1
Grundbegriffe der mathematischen Statistik

Elementare statistische Berechnungen werden schon seit Jahrtausenden durchgeführt. Das arithmetische Mittel aus einer Anzahl von Mess- oder Beobachtungswerten ist schon sehr lange bekannt.

Zuerst entstand die beschreibende Statistik mit dem Sammeln von Daten etwa bei Volkszählungen oder in Krankenregistern und deren Verdichtung in Form von Maßzahlen oder Grafiken. Die mathematische Statistik entwickelte sich ab Ende des 19. Jahrhunderts aufbauend auf der Wahrscheinlichkeitsrechnung. Anfang des 20. Jahrhunderts gehörten vor allem Karl Pearson und Sir Ronald Aymler Fisher zu ihren Pionieren. Das Buch von Fisher (1925) ist ein Meilenstein, in ihm werden die vom Autor mehrere Jahre zuvor entwickelten Grundlagen der Statistik wie die Maximum-Likelihood-Methode und die Varianzanalyse oder Begriffe wie Suffizienz und Effizienz Versuchsanstellern nahegebracht. Ein wichtiges Informationsmaß heißt noch heute Fisher-Information (siehe Abschn. 1.4).

Wir wollen auf die Details der historischen Entwicklung nicht eingehen und verweisen Interessierte auf Stigler (2000). Stattdessen beschreiben wir den heutigen Stand der Theorie. Wir wollen aber nicht vergessen, dass viele Anregungen aus Anwendungen kamen und bringen deshalb auch immer wieder Beispiele.

Die Wahrscheinlichkeitsrechnung ist zwar die Grundlage der mathematischen Statistik, aber viele praktische Probleme, in denen Aussagen über Zufallsvariablen gemacht werden sollen, sind mit der Wahrscheinlichkeitsrechnung allein nicht zu lösen. Das liegt daran, dass über die Verteilungsfunktion der Zufallsvariablen nicht alles bekannt ist und das Problem oft darin besteht, Aussagen über mindestens einen der Parameter einer Verteilungsfunktion zu machen oder dass sogar die Verteilungsfunktion gänzlich unbekannt ist. Die mathematische Statistik wird in vielen einführenden Texten als die Theorie der Auswertung von Versuchen oder Erhebungen betrachtet, d. h., man geht davon aus, dass bereits eine Zufallsstichprobe (nach Abschn. 1.1) vorliegt. Wie man auf optimalem Weg zu dieser Zufallsstichprobe gelangt, bleibt meist unberücksichtigt – dies wird gesondert in der statistischen Versuchsplanung abgehandelt. In den Anwendungen ist es klar, dass man erst den Versuch (die Erhebung) plant und dann, wenn der Versuch durchgeführt wurde, mit der Auswertung beginnt. In der Theorie ist es aber zweckmäßig, zunächst die optimale Auswertung zu ermitteln, um dann für diese den optimalen Versuchsplan zu bestimmen, z. B. den kleinsten Versuchsumfang für eine varianz-

Mathematische Statistik, 1. Auflage. Dieter Rasch und Dieter Schott.

optimale Schätzfunktion. Daher wird hier so verfahren, dass zunächst einmal die optimale Auswertung bestimmt wird und später für diese der Versuchsplan zu erarbeiten ist. Eine Ausnahme bilden dabei die sequentiellen Verfahren, bei denen Planung und Auswertung gemeinsam vorgenommen werden.

Wir müssen uns darüber im Klaren sein, dass es sich bei der Behandlung der mathematischen Statistik einerseits und bei ihrer Anwendung auf konkretes Datenmaterial andererseits um zwei völlig verschiedene Begriffssysteme handelt. In beiden treten oft die gleichen Termini auf, die es genau auseinanderzuhalten gilt. Wir sprechen davon, dass den Begriffen der empirischen Ebene (also denen der Realwelt) Modelle in der Theorie zugeordnet werden.

1.1 Grundgesamtheit und Stichprobe

1.1.1 Konkrete Stichproben und Grundgesamtheiten

In den empirischen Wissenschaften werden ein Merkmal oder auch mehrere Merkmale gleichzeitig (ein Merkmalsvektor) an bestimmten Objekten (oder Individuen) beobachtet. Aus den Beobachtungswerten sind Schlüsse auf die Gesamtheit der Merkmalswerte aller Objekte einer Gesamtheit zu ziehen. Ursache dafür ist, dass es sachliche oder ökonomische Gesichtspunkte gibt, die eine vollständige Erfassung der Merkmale aller Objekte nicht ermöglichen. Hierzu einige Beispiele:

- Die Kosten der Erfassung aller Merkmalswerte stehen in keinem Verhältnis zum Wert der Aussage (z. B. Messung der Körpergröße aller zurzeit lebenden Menschen über 18 Jahren).
- Die Erfassung der Merkmalswerte ist mit der Zerstörung der Objekte verbunden (nicht zerstörungsfreie Werkstoffprüfung wie Reißfestigkeit von Tauen oder Strümpfen).
- Die Gesamtheit der Objekte ist hypothetischer Natur, z. B. weil sie teilweise zum Untersuchungszeitpunkt nicht existieren (wie alle Produkte einer Maschine).

