Kapitel 2

Statistische Experimente, statistische Modelle

2.1 Definitionen

In diesem Kapitel führen wir einige Begriffe ein, und zwar in einer solchen Allgemeinheit, daß sie auch für stochastische Prozesse einsetzbar sind.

Definition 2.1. Es seinen $(\Omega, \mathscr{A})$ ein meßbarer Raum und $\mathscr{P}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$. Dann nennen wir $(\Omega, \mathscr{A}, \mathscr{P})$ ein *statistisches Modell.*
Weiterhin sei X eine Zufallsgröße auf $(\Omega, \mathscr{A})$ mit Werten in einem meßbaren Raum $(E, \mathscr{E})$. Dann heißt $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein *statistisches Experiment* und X wird als eine *mathematische Stichprobe* bezeichnet.

Interpretation: Ein zufälliges Experiment wird gemäß $(\Omega, \mathscr{A}, \mathbb{P})$ mit einem bestimmten $\mathbb{P} \in \mathscr{P}$, das aber unbekannt ist, ausgeführt („wahres $\mathbb{P}$“). Dabei wird die Zufallsgröße X beobachtet, ihre Realisierungen x gehören zu E. x heißt *konkrete Stichprobe*, $(E, \mathscr{E})$ nennt man *Stichprobenraum.*

Durch $\mathbb{P}^X(B) = \mathbb{P}(X \in B),\ B \in \mathscr{E}$, ist ein Wahrscheinlichkeitsmaß $\mathbb{P}^X$ auf $\mathscr{E}$ definiert, die Wahrscheinlichkeitsverteilung von X unter $\mathbb{P}$ (*Stichprobenverteilung*).
Wir setzen $\mathscr{P}^X := \{\mathbb{P}^X : \mathbb{P} \in \mathscr{P}\}$.
$(E, \mathscr{E}, \mathscr{P}^{\mathscr{X}})$ ist ebenfalls ein statistisches Modell.

Definition 2.2. $\mathscr{P}^X$ heißt die *zum statistischen Experiment* $(\Omega, \mathscr{A}, \mathscr{P}, X)$ *gehörende Familie von Stichprobenverteilungen*

In dem genannten Experiment wird nur x beobachtet, $(\Omega, \mathscr{A}, \mathscr{P})$ und X sind Hilfskonstruktionen. $\mathscr{P}^X$ ist eine bekannte Familie, unter welcher wahren Verteilung $\mathbb{P}^X \in \mathscr{P}$ die Stichprobe realisiert wurde, ist unbekannt.
Ziel ist es, aus der Kenntnis von x Schlüsse auf das „wahre“ $\mathbb{P}^X$ zu ziehen und die Güte dieser Rückschlüsse zu bewerten. Häufig indiziert man $\mathscr{P}$ der besseren Handhabung wegen, das heißt, man setzt $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und entsprechend $\mathscr{P}^X = (\mathbb{P}_\vartheta^X,\ \vartheta \in \Theta)$. Θ heißt *Parametermenge* oder *Parameterraum*.

Gilt $\Theta \subseteq \mathbb{R}^k$ für ein $k \geq 1$, so nennt man $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein *parametrisches statistisches Modell* (*k-parametrisches Modell*). Läßt sich Θ dagegen nicht durch endlich viele Parameter beschreiben, so spricht man von einem *nichtparametrischen Modell*.
Im Fall $E = \mathbb{R}^n, \mathscr{E} = \mathscr{B}^n$ ist X ein zufälliger Vektor, x eine seiner Realisierungen. Wie bereits erwähnt, bezeichnet man in diesem Fall X als eine (mathematische) Stichprobe und x im Unterschied dazu als eine konkrete Stichprobe.

Im Fall, daß E ein Funktionenraum ist, bildet $X = (X_t,\ t \in T)$ einen stochastischen Prozess und x eine Realisierung desselben. In diesem Fall ist der Begriff „Stichprobe“ weniger gebräuchlich. Man spricht von *Trajektorien* oder *Pfaden*. Wir werden in dieser Vorlesung in jedem der genannten Fälle X bzw. x als mathematische bzw. konkrete Stichprobe bezeichnen.

2.2 Klassische Statistische Experimente

Die Wahrscheinlichkeitsverteilungen $\mathbb{P}_\vartheta^X,\ \vartheta \in \Theta$, der Stichprobe X bilden eine Ausgangsbasis der mathematischen Statistik.
Auf ihrer Grundlage werden Schätzer für das unbekannte ϑ oder Tests für Hypothesen über ϑ konstruiert und untersucht. Für ihre Beschreibung bedient man sich der sogenannten Likelihoodfunktion, die wir vorerst in zwei Beispielen definieren.

