theta)$ der Erwartungswert von $X_i$. Die theoretischen Momente sind Funktionen des Parameters $\theta$.

Definition 5.2.2. Das $k$-te *empirische Moment* (mit $k \in \mathbb{N}$) der Stichprobe $(x_1, \ldots, x_n)$ ist definiert durch

$$\hat{m}_k = \frac{x_1^k + \ldots + x_n^k}{n}.$$

Zum Beispiel ist $\hat{m}_1$ der empirische Mittelwert $\bar{x}_n$ der Stichprobe.

Die Idee der Momentenmethode besteht darin, die empirischen Momente den theoretischen gleichzusetzen. Dabei sind die empirischen Momente bekannt, denn sie hängen nur von der Stichprobe $(x_1, \ldots, x_n)$ ab. Die theoretischen Momente sind hingegen Funktionen des unbekannten Parameters $\theta$, bzw. Funktionen seiner Komponenten $\theta_1, \ldots, \theta_p$. Um $p$ unbekannte Parameter zu finden, brauchen wir normalerweise $p$ Gleichungen. Wir betrachten also ein System aus $p$ Gleichungen mit $p$ Unbekannten:

$$m_1(\theta_1, \ldots, \theta_p) = \hat{m}_1, \quad \ldots, \quad m_p(\theta_1, \ldots, \theta_p) = \hat{m}_p.$$

Die Lösung dieses Gleichungssystems (falls sie existiert und eindeutig ist) nennt man den *Momentenschätzer* und bezeichnet ihn mit $\hat{\theta}_{\mathrm{ME}}$. Dabei steht "ME" für "Moment Estimator".

Beispiel 5.2.3. Momentenmethode für den Parameter der Bernoulli–Verteilung $\mathrm{Bern}(\theta)$.
In diesem Beispiel betrachten wir eine unfaire Münze. Die Wahrscheinlichkeit $\theta$, dass die Münze bei einem Wurf "Kopf" zeigt, sei unbekannt. Um diesen Parameter zu schätzen, werfen wir die Münze $n = 100$ Mal. Nehmen wir an, dass die Münze dabei $s = 60$ Mal "Kopf" gezeigt hat. Das Problem besteht nun darin, $\theta$ zu schätzen.
Wir betrachten für dieses Problem das folgende mathematische Modell. Zeigt die Münze bei Wurf $i$ Kopf, so setzen wir $x_i = 1$, ansonsten sei $x_i = 0$. Auf diese Weise erhalten wir eine Stichprobe $(x_1, \ldots, x_n) \in \{0, 1\}^n$ mit $x_1 + \ldots + x_n = s = 60$. Wir nehmen an, dass $(x_1, \ldots, x_n)$ eine Realisierung von $n$ unabhängigen Zufallsvariablen $X_1, \ldots, X_n$ mit einer Bernoulli-Verteilung mit Parameter $\theta \in [0, 1]$ ist, d.h.

$$\mathbb{P}_\theta[X_i = 1] = \theta, \quad \mathbb{P}_\theta[X_i = 0] = 1 - \theta.$$

Da wir nur einen unbekannten Parameter haben, brauchen wir nur das erste Moment zu betrachten. Das erste theoretische Moment von $X_i$ ist gegeben durch

$$m_1(\theta) = \mathbb{E}_\theta X_i = 1 \cdot \mathbb{P}_\theta[X_i = 1] + 0 \cdot \mathbb{P}_\theta[X_i = 0] = \theta.$$

Das erste empirische Moment ist gegeben durch

$$\hat{m}_1 = \frac{x_1 + \ldots + x_n}{n} = \frac{s}{n} = \frac{60}{100} = 0.6.$$

Setzen wir beide Momente gleich, so erhalten wir den Momentenschätzer

$$\hat{\theta}_{\mathrm{ME}} = \frac{s}{n} = 0.6.$$

Das Ergebnis ist natürlich nicht überraschend.

Beispiel 5.2.4. Momentenmethode für die Parameter der Normalverteilung $\mathrm{N}(\mu, \sigma^2)$. Sei $(x_1, \ldots, x_n)$ eine Realisierung von unabhängigen und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$, die eine Normalverteilung mit unbekannten Parametern $(\mu, \sigma^2)$ haben. Als Motivation kann etwa Beispiel 5.1.2 dienen. Wir schätzen $\mu$ und $\sigma^2$ mit der Momentenmethode. Da wir zwei Parameter haben, brauchen wir zwei Gleichungen (also Momente der Ordnungen 1 und 2), um diese zu finden. Zuerst berechnen wir die theoretischen Momente. Der Erwartungswert und die Varianz einer $\mathrm{N}(\mu, \sigma^2)$-Verteilung sind gegeben durch

$$\mathbb{E}_{\mu,\sigma^2} X_i = \mu, \quad \operatorname{Var}_{\mu,\sigma^2} X_i = \sigma^2.$$

Daraus ergeben sich die ersten zwei theoretischen Momente:

$$\begin{aligned} m_1(\mu, \sigma^2) &= \mathbb{E}_{\mu,\sigma^2}[X_i] = \mu, \\ m_2(\mu, \sigma^2) &= \mathbb{E}_{\mu,\sigma^2}[X_i^2] = \operatorname{Var}_{\mu,\sigma^2} X_i + (\mathbb{E}_{\mu,\sigma^2}[X_i])^2 = \sigma^2 + \mu^2. \end{aligned}$$

Setzt man die theoretischen und die empirischen Momente gleich, so erhält man das Gleichungssystem

$$\begin{aligned} \frac{x_1 + \ldots + x_n}{n} &= \mu, \\ \frac{x_1^2 + \ldots + x_n^2}{n} &= \sigma^2 + \mu^2. \end{aligned}$$

Dieses Gleichungssystem lässt sich wie folgt nach $\mu$ und $\sigma^2$ auflösen:

$$\begin{aligned} \mu &= \bar{x}_n, \\ \sigma^2 &= \frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2 = \frac{1}{n}\left(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2\right) = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{n-1}{n} s_n^2. \end{aligned}$$

Dabei haben wir die Identität $\sum_{i=1}^n x_i^2 - n\bar{x}_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2$ benutzt (Übung). Somit sind die Momentenschätzer gegeben durch

$$\hat{\mu}_{\mathrm{ME}} = \bar{x}_n, \quad \hat{\sigma}^2_{\mathrm{ME}} = \frac{n-1}{n} s_n^2.$$

Beispiel 5.2.5. Momentenmethode für den Parameter der Poisson–Verteilung $\mathrm{Poi}(\theta)$.
In diesem Beispiel betrachten wir ein Portfolio aus $n$ Versicherungsverträgen. Es sei $x_i \in \{0, 1, \ldots\}$ die Anzahl der Schäden, die der Vertrag $i$ in einem bestimmten Zeitraum erzeugt hat:

| Vertrag | 1 | 2 | 3 | ... | $n$ |
|---|---|---|---|---|---|
| Schäden | $x_1$ | $x_2$ | $x_3$ | ... | $x_n$ |

In der Versicherungsmathematik nimmt man oft an, dass die konkrete Stichprobe $(x_1, \ldots, x_n)$ eine Realisierung von $n$ unabhängigen und identisch verteilten Zufallsvariablen $(X_1, \ldots, X_n)$ ist, die eine Poissonverteilung mit einem unbekannten Parameter $\theta \geq 0$ haben.