Die wenigen praktischen Fälle, in denen alle Objekte einer Gesamtheit beobachtet werden und auf keine umfassendere Gesamtheit geschlossen werden soll, können wir vernachlässigen, für sie benötigt man die mathematische Statistik nicht. Wir gehen also davon aus, dass aus einer Gesamtheit nur eine Teilmenge ausgewählt wird, um das Merkmal (den Merkmalsvektor) zu beobachten von dem auf die gesamte Population geschlossen werden soll. Einen solchen Teil nennen wir (konkrete) Stichprobe (der Objekte). Die Menge der an diesen Objekten gemessenen Merkmalswerte nennen wir (konkrete) Stichprobe der Merkmalswerte. Jedes Objekt der Population soll einen Merkmalswert besitzen (unabhängig davon, ob

wir ihn erfassen oder nicht). Die der Population entsprechende Gesamtheit der Merkmalswerte der Objekte dieser Population nennen wir Grundgesamtheit.

Eine Population und das zu erfassende Merkmal und damit auch die Grundgesamtheit müssen eindeutig definiert sein. Populationen sind vor allem räumlich und zeitlich abzugrenzen. Von einem beliebigen Objekt der Realwelt muss prinzipiell feststehen, ob es zur Population gehört oder nicht. Wir betrachten im Folgenden einige Beispiele:

Population		**Grundgesamtheit**	
A	Färsen einer bestimmten Rasse	A_1	Jahresmilchmenge dieser Färsen
	eines bestimmten Gebietes	A_2	180-Tage-Körpermasse dieser Färsen
	in einem bestimmten Jahr	A_3	Rückenhöhe dieser Färsen
B	Bewohner einer Stadt	B_1	Blutdruck dieser Bewohner um 6:00 Uhr
	an einem bestimmten Tag	B_2	Alter der Bewohner

Es ist einleuchtend, dass Schlüsse von der Stichprobe auf die Grundgesamtheit falsch sein können. Wenn man z. B. aus der Population B die Kinder einer Kindertagesstätte auswählt, ist möglicherweise der Blutdruck, aber ganz sicher das Alter nicht auf die Population verallgemeinerbar. Generell sprechen wir von Merkmalen, sofern diese aber einen bestimmten Einfluss auf die Versuchsergebnisse haben können, nennen wir sie auch Faktoren, die (meist wenigen) Merkmalswerte heißen dann Faktorstufen, die Kombination von Faktorstufen mehrerer Faktoren heißen Faktorstufenkombinationen.

Hinsichtlich aller Faktoren, die das Merkmal in einer Grundgesamtheit beeinflussen können, sollte die Stichprobe „repräsentativ“ sein. Das heißt, in der Stichprobe der Objekte sollte sich die Zusammensetzung der Population widerspiegeln. Das ist aber bei kleinen Stichproben und vielen Faktorstufenkombinationen gar nicht möglich. In Population B gibt es hinsichtlich der Faktoren Alter und Geschlecht schon etwa 200 Faktorstufenkombinationen, die sich unmöglich in einer Stichprobe von 100 Einwohnern widerspiegeln können. Wir empfehlen daher, den Begriff „repräsentative Stichprobe“ nicht zu verwenden, da er nicht sauber definiert werden kann.

Stichproben sollen nicht danach beurteilt werden, welche Elemente sie enthalten, sondern danach, wie sie erhalten (gezogen) wurden. Die Art und Weise, wie eine Stichprobe erhoben wird, heißt Stichprobenverfahren. Es kann entweder auf die Objekte als Merkmalsträger oder auf die Grundgesamtheit der Merkmalswerte (z. B. in einer Datenbank) angewendet werden. Im letzteren Fall entsteht die Stichprobe der Merkmalswerte unmittelbar. Im ersteren Fall muss das Merkmal an den ausgewählten Objekten noch erfasst werden. Beide Vorgehensweisen (nicht unbedingt die entstehenden Stichproben) sind dann identisch, wenn für jedes ausgewählte Objekt der Merkmalswert erfasst wird. Davon gehen wir in diesem Kapitel aus. In zensierten Stichproben ist das nicht der Fall. Eine Stichprobe heißt zensiert, wenn der Merkmalswert nicht an allen Versuchseinheiten erfasst

werden konnte. Bricht man z. B. eine Lebensdauerermittlung (z. B. von elektronischen Bauteilen) nach einer bestimmten Zeit ab, liegen Messwerte für Objekte mit längerer Lebensdauer (als die Beobachtungszeit) nicht vor.

Im Folgenden wird nicht zwischen Stichproben der Objekte und der Merkmalswerte unterschieden, die Definitionen gelten für beide.

