6. Schätzprobleme

6.1. Beispiele

a) In einem Teich befindet sich eine unbekannte Anzahl n von Fischen. Man schätze n (z. B. durch Angeln, markieren, wieder aussetzen und nochmal angeln; vgl. Übung)

b) Weiteres Beispiel: Wie groß ist die Zahl n der Taxis in einer Stadt?
Annahme: die Taxis sind von 1, ..., n durchnummeriert. Man beobachtet m Nummern, z. B. 24, 71, 5, 16, 13 ($m = 5$). Was ist eine vernünftige Schätzung für n?

c) (Medizinische) Versuche: Wie groß ist die (unbekannte) Wahrscheinlichkeit p, dass ein bestimmtes Medikament wirkt? Man testet es an n Personen und beobachtet, dass es bei m wirkt. Was ist eine gute Schätzung für p? (Analog: p = unbekannter Anteil der CDU-Wähler in der Bevölkerung).

6.2 Definition:
Ein *statistisches Modell* ist ein Tripel $(\mathcal{X}, \mathcal{A}, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$, wobei $(\mathcal{X}, \mathcal{A})$ ein Messraum ist, Θ eine (nichtleere) Indexmenge und $\mathbf{P}_\vartheta, \vartheta \in \Theta$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$ sind. $\mathcal{X}$ heißt dabei *Beobachtungsraum* oder *Stichprobenraum*.

Interpretation: Der Beobachtungsraum $\mathcal{X}$ beschreibt die Menge der möglichen Beobachtungen, $\mathbf{P}_\vartheta, \vartheta \in \Theta$ sind die im Modell als möglich angenommenen Verteilungen. Ziel ist es, den wahren Parameter $\vartheta \in \Theta$ oder eine Funktion $\tau(\vartheta)$ aufgrund der Beobachtung $x \in \mathcal{X}$ zu schätzen. Oft ist $\mathcal{X} = \mathbf{R}^n$ oder jedenfalls $\mathcal{X} \subseteq \mathbf{R}^n$.

6.3 Definition:
Sei $(\mathcal{X}, \mathcal{A}, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und $(\Sigma, \mathcal{S})$ ein Messraum.

a) Eine messbare Abbildung von $(\mathcal{X}, \mathcal{A})$ nach $(\Sigma, \mathcal{S})$ heißt *Statistik*.

b) Ist $\tau : \Theta \to \Sigma$ eine Abbildung, so heißt die Statistik $T : \mathcal{X} \to \Sigma$ ein *Schätzer* für τ.

Im Beispiel 6.1.c) ist es naheliegend, $\mathcal{X} = \{0, 1\}^n$ zu wählen,

$$\Theta = [0, 1] \text{ und } \mathbf{P}_\vartheta(\{\underbrace{0100101 \cdots 11011}_{n \text{ Ziffern } \in\{0,1\}}\}) = \vartheta^{\sharp\text{Einsen}}(1 - \vartheta)^{\sharp\text{Nullen}},$$

wobei $x = 0100101 \cdots 111011$ bedeutet, dass das Medikament bei der ersten, dritten, vierten, sechsten $\cdots$ Person nicht wirkte, bei den anderen dagegen schon. Also ist $\mathbf{P}_\vartheta = \otimes_{i=1}^n Q_\vartheta$, wobei $Q_\vartheta(\{1\}) = \vartheta, Q_\vartheta(\{0\}) = 1 - \vartheta$ ist. Bei diesem statistischen Modell wird also von vornherein postuliert, dass alle Versuche unabhängig sind. Wir wählen $(\Sigma, \mathcal{S}) = (\Theta, \mathcal{B}_{[0,1]})$ und $\tau : \Theta \to \Sigma$ als die identische Abbildung.

Ein vernünftiger Schätzer $T : \mathcal{X} \to \Theta$ für das unbekannte $\vartheta = \tau(\vartheta)$ ist in diesem Fall

$$T(x_1, \cdots, x_n) = \frac{\sharp \text{ Einsen unter den } x_1, \cdots x_n}{n}.$$

Wenn man an der Unabhängigkeit der Versuchsergebnisse zweifelt, könnte man auch die Menge *aller* Wahrscheinlichkeitsmaße $(\mathbf{P}_\vartheta)_{\vartheta \in \Theta}$ auf $\mathcal{X} = \{0,1\}^n$ ansetzen. Aber: wie soll man in diesem Fall das wahre ϑ (oder $\mathbf{P}_\vartheta$) aus den Beobachtungen schätzen?