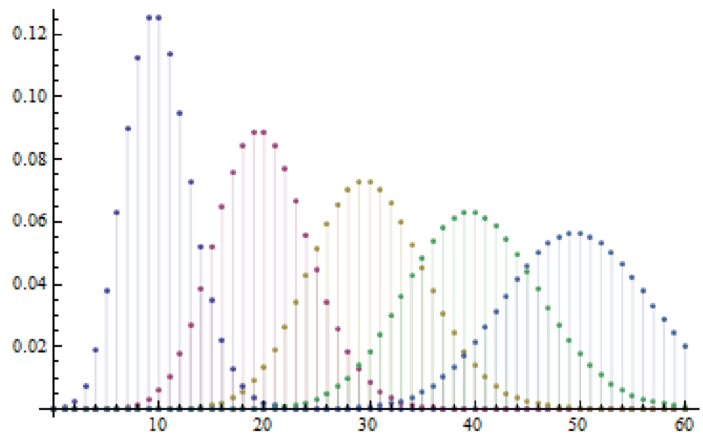


Abbildung 2. Zähldichten der Poissonverteilungen, die zu verschiedenen Werten des Parameters $\theta$ gehören.

Wir schätzen $\theta$ mit der Momentenmethode. Da der Erwartungswert einer $\mathrm{Poi}(\theta)$-Verteilung gleich $\theta$ ist, gilt

$$m_1(\theta) = \mathbb{E}_\theta X_i = \theta.$$

Das erste empirische Moment ist gegeben durch

$$\hat{m}_1(\theta) = \frac{x_1 + \ldots + x_n}{n} = \bar{x}_n.$$

Nun setzen wir die beiden Momente gleich und erhalten den Momentenschätzer

$$\hat{\theta}_{\mathrm{ME}} = \bar{x}_n.$$

## 5.3. Maximum-Likelihood-Methode

Die Maximum-Likelihood-Methode wurde von Carl Friedrich Gauß entdeckt und von Ronald Fisher weiterentwickelt. Die Maximum-Likelihood-Methode ist (wie auch die Momentenmethode) ein Verfahren, um Schätzer für die unbekannten Komponenten des Parametervektors $\theta = (\theta_1, \ldots, \theta_p)$ zu gewinnen. Sei $(x_1, \ldots, x_n)$ eine Stichprobe. Wir werden annehmen, dass entweder alle Verteilungen aus der parametrischen Familie $\{F_\theta : \theta \in \Theta\}$ diskret oder alle Verteilungen absolut stetig sind.

Der diskrete Fall. Seien zuerst die Zufallsvariablen $X_i$ für alle Werte des Parameters $\theta$ diskret. Wir bezeichnen die Zähldichte von $X_i$ mit $h_\theta$. Dann ist die *Likelihood-Funktion* gegeben durch

$$L(\theta) = L(x_1, \dots, x_n; \theta) = \mathbb{P}_\theta[X_1 = x_1, \dots, X_n = x_n].$$

Die Likelihood-Funktion hängt sowohl von der Stichprobe, als auch vom Parameterwert $\theta$ ab, wir werden sie aber hauptsächlich als Funktion von $\theta$ auffassen. Wegen der Unabhängigkeit von $X_1, \dots, X_n$ gilt

$$L(x_1, \dots, x_n; \theta) = \mathbb{P}_\theta[X_1 = x_1] \cdot \ldots \cdot \mathbb{P}_\theta[X_n = x_n] = h_\theta(x_1) \cdot \ldots \cdot h_\theta(x_n).$$

Die Likelihood-Funktion ist somit die Wahrscheinlichkeit, die gegebene Stichprobe $(x_1, \dots, x_n)$ zu beobachten, wobei diese Wahrscheinlichkeit als Funktion des Parameters $\theta$ aufgefasst wird.

Der absolut stetige Fall. Seien nun die Zufallsvariablen $X_i$ für alle Werte des Parameters $\theta$ absolut stetig. Wir bezeichnen die Dichte von $X_i$ mit $h_\theta$. In diesem Fall definieren wir die Likelihood-Funktion wie folgt:

$$L(\theta) = L(x_1, \dots, x_n; \theta) = h_\theta(x_1) \cdot \ldots \cdot h_\theta(x_n).$$

In beiden Fällen besteht die Idee der Maximum-Likelihood-Methode darin, einen Wert von $\theta$ zu finden, der die Likelihood-Funktion maximiert:

$$L(\theta) \to \max .$$

Der *Maximum-Likelihood-Schätzer* (oder der *ML-Schätzer*) ist definiert durch

$$\hat{\theta}_{ML} = \operatorname{argmax}_{\theta \in \Theta} L(\theta).$$

Es kann passieren, dass dieses Maximierungsproblem mehrere Lösungen hat. In diesem Fall muss man eine dieser Lösungen als Schätzer auswählen.

Beispiel 5.3.1. Maximum-Likelihood-Schätzer für den Parameter der Bernoulli-Verteilung $\text{Bern}(\theta)$.
Wir betrachten wieder eine unfaire Münze, wobei die mit $\theta$ bezeichnete Wahrscheinlichkeit von "Kopf" wiederum unbekannt sei. Nach $n = 100$ Würfen habe die Münze $s = 60$ Mal "Kopf" gezeigt. Wir werden nun $\theta$ mit der Maximum-Likelihood-Methode schätzen. Das kann man mit zwei verschiedenen Ansätzen machen, die aber (wie wir sehen werden) zum gleichen Ergebnis führen.

Erstes Modell. Das Ergebnis des Experiments, bei dem die Münze $n$ Mal geworfen wird, können wir in einer Stichprobe $(x_1, \dots, x_n) \in \{0, 1\}^n$ darstellen, wobei $x_i = 1$ ist, wenn die Münze bei Wurf $i$ "Kopf" gezeigt hat, und $x_i = 0$ ist, wenn die Münze bei Wurf $i$ "Zahl" gezeigt hat. Wir modellieren die Stichprobe $(x_1, \dots, x_n)$ als eine Realisierung von unabhängigen und identisch verteilten Zufallsvariablen $X_1, \dots, X_n$, die Bernoulli-verteilt sind mit Parameter $\theta$. Es handelt sich um diskrete Zufallsvariablen und die Zähldichte ist gegeben durch

$$h_\theta(x) = \mathbb{P}_\theta[X_i = x] = \begin{cases} \theta, & \text{falls } x = 1, \\ 1 - \theta, & \text{falls } x = 0, \\ 0, & \text{sonst.} \end{cases}$$

Somit gilt für die Likelihood-Funktion, dass:

$$L(x_1, \dots, x_n; \theta) = \mathbb{P}_\theta[X_1 = x_1, \dots, X_n = x_n] = h_\theta(x_1) \cdot \ldots \cdot h_\theta(x_n) = \theta^s (1 - \theta)^{n-s},$$

wobei $s = x_1 + \ldots + x_n = 60$ ist. Wir maximieren nun $L(\theta)$; siehe Abbildung 3.