Definition 1.1
Ein Stichprobenverfahren ist eine Vorschrift für die Auswahl einer endlichen Teilmenge, genannt Stichprobe, aus einer wohldefinierten endlichen Population (Grundgesamtheit), es heißt zufällig, wenn jedes Element der Grundgesamtheit mit der gleichen Wahrscheinlichkeit p in die Stichprobe gelangen kann. Eine (konkrete) Stichprobe ist das Ergebnis der Anwendung eines Stichprobenverfahrens. Stichproben, die das Ergebnis eines zufälligen Stichprobenverfahrens sind, heißen (konkrete) zufällige Stichproben oder (konkrete) Zufallsstichproben.

In der Stichprobentheorie (siehe z. B. Cochran und Boing 1972; Kauermann und Küchenhoff 2011 oder Quatember 2014) wird eine Vielzahl von zufälligen Stichprobenverfahren zur Verfügung gestellt. Wir verwenden ab jetzt die Begriffe Population und Grundgesamtheit synonym.

1.1.2 Stichprobenverfahren

Bei Zufallsauswahlverfahren unterscheiden wir u. a.:

- die einfache oder reine Zufallsauswahl, bei der jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit hat, in die Stichprobe zu gelangen.
- die geschichtete Auswahl, bei der innerhalb zuvor festgelegter (disjunkter) Klassen eine zufällige Auswahl vorgenommen wird, sie ist nur dann insgesamt zufällig, wenn die Auswahlwahrscheinlichkeiten innerhalb der Klassen proportional zum Umfang der Klassen gewählt werden.
- die Klumpenauswahl, hier wird eine Grundgesamtheit in Gruppen (Klumpen) eingeteilt. Die Auswahl der Untersuchungsobjekte erfolgt nicht unter den Elementen der Grundgesamtheit, sondern unter den (disjunkten) Klumpen. In den ausgewählten Klumpen werden dann alle Elemente erfasst. Sie findet häufig in Form der Flächenstichproben Anwendung. Sie ist nur dann zufällig im Sinne der Definition 1.1, wenn die Klumpen gleich viele Elemente enthalten.
- die mehrstufige Auswahl, sie ist dadurch gekennzeichnet, dass mindestens zwei Auswahlstufen bestehen. Die Grundgesamtheit wird z. B. in zweistufiger Auswahl in Primäreinheiten in Form disjunkter Teilmengen zerlegt. Aus der Menge der Primäreinheiten erfolgt zunächst eine Zufallsauswahl. Aus jeder ausgewählten Primäreinheit erfolgt eine Zufallsauswahl von Untersuchungseinheiten (Sekundäreinheiten). Eine mehrstufige Auswahl ist dann vorteilhaft, wenn die Grundgesamtheit hierarchisch gegliedert ist (Land, Provinzen, Städ-

te in der Provinz). Sie ist nur dann zufällig im Sinne der Definition 1.1, wenn die Primäreinheiten gleich viele Sekundäreinheiten enthalten.
- die stets sequentielle Auswahl; hier liegt der Stichprobenumfang nicht vor Beginn des Auswahlprozesses fest, es wird zunächst eine kleine Stichprobe gezogen und analysiert. Es erfolgt dann eine Entscheidung, ob die vorliegende Information hinreichend ist, z. B. um eine Hypothese abzulehnen oder anzunehmen (siehe Kapitel 3), oder ob mehr Information durch Ziehung einer weiteren Einheit beschafft werden soll.

Sowohl ein zufälliges Stichprobenverfahren als auch eine willkürliche Auswahl aufs Geratewohl können zur gleichen konkreten Stichprobe führen. Ob einer konkreten Stichprobe eine zufällige oder eine willkürliche Auswahl zugrunde liegt, kann nicht anhand dieser Stichprobe beurteilt werden, sondern eben nur anhand des verwendeten Auswahlverfahrens.

Bei der reinen Zufallsauswahl wird Definition 1.1 direkt angewendet, jedes Element einer Grundgesamtheit vom Umfang N wird mit der gleichen Wahrscheinlichkeit p der Grundgesamtheit entnommen. Wir nennen die Anzahl der Elemente in einer Stichprobe den Stichprobenumfang und bezeichnen diesen in der Regel mit n.

Der praktisch wichtige Fall einer reinen Zufallsauswahl ist der, dass entnommene Elemente nicht in die Grundgesamtheit zurückgelegt werden, wie das etwa bei der Ziehung der Lottozahlen der Fall ist. Hier werden in Deutschland $n = 6$ Zahlen aus $N = 49$ gegebenen Zahlen gezogen. Bei einer uneingeschränkten Zufallsstichprobe vom Umfang n haben alle möglichen $\binom{N}{n}$ Teilmengen die gleiche Wahrscheinlichkeit $\frac{1}{\binom{N}{n}}$ dafür, in die Stichprobe zu gelangen.

