Kapitel VI
Grundlagen der Statistik: Schätztheorie für diskrete Verteilungen

In der Wahrscheinlichkeitsrechnung geht es immer um die Berechnung von Wahrscheinlichkeiten, Erwartungswerten, etc. in einem zuvor spezifizierten Modell und damit unter einer zuvor festgelegten, also bekannten W-Verteilung. Diese Annahme einer bekannten W-Verteilung läßt sich in manchen Fällen durch "geometrische" Überlegungen rechtfertigen – z.B. die Laplace-Verteilung bei symmetrischen homogenen Körpern wie Münze, Würfel, etc. In anderen Situationen werden die Ausgangsverteilungen durch Modellannahmen nahegelegt (z.B. die Binomialverteilung durch Bernoulli-Experimente oder die Poisson-Verteilung durch den Poissonschen Grenzwertsatz). In vielen praktischen Problemen ist jedoch bereits die Ausgangsverteilung unbekannt, und es stellt sich die Frage, ob auf der Basis gegebener Beobachtungswerte eines betrachteten Zufallsexperiments Rückschlüsse auf diese Verteilung möglich sind. Dies führt uns geradewegs in die Welt der *Statistik*, deren Aufgabe nicht mehr darin besteht, ein vorgegebenes Modell zu analysieren, sondern vielmehr nach Durchführung oder Realisierung des Experiments das Ergebnis zu nutzen, um aus einer Klasse von Modellen das im geeigneten Sinn plausibelste auszuwählen.

18. Problemstellung

Um die Aufgabenstellung weiter zu erläutern und erste Begriffsbildungen vorzunehmen, beginnen wir mit einem nicht sehr originellen, aber instruktiven Beispiel.

18.1. Beispiel. Eine verbogene Münze wird zu einem "Spiel" genutzt: Gegen eine Gebühr von 1 Euro darf man einmal die Münze werfen und erhält 2 Euro zurück, wenn "Zahl" ("1") erscheint, während bei "Kopf" ("0") der Einsatz verloren ist. Die Modellbildung ist hier offensichtlich: Es liegt ein Bernoulli-Experiment (Ω, p) vor, d.h.

$$\Omega = \{0,1\} \quad \text{und} \quad p(1) = 1 - p(0) = \theta \text{ für ein } \theta \in [0,1]. \tag{18.1}$$

Es interessiert der Gewinn/Verlust, gegeben durch die Zufallsgröße

$$Y(\omega) \;=\; \begin{cases} 1, & \text{falls } \omega = 1 \qquad \text{(2 Euro Auszahlung – 1 Euro Einsatz)} \\ -1, & \text{falls } \omega = 0 \qquad \text{(–1 Euro Einsatz)} \end{cases} .$$

Ein naheliegendes Kriterium für die Entscheidung über die Teilnahme an diesem "Spiel" wird durch den Erwartungswert

$$EY \;=\; 1 \cdot p(1) + (-1) \cdot p(0) \;=\; 2\theta - 1$$

geliefert; für $\theta > 1/2$ ist es "günstig" teilzunehmen, für $\theta < 1/2$ "ungünstig", für $\theta = 1/2$ handelt es sich um ein faires Spiel (Laplace-Experiment).

Das Problem besteht darin, daß der Wert θ unbekannt ist und wir ihn ohne weitere Informationen über das Modell auch nicht bestimmen können. Was also tun, wenn solche Informationen nicht verfügbar sind? Erinnern wir uns dazu an die Aussage des schwachen Gesetzes der großen Zahlen: Sei Y_i der Gewinn/Verlust der i-ten Runde und $X_i \overset{\text{def}}{=} \mathbf{1}_{\{1\}}(Y_i)$ für $i \in \mathbb{N}$. Wird die Münze n-mal unabhängig voneinander geworfen und die Anzahl $S_n \overset{\text{def}}{=} X_1 + ... + X_n$ der Gewinnrunden registriert, so konvergiert $n^{-1}S_n$ für $n \to \infty$ in Wahrscheinlichkeit gegen $EX_1 = P(X_1 = 1) = \theta$. Natürlich ist es praktisch unmöglich, die Münze unendlich oft zu werfen. Deshalb wird man sich mit einer endlichen Anzahl n_0 von Versuchen begnügen müssen und auf der Grundlage der Wurfergebnisse eine Entscheidung über den Wert von θ fällen.