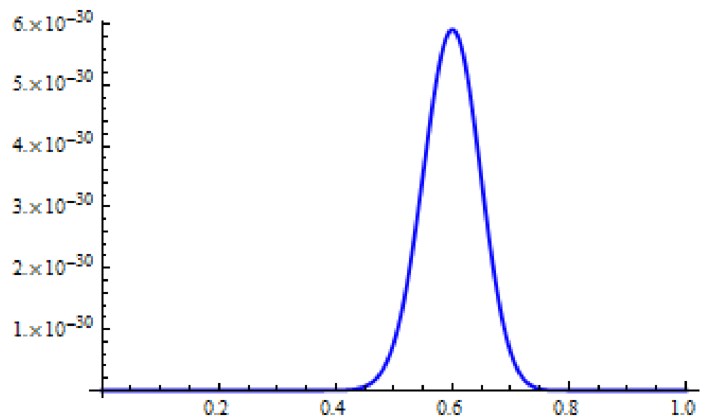


ABBILDUNG 3. Die Likelihood-Funktion $L(\theta) = \theta^{60}(1-\theta)^{40}$, $\theta \in [0,1]$, aus Beispiel 5.3.1, erstes Modell. Das Maximum wird an der Stelle $\theta = 0.6$ erreicht.

Wir benötigen eine Fallunterscheidung. FALL 1. Sei $s = 0$. Dann ist $L(\theta) = (1-\theta)^n$ und somit gilt argmax $L(\theta) = 0$.

FALL 2. Sei $s = n$. Dann ist $L(\theta) = \theta^n$ und somit gilt argmax $L(\theta) = 1$.

FALL 3. Sei nun $s \notin \{0, n\}$. Wir leiten die Likelihood-Funktion nach $\theta$ ab:

$$\frac{d}{d\theta}L(\theta) = s\theta^{s-1}(1-\theta)^{n-s} - (n-s)\theta^s(1-\theta)^{n-s-1} = \left(\frac{s}{\theta} - \frac{n-s}{1-\theta}\right)\theta^s(1-\theta)^{n-s}.$$

Die Ableitung ist gleich 0 an der Stelle $\theta = \frac{s}{n}$. (Das wprde für $s = 0$ und $s = n$ nicht stimmen). Außerdem ist $L$ nichtnegativ und es gilt

$$\lim_{\theta \downarrow 0} L(\theta) = \lim_{\theta \uparrow 1} L(\theta) = 0.$$

Daraus folgt, dass die Stelle $\theta = \frac{s}{n}$ das globale Maximum der Funktion $L(\theta)$ ist.

Die Ergebnisse der drei Fälle können wir nun wie folgt zusammenfassen: Der Maximum-Likelihood-Schätzer ist gegeben durch

$$\hat{\theta}_{ML} = \frac{s}{n} \quad \text{für } s = 0, 1, \ldots, n.$$

Somit ist in unserem Beispiel $\hat{\theta}_{ML} = \frac{60}{100} = 0.6$.

ZWEITES MODELL. In diesem Modell betrachten wir $s = 60$ als eine Realisierung einer binomialverteilten Zufallsvariable $S$ mit Parametern $n = 100$ (bekannt) und $\theta \in [0,1]$ (unbekannt). Somit ist die Likelihood-Funktion

$$L(s;\theta) = \mathbb{P}[S = s] = \binom{n}{s}\theta^s(1-\theta)^{n-s}.$$

Maximierung dieser Funktion führt genauso wie im ersten Modell zu dem Maximum-Likelihood-Schätzer

$$\hat{\theta}_{ML} = \frac{s}{n}.$$

Beispiel 5.3.2. Maximum-Likelihood-Schätzer für den Parameter der Poisson–Verteilung $\mathrm{Poi}(\theta)$.
Sei $(x_1, \ldots, x_n) \in \mathbb{N}_0^n$ eine Realisierung der unabhängigen und mit Parameter $\theta$ Poisson-verteilten Zufallsvariablen $X_1, \ldots, X_n$. Wir schätzen $\theta$ mit der Maximum-Likelihood-Methode.

Die Zähldichte der Poissonverteilung $\mathrm{Poi}(\theta)$ ist gegeben durch

$$h_\theta(x) = e^{-\theta} \frac{\theta^x}{x!}, \quad x = 0, 1, \ldots.$$

Dies führt zu folgender Likelihood-Funktion

$$L(x_1, \ldots, x_n; \theta) = e^{-\theta} \frac{\theta^{x_1}}{x_1!} \cdot \ldots \cdot e^{-\theta} \frac{\theta^{x_n}}{x_n!} = e^{-\theta n} \frac{\theta^{x_1 + \ldots + x_n}}{x_1! \cdot \ldots \cdot x_n!}.$$

An Stelle der Likelihood-Funktion ist es in diesem Falle einfacher, die sogenannte *log-Likelihood-Funktion* zu betrachten:

$$\log L(\theta) = -\theta n + (x_1 + \ldots + x_n) \log \theta + \log(x_1! \ldots x_n!).$$

Nun wollen wir einen Wert von $\theta$ finden, der diese Funktion maximiert. Für $x_1 = \ldots = x_n = 0$ ist dieser Wert offenbar $\theta = 0$. Seien nun nicht alle $x_i$ gleich 0. Die Ableitung von $\log L(\theta)$ ist gegeben durch

$$\frac{d}{d\theta} \log L(\theta) = -n + \frac{x_1 + \ldots + x_n}{\theta}.$$

Die Ableitung ist gleich 0 an der Stelle $\theta = \bar{x}_n$. (Das ist im Falle, wenn alle $x_i$ gleich 0 sind, falsch, denn dann wäre die Ableitung an der Stelle 0 gleich $-n$). Um zu sehen, dass $\theta = \bar{x}_n$ tatsächlich das globale Maximum der Funktion $\log L(\theta)$ ist, kann man wie folgt vorgehen. Es gilt offenbar $\frac{d}{d\theta} \log L(\theta) > 0$ für $0 \le \theta < \bar{x}_n$ und $\frac{d}{d\theta} \log L(\theta) < 0$ für $\theta > \bar{x}_n$. Somit ist die Funktion $\log L(\theta)$ strikt steigend auf $[0, \bar{x}_n)$ und strikt fallend auf $(\bar{x}_n, \infty)$. Die Stelle $\bar{x}_n$ ist also tatsächlich das globale Maximum. Der Maximum-Likelihood-Schätzer ist somit

$$\hat{\theta}_{ML} = \bar{x}_n = \frac{x_1 + \ldots + x_n}{n}.$$

Nun betrachten wir einige Beispiele zur Maximum-Likelihood-Methode im Falle der absolut stetigen Verteilungen.