Ob eine Stichprobe eine Zufallsstichprobe ist oder nicht, kann man ihr – wie gesagt – nicht ansehen. Man muss vielmehr das Verfahren betrachten, mit dem sie gezogen wurde. Allerdings wird man sofort misstrauisch, wenn extreme Stichproben auftreten. Wird aus einer Grundgesamtheit mit 10 000 Losen der Hauptgewinn während des Kaufs eines speziellen Loses gezogen, so ist das schon ungewöhnlich, kann aber, wie man im Volksmund sagt, schon mit rechten Dingen zugegangen sein, in unserer Terminologie also das Ergebnis eines Zufallsstichprobenverfahrens sein. Zieht dieselbe Person an drei aufeinanderfolgenden Verlosungen den Hauptgewinn und stellt sich dann noch heraus, dass es sich um den Bruder des Losverkäufers handelt, stellen sich berechtigte Zweifel ein. Wir weigern uns, Ereignisse mit solch geringer Wahrscheinlichkeit zu akzeptieren und vermuten, dass das zugrunde gelegte Modell falsch ist. In diesem Fall nehmen wir an, dass kein Zufallsstichprobenverfahren zugrunde gelegt wurde und Betrug im Spiel ist. Trotzdem besteht eine ganz geringe Wahrscheinlichkeit für dieses Ereignis als Ergebnis eines Zufallsstichprobenverfahrens, nämlich 1/1 000 000 000 000.

Nebenbei bemerkt bildet diese Art, Modelle (Sachverhalte) zu verwerfen, unter denen ein beobachtetes Ereignis eine sehr kleine Wahrscheinlichkeit hat und statt dessen solche Modelle zu akzeptieren, bei denen die Wahrscheinlichkeit dieses Ereignisses größer ist, die Basis für die statistischen Tests in Kapitel 3.

Bei einer Zufallsstichprobe mit Zurücklegen wird auch ein reines Stichprobenverfahren verwendet, also jedes Element hat die gleiche Wahrscheinlichkeit mit Zurücklegen gezogen zu werden. Es wird jedes gezogene und beobachtete Element in die Grundgesamtheit zurückgelegt, bevor das nächste Element gezogen wird. Das geht nur bei zerstörungsfreier Beobachtung, d. h. in solchen Fällen, bei denen sich die Stichprobeneinheit durch die Beobachtung nicht verändert (Beispiele, bei denen ein Zurücklegen nicht möglich ist, sind Zerreißproben, Untersuchungen an geschlachteten Tieren, Fällen von Bäumen, Abernten u. a.). Dieses Verfahren heißt einfaches Zufallsstichprobenverfahren mit Zurücklegen. Bei Zufallsstichprobenverfahren ohne Zurücklegen erhält man $n < N$ verschiedene Elemente, beim Zufallsstichprobenverfahren mit Zurücklegen kann dasselbe Element mehrfach in der Stichprobe auftauchen und es kann auch $n > N$ sein.

Eine mitunter praktisch einfacher zu realisierende Methode ist die systematische Auswahl mit Zufallsstart. Sie ist anwendbar, wenn die Elemente der endlichen Auswahlgrundlage von 1 bis N durchnummeriert sind und die Folge nicht mit dem Merkmal zusammenhängt. Wenn N/n ganzzahlig ist, wählt man zufällig eine Zahl i zwischen 1 und N/n aus und bildet die Stichprobe aus den Elementen $i, N/n + i, 2N/n + i, \ldots, (n-1)N/n + i$. Näheres hierzu und was zu tun ist, wenn N/n nicht ganzzahlig ist, findet man bei (Rasch *et al.*, 2008), Verfahren 1/31/1210.

Die oben erwähnte geschichtete Auswahl bietet sich dann an, wenn die Grundgesamtheit vom Umfang N auf inhaltlich relevante Weise in s Teilgesamtheiten vom Umfang $N_1, N_2, \ldots, N_s$ zerfällt. Insbesondere kann die Grundgesamtheit gelegentlich nach den Stufen eines vermuteten Störfaktors in solche Teilgesamtheiten unterteilt werden. Man bezeichnet diese Teilgesamtheiten als Schichten. Will man aus dieser Grundgesamtheit Stichproben vom Umfang n erheben, so muss man bei einem uneingeschränkten Zufallsstichprobenverfahren befürchten, dass nicht alle Schichten überhaupt bzw. zumindest nicht in angemessener Weise berücksichtigt werden. Dann ist es besser, ein geschichtetes Zufallsstichprobenverfahren durchzuführen. Man erhebt dabei jeweils Teilstichproben vom Umfang n_i ($i = 1, 2, \ldots, s$) aus der i-ten Schicht. Die Teilstichproben werden aus der jeweiligen Schicht nach einem reinen Zufallsstichprobenverfahren gezogen. Dies entspricht, wenn n_i/n proportional zu N_i/N gewählt wird, auch insgesamt einem Zufallsstichprobenverfahren.