Immerhin liefert die Tschebyschev-Ungleichung für den Fall, daß n_0 unabhängige Versuchswiederholungen durchgeführt werden, die Abschätzung

$$P\Big(|n_0^{-1}S_{n_0} - \theta| \geq \varepsilon\Big) \;\leq\; \frac{\theta(1-\theta)}{n_0\varepsilon^2} \;\leq\; \frac{1}{4n_0\varepsilon^2}$$

(wegen $x(1-x) \leq 1/4$ für alle $x \in [0,1]$). Man hat also eine Schranke für die Wahrscheinlichkeit einer Abweichung von mehr als ε der relativen Häufigkeit von dem wahren, aber unbekannten Wert θ.

Nach diesem Beispiel wollen wir die Problemstellung im Rahmen eines allgemeinen mathematischen Modells erfassen und formalisieren.

18.2. Generalvoraussetzung. Es sei $\mathcal{P}$ eine nichtleere Familie von diskreten W-Verteilungen über Ω und $X = (X_1, ..., X_n)$ ein Zufallsvektor $X : \Omega \to \mathbb{R}^n$. Die induzierte Klasse

$$\mathcal{P}^X \;\overset{\text{def}}{=}\; \{P^X : P \in \mathcal{P}\}$$

von W-Verteilungen sei durch einen reellwertigen *Parameter* θ parametrisiert:

$$\mathcal{P}^X \;=\; \{P_\theta^X : \theta \in \Theta\}.$$

Die Menge $\Theta \subset \mathbb{R}$ der möglichen Parameterwerte heißt auch *Parameterraum.*

Im obigen Beispiel ist im Fall von n Spielrunden $X = (X_1, ..., X_n)$ der Zufallsvektor, der aus den Ergebnissen $X_1, ..., X_n$ der n Wurfergebnisse besteht. Wir dürfen davon ausgehen, daß $X_1, ..., X_n$ stochastisch unabhängig und jeweils die Zähldichte aus (18.1) mit unbekanntem $\theta \in \Theta \overset{\text{def}}{=} [0,1]$ besitzen. Damit gilt für alle $x_1, ..., x_n \in \{0,1\}$ und mit $s_n \overset{\text{def}}{=} \sum_{i=1}^n x_i$

$$P^X(\{(x_1, ..., x_n)\}) \;=\; = \;\theta^{s_n}(1-\theta)^{n-s_n},$$

d.h. die Klasse $\mathcal{P}^X$ kann durch den Parameter $\theta \in \Theta = [0,1]$ parametrisiert werden.

Das weitere Vorgehen hängt von der jeweiligen Zielsetzung ab: Will man eine Aussage über den genauen Wert von θ – allgemein über einen Funktionswert $\gamma(\theta)$ – machen, so spricht man von einem *Schätzproblem.* Eine *Schätzfunktion* ist dann eine Vorschrift, die jedem möglichen Beobachtungswert $(x_1, ..., x_n)$ denjenigen Parameterwert aus Θ – bzw. $\gamma(\Theta)$ – zuordnet, der aufgrund dieser Information als *Schätzwert* angesehen wird.

18.3. Definition. Es seien $\mathcal{P}^X = \{P_\theta^X : \theta \in \Theta\}$ eine Familie von (diskreten) Verteilungen von $X : \Omega \to \mathbb{R}^n$ und $\gamma : \Theta \to \mathbb{R}$ eine Abbildung. Eine *Schätzfunktion* oder kurz *Schätzer* (für $\gamma(\theta)$) ist eine Abbildung $g : \mathbb{R}^n \to \mathbb{R}$.

Gemäß der obigen Motivation wird $g(X_1(\omega), ..., X_n(\omega))$ als *Schätzwert* für das "wahre" $\gamma(\theta)$ verwendet. Natürlich gibt es eine Fülle von Schätzern, in unserem Ausgangsbeispiel 18.1 etwa für θ

$$g_1 : \; (x_1, ..., x_n) \mapsto \overline{x}_n \;\overset{\text{def}}{=}\; \frac{1}{n}\sum_{i=1}^n x_i,$$

$$g_2 : \; (x_1, ..., x_n) \mapsto \frac{1}{2},$$

d.h. bei Verwendung von g_2 wird ungeachtet der Versuchsergebnisse daran geglaubt, daß das Spiel fair ist. Auch möglich wäre der offenbar wenig sinnvolle Schätzer

$$g_3 : \; (x_1, ..., x_n) \mapsto x_1,$$

der nur die erste Beobachtung berücksichtigt und nur zu den Schätzwerten 0 oder 1 führen kann, usw. In der Tat ist die Menge der Funktionen $g : \mathbb{R}^n \to \mathbb{R}$ überabzählbar.