Beispiel 5.3.3. Maximum-Likelihood-Schätzer für den Endpunkt der Gleichverteilung $\mathrm{U}[0, \theta]$.
Stellen wir uns vor, dass jemand in einem Intervall $[0, \theta]$ zufällig, gleichverteilt und unabhängig voneinander $n$ Punkte $x_1, \ldots, x_n$ ausgewählt und markiert hat. Uns werden nun die Positionen der $n$ Punkte gezeigt, nicht aber die Position des Endpunktes $\theta$; siehe Abbildung 4. Wir sollen $\theta$ anhand der Stichprobe $(x_1, \ldots, x_n)$ rekonstruieren.

Abbildung 4. Rote Kreise zeigen eine Stichprobe vom Umfang $n = 7$, die gleichverteilt auf einem Intervall $[0, \theta]$ ist. Die schwarzen Kreise zeigen die Endpunkte des Intervalls. Die Position des rechten Endpunktes soll anhand Stichprobe geschätzt werden.

Der Parameterraum ist hier $\Theta = \{\theta > 0\} = (0, \infty)$. Wir modellieren $(x_1, \ldots, x_n)$ als Realisierungen von unabhängigen und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$, die gleichverteilt auf einem Intervall $[0, \theta]$ sind. Die Zufallsvariablen $X_i$ sind somit absolut stetig und ihre Dichte ist gegeben durch

$$h_\theta(x) = \begin{cases} \frac{1}{\theta}, & \text{falls } x \in [0, \theta], \\ 0, & \text{falls } x \notin [0, \theta]. \end{cases}$$

Das führt zu folgender Likelihood-Funktion

$$L(x_1, \ldots, x_n; \theta) = h_\theta(x_1) \cdot \ldots \cdot h_\theta(x_n) = \frac{1}{\theta^n} \mathbb{1}_{x_1 \in [0,\theta]} \cdot \ldots \cdot \mathbb{1}_{x_n \in [0,\theta]} = \frac{1}{\theta^n} \mathbb{1}_{x_{(n)} \le \theta}.$$

Dabei ist $x_{(n)} = \max\{x_1, \ldots, x_n\}$ die maximale Beobachtung dieser Stichprobe. Der Graph der Likelihood-Funktion ist auf Abbildung 5 zu sehen.

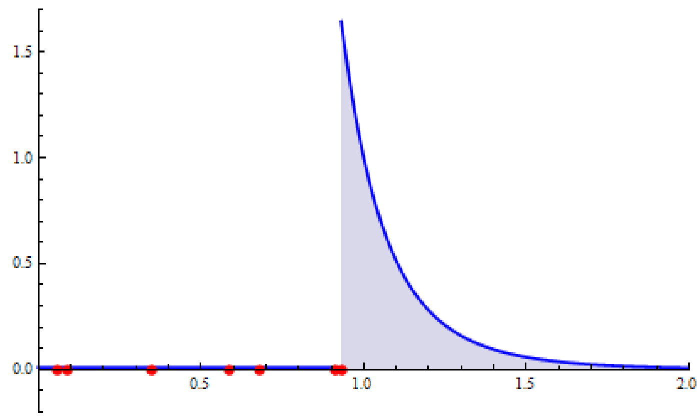


Abbildung 5. Maximum-Likelihood-Schätzung des Endpunktes der Gleichverteilung. Die roten Punkte zeigen die Stichprobe. Die blaue Kurve ist die Likelihood-Funktion $L(\theta)$.

Die Funktion $L(\theta)$ ist 0 solange $\theta < x_{(n)}$, und monoton fallend für $\theta > x_{(n)}$. Somit erhalten wir den Maximum-Likelihood-Schätzer

$$\hat{\theta}_{ML} = \operatorname*{argmax}_{\theta > 0} L(\theta) = x_{(n)}.$$

Der Maximum-Likelihood-Schätzer in diesem Beispiel ist also das Maximum der Stichprobe. Es sei bemerkt, dass dieser Schätzer den wahren Wert $\theta$ immer unterschätzt, denn die maximale Beobachtung $x_{(n)}$ ist immer kleiner als der wahre Wert des Parameters $\theta$.

Aufgabe 5.3.4. Bestimmen Sie den Momentenschätzer im obigen Beispiel und zeigen Sie, dass er nicht mit dem Maximum-Likelihood-Schätzer übereinstimmt.

Beispiel 5.3.5. Maximum-Likelihood-Schätzer für die Parameter der Normalverteilung $N(\mu, \sigma^2)$. Es sei $(x_1, \ldots, x_n)$ eine Realisierung von unabhängigen und mit Parametern $\mu, \sigma^2$ normalverteilten Zufallsvariablen $X_1, \ldots, X_n$. Wir schätzen $\mu$ und $\sigma^2$ mit der Maximum-Likelihood-Methode. Die Dichte von $X_i$ ist gegeben durch

$$h_{\mu,\sigma^2}(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right), \quad t \in \mathbb{R}.$$

Dies führt zu folgender Likelihood-Funktion:

$$L(\mu, \sigma^2) = L(x_1, \dots, x_n; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Die log-Likelihood-Funktion sieht folgendermaßen aus:

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2.$$

Wir bestimmen das Maximum dieser Funktion. Sei zunächst $\sigma^2$ fest. Wir betrachten die Funktion $\log L(\mu, \sigma^2)$ als Funktion von $\mu$ und bestimmen das Maximum dieser Funktion. Wir leiten nach $\mu$ ab:

$$\frac{\partial \log L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu).$$

Die Ableitung ist gleich 0 an der Stelle $\mu = \bar{x}_n$. Für $\mu < \bar{x}_n$ ist die Ableitung positiv (und somit die Funktion steigend), für $\mu > \bar{x}_n$ ist die Ableitung negativ (und somit die Funktion fallend). Also wird bei festem $\sigma^2$ an der Stelle $\mu = \bar{x}_n$ das globale Maximum erreicht. Nun machen wir auch $s := \sigma^2$ variabel. Wir betrachten die Funktion

$$\log L(\bar{x}_n, s) = -\frac{n}{2}\log(2\pi s) - \frac{1}{2s}\sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Falls alle $x_i$ gleich sind, wird das Maximum an der Stelle $s = 0$ erreicht. Es seien nun nicht alle $x_i$ gleich. Wir leiten nach $s$ ab:

$$\frac{\partial \log L(\bar{x}_n, s)}{\partial s} = -\frac{n}{2s} + \frac{1}{2s^2}\sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Die Ableitung ist gleich 0 an der Stelle

$$s = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{n-1}{n} s_n^2 =: \tilde{s}_n^2.$$

(Würden alle $x_i$ gleich sein, so würde das nicht stimmen, denn an der Stelle 0 existiert die Ableitung nicht). Für $s < \tilde{s}_n^2$ ist die Ableitung positiv (und die Funktion somit steigend), für $s > \tilde{s}_n^2$ ist die Ableitung negativ (und die Funktion somit fallend). Somit wird an der Stelle $s = \tilde{s}_n^2$ das globale Maximum der Funktion erreicht. Wir erhalten somit die folgenden Maximum-Likelihood-Schätzer:

$$\hat{\mu}_{\mathrm{ML}} = \bar{x}_n, \quad \hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Im nächsten Beispiel betrachten wir die sogenannte *Rückfangmethode* (Englisch: *capture-recapture method*) zur Bestimmung der Größe einer Population.