Während beim geschichteten Zufallsstichprobenverfahren aus jeder Teilmenge Elemente erhoben werden, werden bei der mehrstufigen Auswahl wie oben beschrieben auf jeder Stufe zufällig Teilmengen oder Elemente entnommen. Im zweistufigen Fall bestehe die Grundgesamtheit aus k disjunkten Teilmengen vom Umfang N_0, den Primäreinheiten. Es wird nun vorausgesetzt, dass sich die Merkmalswerte zwischen den Primäreinheiten nur zufällig unterscheiden, sodass nicht aus allen Primäreinheiten Elemente entnommen werden müssen. Ist der gewünschte Stichprobenumfang $n = rn_0$ mit $r < k$, so wählt man zunächst nach einem reinen Zufallsstichprobenverfahren r der k Primäreinheiten aus. Aus jeder der r in der ersten Stufe erhobenen Primäreinheiten wählt man in der zweiten Stufe nach einem reinen Zufallsstichprobenverfahren je n_0 Objekte (Sekundär-

Tab. 1.1 Anzahl K möglicher Stichproben für verschiedene Stichprobenverfahren.

Stichprobenverfahren	Anzahl K möglicher Stichproben
Reine Zufallsauswahl	$K = \binom{1000}{100} > 10^{140}$
Systematische Auswahl mit Zufallsstart $k = 10$	$K = 10$
Geschichtetes Stichprobenverfahren, $k = 20$, $N_i = 50$, $i = 1, \dots, 20$	$K = \left[\binom{50}{5}\right]^{20} = 2{,}118\,76 \cdot 10^{120}$
Geschichtetes Stichprobenverfahren, $k = 10$, $N_i = 100$, $i = 1, \dots, 10$	$K = \left[\binom{100}{10}\right]^{10} = 1{,}731\,030\,9 \cdot 10^{130}$
Geschichtetes Stichprobenverfahren, $k = 5$, $N_i = 200$, $i = 1, \dots, 5$	$K = \left[\binom{200}{20}\right]^{5} = 1{,}613\,587\,8 \cdot 10^{135}$
Geschichtetes Stichprobenverfahren, $k = 2$, $N_1 = 400$, $N_2 = 600$	$K = \binom{400}{40} \cdot \binom{600}{60} = 5{,}466\,241\,4 \cdot 10^{138}$
Zweistufiges Verfahren $k = 20$, $N_0 = 50$, $r = 4$	$K = \binom{20}{4} \cdot \binom{50}{25} = 6{,}124\,593\,9 \cdot 10^{17}$
Zweistufiges Verfahren $k = 20$, $N_0 = 50$, $r = 5$	$K = \binom{20}{5} \cdot \binom{50}{20} = 7{,}306\,913\,1 \cdot 10^{17}$
Zweistufiges Verfahren $k = 10$, $N_0 = 100$, $r = 2$	$K = \binom{10}{2} \cdot \binom{100}{50} = 4{,}540\,110\,5 \cdot 10^{30}$
Zweistufiges Verfahren $k = 10$, $N_0 = 100$, $r = 4$	$K = \binom{10}{4} \cdot \binom{100}{25} = 5{,}092\,904\,7 \cdot 10^{25}$
Zweistufiges Verfahren $k = 5$, $N_0 = 200$, $r = 2$	$K = \binom{5}{2} \cdot \binom{200}{50} = 4{,}538\,583\,8 \cdot 10^{48}$
Zweistufiges Verfahren $k = 2$, $N_0 = 500$, $r = 1$	$K = \binom{2}{1} \cdot \binom{500}{100} > 10^{100}$

einheiten) aus. Die Anzahl der möglichen Stichproben beträgt $\binom{k}{r} \cdot \binom{N_0}{n_0}$ und entsprechend Definition 1.1 gelangt jedes Element der Grundgesamtheit mit der gleichen Wahrscheinlichkeit $p = \frac{r}{k} \cdot \frac{n_0}{N_0}$ in die Stichprobe.

Beispiel 1.1
Aus einer Grundgesamtheit mit $N = 1000$ Objekten soll eine Zufallsstichprobe ohne Zurücklegen vom Umfang $n = 100$ gezogen werden. Tabelle 1.1 gibt für verschiedene Verfahren die Anzahl der Stichproben an, die Wahrscheinlichkeit der Auswahl ist für jedes Objekt $p = 0{,}1$.

1.2 Mathematische Modelle für Grundgesamtheit und Stichprobe

In der mathematischen Statistik werden Begriffe definiert, die als Modelle (Verallgemeinerungen) für die in der Empirie gebräuchlichen Begriffe verwendet wer-

den. Der Grundgesamtheit, die einer Häufigkeitsverteilung der Merkmalswerte entspricht, wird als Modell die Wahrscheinlichkeitsverteilung gegenübergestellt. Die durch ein Zufallsstichprobenverfahren entstandene konkrete Stichprobe wird durch die realisierte (theoretische) Zufallsstichprobe modelliert. Diese Modellvorstellungen sind dann adäquat, wenn der Umfang N der Grundgesamtheit sehr groß im Vergleich zum Umfang n der Stichprobe ist.