Generelles Problem:

Θ zu groß:	eine vernünftige Schätzung von ϑ ist wg. in Relation zur Größe von Θ zu geringer Datenmenge problematisch.
Θ zu klein:	Gefahr, dass das "wahre" ϑ nicht in Θ liegt.

Bemerkung:
In 6.3. wird nichts über die Güte eines Schätzers ausgesagt. Er kann auch völlig unsinnig sein. Unser Ziel ist es, Gütekriterien für Schätzer zu finden und gute Schätzer zu konstruieren. Grob gesprochen wird man einen Schätzer T für τ dann als "gut" bezeichnen, wenn für alle $\vartheta \in \Theta$ $T(x)$ mit großer Wahrscheinlichkeit (bzgl. $\mathbf{P}_\vartheta$) einen Wert nahe bei $\tau(\vartheta)$ annimmt. Ein Gütekriterium ist:

6.4. Definition:
Sei $\Sigma \subset \mathbf{R}^k$. Ein Schätzer T heißt *erwartungstreu* oder *unverzerrt* (englisch *unbiased*) für $\tau(\vartheta)$, wenn

$$\mathbf{E}_\vartheta(T(X)) = \tau(\vartheta) \text{ für alle } \vartheta \in \Theta.$$

Dabei sei $\mathbf{E}_\vartheta$ der Erwartungswert bezüglich des Wahrscheinlichkeitsmaßes $\mathbf{P}_\vartheta$ (im Fall $k > 1$ ist die Gleichheit komponentenweise zu verstehen). $X : \mathcal{X} \to \mathcal{X}$ steht dabei für die identische Abbildung (= Beobachtung).

Allgemein heißt

$$b(\vartheta, T) := \mathbf{E}_\vartheta(T(X)) - \tau(\vartheta)$$

Verzerrung (engl.: *bias*) des Schätzers T für τ.

Beispiel 6.1.c), Fortsetzung:
Wir zeigen, dass der betrachtete Schätzer $T(x_1, \cdots, x_n) = \frac{\sharp\text{Einsen unter den } x_1, \cdots, x_n}{n}$ erwartungstreu für $\tau(\vartheta) := \vartheta$ ist:

$$\mathbf{E}_\vartheta T(X_1, \cdots, X_n) = \mathbf{E}_\vartheta \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}_\vartheta X_i = \frac{1}{n} \sum_{i=1}^n \vartheta = \vartheta.$$

6.5. Beispiel:
Bei einer Warenlieferung von 1000 Einheiten seien M defekt (M unbekannt und zu schätzen).
Es werden 15 Einheiten getestet. Sinnvolles statistisches Modell:
$\mathcal{X} = \{0, \cdots, 15\}$, $x = \sharp$defekte Teile in Stichprobe
$\Theta = \{0, \cdots, 1000\}$. $\tau(\vartheta) = \vartheta(= M)$. Sinnvoll erscheint

$$T(x) = \frac{x}{15} \cdot 1000.$$

Ist T erwartungstreu?

Offenbar ist im Fall von $\vartheta = M$ defekten Teilen die Anzahl der defekten Teile in der Stichprobe hypergeometrisch mit Parameter (15, 100, M)–verteilt, also

$$\mathbf{E}_\vartheta T(X) = \mathbf{E}_\vartheta \left(\frac{X}{15} \cdot 1000 \right) = \frac{1000}{15} \mathbf{E}_\vartheta(X) = \frac{1000}{15} \frac{15 \cdot \vartheta}{1000} = \vartheta.$$

Es gibt Fälle, bei denen zwar erwartungstreue Schätzer existieren, diese aber unsinnig sind (obwohl es vernünftige nicht–erwartungstreue Schätzer gibt).