Beispiel 5.3.6. In einem Teich befinden sich $n$ Fische, wobei $n$ (die Populationsgröße) unbekannt sei. Um die Populationsgröße $n$ zu schätzen, kann man wie folgt vorgehen. Im ersten Schritt ("capture") werden aus dem Teich $n_1$ (eine bekannte Zahl) Fische gefangen und markiert. Danach werden die $n_1$ Fische wieder in den Teich zurückgeworfen. Im zweiten Schritt ("recapture") werden $k$ Fische ohne Zurücklegen gefangen. Unter diesen $k$ Fischen seien $k_1$ markiert und $k - k_1$ nicht markiert.

Anhand dieser Daten kann man $n$ wie folgt schätzen. Man setzt den Anteil der markierten Fische unter den gefangenen Fischen dem Anteil der markierten Fische unter allen Fischen gleich:

$$\frac{k_1}{k} = \frac{n_1}{n}.$$

Aus dieser Gleichung ergibt sich der folgende Schätzer für die Populationsgröße:

$$\hat{n} = \frac{n_1 k}{k_1}.$$

Nun werden wir die Maximum-Likelihood-Methode anwenden und schauen, ob sie den gleichen Schätzer liefert. Die Anzahl $k_1$ der markierten Fische unter den $k$ gefangenen Fischen betrachten wir als eine Realisierung der Zufallsvariable $X$ mit einer hypergeometrischen Verteilung. Die Likelihood-Funktion ist somit gegeben durch

$$L(k_1; n) = \mathbb{P}[X = k_1] = \frac{\binom{n_1}{k_1} \cdot \binom{n-n_1}{k-k_1}}{\binom{n}{k}}.$$

Die Frage ist nun, für welches $n$ diese Funktion maximal ist. Dabei darf $n$ nur Werte $\{0, 1, 2, \ldots\}$ annehmen. Um dies herauszufinden, betrachten wir die folgende Funktion:

$$R(n) = \frac{L(k_1; n)}{L(k_1; n-1)} = \frac{\binom{n_1}{k_1} \cdot \binom{n-n_1}{k-k_1} \cdot \binom{n-1}{k}}{\binom{n}{k} \cdot \binom{n_1}{k_1} \cdot \binom{n-1-n_1}{k-k_1}} = \frac{(n-k) \cdot (n-n_1)}{n \cdot (n - n_1 - k + k_1)}.$$

Eine elementare Rechnung zeigt:

(1) für $n < \hat{n}$ ist $R(n) < 1$;
(2) für $n > \hat{n}$ ist $R(n) > 1$;
(3) für $n = \hat{n}$ ist $R(n) = 1$.

Dabei benutzen wir die Notation $\hat{n} = \frac{n_1 k}{k_1}$. Daraus folgt, dass die Likelihood-Funktion $L(n)$ für $n < \hat{n}$ steigt und für $n > \hat{n}$ fällt. Ist nun $\hat{n}$ keine ganze Zahl, so wird das Maximum von $L(n)$ an der Stelle $n = [\hat{n}]$ erreicht. Ist aber $\hat{n}$ eine ganze Zahl, so gibt es zwei Maxima an den Stellen $n = \hat{n}$ und $n = \hat{n} - 1$. Dabei sind die Werte von $L(n)$ an diesen Stellen gleich, denn $R(\hat{n}) = 1$. Dies führt zum folgenden Maximum-Likelihood-Schätzer:

$$\hat{n}_{\mathrm{ML}} = \begin{cases} \left[\frac{n_1 k}{k_1}\right], & \text{falls } \frac{n_1 k}{k_1} \notin \mathbb{Z}, \\ \frac{n_1 k}{k_1} \text{ oder } \frac{n_1 k}{k_1} - 1, & \text{falls } \frac{n_1 k}{k_1} \in \mathbb{Z}. \end{cases}$$

Im zweiten Fall ist der Maximum-Likelihood-Schätzer nicht eindeutig definiert. Der Maximum-Likelihood-Schätzer $\hat{n}_{\mathrm{ML}}$ unterscheidet sich also nur unwesentlich vom Schätzer $\hat{n}$.

### 5.4. Bayes-Methode

Für die Einführung des Bayes-Schätzers muss das parametrische Modell etwas modifiziert werden. Um die Bayes-Methode anwenden zu können, werden wir zusätzlich annehmen, dass der Parameter $\theta$ selber eine Zufallsvariable mit einer gewissen (und bekannten) Verteilung ist. Wir betrachten zuerst ein Beispiel.

Beispiel 5.4.1. Eine Versicherung teile die bei ihr versicherten Autofahrer in zwei Kategorien: Typ 1 und Typ 2 (z.B. nach dem Typ des versicherten Fahrzeugs) ein. Die Wahrscheinlichkeit, dass ein Autofahrer vom Typ 1 (bzw. Typ 2) pro Jahr einen Schaden meldet, sei $\theta_1 = 0.4$ (bzw. $\theta_2 = 0.1$). Nun betrachten wir einen Autofahrer von einem unbekannten Typ, der in $n = 10$ Jahren $s = 2$ Schäden hatte. Können wir den Typ dieses Autofahrers raten (schätzen)?

Der Parameterraum ist in diesem Fall $\Theta = \{\theta_1, \theta_2\}$. Es sei $S$ die Zufallsvariable, die die Anzahl der Schäden modelliert, die ein Autofahrer in $n = 10$ Jahren meldet. Unter $\theta = \theta_1$ (also für Autofahrer vom Typ 1) gilt $S \sim \text{Bin}(n, \theta_1)$. Unter $\theta = \theta_2$ (also für Autofahrer vom Typ 2) ist $S \sim \text{Bin}(n, \theta_2)$. Dies führt zur folgenden Likelihood-Funktion:

$$L(s;\theta_1) = \mathbb{P}_{\theta_1}[S = s] = \binom{n}{s}\theta_1^s(1-\theta_1)^{n-s} = \binom{10}{2} \cdot 0.4^2 \cdot 0.6^8 = 0.1209,$$
$$L(s;\theta_2) = \mathbb{P}_{\theta_2}[S = s] = \binom{n}{s}\theta_2^s(1-\theta_2)^{n-s} = \binom{10}{2} \cdot 0.1^2 \cdot 0.9^8 = 0.1937.$$

Wir können nun die Maximum-Likelihood-Methode anwenden, indem wir $L(\theta_1)$ mit $L(\theta_2)$ vergleichen. Es gilt $L(\theta_2) > L(\theta_1)$ und somit handelt es sich vermutlich um einen Autofahrer vom Typ 2.