Andererseits sind aber offensichtlich viele Schätzfunktionen wertlos. Man will $\gamma(\theta)$ ja nicht irgendwie schätzen, sondern möglichst gut, d.h. möglichst genau. Ein *"idealer Schätzer"* wäre sicherlich ein solcher, der für jedes $\theta \in \Theta$ und jeden Beobachtungswert $(x_1, ..., x_n)$ den wahren Wert $\gamma(\theta)$ angibt. Leider kann es einen derartigen Schätzer nur in Ausnahmenfällen geben: Da in unserem Ausgangsbeispiel nur 2^n verschiedene Beobachtungswerte möglich sind, kann jeder Schätzer $g : \{0,1\}^n \to \Theta = [0,1]$ nur maximal 2^n Funktionswerte haben. Überabzählbar viele Parameter $\theta \in [0,1]$ können somit gar nicht angenommen werden und werden deshalb mit Sicherheit falsch geschätzt. Da also ein idealer Schätzer i.a. nicht existiert, muß man sich mit weniger zufriedengeben. Das Ziel muß darin bestehen, einen besten

Schätzer auszuwählen bezüglich eines *erfüllbaren Gütekriteriums.* Zwei Ansätze werden im folgenden vorgestellt.

19. Das Maximum-Likelihood-Prinzip

Gegeben eine Familie $\mathcal{P}^X = \{P_\theta^X : \theta \in \Theta\}$ diskreter W-Verteilungen, soll – auf der Grundlage des beobachteten Wertes von X – der unbekannte Parameter θ geschätzt werden. Die Funktion

$$\mathbb{L}(\theta|x) \stackrel{\text{def}}{=} P_\theta^X(\{x\}) \;=\; P_\theta(X = x) \tag{19.1}$$

heißt *Likelihood-Funktion.* Bei festem θ ist $\mathbb{L}(\theta|\cdot)$ offenbar nichts anderes als die Zähldichte von P_θ^X. In der häufig vorliegenden Situation einer n-fachen Versuchswiederholung ist $X = (X_1, ..., X_n)$ mit unabhängigen, identisch verteilten Zufallsgrößen $X_1, ..., X_n$. Bezeichnet p_θ deren Zähldichte, also $p_\theta(y) = P_\theta(X_i = y)$ für alle $y \in \mathbb{R}$, $\theta \in \Theta$ und $i = 1, ..., n$, so gilt offenbar vermöge der Unabhängigkeit

$$\mathbb{L}(\theta|x) = \; P_\theta(X = x) \;=\; P_\theta(X_1 = x_1, ..., X_n = x_n) \;=\; \prod_{i=1}^{n} p_\theta(x_i) \tag{19.2}$$

für alle $x = (x_1, ..., x_n) \in \mathbb{R}^n$ und $\theta \in \Theta$.

Das *Maximum-Likelihood-Prinzip* basiert auf der folgenden einfachen Idee: Wähle, wenn möglich, zu gegebener Beobachtung x denjenigen Parameterwert $\hat{\theta}(x) \in \Theta$ als Schätzwert für θ, unter dem die Wahrscheinlichkeit für das Auftreten von x maximal wird, d.h.

$$P_{\hat{\theta}(x)}(X = x) \;=\; \max_{\theta \in \Theta} P_\theta(X = x) \tag{19.3}$$

oder gleichbedeutend

$$\mathbb{L}(\hat{\theta}(x)|x) \;=\; \max_{\theta \in \Theta} \mathbb{L}(\theta|x). \tag{19.3'}$$

Auch wenn dieses Prinzip daran scheitern kann, daß ein $\hat{\theta}(x)$ nicht für jeden möglichen Beobachtungswert existiert, besticht es durch seine intuitiv einleuchtende Philosophie.

19.1. Definition. Existiert $\hat{\theta}(x)$ gemäß (19.3) für jedes x, so heißt $\hat{\theta}$ der *Maximum-Likelihood-Schätzer (MLS)* für θ.

Betrachten wir als nächstes einige Beispiele:

19.2. Beispiel. ***(Verbogene Münze, Fortsetzung von 18.1)*** In dieser Situation ist $X = (X_1, ..., X_n)$ mit stochastisch unabhängigen, identisch $B(1, \theta)$-verteilten Komponenten.