6.6. Beispiel:
Sei $\mathcal{X} = \mathbf{N}_0$, $\Theta = (0, 1)$, $\mathbf{P}_\vartheta =$ Poissonverteilung mit
Parameter $\lambda = -\frac{1}{2} \ln \vartheta$.
Behauptung: $T(x) = (-1)^x$ ist der einzige erwartungstreue Schätzer für ϑ, aber unsinnig.
Beweis:

i) T *ist* erwartungstreu:
$\mathbf{E}_\vartheta \left((-1)^X \right) = \sum_{k=0}^{\infty} (-1)^k \frac{\left(-\frac{1}{2} \ln \vartheta\right)^k}{k!} \underbrace{e^{\frac{1}{2} \ln \vartheta}}_{\sqrt{\vartheta}} = \sqrt{\vartheta} \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2} \ln \vartheta\right)^k}{k!} = \sqrt{\vartheta}\sqrt{\vartheta} = \vartheta.$

ii) T ist der *einzige* erwartungstreue Schätzer:
Sei S auch erwartungstreu, d. h. $\forall \vartheta \in (0, 1)$ gilt
$e^{-2\lambda} = \vartheta = \mathbf{E}_\vartheta S(X) = \sum_{k=0}^{\infty} S(k) \frac{\lambda^k}{k!} e^{-\lambda}$ d.h. $\forall \lambda \in (0, \infty)$ gilt $e^{-\lambda} = \sum_{k=0}^{\infty} S(k) \frac{\lambda^k}{k!}$
Nach dem Identitätssatz für Potenzreihen folgt $S(k) = (-1)^k \; \forall k \in \mathbf{N}_0$.

iii) T ist *unsinnig*:
T nimmt nur die Werte 1 und -1 an, die gar nicht in Θ liegen!

Bemerkung:
Es gibt auch Fälle, in denen überhaupt kein erwartungstreuer Schätzer existiert.

Wenn man zwei erwartungstreue Schätzer T_1 und T_2 für dasselbe τ hat, dann ist es naheliegend, T_1 dem Schätzer T_2 dann vorzuziehen, wenn er für *alle* ϑ eine kleinere Varianz als T_2 aufweist (dann liegt nämlich T_1 im Mittel näher bei $\tau(\vartheta)$ als T_2 und zwar *unabhängig* von ϑ).

6.7. Definition:
Im statistischen Modell $(\mathcal{X}, \mathcal{A}, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ seien T_1 und T_2 erwartungstreue Schätzer für $\tau : \Theta \to \mathbf{R}$ $(k = 1!)$.
Gilt

$$V_\vartheta(T_1) \leq V_\vartheta(T_2) \;\forall \vartheta \in \Theta,$$

dann heißt T_1 *gleichmäßig besser* als T_2. Gilt

$$V_\vartheta(T_1) \leq V_\vartheta(T) \;\forall \vartheta \in \Theta$$

für *alle* erwartungstreuen Schätzer T für τ, so heißt T_1 *gleichmäßig bester erwartungstreuer Schätzer* (oder "UMVU" für *uniformly minimum variance unbiased*) für τ. (V_ϑ: Varianz bzgl. $\mathbf{P}_\vartheta$).

Bemerkung:
In vielen Situationen gibt es gar keinen UMVU-Schätzer (auch in vielen solchen, in denen erwartungstreue Schätzer existieren). Der Schätzer T in Beispiel 6.6. ist natürlich — als einziger erwartungstreuer Schätzer für θ — UMVU.

6.8. Beispiel:
Sei $\mathcal{X} = \mathbf{R}^n$ mit (mit Borel–σ–Algebra), Θ = Menge aller Wahrscheinlichkeitsmaße auf $\mathbf{R}$ mit existierendem 2. Moment. Sei $\tau(\vartheta) := \mu_\vartheta$ der Erwartungswert und $\sigma^2(\vartheta)$ die Varianz von
ϑ (also $\mu_\vartheta := \int\limits_{-\infty}^{\infty} x d\vartheta(x), \cdots$). Weiter sei $\mathbf{P}_\vartheta = \vartheta^{\otimes n}$ d. h. die Verteilung von n unabhängigen gemäß ϑ verteilten (reellwertigen) Zufallsgrößen.