Es sei $(C, \mathscr{C})$ ein meßbarer Raum. Wir setzen $E = C^n, \mathscr{E} = \mathscr{C}^{\otimes n}$ und $X = (X_1, X_2, \ldots, X_n)$, wobei $X_k : (\Omega, \mathscr{A}) \longrightarrow (C, \mathscr{C})$ für jedes $k = 1, \ldots, n$ eine Zufallsgröße ist.
Es sei weiterhin $\mathscr{P}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$, $\mathscr{P} = \{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$. Unter folgenden Voraussetzungen

Voraussetzung 1 Für jedes $\mathbb{P}_\vartheta \in \mathscr{P}$ sind die Zufallsgrößen $X_1, \ldots, X_n$ identisch verteilt, das heißt,

$$\mathbb{P}_\vartheta(X_k \in B) = \mathbb{P}_\vartheta(X_1 \in B), \quad B \in \mathscr{C}, \quad k = 1, \ldots, n.$$

Voraussetzung 2 Für jedes $\mathbb{P}_\vartheta \in \mathscr{P}$ sind die $X_1, \ldots, X_n$ unter $\mathbb{P}_\vartheta$ in ihrer Gesamtheit unabhängig, d.h.

$$\mathbb{P}_\vartheta(X \in B_1 \times \ldots \times B_n) = \prod_{i=1}^{n} \mathbb{P}_\vartheta(X_i \in B_i).$$

gilt dann

$$\mathbb{P}_\vartheta^X(B_1 \times \ldots \times B_n) = \prod_{i=1}^{n} \mathbb{P}_\vartheta(X_1 \in B_i) = \prod_{i=1}^{n} \mathbb{P}_\vartheta^{X_1}(B_i) \tag{2.1}$$

Durch $\mathbb{P}_\vartheta^X$ ist ein Wahrscheinlichkeitsmaß auf $\mathscr{C}^{\otimes n}$ definiert, das wir ebenfalls mit $\mathbb{P}_\vartheta^X$ bezeichnen.

Beispiel 2.1 (diskreter Fall). Es seien $C = \{c_1, c_2, \ldots\}$, $\mathscr{C} = \mathscr{P}(C)$ die Potenzmenge von C, $E = C^n$ und $\mathscr{E} = \mathscr{C}^{\otimes n}$. X_1 nehme nur Werte aus C an, d.h. X_1 habe eine diskrete Verteilung mit

$$\mathbb{P}_\vartheta(X_1 = c_k) = p_k(\vartheta), \quad p_k(\vartheta) \geq 0, \quad k = 1, 2, \ldots, \quad \sum_{k=1}^{\infty} p_k(\vartheta) = 1.$$

Dann hat auch $X = (X_1, X_2, \ldots, X_n)$ eine diskrete Verteilung und es gilt:

$$\mathbb{P}_\vartheta^X(\,(c_{i_1}, \ldots, c_{i_n})\,) = \prod_{k=1}^{n} p_\vartheta(c_{i_k}) = \mathbb{P}_\vartheta(X_1 = c_{i_1}, \ldots, X_n = c_{i_n}) =: L_n(x, \vartheta), \quad x = (c_{i_1}, \ldots, c_{i_n})$$

Diese Funktion L_n heißt *Likelihoodfunktion* des statistischen Experiments $(\Omega, \mathscr{A}, \mathscr{P}, X)$.

Beispiel 2.2 (stetiger Fall). Es seien $C = \mathbb{R}$, $E = \mathbb{R}^n$ und $\mathscr{E} = \mathscr{B}(\mathbb{R}^n)$. X_1 besitze eine Dichte $f_\vartheta(x)$, d. h. es gelte

$$\mathbb{P}_\vartheta^X(D) = \int_D f_\vartheta(s)\, ds, \quad D \in \mathscr{B}.$$

Dann besitzt auch $X = (X_1, \ldots, X_n)$ eine Dichte $f_\vartheta^X(x_1, \ldots, x_n) = \prod_{i=1}^n f_\vartheta(x_i)$ und es gilt

$$\mathbb{P}_\vartheta^X(B_1 \times \ldots \times B_n) = \int_{B_1} \cdots \int_{B_n} f_\vartheta(x_1) \cdot \ldots \cdot f_\vartheta(x_n)\, dx_1 \ldots dx_n.$$

In diesem Fall bezeichnet man $L_n(x; \vartheta) := \prod_{k=1}^n f_\vartheta(x_k)$ als *Likelihoodfunktion* des statistischen Experiments $(\Omega, \mathscr{A}, \mathscr{P}, X)$.