Definition 1.2
Eine n-dimensionale Zufallsvariable

$$\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}, \quad n \geq 1$$

mit den Komponenten $\boldsymbol{y}_i$ heißt Zufallsstichprobe, wenn

- die $\boldsymbol{y}_i$ die gleiche durch die Verteilungsfunktion $F(y_i, \theta) = F(y, \theta)$ charakterisierte Verteilung mit dem Parameter(vektor) $\theta \in \Omega \subseteq R^p$ haben und
- die $\boldsymbol{y}_i$ voneinander stochastisch unabhängig sind, sodass für die Verteilungsfunktion $F(Y, \theta)$ von $\boldsymbol{Y}$

$$F(Y, \theta) = \prod_{i=1}^{n} F(y_i, \theta), \theta \in \Omega \subseteq R^p$$

gilt.

Die Werte $Y = (y_1, y_2, \dots, y_n)^{\mathrm{T}}$ einer Zufallsstichprobe $\boldsymbol{Y}$ heißen Realisationen. Die Gesamtheit $\{Y\}$ aller möglichen Realisationen von $\boldsymbol{Y}$ heißt Stichprobenraum.

In diesem Buch werden Zufallsvariablen fett gedruckt, und der Stichprobenraum $\{Y\}$ liegt stets im n-dimensionalen euklidischen Raum, d. h. $\{Y\} \subset R^n$.

Die Funktion

$$L(Y, \theta) = \begin{cases} f(Y, \theta) = \frac{\partial F(Y, \theta)}{\partial Y}, & \text{für kontinuierliche } y \\ p(Y, \theta), & \text{für diskrete } y \end{cases}$$

mit der Wahrscheinlichkeitsfunktion $p(Y, \theta)$ bzw. der Dichtefunktion $f(Y, \theta)$ bei gegebenem Y als Funktion von θ heißt Likelihood-Funktion.

Das Wort Zufallsstichprobe kann nun folgendes bedeuten:

- Zufallsstichprobe als Zufallsvariable $\boldsymbol{Y}$ nach Definition 1.2,
- (konkrete) Zufallsstichprobe als Teilmenge einer Population (Grundgesamtheit), die nach einem Zufallsstichprobenverfahren ausgewählt wurde.

Die Realisationen Y einer Zufallsstichprobe $\boldsymbol{Y}$ werden wir dagegen stets realisierte Zufallsstichprobe nennen.

Eine Zufallsstichprobe $\boldsymbol{Y}$ ist das mathematische Modell des reinen Zufallsstichprobenverfahrens, konkrete Zufallsstichprobe und realisierte Zufallsstichprobe entsprechen einander auch in der Symbolik.

Wir beschreiben in diesem Buch das „klassische“ Vorgehen, nach dem $\boldsymbol{Y}$ mit der Verteilungsfunktion $F(Y, \theta)$ mit dem festen (nicht zufälligen) Parameter $\theta \in \Omega \subseteq R^p$ verteilt ist. Daneben gibt es das Bayessche Vorgehen, bei dem man ein zufälliges $\boldsymbol{\theta}$ annimmt, das mit einer a-priori-Verteilung mit einem als bekannt vorausgesetztem Parameter φ verteilt ist. Beim empirischen Bayesschen Vorgehen wird die a-priori-Verteilung aus bereits ermittelten Daten geschätzt.

1.3 Suffizienz und Vollständigkeit

Eine Zufallsgröße enthält gewisse Informationen über die Verteilung und deren Parameter. Vor allem für große n (etwa $n > 100$) möchte man die Elemente der Zufallsstichprobe so verdichten, dass möglichst wenige neue Zufallsvariablen möglichst viel von dieser Information enthalten. Diese unklar formulierte Wunschvorstellung soll jetzt schrittweise bis zum Konzept der minimal suffizienten Maßzahl präzisiert werden. Zunächst wiederholen wir hier die Definition einer Exponentialfamilie.

Die Verteilung einer Zufallsvariablen $\boldsymbol{y}$ mit dem Parametervektor $\theta = (\theta_1, \theta_2, \dots, \theta_p)^{\mathrm{T}}$ gehört zu einer k-parametrischen Exponentialfamilie, wenn ihre Likelihood-Funktion in der Form

$$f(y, \theta) = h(y)\mathrm{e}^{\sum_{i=1}^{k} \eta_i(\theta) \cdot T_i(y) - B(\theta)}$$

geschrieben werden kann, wobei folgendes gilt:

- η_i und B sind reelle Funktionen von θ und B hängt nicht von y ab.
- Die Funktion $h(y)$ ist nichtnegativ und hängt nicht von θ ab.

Die Exponentialfamilie ist in kanonischer Form mit den sogenannten natürlichen Parametern η_i, falls ihre Elemente als

$$f(y, \eta) = h(y)\mathrm{e}^{\sum_{i=1}^{k} \eta_i \cdot T_i(y) - A(\eta)} \quad \text{mit} \quad \eta = (\eta_1, \dots, \eta_k)^{\mathrm{T}}$$

geschrieben werden können.