Behauptung: $T_1(x_1, \cdots, x_n) := \frac{1}{n} \sum\limits_{i=1}^{n} x_i$ ist UMVU für $\tau(\vartheta) = \mu_\vartheta$.

i) T_1 ist erwartungstreu:
$\mathbf{E}_\vartheta T_1(X_1, \cdots, X_n) = \frac{1}{n} \sum\limits_{i=1}^{n} \mathbf{E}_\vartheta(X_i) = \frac{1}{n} \sum\limits_{i=1}^{n} \mu_\vartheta = \mu_\vartheta.$

ii) $V_\vartheta(T_1(X_1, \cdots, X_n)) = V_\vartheta(\frac{1}{n} \sum\limits_{i=1}^{n} X_i) = \frac{1}{n}\sigma^2(\vartheta).$

iii) Zu zeigen: Für jede Funktion $T : \mathbf{R}^n \to \mathbf{R}$ mit $\mathbf{E}_\vartheta(T(X_1, \cdots X_n)) = \mu_\vartheta$ gilt
$V_\vartheta(T(X_1, \cdots, X_n)) \geq \frac{1}{n}\sigma^2(\vartheta).$

Wir zeigen dies der Einfachheit halber nur für $n = 1$ und $n = 2$.

(α) $n = 1$:
Sei $y \in \mathbf{R}$ und δ_y das Punktmaß auf y d. h. $\delta_y(A) = \begin{cases} 1 & y \in A \\ 0 & y \notin A \end{cases}$.
Wenn $T : \mathbf{R} \to \mathbf{R}$ erwartungstreu ist, dann muß insbesondere für jedes $y \in \mathbf{R}$ $\mathbf{E}_{\delta_y}(T(X)) = y$ gelten, d. h. $T(y) = y$, d. h. es gibt für $n = 1$ nur einen erwartungstreuen Schätzer, nämlich $T = T_1$. Dieser ist (da er ein endliches 2. Moment hat) UMVU.

(β) $n = 2$:
Sei $x < y$ fest und für $p \in [0, 1]$ das Wahrscheinlichkeitsmaß ϑ_p definiert als $\vartheta_p(\{x\}) = p,\ \vartheta_p(\{y\}) = q = 1 - p$. Wenn $T : \mathbf{R}^2 \to \mathbf{R}$ erwartungstreu ist, dann folgt

$$\begin{aligned} \mathbf{E}_{\vartheta_p} T(X_1, X_2) &= p^2 T(x, x) + p(1-p)T(x, y) + (1-p)pT(y, x) + (1-p)^2 T(y, y) \\ &\underset{\text{wg.Erw.treue}}{\overset{!}{=}} px + (1-p)y \end{aligned}$$

Da die letzte Identität für *alle* $p \in [0, 1]$ gilt, erhält man durch Koeffizientenvergleich

$$T(x, x) = x\ ,\ T(y, y) = y \text{ und } T(x, y) + T(y, x) = x + y.$$

Hier gibt es übrigens zahlreiche erwartungstreue Schätzer, neben T_1 z. B. noch $T_2(x, y) = \frac{1}{3}x + \frac{2}{3}y$ aber auch gewisse nichtlineare T.

Es folgt für beliebiges $\vartheta \in \Theta$

$$\begin{aligned} V_\vartheta\left(T(X_1, X_2)\right) &= \mathbf{E}_\vartheta\left((T(X_1, X_2) - \mu_\vartheta)^2\right) \\ &= \frac{1}{2}\left[\mathbf{E}_\vartheta\left((T(X_1, X_2) - \mu_\vartheta)^2\right) + \mathbf{E}_\vartheta\left((T(X_2, X_1) - \mu_\vartheta)^2\right)\right] \\ &= \frac{1}{2}\mathbf{E}_\vartheta\left(T(X_1, X_2)^2 + T(X_2, X_1)^2\right) - \mu_\vartheta^2 \\ &\geq \frac{1}{4}\mathbf{E}_\vartheta\left(T(X_1, X_2) + T(X_2, X_1)\right)^2 - \mu_\vartheta^2 \\ &= \frac{1}{4}\mathbf{E}_\vartheta\left((X_1 + X_2)^2\right) - \mu_\vartheta^2 = \frac{1}{4}(2\mathbf{E}_\vartheta(X^2) + 2\mu_\vartheta^2) - \mu_\vartheta^2 = \frac{1}{2}\sigma^2(\vartheta). \end{aligned}$$