Interpretation: Die Stichprobe $X = (X_1, \dots, X_n)$ modelliert n voneinander unabhängige, unter gleichartigen Bedingungen ausgeführte zufällige Experimente, bei denen jeweils $X_1, X_2, \dots, X_n$ beobachtet wird. Wir definieren $Q_\vartheta := \mathbb{P}_\vartheta^{X_1}$.
Man sagt, X sei eine *mathematische Stichprobe aus einer nach Q_ϑ verteilten Grundgesamtheit.*
Jede ihrer Realisierungen x nennt man eine *konkrete Stichprobe aus einer nach Q_ϑ verteilten Grundgesamtheit.*
Bezeichnung: Klassisches statistisches Experiment.

Wir kehren zurück zu allgemeinen statistischen Experimenten.

Definition 2.3. Es seien $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $(E, \mathscr{E})$ und H eine meßbare Abbildung von $(E, \mathscr{E})$ in einen meßbaren Raum $(F, \mathscr{F})$. H heißt eine *Stichprobenfunktion*. Insbesondere ist $(\Omega, \mathscr{A}, \mathscr{P}, H \circ X)$ ein statistisches Experiment mit dem Stichprobenraum $(F, \mathscr{F})$.

Im allgemeinen geht bei dieser Abbildung H Information verloren (Datenreduktion), andererseits kann $H(x)$ einfacher und übersichtlicher sein als x.

Setzt man in H die Zufallsgröße X ein, d. h. $H(X) = H\big((X_1, \dots, X_n)\big)$, so erhält man eine neue Zufallsgröße, sie hat die Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta^H(B) = \mathbb{P}_\vartheta(H(X) \in B), \quad B \in \mathscr{F}$. Diese Wahrscheinlichkeitsverteilung wird unter anderem zum Studium der Eigenschaften der Stichprobe x in ihrem Verhältnis zum wahren Parameter ϑ herangezogen. Die Berechnung der Verteilungen von Stichprobenfunktionen $H(X)$ gehört zu den wesentlichen Aufgaben der mathematischen Statistik.

Anstelle von Stichprobenfunktion verwendet man auch einfach die Bezeichnung *Statistik*.

Wir nehmen an, daß $\mathscr{P}$ die Form $\mathscr{P} = \{\mathbb{P}_\vartheta,\ \vartheta \in \Theta\}$ mit irgendeiner Parametermenge Θ hat. Eine Grundaufgabe der Statistik ist es, von der Beobachtung x auf den Parameter ϑ bzw. eine Funktion $\gamma(\vartheta)$ zu schließen. Häufig möchte man ϑ bzw. $\gamma(\vartheta)$ mit möglichst großer Genauigkeit bestimmen, man sagt „schätzen“.

Definition 2.4. Es seien γ eine Abbildung von Θ in eine Menge Γ, $\mathscr{A}_\Gamma$ eine σ-Algebra von Teilmengen von Γ und g eine meßbare Abbildung von $(E, \mathscr{E})$ in $(\Gamma, \mathscr{A}_\Gamma)$.
Dann heißt g eine *Schätzfunktion*, $g(X)$ ein *Schätzer* und $g(x)$ ein *Schätzwert* für $\gamma(\vartheta)$.

Jeder Schätzer ist also auch eine Stichprobenfunktion.

2.3 Ein Beispiel aus der klassischen Mathematischen Statistik

Es sei X_0 eine reellwertige Zufallsgröße mit unbekannter Verteilungsfunktion F. Zu schätzen sei der Erwartungswert

$$E(X_0) = \int_{\mathbb{R}} x\, dF(x) =: m_F.$$

Da man über F keine Vorinformation hat, setzt man

$$\Theta = \{F \,:\, F \text{ Verteilungsfunktion auf } (\mathbb{R}, \mathscr{B}) \text{ mit } |m_F| < \infty\}$$

Die Problemstellung legt nahe $\gamma(F) = m_F$ zu setzen.

Vorausgesetzt werde ferner, daß eine n-elementige Stichprobe $x = (x_1, x_2, \ldots, x_n)$ vorliegt, die aus n voneinander unabhängigen unter gleichartigen Bedingungen durchgeführten Versuchen gewonnen wurden. Dabei wird beim k-ten Versuch, $k = 1, \ldots, n$ registriert, welchen Wert die Zufallsgröße X_0 annimmt. Intuitiv verwenden wir als Schätzwert $g(x)$ für $\gamma(F) = m_F$ den Wert

$$g(x) = \overline{x}_n = \frac{1}{n} \sum_{k=1}^{n} x_k.$$

Die eben getroffenen Voraussetzungen legen es nahe, eine mathematische Stichprobe $X = (X_1, X_2, \ldots, X_n)$ zu betrachten, die aus n voneinander unabhängigen und identisch wie X_0 verteilten Zufallsgrößen besteht.
Als Schätzer für m_F ergibt sich

$$g(X) = \overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Zur Illustration typischer Aussagen der Mathematischen Statistik stellen wir eine Reihe von Eigenschaften dieses Schätzers zusammen.