Sei nun zusätzlich bekannt, dass 90% aller Autofahrer vom Typ 1 und somit nur 10% vom Typ 2 seien. Mit dieser zusätzlichen Vorinformation ist es natürlich, den Parameter $\theta$ als eine Zufallsvariable zu modellieren. Die Zufallsvariable $\theta$ nimmt zwei Werte $\theta_1$ und $\theta_2$ an und die Wahrscheinlichkeiten dieser Werte sind

$$q(\theta_1) := \mathbb{P}[\theta = \theta_1] = 0.9 \text{ und } q(\theta_2) := \mathbb{P}[\theta = \theta_2] = 0.1.$$

Die Verteilung von $\theta$ nennt man auch die *a-priori-Verteilung*. Wie ist nun die Anzahl der Schäden $S$ verteilt, die ein Autofahrer von einem unbekannten Typ in $n$ Jahren meldet? Die Antwort erhält man mit der Formel der totalen Wahrscheinlichkeit:

$$\begin{aligned}\mathbb{P}[S = s] &= \mathbb{P}[\theta = \theta_1] \cdot \mathbb{P}[S = s|\theta = \theta_1] + \mathbb{P}[\theta = \theta_2] \cdot \mathbb{P}[S = s|\theta = \theta_2] \\ &= q(\theta_1)\binom{n}{s}\theta_1^s(1-\theta_1)^{n-s} + q(\theta_2)\binom{n}{s}\theta_2^s(1-\theta_2)^{n-s}.\end{aligned}$$

Es sei bemerkt, dass die Zufallsvariable $S$ *nicht* binomialverteilt ist. Vielmehr ist die Verteilung von $S$ eine Mischung aus zwei verschiedenen Binomialverteilungen. Man sagt auch das $S$ *bedingt* binomialverteilt ist:

$$S|\{\theta = \theta_1\} \sim \text{Bin}(n, \theta_1) \text{ und } S|\{\theta = \theta_2\} \sim \text{Bin}(n, \theta_2).$$

Nun betrachten wir einen Autofahrer von einem unbekannten Typ, der $s = 2$ Schäden gemeldet hat. Die Wahrscheinlichkeit, dass 2 Schäden gemeldet werden, können wir mit der obigen Formel bestimmen:

$$\mathbb{P}[S = 2] = 0.9 \cdot 0.1209 + 0.1 \cdot 0.1937 = 0.1282.$$

Die *a-posteriori-Verteilung* von $\theta$ ist die Verteilung von $\theta$ gegeben die Information, dass $S = 2$. Zum Beispiel ist die a-posteriori-Wahrscheinlichkeit von $\theta = \theta_1$ definiert als die bedingte

Wahrscheinlichkeit, dass $\theta = \theta_1$, gegeben, dass $S = 2$. Um die a-posteriori-Verteilung zu berechnen, benutzen wir die Bayes-Formel:

$$q(\theta_1|s) := \mathbb{P}[\theta = \theta_1|S = s] = \frac{\mathbb{P}[\theta = \theta_1 \cap S = s]}{\mathbb{P}[S = s]} = \frac{\mathbb{P}[\theta = \theta_1] \cdot \mathbb{P}[S = s|\theta = \theta_1]}{\mathbb{P}[S = s]}.$$

Mit den oben berechneten Werten erhalten wir, dass

$$q(\theta_1|2) = \frac{0.9 \cdot 0.1209}{0.1282} = 0.8486.$$

Die a-posteriori-Wahrscheinlichkeit von $\theta = \theta_2$ kann analog berechnet werden. Es geht aber auch einfacher:

$$q(\theta_2|2) = 1 - q(\theta_1|s) = 0.1513.$$

Nun können wir die a-posteriori Wahrscheinlichkeiten vergleichen. Da $q(\theta_1|2) > q(\theta_2|2)$, handelt es sich vermutlich um einen Autofahrer vom Typ 1.

Bemerkung 5.4.2. Das Wort “a priori” steht für “vor dem Experiment”, das Wort “a posteriori” steht für “nach dem Experiment”.

Nun beschreiben wir die allgemeine Form der Bayes–Methode.

Bayes-Methode im diskreten Fall. Zuerst betrachten wir den Fall, dass $\theta$ eine diskrete Zufallsvariable ist. Die möglichen Werte für $\theta$ seien $\theta_1, \theta_2, \ldots$. Die Verteilung von $\theta$ (die auch die *a-priori-Verteilung* genannt wird) sei bekannt:

$$q(\theta_i) := \mathbb{P}[\theta = \theta_i], \quad i = 1, 2, \ldots.$$

Seien $(X_1, \ldots, X_n)$ Zufallsvariablen mit der folgenden Eigenschaft: Gegeben, dass $\theta = \theta_i$, sind die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit Zähldichte/Dichte $h_{\theta_i}(x)$. Es sei bemerkt, dass die Zufallsvariablen $X_1, \ldots, X_n$ nicht unabhängig, sondern lediglich bedingt unabhängig sind. Es werde nun eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ beobachtet.
Die *a-posteriori-Verteilung* von $\theta$ ist die bedingte Verteilung von $\theta$ gegeben die Information, dass $X_1 = x_1, \ldots, X_n = x_n$, d.h.

$$q(\theta_i|x_1, \ldots, x_n) := \mathbb{P}[\theta = \theta_i|X_1 = x_1, \ldots, X_n = x_n], \quad i = 1, 2, \ldots.$$

Hier nehmen wir der Einfachheit halber, dass die Zufallsvariablen $X_1, \ldots, X_n$ diskret sind. Diese Wahrscheinlichkeit berechnet man mit der Bayes–Formel:

$$\begin{aligned} q(\theta_i|x_1, \ldots, x_n) &= \mathbb{P}[\theta = \theta_i|X_1 = x_1, \ldots, X_n = x_n] \\ &= \frac{\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n, \theta = \theta_i]}{\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n]} \\ &= \frac{\mathbb{P}[\theta = \theta_i] \cdot \mathbb{P}[X_1 = x_1, \ldots, X_n = x_n|\theta = \theta_i]}{\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n]} \\ &= \frac{q(\theta_i)h_{\theta_i}(x_1) \ldots h_{\theta_i}(x_n)}{\sum_j q(\theta_j)h_{\theta_j}(x_1) \ldots h_{\theta_j}(x_n)}. \end{aligned}$$

Wir haben dabei angenommen, dass $X_i$ diskret sind, die Endformel macht aber auch für absolut stetige Variablen $X_i$ Sinn.