Warnung: Verkleinert man in Beispiel 6.8 die Menge Θ, dann kann T_1 die Eigenschaft UMVU zu sein, verlieren. Als Beispiel betrachte man $\widetilde{\Theta} = \{\mathbf{P}, Q\}$ mit $\mathbf{P}(\{0\}) = \mathbf{P}(\{1\}) = 1/2$ und $Q(\{2\}) = Q(\{3\}) = 1/2$. Sei $n = 1$. Dann ist offensichtlich $\widetilde{T}$ def. als

$$\widetilde{T}(x) = \begin{cases} 0,5 & \text{wenn} \quad x \in \{0, 1\} \\ 2,5 & \text{wenn} \quad x \in \{2, 3\} \\ \text{egal} & \text{sonst} \end{cases}$$

UMVU, da $V_\vartheta(\widetilde{T}) = 0$ für alle (d. h. beide) $\vartheta \in \widetilde{\Theta}$, während T_1 *nicht* UMVU ist!

6.9. Nachteile des Gütekriteriums "Erwartungstreue":

a) u. U. existiert kein erwartungstreuer Schätzer

b) selbst wenn, kann sogar der beste erwartungstreue Schätzer unsinnig sein (Beispiel 6.6)

c) Keine Invarianz unter Transformationen, d. h. ist T erwartungstreu für $\tau(\vartheta)$ und $f : \mathbf{R}^k \to \mathbf{R}^\ell$, dann ist i. allg. $f(T)$ *nicht* erwartungstreu für $f(\tau(\vartheta))$!

Ein Beispiel zu c): $n = k = \ell = 1,\ f(x) = x^2,\ (\mathcal{X}, \mathcal{A}, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ wie in Beispiel 6.8.

$T_1(x) = x$ ist erwartungstreu aber

$$\mathbf{E}_\vartheta\left((T_1(X))^2\right) = \mathbf{E}_\vartheta(X^2) > (\mathbf{E}_\vartheta X)^2 = \mu_\vartheta^2 = f\left(\mathbf{E}_\vartheta(X)\right)$$

$\forall\vartheta \in \Theta$ mit $\sigma^2(\vartheta) > 0$!

Mögliche Gütekriterien für Schätzer:

6.10 Definition: $R(\vartheta, T) := \mathbf{E}_\vartheta(|T(X) - \tau(\vartheta)|^2)$ heißt *Risiko* des Schätzers T an der Stelle ϑ.

Wünschenswert ist ein solches T, für das $R(\vartheta, T)$ möglichst für alle $\vartheta \in \Theta$ klein ist. Im Fall $k = 1$ sahen wir, dass ein UMVU–Schätzer T_1 ein solcher erwartungstreuer Schätzer ist, dass $R(\vartheta, T_1) \le R(\vartheta, T)$ für *alle* $\vartheta \in \Theta$ und *alle* erwartungstreuen Schätzer T gilt.

Mögliche Definitionen von *Optimalität* könnten sein:

a) optimal ist T_1 für $\tau(\vartheta)$ dann, wenn
$\sup\limits_\vartheta R(\vartheta, T_1) \le \sup\limits_\vartheta R(\vartheta, T)$
für *alle* Schätzer T gilt. Ein solches T_1 heißt *Minimax–Schätzer*, da er das maximale Risiko minimiert.

b) optimal ist T_1 für $\tau(\vartheta)$ dann, wenn es den Ausdruck
$r(T) := \int\limits_\Theta R(\vartheta, T) d\mu(\vartheta)$
minimiert, wobei μ ein vorgegebenes Maß auf Θ (mit einer geeigneten σ–Algebra) ist. Solche Schätzer heißen *Bayes–Schätzer*. Sie hängen natürlich i. allg. von der Wahl von μ ab. μ wird oft als *à–priori–Verteilung* bezeichnet. Bayes–Schätzer minimieren also das (bezüglich μ) mittlere Risiko.