Aussage: Es gelten folgende Eigenschaften

a) $E_F(\overline{X}_n) = m_F$, man sagt, $\overline{X}_n$ ist ein *erwartungstreuer Schätzer*

b) $D_F^2\,\overline{X}_n = E_F\big((\overline{X}_n - m_F)^2\big) = \frac{D_F^2 X_0}{n}$, falls $\sigma^2 := D_F^2 X_0 < \infty$.
Insbesondere gilt

$$\mathbb{P}_F(|\overline{X}_n - m_F| \geq a) \leq \frac{\sigma_F^2}{na^2} \qquad a > 0, \quad n \geq 1$$

(Schwaches Gesetz der großen Zahlen). Man sagt, daß der Schätzer $\overline{X}_n$ *konsistent* ist.

c) $\lim_{n\to\infty} \overline{X}_n = m_F \quad \mathbb{P}_F$-fast sicher (Starkes Gesetz der großen Zahlen)

d) Angenommen,

$$I_F := \{t \in \mathbb{R} \,:\, E_F(e^{tX}) < \infty\}$$

ist eine Umgebung von 0. h_F sei die Cramertransformierte von F (sieh Übungen), ϵ sei irgendeine positive Zahl. Dann ist

$$P_F(\overline{X}_n \geq m_F+\epsilon) \leq \exp\{-nh_F(m_F+\epsilon)\} \text{ und } P_F(\overline{X}_n \leq m_F-\epsilon) \leq \exp\{-nh_F(m_F-\epsilon)\},$$

das bedeutet insbesondere, daß $P_F(\,|\overline{X}_n - m_F| \geq \epsilon)$ exponentiell schnell gegen 0 konvergiert.

e) Der Zentrale Grenzwertsatz besagt, daß

$$P_F(\,|\overline{X}_n - m_F| > \epsilon) = P_F\big(\,|\overline{X}_n - m_F|\frac{\sqrt{n}}{\sigma} > \epsilon\frac{\sqrt{n}}{\sigma}\big) \approx 2\Big(1 - \Phi(\epsilon\,\sigma_F^2\sqrt{n})\Big).$$

Beide Methoden d) und e) führen zur Abschätzung der Genauigkeit der Approximation von m_F durch $\overline{X}_n$.

Die erste Methode liefert:

> Die Wahrscheinlichkeit, daß man sich irrt, wenn man sagt, m_F befinde sich in $(-\infty, \overline{X}_n+\epsilon)$ und in $(\overline{X}_n-\epsilon, \infty)$ konvergiert mit wachsendem n exponentiell schnell gegen Null.

Die zweite Methode liefert:

> Es wird ein $\alpha \in (0,1)$ fixiert und man erhält für große n Vertrauens-Intervalle
>
> $$\Big(-\infty, \overline{X}_n + \frac{\sigma_F}{\sqrt{n}}q_{1-\alpha}\Big],\ \Big[\overline{X}_n - \frac{\sigma_F}{\sqrt{n}}q_{1-\alpha}, \infty\Big),\ \Big[\overline{X}_n - \frac{\sigma_F}{\sqrt{n}}q_{1-\frac{\alpha}{2}}, \overline{X}_n + \frac{\sigma_F}{\sqrt{n}}q_{1-\frac{\alpha}{2}}\Big)$$
>
> des Niveaus $1-\alpha$, von denen man sagen kann, daß m_F mit einer Irrtumswahrscheinlichkeit nahe bei α in denen liegt.

2.4 Empirische Schätzer

Klassischer Fall:
Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $E = \mathbb{R}^n$, $\mathscr{E} = \mathscr{B}^n$ und $X = (X_1, X_2, \ldots, X_n)$ bestehe aus reellwertigen, unabhängigen und identisch verteilten Zufallsgrößen $X_k,\ k = 1, \ldots, n$ mit Verteilungsfunktion F. Die Familie $\mathscr{P}$ sei parametrisiert: $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta),\ \Theta \subseteq \mathbb{R}^k$.

Empirische Verteilungsfunktion

Es seien

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_n \leq x\}}, \qquad m_l(F) = \int_{\mathbb{R}} x^l \, dF(x), \quad l \in \mathbb{N}$$

Diese Verteilungsfunktion $\hat{F}_n(\,\cdot\,)$ gehört zur gleichmäßigen Verteilung auf $(X_1, X_2, \ldots, X_n)$. Sie hat $m_l(\hat{F}_n) = \frac{1}{n} \sum_{k=1}^{n} X_k^l$ als l-tes Moment.