Die Zähldichte p_θ der X_i hat die Form $p_\theta(y) = \theta^y(1-\theta)^{1-y}\mathbf{1}_{\{0,1\}}(y)$. Daher gilt gemäß (19.2) für alle $x = (x_1, ..., x_n) \in \{0,1\}^n$

$$\mathbb{L}(\theta|x) \; = \; \theta^{s_n}(1-\theta)^{n-s_n},$$

wobei $s_n \stackrel{\text{def}}{=} \sum_{i=1}^n x_i$. Für $x \notin \{0,1\}^n$ ist die Likelihoodfunktion identisch 0 und somit trivial. In diesem Fall kann man $\hat{\theta}(x)$ beliebig wählen, was allerdings aus statistischer Sicht keine Bedeutung hat, weil die Beobachtung mit Wahrscheinlichkeit aus $\{0,1\}^n$ stammt. Man betrachtet das Maximierungsproblem folglich immer nur auf der Teilmenge $\mathfrak{X}$ von $\mathbb{R}^n$ (dem eigentlichen Stichprobenraum), auf dem $\mathcal{P}^X$ positive Masse besitzt, d.h.

$$\mathfrak{X} \; = \; \{x \in \mathbb{R}^n : P_\theta(X = x) > 0 \text{ für mindestens ein } \theta \in \Theta\}.$$

Dies werden wir in den nachfolgenden Beispielen ohne nochmaligen Hinweis immer so handhaben.

Da $\mathbb{L}(\cdot|x)$ in θ differenzierbar ist, bestimmen wir das Maximum durch Differentiation nach θ. Da dies einfacher ist für die *Log-Likelihood-Funktion*

$$\ell(\theta|x) \; \stackrel{\text{def}}{=} \; \log \mathbb{L}(\theta|x) \; = \; s_n \log\theta + (n - s_n)\log(1-\theta),$$

und diese Funktion wegen der strengen Monotonie des Logarithmus ihr Maximum an derselben Stelle annimmt, gehen wir zu $\ell(\theta|x)$ über und erhalten:

$$\frac{d}{d\theta}\ell(\theta|x) \; = \; \frac{s_n}{\theta} - \frac{n - s_n}{1-\theta}.$$

Die sogenannte *Likelihood-Gleichung*

$$\frac{d}{d\theta}\ell(\theta|x) \; = \; 0$$

ergibt hier demnach (für $\theta \in (0,1)$)

$$\frac{s_n}{\theta} - \frac{n-s_n}{1-\theta} \; = \; \frac{(1-\theta)s_n - \theta n + \theta s_n}{\theta(1-\theta)} \; = \; 0$$

und folglich als Lösung

$$\hat{\theta}(x) \; = \; \frac{s_n}{n}.$$

Da $\lim_{\theta\downarrow 0} \ell(\theta|x) = \lim_{\theta\uparrow 1} \ell(\theta|x) = -\infty$, muß in $\hat{\theta}(x)$ ein Maximum vorliegen. Der MLS für θ lautet also $\hat{\theta}(x) = \frac{s_n}{n} = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{1\}}(y_i)$, ein sehr plausibles Ergebnis.

19.3. Beispiel. ***(Binomialverteilung)*** Die Beobachtung bestehe diesmal aus n unabhängigen jeweils $B(m,\theta)$-verteilten Zufallsgrößen, wobei $m \in \mathbb{N}$ bekannt, $\theta \in [0,1]$ aber unbekannt ist. Wir suchen den MLS für θ. Die Likelihood-Funktion hat hier die Gestalt

$$\mathbb{L}(\theta|x) \; = \; \left(\prod_{i=1}^n \binom{m}{x_i}\right)\theta^{s_n}(1-\theta)^{mn-s_n}$$

für $x = (x_1, ..., x_n) \in \{0, ..., m\}^n = \mathfrak{X}$, wobei wieder $s_n = x_1 + ... + x_n$. Es folgt für die Log-Likelihood-Funktion

$$\ell(\theta|x) \;=\; \sum_{i=1}^{n} \log\left(\binom{m}{x_i}\right) \;+\; s_n \log\theta \;+\; (mn - s_n)\log(1-\theta).$$

Sie hat bis auf den ersten von θ unabhängigen Term im wesentlichen dieselbe Gestalt wie im vorherigen Beispiel. Es ergibt sich durch analoges Vorgehen

$$\hat{\theta}(x) \;=\; \frac{s_n}{mn}.$$

Dies läßt sich im übrigen auch dadurch einsehen, daß ja die $B(m, \theta)$-Verteilung gerade die Anzahl von Erfolgen (Einsen) beim m-fachen Münzwurf angibt. Die n-fache Erhebung von Daten gemäß einer $B(m, \theta)$-Verteilung entspricht also der mn-fachen Durchführung eines Münzwurfs, wenn man die Würfe in Blöcken der Länge m durchführt und die Daten als beobachtete Anzahl von Erfolgen in diesen Blöcken interpretiert. Wir hatten aber in 19.2 gesehen, daß beim n-fachen Münzwurf zur Bestimmung des MLS für θ ohnehin nur die Anzahl der Erfolge benötigt wird, und diese läßt sich auch aus den geblockten Daten durch Summation ermitteln.