In der Bayes–Statistik schreibt man oft $A(t) \propto B(t)$, wenn es eine Konstante $C$ (die von $t$ nicht abhängt) mit $A(t) \propto C \cdot B(t)$ gibt. Das Zeichen $\propto$ steht also für die Proportionalität von Funktionen. Die Formel für die a-posteriori-Zähldichte von $\theta$ kann man dann auch wie folgt schreiben:

$$q(\theta_i|x_1,\ldots,x_n) \propto q(\theta_i)h_{\theta_i}(x_1)\ldots h_{\theta_i}(x_n).$$

Die a-posteriori-Zähldichte $q(\theta_i|x_1,\ldots,x_n)$ ist somit proportional zur a-priori-Zähldichte $q(\theta_i)$ und zur Likelihood-Funktion $L(x_1,\ldots,x_n;\theta_i) = h_{\theta_i}(x_1)\ldots h_{\theta_i}(x_n)$.

Nach der Anwendung der Bayes-Methode erhalten wir als Endergebnis die a-posteriori-Verteilung des Parameters $\theta$. Oft möchte man allerdings das Endergebnis in Form einer Zahl haben. In diesem Fall kann man z. B. folgendermaßen vorgehen: Der *Bayes–Schätzer* wird definiert als der Erwartungswert der a-posteriori-Verteilung:

$$\hat{\theta}_{\text{Bayes}} = \sum_i \theta_i q(\theta_i|x_1,\ldots,x_n).$$

Alternativ kann man den Bayes-Schätzer auch als den Median der a-posteriori-Verteilung definieren.

Bayes-Methode im absolut stetigen Fall. Sei nun $\theta$ eine absolut stetige Zufallsvariable (bzw. Zufallsvektor) mit Werten in $\mathbb{R}^p$ und einer Dichte $q(\tau)$. Dabei bezeichnen wir mit $\tau \in \mathbb{R}^p$ mögliche Werte von $\theta$. Die Dichte $q(\tau)$ wird auch die a-priori-Dichte genannt. Seien $(X_1,\ldots,X_n)$ Zufallsvariablen mit der folgenden Eigenschaft: Gegeben, dass $\theta = \tau$ sind die Zufallsvariablen $X_1,\ldots,X_n$ unabhängig und identisch verteilt mit Zähldichte/Dichte $h_\tau(x)$. Sei $(x_1,\ldots,x_n)$ eine Realisierung von $(X_1,\ldots,X_n)$. Die a-posteriori-Verteilung von $\theta$ ist die bedingte Verteilung von $\theta$ gegeben die Information, dass $X_1 = x_1,\ldots,X_n = x_n$. Indem wir in der Formel aus dem diskreten Fall die Zähldichte von $\theta$ durch die Dichte von $\theta$ ersetzen, erhalten wir die folgende Formel für die a-posteriori-Dichte von $\theta$:

$$q(\tau|x_1,\ldots,x_n) = \frac{q(\tau)h_\tau(x_1)\ldots h_\tau(x_n)}{\int_{\mathbb{R}^p} q(t)h_t(x_1)\ldots h_t(x_n)dt}.$$

Das können wir auch wie folgt schreiben:

$$q(\tau|x_1,\ldots,x_n) \propto q(\tau)h_\tau(x_1)\ldots h_\tau(x_n).$$

Die a-posteriori-Dichte $q(\tau|x_1,\ldots,x_n)$ ist somit proportional zur a-priori-Dichte $q(\tau)$ und zur Likelihood-Funktion $L(x_1,\ldots,x_n;\tau) = h_\tau(x_1)\ldots h_\tau(x_n)$.
Genauso wie im diskreten Fall ist der Bayes-Schätzer definiert als der Erwartungswert der a-posteriori-Verteilung, also

$$\hat{\theta}_{\text{Bayes}} = \int_{\mathbb{R}^p} \tau q(\tau|x_1,\ldots,x_n)d\tau.$$

Aufgabe 5.4.3. Zeigen Sie, dass im diskreten Fall (bzw. im stetigen Fall) $q(\tau|x_1,\ldots,x_n)$ als Funktion von $\tau$ tatsächlich eine Zähldichte (bzw. eine Dichte) ist.

Beispiel 5.4.4. Ein Unternehmen möchte ein neues Produkt auf den Markt bringen. Die a-priori Information sei, dass der Marktanteil $\theta$ bei ähnlichen Produkten in der Vergangenheit immer zwischen 0.1 und 0.3 lag. Da keine weiteren Informationen über die Verteilung von $\theta$

vorliegen, kann man z.B. die Gleichverteilung auf $[0.1, 0.3]$ als die a-priori-Verteilung von $\theta$ ansetzen. Die a-priori-Dichte für den Marktanteil $\theta$ ist somit

$$q(\tau) = \begin{cases} 5, & \text{falls } \tau \in [0.1, 0.3], \\ 0, & \text{sonst.} \end{cases}$$

Man kann nun den a-priori-Schätzer für den Marktanteil z.B. als den Erwartungswert dieser Verteilung berechnen:

$$\hat{\theta}\text{apr} = \mathbb{E}\theta = \int_{\mathbb{R}} \tau q(\tau) d\tau = 0.2.$$

Außerdem seien $n$ Kunden befragt worden, ob sie das neue Produkt kaufen würden. Sei $x_i = 1$, falls der $i$-te Kunde die Frage bejaht und 0, sonst. Es sei $s = x_1 + \ldots + x_n$ die Anzahl der Kunden in dieser Umfrage, die das neue Produkt kaufen würden. Wir könnten nun den Marktanteil des neuen Produkts z.B. mit der Momentenmethode (Beispiel 5.2.3) oder mit der Maximum-Likelihood-Methode (Beispiel 5.3.1) schätzen:

$$\hat{\theta}_{\text{ME}} = \hat{\theta}_{\text{ML}} = \frac{s}{n}.$$

Dieser Schätzer ignoriert allerdings die a-priori Information. Mit der Bayes-Methode können wir einen Schätzer konstruieren, der sowohl die a-priori Information, als auch die Befragung berücksichtigt. Wir betrachten $(x_1, \ldots, x_n)$ als eine Realisierung der Zufallsvariablen $(X_1, \ldots, X_n)$. Wir nehmen an, dass bei einem gegebenen $\theta$ die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig und mit Parameter $\theta$ Bernoulli–verteilt sind:

$$q_\theta(0) := \mathbb{P}_\theta[X_i = 0] = 1 - \theta, \quad q_\theta(1) := \mathbb{P}_\theta[X_i = 1] = \theta.$$

Die Likelihood-Funktion ist

$$L(x_1, \ldots, x_n; \tau) = h_\tau(x_1) \ldots h_\tau(x_n) = \tau^s (1-\tau)^{n-s},$$

wobei $s = x_1 + \ldots + x_n$. Die a-posteriori-Dichte von $\theta$ ist proportional zu $q(\tau)$ und $L(x_1, \ldots, x_n; \tau)$ und ist somit gegeben durch

$$q(\tau | x_1, \ldots, x_n) = \begin{cases} \frac{5\tau^s(1-\tau)^{n-s}}{\int_{0.1}^{0.3} 5t^s(1-t)^{n-s} dt}, & \text{für } \tau \in [0.1, 0.3], \\ 0, & \text{sonst.} \end{cases}$$

Es sei bemerkt, dass die a-posteriori-Dichte (genauso wie die a-priori-Dichte) außerhalb des Intervalls $[0.1, 0.3]$ verschwindet. Wir können nun den Bayes–Schätzer für den Marktanteil $\theta$ bestimmen:

$$\hat{\theta}_{\text{Bayes}} = \int_{0.1}^{0.3} \tau q(\tau | x_1, \ldots, x_n) d\tau = \frac{\int_{0.1}^{0.3} \tau^{s+1} (1-\tau)^{n-s} d\tau}{\int_{0.1}^{0.3} t^s (1-t)^{n-s} dt}.$$

Der Bayes-Schätzer liegt im Intervall $[0.1, 0.3]$ (denn außerhalb dieses Intervalls verschwindet die a-posteriori-Dichte) und widerspricht somit der a-priori Information nicht.