Aussage 2.1 (Hauptsatz der mathematischen Statistik). *Es sei $(X_n,\, n \geq 1)$ eine Folge unabhängiger, identisch nach F verteilter Zufallsgrößen mit Werten in $\mathbb{R}^k$.*
Dann konvergiert die Folge $\big(\hat{F}_n(\,\cdot\,, \omega)\big)$ schwach gegen $F(\,\cdot\,)$ (d.h. $\hat{F}_n(f) \longrightarrow F(f), \forall f \in \mathbb{C}_b$). Ist $k = 1$, so erfolgt die Konvergenz $\mathbb{P}$-f.s. gleichmäßig, d.h.

$$\sup_{x \in \mathbb{R}} \mid \hat{F}_n(x, \omega) - F(x) \mid \longrightarrow 0 \qquad \textit{für } \mathbb{P}\textit{-f.a. } \omega \in \Omega.$$

Beweis:

s. Dacunha-Castelle, Duflo I, 4.4 □

Eine Konstruktionsmethode für Schätzer: Empirische Schätzer

Ist $\gamma(F)$ die zu schätzende Größe, so verwendet man $\gamma(\hat{F}_n)$ als Schätzer (sofern γ auf den Treppenfunktionen definiert ist, bzw. Sinn macht).

a) Momentenmethode
Zu schätzen sind der Erwartungswert $\mu = m_1(F) = \int_{\mathbb{R}} x \, dF(x)$ und die Streuung $\sigma_F^2 = m_2(F) - \big(m_1(F)\big)^2$.

Wir wenden die Abbildungen $F \longrightarrow m_1(F)$ und $F \longrightarrow \sigma_F^2$ auf $\hat{F}_n$ an und erhalten die „Momentenschätzer"

$$\hat{\mu} = m_1(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} X_k = \bar{X}_n, \qquad \text{(erwartungstreu)}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big)^2 = \frac{1}{n}\sum_{k=1}^{n}\big(X_k - \bar{X}_k\big)^2.$$

Allgemeiner: Man berechne die Momente $m_l(F_\vartheta) = \mathbb{E}_\vartheta(X_1^l)$, ersetze $m_l(F_\vartheta)$ durch $m_l(\hat{F}_n)$ und löse die Gleichungen nach ϑ (bzw. nach $\gamma(\vartheta)$) auf. Im Ergebnis erhält man einen Schätzer $\hat{\vartheta}_n$ bzw. $\hat{\gamma}_n$ für ϑ bzw. $\gamma(\vartheta)$, einen sogenannten Momentenschätzer.

Beispiel 2.3. Es sei F_λ die Verteilungsfunktion der Exponentialverteilung mit Parameter $\lambda > 0$.
In diesem Fall gilt:

$$m_1(F) = \lambda \int_0^\infty x e^{-\lambda x}\, dx = \frac{1}{\lambda}, \quad \bar{X}_n = \widehat{\left(\frac{1}{\lambda}\right)}, \quad \text{wobei } \gamma(F_\lambda) = \frac{1}{\lambda}\,.$$

$\bar{X}_n$ ist ein erwartungstreuer Schätzer für $\gamma(\lambda) = \frac{1}{\lambda}$. Ein möglicher Schätzer für λ wäre zum Beispiel $\dfrac{1}{\bar{X}_n}$. Es wird sich aber herausstellen, daß dieser Schätzer nicht erwartungstreu ist.

Wir kehren zur Schätzung von σ_F^2 zurück:

$$\hat{\sigma}_F^2 := \frac{1}{n}\sum_{k=1}^{n}\big(X_k - \bar{X}_k\big)^2 \qquad \text{ist die Momentenschätzung für } \sigma_F^2.$$

Es gilt:

$$\begin{aligned}
E_F(\hat{\sigma}_F^2) &= \frac{1}{n} E_F\left(\left(\sum_{k=1}^{n} X_k^2\right) - n\bar{X}_n^2\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} E_F X_k^2 - E_F \bar{X}_n^2 \\
&= \frac{1}{n}\big(\big(\sigma_F^2 + \mu_F^2\big)\, n\big) - \frac{\sigma_F^2}{n} - \mu^2 \\
&= \sigma_F^2 + \mu_F^2 - \frac{\sigma_F^2}{n} - \mu^2 \\
&= \sigma_F^2 \left(\frac{n-1}{n}\right)
\end{aligned}$$

Also ist $\hat{\sigma}_F^2$ nicht erwartungstreu, man unterschätzt σ_F^2 regelmäßig. Aber

$$S_n^2 = \frac{1}{n-1}\hat{\sigma}_F^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}_n\right)^2$$

ist eine erwartungstreue Schätzung für σ_F^2.