19.4. Beispiel. ***(Geometrische Verteilung)*** Besteht die Beobachtung aus n unabhängigen, jeweils geometrisch verteilten Zufallsgrößen, lautet die Likelihood-Funktion

$$\mathbb{L}(\theta|x) \;=\; \theta^n (1-\theta)^{s_n}$$

also die Log-Likelihood-Funktion

$$\ell(\theta|x) \;=\; n\log\theta \;+\; s_n \log(1-\theta).$$

Es ergibt sich leicht

$$\hat{\theta}(x) \;=\; \frac{n}{s_n + n} \;=\; \frac{1}{\frac{s_n}{n} + 1}$$

als MLS für θ.

Unser letztes Beispiel unterscheidet sich von den vorherigen darin, daß der unbekannte Parameter aus einer diskreten Menge stammt und folglich die Bestimmung des MLS nicht vermöge Differentiation gefunden werden kann.

19.5. Beispiel. ***(Laplace-Verteilungen)*** In einer Urne befinde sich eine unbekannte Anzahl θ Kugeln, die mit 1 bis θ durchnumeriert seien. Auf der Grundlage von n Ziehungen mit Zurücklegen soll ein MLS für θ bestimmt werden. Die Beobachtung besteht hier aus n unabhängigen, jeweils auf $\{1, ..., \theta\}$ Laplace-verteilten Zufallsgrößen $X_1, ..., X_n$, so daß

$$\mathbb{L}(\theta|x) \;=\; \begin{cases} \dfrac{1}{\theta^n}, & \text{falls } (x_1, ..., x_n) \in \{1, ..., \theta\}^n \\ 0, & \text{sonst} \end{cases} \;=\; \frac{1}{\theta^n} \mathbf{1}_{\{1,...,\theta\}}(\max_{1 \le i \le n} x_i).$$

Da θ^{-n} monoton fällt in θ, lautet der MLS für θ offenbar

$$\hat{\theta}(x) \;=\; \inf\{\theta : \mathbf{1}_{\{1,\ldots,\theta\}}(\max_{1\le i\le n} x_i) = 1\} \;=\; \max_{1\le i\le n} x_i$$

für alle $x = (x_1, \ldots, x_n) \in \{1, \ldots, \theta\}^n$.

Eine nützliche Eigenschaft von MLS ist ihre Invarianz unter Parametertransformationen. Zur Präzisierung dieser Feststellung sei $\gamma : \Theta \to \Theta'$ eine Parameterfunktion. Zum Schätzen von $\gamma(\theta)$ (bei Beobachtung von $X = x$) mit Hilfe der Maximum-Likelihood-Methode definieren wir zunächst die von γ *induzierte Likelihood-Funktion* $\mathbb{L}_\gamma(\cdot|x) : \gamma(\Theta) \to [0, \infty)$ durch

$$\mathbb{L}_\gamma(\eta|x) \;=\; \sup_{\theta:\gamma(\theta)=\eta} \mathbb{L}(\theta|x). \tag{19.2}$$

Ein Schätzer $\hat{\eta} : \mathfrak{X} \to \gamma(\Theta)$ für $\eta = \gamma(\theta)$ heißt MLS für η, falls dieser $\mathbb{L}_\gamma(\cdot|x)$ für jedes $x \in \mathfrak{X}$ maximiert. Die naheliegende Frage, ob $\hat{\eta} = \gamma(\hat{\theta})$ dann einen MLS für η bildet, beantwortet

19.6. Satz. *Sei $\hat{\theta}$ ein MLS für θ und $\gamma : \Theta \to \Theta'$ eine Parameterfunktion. Dann ist $\hat{\eta} = \gamma(\hat{\theta})$ ein MLS für $\eta = \gamma(\theta)$.*