6.11. Konsistenz:
Wir betrachten die folgende spezielle Situation: Θ sei irgendeine (nichtleere) Menge von Wahrscheinlichkeitsmaßen auf $\mathbf{R}$ und $X_1, X_2, X_3, \cdots$ sei eine Folge von u.i.v. reellen Zufallsgrößen mit unbekannter Verteilung $\vartheta \in \Theta$. Sei $\tau : \Theta \to \mathbf{R}$ gegeben. Für jedes $n \in \mathbf{N}$ sei ein Schätzer $T_n : \mathbf{R}^n \to \mathbf{R}$ für τ gesucht. Von einer vernünftigen Schätzfolge $T_n, n \in \mathbf{N}$ erwartet man, dass $T_n(X_1, \cdots X_n)$ für $n \to \infty$ (d. h. Stichprobenumfang $\to \infty$) bei beliebigem zugrundeliegenden wahren $\vartheta \in \Theta$ gegen das wahre $\tau(\vartheta)$ konvergiert. Da $T_n(X_1, \cdots, X_n), n \in \mathbf{N}$ eine Folge von Zufallsvariablen ist, gibt es unterschiedliche Konvergenzbegriffe:

Definition:
a) Die Schätzfolge $T_n, n \in \mathbf{N}$ heißt *stark konsistent*, wenn für *alle* $\vartheta \in \Theta$ gilt
$\mathbf{P}_\vartheta \left(T_n(X_1, X_2, \cdots, X_n) \to \tau(\vartheta)\right) = 1$
(wobei $\mathbf{P}_\vartheta$ die Verteilung von $X_1, X_2, \cdots$ ist, wenn $X_1, X_2, \cdots$ u.i.v. mit Verteilung ϑ).
b) Die Schätzfolge $T_n, n \in \mathbf{N}$ heißt *schwach konsistent* (oft auch einfach *konsistent*), wenn für *alle* $\vartheta \in \Theta$ und alle $\varepsilon > 0$ gilt:
$\lim_{n\to\infty} \mathbf{P}_\vartheta \left(|T_n(X_1, X_2, \cdots, X_n) - \tau(\vartheta)| \geq \varepsilon\right) = 0.$

Bemerkung:
Man beachte, dass die Begriffe *stark* und *schwach* hier dieselbe Bedeutung wie bei den Gesetzen der großen Zahlen haben.

6.12. Beispiel:
Ähnlich wie im Beispiel 6.8. sei Θ die Menge aller Wahrscheinlichkeitsmaße auf $\mathbf{R}$ mit existierendem Erwartungswert. Sei $\tau(\vartheta) := \mu_\vartheta$ der Erwartungswert von ϑ. $X_1, X_2, X_3, \cdots$ seien u.i.v. mit Verteilung $\vartheta \in \Theta$. Die Schätzfolge

$$T_n(x_1, \cdots, x_n) := \frac{1}{n} \sum_{i=1}^{n} x_i$$

ist stark konsistent, denn nach dem starken Gesetz der großen Zahlen gilt für jedes $\vartheta \in \Theta$

$$\mathbf{P}_\vartheta \left(T_n(X_1, \cdots, X_n) \to \tau(\vartheta) \text{ für } n \to \infty\right) = 1.$$

6.13. Maximum Likelihood Schätzung:

Definition:
Sei $\Theta \subset \mathbf{R}^k$ und $\Sigma = \Theta$. Ein Schätzer $T : \mathcal{X} \to \Theta$ heißt Maximum–Likelihood–Schätzer (ML), falls entweder

a) $\mathcal{X}$ diskret (d. h. höchstens abzählbar) und $T(x)$ so gewählt ist, dass $\mathbf{P}_{T(x)}(\{x\}) = \sup\limits_{\vartheta\in\Theta} \mathbf{P}_\vartheta(\{x\})$, $x \in \mathcal{X}$ oder

b) $\mathcal{X} = \mathbf{R}^n$ und alle $\mathbf{P}_\vartheta, \vartheta \in \Theta$ haben eine Dichte f_ϑ und $f_{T(x)}(x) = \sup\limits_{\vartheta\in\Theta} f_\vartheta(x)$, $x \in \mathcal{X}$.