Für das Beispiel 2.3 gilt dann:
S_n^2 ist eine erwartungstreue Schätzung für $\frac{1}{\lambda^2}$, $\sqrt{S_n^2}$ ist eine Schätzung für $\frac{1}{\lambda}$.

Beispiel 2.4. Es sei $X = (X_!, X_2, \ldots, X_n)$ eine klassische mathematische Stichprobe aus einer Grundgesamtheit, die eine gemischte Poissonverteilung besitzt:

$$\mathbb{P}_\vartheta(X_1 = k) = a \cdot \frac{\lambda_1^k}{k!}e^{-\lambda_1} + (1-a)\,\frac{\lambda_2^k}{k!}e^{-\lambda_2}, \qquad k \geq 0$$

mit $\vartheta = (a, \lambda_1, \lambda_2)\,, \quad a \in (0,1)\,, \lambda_1, \lambda_2 > 0$. Die entsprechende Verteilungsfunktion werde mit F_ϑ bezeichnet.
Für die momenterzeugende Funktion

$$\varphi_\vartheta(s) := \mathbb{E}_\vartheta\left(s^{X_1}\right) = \sum_{k\geq 0}\mathbb{P}_\vartheta(X_1 = k)s^k = ae^{\lambda_1(s-1)} + (1-a)e^{\lambda_2(s-1)}, \quad s \in [0,1]$$

gilt:

$$\begin{aligned}
T_1(F_\vartheta) &:= \varphi_\vartheta'(1) = \mathbb{E}_\vartheta\, X_1 = a\lambda_1 + (1-a)\lambda_2\\
T_2(F_\vartheta) &:= \varphi_\vartheta''(1) = \mathbb{E}_\vartheta\, X(X-1) = a\lambda_1^2 + (1-a)\lambda_2^2\\
T_3(F_\vartheta) &:= \varphi_\vartheta'''(1) = \mathbb{E}_\vartheta\, X(X-1)(X-2) = a\lambda_1^3 + (1-a)\lambda_2^3
\end{aligned}$$

Wir definieren die entsprechenden empirischen Momente

$$\begin{aligned}
T_1(\hat{F}_n) =& \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad T_2(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} X_k(X_k - 1),\\
T_3(\hat{F}_n) =& \frac{1}{n}\sum_{k=1}^{n} X_k(X_k-1)(X_k-2).
\end{aligned}$$

Ist $x = (x_1, x_2, \ldots, x_n)$ eine konkrete Stichprobe aus der nach F_ϑ verteilten Grundgesamtheit, so erhält man folgende Gleichungen, aus denen sich die empirischen Schätzer

$\hat{a}, \hat{\lambda}_1, \hat{\lambda}_2$ für $\vartheta = (a, \lambda_1, \lambda_2)$ berechnen lassen:

$$\begin{aligned}
\hat{a}\hat{\lambda}_1 + (1-\hat{a})\,\hat{\lambda}_2 &= T_1(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} x_k \\
\hat{a}\hat{\lambda}_1^2 + (1-\hat{a})\,\hat{\lambda}_2^2 &= T_2(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} x_k(x_k-1) \\
\hat{a}\hat{\lambda}_1^3 + (1-\hat{a})\,\hat{\lambda}_2^3 &= T_3(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} x_k(x_k-1)(x_k-2).
\end{aligned}$$

b) Schätzung der Schranken des Trägers von F:

$$\begin{aligned}
m &= \sup\{a \in \mathbb{R} \,:\, F(a) = 0\}, & M &= \inf\{a \in \mathbb{R} \,:\, F(a) = 1\} \\
\hat{m}_n &= \min\{X_k,\ k = 1, \ldots, n\}, & \hat{M}_n &= \max\{X_k,\ k = 1, \ldots, n\}
\end{aligned}$$

2.5 Eigenschaften von Schätzern

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $(E, \mathscr{E})$ und es sei $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$. Weiterhin sei γ wie oben eine meßbare Funktion von Θ in Γ und $g(X)$ ein Schätzer für $\gamma(\vartheta)$.
Der Einfachheit halber nehmen wir an, $\Gamma \subseteq \mathbb{R}$. Zur Beurteilung der Güte des Schätzers $g(X)$ definieren wir die *Risikofunktion*

$$R(g,\, \vartheta) := \mathbb{E}_\vartheta \left(g(X) - \gamma(\vartheta)\right)^2 \quad \vartheta \in \Theta$$