Wir gehen von einer Verteilungsfamilie $(P_\theta, \theta \in \Omega)$ von Zufallsvariablen $\boldsymbol{y}$ mit der Verteilungsfunktion $F(y, \theta)$, $\theta \in \Omega$ aus. Die Realisationen $Y = (y_1, \dots, y_n)^{\mathrm{T}}$ der Zufallsstichprobe

$$\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$$

mit wie $\boldsymbol{y}$ verteilten Komponenten liegen im Stichprobenraum $\{Y\}$. Nach Definition 1.2 ist mit $F(y, \theta)$ auch die Verteilungsfunktion $F(Y, \theta)$ einer Zufallsstichprobe $\boldsymbol{Y}$ eindeutig festgelegt.

Definition 1.3
Eine messbare Abbildung $\boldsymbol{M} = M(\boldsymbol{Y}) = [M_1(\boldsymbol{Y}), \dots, M_r(\boldsymbol{Y})]^{\mathrm{T}}$, $r \leq n$ von $\{Y\}$ auf einen Raum $\{M\}$, die nicht von $\theta \in \Omega$ abhängt, heißt (statistische) Maßzahl oder auch Statistik.

Definition 1.4
Eine Maßzahl $\boldsymbol{M}$ heißt suffizient oder erschöpfend bezüglich einer Verteilungsfamilie $(P_\theta, \theta \in \Omega)$ bzw. bezüglich $\theta \in \Omega$, falls die bedingte Verteilung einer Zufallsstichprobe $\boldsymbol{Y}$ bei gegebenem $\boldsymbol{M} = M(\boldsymbol{Y}) = M(Y)$ von θ unabhängig ist.

Beispiel 1.2
Die Komponenten einer Zufallsstichprobe $\boldsymbol{Y}$ mögen einer Zweipunktverteilung mit den Werten 1 und 0 folgen. Dabei sei $P(\boldsymbol{y}_i = 1) = p$ und $P(\boldsymbol{y}_i = 0) = 1 - p$ mit $0 < p < 1$. Dann ist $\boldsymbol{M} = M(\boldsymbol{Y}) = \sum_{i=1}^{n} \boldsymbol{y}_i$ suffizient bezüglich $\theta \in (0, 1) = \Omega$. Um das zu zeigen, müssen wir nachweisen, dass $P(\boldsymbol{Y} = Y | \sum_{i=1}^{n} \boldsymbol{y}_i = M)$ von p unabhängig ist. Nun ist $P(\boldsymbol{Y} = Y|M) = \frac{P(\boldsymbol{Y}=Y, \boldsymbol{M}=M)}{P(\boldsymbol{M}=M)}$, $M = 0, 1, \dots, n$. Aus der Wahrscheinlichkeitsrechnung wissen wir, dass $\boldsymbol{M} = M(\boldsymbol{Y}) = \sum_{i=1}^{n} \boldsymbol{y}_i$ binomialverteilt ist mit den Parametern n und p, also gilt:

$$P(\boldsymbol{M} = M) = \binom{n}{M} p^M (1-p)^{n-M}, \quad M = 0, 1, \dots, n$$

Ferner ist mit $y_i = 0$ oder $y_i = 1$ und $A(M) = \{Y | M(Y) = M\}$

$$\begin{aligned} P[\boldsymbol{Y} = Y, M(\boldsymbol{Y}) = M] &= P(\boldsymbol{y}_1 = y_1, \dots, \boldsymbol{y}_n = y_n) I_{A(M)}(Y) \\ &= \prod_{i=1}^{n} \left(\binom{1}{y_i} p^{y_i} (1-p)^{1-y_i} \right) I_{A(M)}(Y) \\ &= p^{\sum_{i=1}^{n} y_i} (1-p)^{n - \sum_{i=1}^{n} y_i} I_{A(M)}(Y) \\ &= p^M (1-p)^{n-M} \end{aligned}$$

Daher ist $P(\boldsymbol{Y} = Y|M) = \frac{1}{\binom{n}{M}}$ und das ist unabhängig von p.

Auf diese Weise ist der Nachweis der Suffizienz recht mühsam, er gelingt aber auch für kontinuierliche Verteilungen, wie das nächste Beispiel zeigt.

Beispiel 1.3
Die Komponenten $\boldsymbol{y}_i$ einer Zufallsstichprobe $\boldsymbol{Y}$ vom Umfang n seien nach $N(\mu, 1)$ mit Erwartungswert μ und Varianz $\sigma^2 = 1$ verteilt. Dann ist $\boldsymbol{M} = \sum \boldsymbol{y}_i$ suffizient bezüglich $\mu \in R^1 = \Omega$. Um das zu zeigen, vermerken wir zunächst, dass $\boldsymbol{Y}$ nach $N(\mu e_n, E_n)$ verteilt ist. Nun führen wir die eineindeutige Transformation

$$\boldsymbol{Z} = A\boldsymbol{Y} = (\boldsymbol{z}_1 = \sum \boldsymbol{y}_i, \boldsymbol{y}_2 - \boldsymbol{y}_1, \dots, \boldsymbol{y}_n - \boldsymbol{y}_1) \quad \text{mit} \quad A = \begin{pmatrix} 1 & \mathrm{e}_{n-1}^{\mathrm{T}} \\ -e_{n-1} & E_{n-1} \end{pmatrix}$$