Beweis: Die Behauptung ergibt sich leicht aus der Gleichungskette

$$\begin{aligned} \mathbb{L}_\gamma(\gamma(\hat{\theta})|x) \;&=\; \sup_{\theta:\gamma(\theta)=\gamma(\hat{\theta})} \mathbb{L}(\theta|x) \;=\; \mathbb{L}(\hat{\theta}|x) \;=\; \sup_{\theta\in\Theta} \mathbb{L}(\theta|x) \\ &=\; \sup_{\eta\in\gamma(\Theta)} \sup_{\theta:\gamma(\theta)=\eta} \mathbb{L}(\theta|x) \;=\; \sup_{\eta\in\gamma(\Theta)} \mathbb{L}_\gamma(\eta|x). \end{aligned}$$ $\diamondsuit$

19.7. Beispiel. ***(Geometrische Verteilung, Fortsetzung von 19.4)*** Der Mittelwert der geometrischen Verteilung mit Parameter θ lautet bekanntlich $\gamma(\theta) = \frac{1-\theta}{\theta}$ (☞ 6.4(c)). Gemäß Satz 19.6 erhalten wir nun ohne weitern Aufwand $\gamma(\hat{\theta}(x)) = \frac{s_n}{n}$ als MLS für $\gamma(\theta)$.

20. Erwartungstreue Schätzer

Wir kommen nun zu einem anderen Schätzprinzip, das darauf basiert, die Klasse der Schätzer, in der man nach einem besten sucht, einzuschränken.

20.1. Definition. Es sei $X : \Omega \to \mathbb{R}^n$ ein diskreter Zufallsvektor und $\mathcal{P}^X = \{P_\theta^X : \theta \in \Theta\}$ die Familie der möglichen Verteilungen von X, wobei $\Theta \subset \mathbb{R}$. Sei ferner $\gamma : \Theta \to \mathbb{R}$ eine Parameterfunktion. Ein Schätzer g für $\gamma(\theta)$ heißt *erwartungstreu*, wenn

$$E_\theta g(X) \;\overset{\text{def}}{=}\; E_{P_\theta^X} g \;=\; \sum_{x\in\mathbb{R}^n} g(x) P_\theta(X = x) \;=\; \gamma(\theta) \tag{20.1}$$

für alle $\theta \in \Theta$ gilt. Existiert bei gegebener Familie $\mathcal{P}^X$ ein solcher Schätzer, so heißt $\gamma(\theta)$ *erwartungstreu schätzbar.*

Ein erwartungstreuer Schätzer besitzt also die Eigenschaft, zumindest im Mittel stets richtig zu schätzen. Betrachten wir nochmals die Schätzer g_1, g_2 und g_3 für θ zu Beispiel 18.1 (☞ nach 18.3). Offensichtlich gilt sowohl

$$E_\theta g_1(X) \;=\; E_{P_\theta}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{1\}}(X_i)\right) \;=\; \frac{1}{n}\sum_{i=1}^{n} P_\theta(X_i = 1) \;=\; \theta$$

als auch

$$E_\theta g_3(X) \;=\; E(\mathbf{1}_{\{1\}}(X_i)) \;=\; P_\theta(X_1 = 1) \;=\; \theta$$

für alle $\theta \in [0,1]$, d.h. g_1 und g_3 sind erwartungstreu für θ. Dagegen gilt für $\theta \neq 1/2$

$$E_\theta g_2(X) \;=\; \frac{1}{2} \;\neq\; \theta,$$

d.h., g_2 ist nicht erwartungstreu. Die in den Beispielen 19.2, 19.3 und 19.7 erhaltenen MLS sind jeweils erwartungstreu, wie man leicht nachprüft.

Gegeben einen diskreten Zufallsvektor $X : \Omega \to \mathbb{R}^n$ mit Komponenten $X_1, ..., X_n$ und Verteilung aus der Menge $\mathcal{P}^X = \{P_\theta^X : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, sprechen wir im folgenden kurz von einem *UIV-Modell*, wenn $X_1, ..., X_n$ unter jedem P_θ stochastisch unabhängig und identisch verteilt sind, wobei "UIV" als Abkürzung für diese Eigenschaften steht.

Der anschließende, sehr einfache Satz zeigt, daß in einem UIV-Modell das *Stichprobenmittel* $\overline{x}_n \overset{\text{def}}{=} n^{-1}(x_1 + ... + x_n) = n^{-1}s_n$ immer einen erwartungstreuen Schätzer für den Mittelwert der X_i bildet, sofern dieser für jedes θ existiert.

20.2. Satz. *Gegeben sei ein UIV-Modell $(X, \mathcal{P}^X)$ mit*

$$\mu(\theta) \;\overset{\text{def}}{=}\; E_\theta X_1 \;<\; \infty$$

für alle $\theta \in \Theta$. Dann ist $\hat{\mu}(x) = \overline{x}_n$ ein erwartungstreuer Schätzer für $\mu(\theta)$.