$R(g,\, \vartheta)$ ist also die mittlere quadratische Abweichung des Schätzers $g(X)$ von dem zu schätzenden Wert $g(\vartheta)$, wenn $\mathbb{P}_\vartheta$ die zugrundeliegende Wahrscheinlichkeitsverteilung ist.
Man wähle die Schätzfunktion $g(\cdot)$ so, daß $R(g,\, \vartheta)$ möglichst minimal wird.
Unter der Voraussetzung, daß $R(g,\, \vartheta) < \infty, \quad \forall\, \vartheta \in \Theta$ definiere:

Definition 2.5. Ein Schätzer $h(X)$ für die Funktion $\gamma(\vartheta)$ heißt *besser als* $g(X)$, falls gilt:

$$R(h,\, \vartheta) \leq R(g,\, \vartheta) \text{ für alle } \vartheta \in \Theta \text{ und } R(h,\, \vartheta) < R(g,\, \vartheta) \text{ für mindestens ein } \vartheta \in \Theta.$$

Wenn es zu einem gegebenen Schätzer $g(X)$ für $\gamma(\vartheta)$ einen besseren Schätzer $h(X)$ für $\gamma(\vartheta)$ gibt, so nennt man $g(X)$ *nicht zulässig* (als Schätzer für $\gamma(\vartheta)$).
$g(X)$ heißt *zulässiger Schätzer* für $\gamma(\vartheta)$, falls es keinen besseren Schätzer für $\gamma(\vartheta)$ gibt.

Es ist vernünftig, sich auf zulässige Schätzer zu beschränken.

Definition 2.6. Ein Schätzer $g^*(X)$ für $\gamma(\vartheta)$ heißt *optimal*, falls

$$R(g^*,\, \vartheta) = \inf_g R(g, \vartheta) \qquad \text{für alle } \vartheta \in \Theta,$$

wobei das Infimum über alle (zulässigen) Schätzer $g(X)$ für $\gamma(\vartheta)$ gebildet wird.

Im allgemeinen gibt es keinen optimalen Schätzer für $\gamma(\vartheta)$!

Begründung: Für jedes fest gewählte $\vartheta_0 \in \Theta$ ist $\inf_g R(g, \vartheta_0) = 0$, da der Schätzer $g(X) \equiv \gamma(\vartheta_0)$ unter allen konkurrierenden Schätzern vorkommt.
Dieser Schätzer ist sehr gut, wenn $\vartheta = \vartheta_0$ der wahre Parameter ist, aber für andere ϑ allerdings nicht.

Wir verfolgen unser Ziel, eine vernünftige Schätzfunktion g zu finden, die das mittlere quadratische Risiko möglichst klein hält, durch folgende Überlegungen:
Es gilt

$$\begin{array}{rccc} R(g,\, \vartheta) = & \mathbb{E}_\vartheta \Big(g(X) - \mathbb{E}_\vartheta \big(g(X)\big)\Big)^2 & + \mathbb{E}_\vartheta \Big(\, \mathbb{E}_\vartheta \big(g(X)\big) - \gamma(\vartheta)\Big)^2 & \\ =: & D^2_\vartheta \, g(X) & + b_{g,\gamma}(\vartheta), \qquad \vartheta \in \Theta & \\ \text{Risikofunktion} = & \text{zufallsbedingte Streuung} & + \text{Verzerrung}^2. & \end{array} \tag{2.2}$$

Die Größe $b_{g,\gamma}(\vartheta)$ heißt *Verzerrung* oder *Bias* des Schätzers $g(X)$ bezüglich $\gamma(\vartheta)$.
Wenn man das Risiko $R(\cdot)$ für alle $\vartheta \in \Theta$ minimieren will, ist es also vernünftig, unter den erwartungstreuen (unverzerrten) Schätzern, d.h. Schätzern mit $\mathbb{E}_\vartheta \big(g(X)\big) = \gamma(\vartheta),\ \vartheta \in \Theta$ zu suchen.
Wir beschränken uns deshalb darauf, unverzerrte Schätzer mit möglichst kleiner Streuung zu suchen.

Angenommen $H(X)$ ist eine Stichprobenfunktion und $g(X)$ ist ein Schätzer für $\gamma(\vartheta)$. Eine ähnliche Rechnung wie in (2.2) führt auf

$$R(g,\, \vartheta) = \mathbb{E}_\vartheta \Big(g(X) - \mathbb{E}_\vartheta \big(g(X) \mid H(X)\big)\Big)^2 + \mathbb{E}_\vartheta \Big(\, \mathbb{E}_\vartheta \big(g(X) \mid H(X)\big) - \gamma(\vartheta)\Big)^2$$

Definition 2.7. Die Stichprobenfunktion $H(X)$ heißt eine *suffiziente* oder *erschöpfende* Statistik, falls die Wahrscheinlichkeitsverteilung $\mathbb{P}^X_\vartheta \big(\cdot \mid H(X)\big)$ nicht von ϑ abhängt.