Nehmen wir nun an, wir möchten ein Bayes-Modell konstruieren, in dem wir z.B. Bernoulli-verteilte Zufallsvariablen mit einem Parameter $\theta$ betrachten, der selber eine Zufallsvariable ist. Wie sollen wir die a-priori-Verteilung von $\theta$ wählen? Es wäre schön, wenn die a-posteriori Verteilung eine ähnliche Form haben würde, wie die a-priori-Verteilung. Wie man das erreicht, sehen wir im nächsten Beispiel.

Beispiel 5.4.5. (Bernoulli-Beta-Modell.)
Bei einem gegebenen $\theta \in [0,1]$ seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen, die Bernoulli–verteilt mit Parameter $\theta$ sind. Somit gilt

$$h_\theta(0) = 1 - \theta, \quad h_\theta(1) = \theta.$$

Die a-priori-Verteilung von $\theta$ sei die Betaverteilung $\text{Beta}(\alpha, \beta)$. Somit ist die a-priori-Dichte von $\theta$ gegeben durch

$$q(\tau) = \frac{1}{B(\alpha,\beta)} \tau^{\alpha-1}(1-\tau)^{\beta-1} \propto \tau^{\alpha-1}(1-\tau)^{\beta-1}, \quad \tau \in [0,1].$$

Es werde nun eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ beobachtet. Die Likelihood-Funktion ist

$$L(x_1, \ldots, x_n; \tau) = h_\tau(x_1) \ldots h_\tau(x_n) = \tau^s (1-\tau)^{n-s}, \quad \tau \in [0,1],$$

wobei $s = x_1 + \ldots + x_n$. Für die a-posteriori-Dichte von $\theta$ gilt somit

$$q(\tau | x_1, \ldots, x_n) \propto q(\tau) L(x_1, \ldots, x_n; \tau) \propto \tau^{\alpha+s-1}(1-\tau)^{\beta+n-s-1}, \quad \tau \in [0,1].$$

In dieser Formel haben wir die multiplikative Konstante nicht berechnet. Diese muss aber so sein, dass die a-posteriori-Dichte tatsächlich eine Dichte ist, also

$$q(\tau | x_1, \ldots, x_n) = \frac{1}{\mathrm{B}(\alpha+s, \beta+n-s)} \tau^{\alpha+s-1}(1-\tau)^{\beta+n-s-1}, \quad \tau \in [0,1].$$

Somit ist die a-posteriori-Verteilung von $\theta$ eine Betaverteilung:

$$\text{Beta}(\alpha+s, \beta+n-s).$$

Die a-posteriori-Verteilung stammt also aus derselben Betafamilie, wie die a-priori-Verteilung, bloß die Parameter sind anders. Der Bayes-Schätzer für $\theta$ ist der Erwartungswert der a-posteriori-Betaverteilung:

$$\hat{\theta}_{\text{Bayes}} = \frac{\alpha+s}{\alpha+\beta+n}.$$

Weitere Beispiele von Bayes–Modellen, in denen die a-posteriori-Verteilung zur selben Verteilungsfamilie gehört, wie die a-priori-Verteilung, finden sich in folgenden Aufgaben.

Aufgabe 5.4.6 (Poisson-Gamma-Modell). Bei einem gegebenen Wert des Parameters $\lambda > 0$ seien die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig und Poisson-verteilt mit Parameter $\lambda$. Dabei wird für $\lambda$ eine a-priori-Gammaverteilung mit (deterministischen und bekannten) Parametern $b > 0$, $\alpha > 0$ angenommen, d.h.

$$q(\lambda) = \frac{b^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda} \text{ für } \lambda > 0.$$

Man beobachtet nun eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$. Bestimmen Sie die a-posteriori-Verteilung von $\lambda$ und den Bayes-Schätzer $\hat{\lambda}_{\text{Bayes}}$.

Aufgabe 5.4.7 (Geo-Beta-Modell). Bei einem gegebenen Wert des Parameters $p \in (0,1)$ seien die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig und geometrisch verteilt mit Parameter $p$. Dabei wird für $p$ eine a-priori Betaverteilung mit (deterministischen und bekannten) Parametern $\alpha > 0$, $\beta > 0$ angenommen. Man beobachtet eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$. Bestimmen Sie die a-posteriori-Verteilung von $p$ und den Bayes-Schätzer $\hat{p}_{\text{Bayes}}$.

Aufgabe 5.4.8 (A-priori-Verteilung für den Erwartungswert einer Normalverteilung bei bekannter Varianz). Bei einem gegebenen Wert des Parameters $\mu \in \mathbb{R}$ seien die Zufallsvariablen $X_1, \dots, X_n$ unabhängig und normalverteilt mit Parametern $(\mu, \sigma^2)$, wobei $\sigma^2$ bekannt sei. Dabei wird für $\mu$ eine a-priori Normalverteilung mit (deterministischen und bekannten) Parametern $\mu_0 \in \mathbb{R}$, $\sigma_0^2 > 0$ angenommen. Man beobachtet eine Realisierung $(x_1, \dots, x_n)$ von $(X_1, \dots, X_n)$. Bestimmen Sie die a-posteriori-Verteilung von $\mu$ und den Bayes-Schätzer $\hat{\mu}_{\text{Bayes}}$.

Aufgabe 5.4.9 (A-priori-Verteilung für die Varianz einer Normalverteilung bei bekanntem Erwartungswert). Bei einem gegebenen Wert des Parameters $\tau \in \mathbb{R}$ seien die Zufallsvariablen $X_1, \dots, X_n$ unabhängig und normalverteilt mit Parametern $(\mu, \sigma^2)$, wobei $\mu$ bekannt sei. Dabei wird für $\sigma^2$ eine a-priori inverse Gammaverteilung mit (deterministischen und bekannten) Parametern $b > 0$, $\alpha > 0$ angenommen. Das heißt, es wird angenommen, dass $\tau := 1/\sigma^2$ Gammaverteilt mit Parametern $b$ und $\alpha$ ist. Man beobachtet eine Realisierung $(x_1, \dots, x_n)$ von $(X_1, \dots, X_n)$. Bestimmen Sie die a-posteriori-Verteilung von $\tau = 1/\sigma^2$.