Beweis: Die Behauptung folgt sofort aus der Rechnung

$$E_\theta \hat{\mu}(X) \;=\; E_\theta \overline{X}_n \;=\; \frac{1}{n}\sum_{i=1}^{n} E_\theta X_i \;=\; E_\theta X_1 \;=\; \mu(\theta)$$

für alle $\theta \in \Theta$. ♢

Da in vielen Anwendungen auch die Varianz der Beobachtungen von Bedeutung ist, geben wir als nächstes ein entsprechendes Resultat für diese Größe.

20.3. Satz. *Gegeben sei ein UIV-Modell* $(X, \mathcal{P}^X)$ *mit*

$$\sigma^2(\theta) \stackrel{\text{def}}{=} Var_\theta X_1 \ < \ \infty$$

für alle $\theta \in \Theta$. *Dann ist die Stichprobenvarianz*

$$\hat{\sigma}^2(x) \stackrel{\text{def}}{=} \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2$$

ein erwartungstreuer Schätzer für $\sigma^2(\theta)$.

Beweis: Zunächst notieren wir, daß

$$\begin{aligned} \hat{\sigma}^2(x) &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu(\theta))^2 - \frac{2}{n-1} \sum_{i=1}^{n} (x_i - \mu(\theta))(\overline{x}_n - \mu(\theta)) + \frac{n}{n-1} (\overline{x}_n - \mu(\theta))^2 \\ &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu(\theta))^2 - \frac{n}{n-1} (\overline{x}_n - \mu(\theta))^2 \end{aligned}$$

für alle $\theta \in \Theta$ gilt. Dies liefert unter Beachtung von $E_\theta \overline{X}_n = \mu(\theta)$ und $\mathrm{Var}_\theta \overline{X}_n = \frac{\sigma^2(\theta)}{n}$

$$\begin{aligned} E_\theta \hat{\sigma}^2(X) &= \frac{1}{n-1} \sum_{i=1}^{n} E_\theta (X_i - \mu(\theta))^2 - \frac{n}{n-1} E_\theta (\overline{X}_n - \mu(\theta))^2 \\ &= \frac{1}{n-1} \sum_{i=1}^{n} \mathrm{Var}_\theta X_i - \frac{n}{n-1} \mathrm{Var}_\theta \overline{X}_n \\ &= \frac{n}{n-1} \sigma^2(\theta) - \frac{1}{n-1} \sigma^2(\theta) \\ &= \sigma^2(\theta) \end{aligned}$$

für alle $\theta \in \Theta$, was zu beweisen war. $\diamondsuit$

Schränken wir uns bei der Suche nach Schätzern für $\gamma(\theta)$ auf die Klasse der erwartungstreuen ein, brauchen wir ein Kriterium, um diese vergleichen zu können. Dazu definieren wir für jeden solchen Schätzer g dessen sogenannte *Risikofunktion*

$$R(\theta, g) \ = \ E_\theta (g(X) - \gamma(\theta))^2, \tag{20.2}$$

die offenbar die mittlere quadratische Abweichung des Schätzers unter jedem P_θ vom zu schätzenden Wert $\gamma(\theta)$ angibt. Aufgrund der Erwartungstreue erhalten wir sofort:

20.4. Lemma. *Für jeden erwartungstreuen Schätzer gilt*

$$R(\theta, g) \ = \ Var_\theta g(X)$$

für alle $\theta \in \Theta$.

Beweis: Da $E_\theta g(X) = \gamma(\theta)$ für alle $\theta \in \Theta$, folgt

$$R(\theta, g) \;=\; E_\theta(g(X) - E_\theta g(X))^2 \;=\; \mathrm{Var}_\theta g(X). \qquad \diamondsuit$$

20.5. Definition. Ein erwartungstreuer Schätzer g^* (für $\gamma(\theta)$) heißt *gleichmäßig bester erwartungstreuer Schätzer (GBES)*, wenn er unter allen erwartungstreuen Schätzern die gleichmäßig kleinste Varianz besitzt, d.h., wenn

$$\mathrm{Var}_\theta g^*(X) \;\le\; \mathrm{Var}_\theta g(X)$$

für alle $\theta \in \Theta$ und alle erwartungstreuen Schätzer g gilt.