Interpretation:
Wähle im Fall der Beobachtung $x \in \mathcal{X}$ unter allen zur Konkurrenz zugelassenen Verteilungen diejenige aus, für die die Beobachtung x maximale Wahrscheinlichkeit (bzw. Dichte) hat d. h. die "plausibelste Erklärung" für die Beobachtung.

Bemerkung:
$\vartheta \mapsto \mathbf{P}_\vartheta(x)$ bzw. $\vartheta \mapsto f_\vartheta(x)$ heißen *Likelihoodfunktion* (bei gegebener Beobachtung $x \in \mathcal{X}$).

6.14. Beispiel:
$\Theta = \mathbf{R}, \mathcal{X} = \mathbf{R}^n, \mathbf{P}_\vartheta =$ Verteilung von $(X_1, \cdots, X_n)$ mit $X_1, \cdots, X_n$ u.i. $\mathcal{N}(\vartheta, \sigma^2)$– verteilt, wobei $\sigma^2 > 0$ bekannt (gegeben) sei (der Ausdruck $\mathcal{N}(\vartheta, \sigma^2)$ wird in der Vorlesung erklärt). Wie sieht der ML–Schätzer T_{ML} für ϑ aus?
Für jede Beobachtung $x = (x_1, \cdots, x_n) \in \mathbf{R}^n$ ist dasjenige ϑ gesucht, welches die Dichte von $(X_1, \cdots, X_n)$ an der Stelle x maximiert, also das Maximum von

$$f_\vartheta(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \vartheta)^2}{2\sigma^2}\right).$$

Äquivalent: man finde das ϑ, welches $\log f_\vartheta(x)$ maximiert.

$$\frac{\partial}{\partial\vartheta}(\log f_\vartheta(x)) = \sum_{i=1}^n \frac{\partial}{\partial\vartheta}\left(-\frac{(x_i - \vartheta)^2}{2\sigma^2}\right) = \sum_{i=1}^n \frac{x_i - \vartheta}{\sigma^2} \stackrel{!}{=} 0.$$

Das gesuchte ϑ ist also

$$\hat{\vartheta}_{ML}(x) = T_{ML}(x) = \frac{1}{n}\sum_{i=1}^n x_i \; .$$

6.15. Beispiel:
Ähnlich wie 6.14. nur seien $\mu = \mathbf{E}X$ und σ^2 *beide* unbekannt. Es ist $\Theta = \mathbf{R} \times (0, \infty), \mathcal{X} = \mathbf{R}^n, \mathbf{P}_\vartheta =$ Verteilung von $(X_1, \cdots, X_n)$ mit $X_1, \cdots, X_n$ u.i. $\mathcal{N}(\mu, \sigma^2)$–verteilt ; $\vartheta := (\mu, \sigma^2)$.
Es gilt

$$f_{\mu,\sigma^2}(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Gesucht ist das Paar $\vartheta = (\mu, \sigma^2)$, welches bei gegebener Beobachtung $x = (x_1, \cdots, x_n)$ $f_{\mu,\sigma^2}(x)$ (oder äquivalent $\log f_{\mu,\sigma^2}(x)$) maximiert.

$$\frac{\partial}{\partial\mu}(\log f_{\mu,\sigma^2}(x)) = \sum_{i=1}^{n} \frac{x_i-\mu}{\sigma^2} = 0 \text{ d.h. } \hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i \; .$$

$$\frac{\partial}{\partial\sigma^2}(\log f_{\mu,\sigma^2}(x)) = \sum_{i=1}^{n}\left[-\frac{1}{2}\frac{\partial}{\partial\sigma^2}\log(2\pi\sigma^2) - \frac{\partial}{\partial\sigma^2}\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$= \sum_{i=1}^{n} -\frac{1}{2\sigma^2} + \frac{(x_i-\mu)^2}{2\sigma^4} \overset{!}{=} 0 \Rightarrow \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu}_{ML})^2,$$

also

$$T_{ML}(x) = \left(\hat{\mu}_{ML},\ \hat{\sigma}^2_{ML}\right) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i,\ \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2\right)$$

Bemerkung:
$\hat{\sigma}^2_{ML}$ ist (in der Situation von 6.15.) *nicht* erwartungstreu. Man prüft aber leicht nach, dass $T(x) := \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2$ im Fall $n \geq 2$ die Varianz σ^2 erwartungstreu schätzt.