Wir kommen auf diesen Begriff später ausführlich zurück.

Ist $H(X)$ eine suffiziente Statistik, so ist $g_1(X) := \mathbb{E}_\vartheta\left(g(X) \mid H(X)\right)$ ein neuer Schätzer für $\gamma(\vartheta)$ mit kleinerem Risiko ($g_1(X)$ hängt nicht von ϑ ab):

$$R(g_1,\, \vartheta) = \mathbb{E}_\vartheta\left(g_1(X) - \gamma(\vartheta)\right)^2 \leq R(g,\, \vartheta), \quad \vartheta \in \Theta.$$

Da $g_1(X)$ eine Funktion von $H(X)$ ist, genügt es also, als Schätzer für $\gamma(\vartheta)$ lediglich Funktionen suffizienter Statistiken in Betracht zu ziehen.

Beispiel einer suffizienten Statistik

Ein zufälliger Versuch möge nur die Ausgänge 0 und 1 besitzen, wobei 1 mit der unbekannten Wahrscheinlichkeit $\vartheta \in (0,\, 1) =: \Theta$ als Ergebnis auftritt. Zur Schätzung von ϑ verschafft man sich eine Stichprobe $(i_1, i_2, \ldots, i_n)$ aus n unabhängig und unter gleichartigen Bedingungen durchgeführten Versuchen.
Zur mathematischen Modellierung dieses Sachverhalts führen wir ein:

$$\begin{aligned}
&\Omega = \left\{\omega = (i_1, i_2, \ldots, i_n) \,:\, i_k \in \{0\,, 1\},\ k = 1, 2, \ldots, n\right\}, \quad \mathscr{A} = \wp(\Omega),\\
&X = (X_1, X_2, \ldots, X_n) \text{ mit } X_k(\omega) = i_k, \quad k = 1, \ldots, n \text{ für } \omega = (i_1, i_2, \ldots, i_n) \in \Omega,\\
&S_n = \sum_{k=1}^{n} X_k \text{ und } \mathbb{P}_\vartheta(\{\omega\}) = \vartheta^{\sum_{k=1}^n i_k}(1-\vartheta)^{n-\sum_{k=1}^n i_k}, \quad \omega \in \Omega.
\end{aligned}$$

Bezüglich jedem $\mathbb{P}_\vartheta$ sind $X_1, X_2, \ldots, X_n$ unabhängige und identisch verteilte Zufallsgrößen mit $\mathbb{P}_\vartheta(X_k = i) = \vartheta^i(1-\vartheta)^{1-i}, \quad i \in \{0\,, 1\}$.

Aussage 2.2. $S_n = \sum_{k=1}^n X_k$ *ist eine suffiziente Statistik.*

Beweis:

Es gilt:

$$\mathbb{P}_\vartheta\left(\{\omega\} \mid S_n = m\right) = \begin{cases} \dfrac{1}{\binom{n}{m}}, & \text{falls } \omega = (i_1, i_2, \ldots, i_n) \text{ mit } \sum_{k=1}^n i_k = m \\ 0, & \text{falls } \omega = (i_1, i_2, \ldots, i_n) \text{ mit } \sum_{k=1}^n i_k \neq m \end{cases}$$

□

Zur Schätzung von ϑ beschränken wir uns also auf Funktionen der Form $g(S_n)$.
Soll $g(S_n)$ erwartungstreu sein, so erhalten wir aus der Forderung

$$\mathbb{E}_\vartheta\, g(S_n) = \vartheta, \quad \vartheta \in [0, 1]$$

die Gleichungen

$$\sum_{m=0}^{n} g(m)\binom{n}{m}\vartheta^m(1-\vartheta)^{n-m} = \vartheta, \quad \vartheta \in [0,1],$$

woraus sich durch Koeffizientenvergleich

$$g(m) = \frac{m}{n}, \quad m = 0,1,\ldots,n, \quad \text{d.h. } \hat{\vartheta}_n = \frac{S_n}{n}$$

ergibt. Wir werden später zeigen, daß diese Schätzung die minimale Streuung unter allen erwartungstreuen Schätzungen für ϑ hat.
Es gilt:

$$\mathbb{P}_\vartheta\big(\{\omega\}\big) = \sum_{m=0}^{n} \mathbb{P}\big(\{\omega\} \mid S_n = m\big)\,\mathbb{P}_\vartheta(S_n = m)$$

Die Abhängigkeit des Maßes $\mathbb{P}_\vartheta$ von ϑ auf $\mathscr{A}$ wird also allein durch die Abhängigkeit von $\mathbb{P}_\vartheta^{S_n}$ von ϑ vermittelt.