Im folgenden befassen wir uns mit der Frage nach der Existenz und Eindeutigkeit solcher Schätzer, wobei sich die Eindeutigkeit relativ leicht mittels des folgenden Resultats ergibt. Ein Schätzer $h : \mathbb{R}^n \to \mathbb{R}$ heißt *quadratisch integrierbar*, falls $\mathrm{Var}_\theta h(X) < \infty$ für alle $\theta \in \Theta$, und *erwartungstreuer Nullschätzer*, falls $E_\theta h(X) = 0$ für alle $\theta \in \Theta$ gilt.

20.6. Satz (Kovarianzmethode von Rao). *Es sei g^* ein quadratisch integrierbarer, für $\gamma(\theta)$ erwartungstreuer Schätzer. Genau dann ist g^* ein gleichmäßig bester erwartungstreuer Schätzer für $\gamma(\theta)$, wenn*

$$Cov_\theta(g^*(X), h(X)) \;=\; 0 \tag{20.3}$$

für alle $\theta \in \Theta$ und alle quadratisch integrierbaren erwartungstreuen Nullschätzer h gilt.

Beweis: "$\Rightarrow$" Sei h irgendein quadratisch integrierbarer erwartungstreuer Nullschätzer. Dann ist $g_\alpha \overset{\text{def}}{=} g^* + \alpha h$ für jedes $\alpha \in \mathbb{R}$ erwartungstreu für $\gamma(\theta)$, denn

$$E_\theta g_\alpha(X) \;=\; E_\theta g^*(X) + \alpha E_\theta h(X) \;=\; \gamma(\theta)$$

für alle θ. Außerdem hat $g_\alpha(X)$ unter jedem P_θ endliche Varianz, da dies für $g^*(X)$ und $h(X)$ der Fall ist. Unter Benutzung der Optimalität von g^* und der Varianzformel (7.7) erhalten wir nun

$$\begin{aligned} 0 \;&\le\; \mathrm{Var}_\theta g_\alpha(X) - \mathrm{Var}_\theta g^*(X) \\ &=\; \mathrm{Var}_\theta g^*(X) + 2\alpha \mathrm{Cov}_\theta(g^*(X), h(X)) + \alpha^2 \mathrm{Var}_\theta h(X) - \mathrm{Var}_\theta g^*(X) \\ &=\; 2\alpha E_\theta g^*(X)h(X) + \alpha^2 \mathrm{Var}_\theta h(X) \end{aligned}$$

für alle $\theta \in \Theta$ und *alle* $\alpha > 0$. Dies ist aber nur im Fall $E_\theta g^*(X)h(X) = 0$ möglich, weil andernfalls die rechte Seite für ein betragsmäßig hinreichend kleines, je nach Vorzeichen von $E_\theta g^*(X)h(X)$ links oder rechts von 0 liegendes α negativ ist.

"$\Leftarrow$" Gegeben einen beliebigen erwartungstreuen Schätzer g für $\gamma(\theta)$ mit $\mathrm{Var}_\theta g(X) < \infty$ für alle θ, benutzen wir für diese Schlußrichtung, daß $h \overset{\text{def}}{=} g - g^*$ ein quadratisch integrierbarer

Nullschätzer ist. Mit Hilfe von (20.3) ergibt sich deshalb

$$\begin{aligned} \mathrm{Var}_\theta g(X) &= \mathrm{Var}_\theta(g^*(X) + h(X)) \\ &= \mathrm{Var}_\theta g^*(X) + \mathrm{Var}_\theta h(X) \;\geq\; \mathrm{Var}_\theta g^*(X) \end{aligned}$$

und folglich die Optimalität von g^*. ♢

20.7. Eindeutigkeitssatz. *Es seien g^* und g_* zwei quadratisch integrierbare, gleichmäßig beste erwartungstreue Schätzer für $\gamma(\theta)$, d.h.*

$$Var_\theta g^*(X) \;=\; Var_\theta g_*(X) \tag{20.4}$$

für alle $\theta \in \Theta$. Dann gilt

$$P_\theta(g_*(X) = g^*(X)) \;=\; 1 \tag{20.5}$$

für alle $\theta \in \Theta$.

Beweis: Mittels des vorherigen Satzes in Kombination mit erhalten wir

$$\mathrm{Var}_\theta g^*(X) \;=\; \mathrm{Var}_\theta g_*(X) + \mathrm{Var}_\theta(g^*(X) - g_*(X)) \;=\; \mathrm{Var}_\theta g_*(X)$$

also $\mathrm{Var}_\theta(g^*(X) - g_*(X)) = 0$ für alle $\theta \in \Theta$, was insbesondere (20.5) impliziert, da $E_\theta(g^*(X) - g_*(X)) = 0$ für alle $\theta \in \Theta$. ♢