Bemerkungen (ohne Beweis):
a) ML–Schätzer sind invariant unter funktionalen Transformationen (vgl. 6.9.) d.h. ist T ML für ϑ, dann ist $g(T)$ ML für $g(\vartheta)$ falls g injektiv ist.
b) Man kann zeigen, dass unter nicht allzu starken Bedingungen ML– Schätzer für $n \to \infty$ gewisse Optimalitätseigenschaften besitzen.

6.16. Konfidenzbereiche:
Wenn man einen Schätzer für τ gefunden hat, dann ist man meist an Fehlerabschätzungen interessiert. Man will dazu einen Bereich um den Schätzwert angeben, von dem man mit großer Sicherheit sagen kann, dass er das wahre $\tau(\vartheta)$ enthält.

Definition:
Sei $(\mathcal{X}, \mathcal{A}, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $\tau : \Theta \to \mathbf{R}^k$ und $\alpha \in [0,1]$. Eine Abbildung $C : \mathcal{X} \to \mathcal{B}_k$ (= Borelmengen von $\mathbf{R}^k$) heißt *Konfidenzbereich für τ zum Niveau α* falls $\mathbf{P}_\vartheta(\{x : C(x) \ni \tau(\vartheta)\}) \geq \alpha \ \ \forall\vartheta \in \Theta$ gilt.

Ist im Spezialfall $k = 1$ $C(x)$ für jedes x ein Intervall, dann heißt C auch *Konfidenzintervall für τ zum Niveau α.*

Bemerkung:
Typische Werte für α sind $\alpha = 0,90,\ \alpha = 0,95$ oder $\alpha = 0,99$.

Warnung:
Hat man ein konkretes $x_0 \in \mathcal{X}$ beobachtet, so sollte man die Sprechweise "$\tau(\vartheta)$ liegt mit

der Mindestwahrscheinlichkeit α in $C(x_0)$" vermeiden, denn $\tau(\vartheta)$ ist ja gar nicht zufällig!

Bemerkung:
Bei gegebenem α ist man an möglichst *kleinen* Mengen $C(x)$ interessiert. $C(x) \equiv \mathbf{R}^k$ geht immer, ist aber uninteressant.

Beispiel (vgl. 6.14):
Zu vorgegebenem $\alpha \in (0,1)$ finde man in der Situation von Beispiel 6.14. ein Konfidenzintervall für $\tau(\vartheta) = \vartheta = \mu$ zum Niveau α!
Für jedes $\vartheta = \mu$ ist $Y := \frac{1}{\sigma\sqrt{n}} \left(\sum_{i=1}^n X_i - n\mu\right)$ $\mathcal{N}(0,1)$–verteilt.

Man kann $q_1, q_2 \in \mathbf{R}$ finden, so dass $\Phi(q_2) - \Phi(q_1) = \alpha$ gilt (z.B. so dass $q_1 = -q_2$), wobei Φ die Verteilungsfunktion der Standard-Normalverteilung ist. Nun gilt

$$q_1 \leq Y \leq q_2 \;\Leftrightarrow\; \frac{\sum\limits_{i=1}^n X_i - q_2\sigma\sqrt{n}}{n} \leq \mu \leq \frac{\sum\limits_{i=1}^n X_i - q_1\sigma\sqrt{n}}{n} \quad ,$$

also hat auch das rechte Ereignis (für jedes $\mathbf{P}_\vartheta$) Wahrscheinlichkeit α, d.h.

$$C(x) := \left[\frac{1}{n}\Big(\sum_{i=1}^n x_i - q_2\sigma\sqrt{n}\Big), \frac{1}{n}\Big(\sum_{i=1}^n x_i - q_1\sigma\sqrt{n}\Big)\right]$$

ist ein Konfidenzintervall für $\tau(\vartheta) = \vartheta = \mu$ zum Niveau α.