durch, es gilt $|A| = n$. Wir schreiben $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{Z}_2) = (\sum \boldsymbol{y}_i, \boldsymbol{y}_2 - \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n - \boldsymbol{y}_1)$ und sehen, dass

$$\operatorname{cov}(\boldsymbol{Z}_2, \boldsymbol{z}_1) = \operatorname{cov}(\boldsymbol{Z}_2, \boldsymbol{M}) = \operatorname{cov}\left((-e_{n-1}, E_{n-1})\boldsymbol{Y}, \mathrm{e}_n^{\mathrm{T}}\boldsymbol{Y}\right) = 0_{n-1}$$

gilt.

Wegen der Normalverteilungsannahme sind damit $\boldsymbol{M}$ und $\boldsymbol{Z}_2$ stochastisch unabhängig. Damit sind $\boldsymbol{Z}_2$, aber auch $\boldsymbol{Z}_2|\boldsymbol{M}$ und auch $\boldsymbol{Z}|\boldsymbol{M}$ von μ unabhängig. Wegen der Eineindeutigkeit der Abbildung $\boldsymbol{Z} = A\boldsymbol{Y}$ ist auch $\boldsymbol{Y}|\boldsymbol{M}$ von μ unabhängig und damit ist $\boldsymbol{M} = \sum \boldsymbol{y}_i$ suffizient bezüglich $\mu \in R^1$. Mit $\boldsymbol{M} = \sum \boldsymbol{y}_i$ und einer reellen Zahl $c \neq 0$ ist stets auch $c\boldsymbol{M}$ also z. B. $\frac{1}{n}\sum \boldsymbol{y}_i = \bar{\boldsymbol{y}}$ suffizient.

Die Suffizienz spielt nun aber in der mathematischen Statistik eine so große Rolle, dass wir einfachere Methoden zum Nachweis der Suffizienz und vor allem zum Auffinden suffizienter Maßzahlen benötigen. Der nachfolgende Satz hilft uns da weiter.

Satz 1.1 Zerlegungssatz
Gegeben sei eine Verteilungsfamilie $(P_\theta, \theta \in \Omega)$ einer Zufallsstichprobe $\boldsymbol{Y}$, die von einem endlichen Maß ν dominiert wird. Die Maßzahl $M(\boldsymbol{Y})$ ist genau dann bezüglich θ suffizient, wenn die Radon-Nikodyn-Dichte f_θ von P_θ bezüglich ν in der Form

$$f_\theta(Y) = g_\theta[M(Y)]h(Y) \tag{1.1}$$

ν- fast überall geschrieben werden kann, wobei gilt: die ν-integrierbare Funktion g_θ ist nichtnegativ und messbar, h ist nichtnegativ und $h(Y) = 0$ nur für eine P_θ-Nullmenge.

Der allgemeine Beweis stammt von Halmos und Savage (1949), man findet ihn auch z. B. bei Bahadur (1955) oder Lehmann (1959).

Wir beschäftigen uns in diesem Buch nur mit diskreten und kontinuierlichen Wahrscheinlichkeitsverteilungen, die die Voraussetzungen dieses Satzes erfüllen. Den Beweis des Satzes für solche Verteilungen gibt Rasch (1995). Wir verzichten hier auf dessen Wiederholung.

Für diskrete Verteilungen bedeutet dieser Satz, dass die Wahrscheinlichkeitsfunktion die Form

$$p(Y, \theta) = g[M(Y), \theta]h(Y) \tag{1.2}$$

hat. Für kontinuierliche Verteilungen hat die Dichtefunktion die Form

$$f(Y, \theta) = g[M(Y), \theta]h(Y) \tag{1.3}$$

Korollar 1.1
Ist die Verteilungsfamilie $(P^*(\theta), \theta \in \Omega)$ der Zufallsvariablen $\boldsymbol{y}$ eine k-parametrische Exponentialfamilie mit natürlichem Parameter η und der Likelihood-Funktion

$$L^*(y, \eta) = h^*(y) \mathrm{e}^{\sum_{j=1}^{k} M_j^*(y) - A(\eta)} \tag{1.4}$$

so ist mit der Zufallsstichprobe $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^{\mathrm{T}}$

$$M(\boldsymbol{Y}) = \left(\sum_{i=1}^{n} M_1^*(\boldsymbol{y}_i), \ldots, \sum_{i=1}^{n} M_k^*(\boldsymbol{y}_i) \right)^{\mathrm{T}} \tag{1.5}$$

suffizient bezüglich θ.

Beweis: Es gilt

$$L(y, \eta) = \prod_{i=1}^{n} h^*(y_i) \mathrm{e}^{\sum_{j=1}^{k} \eta_j \sum_{i=1}^{n} M_j^*(y_i) - nA(\eta)} \tag{1.6}$$

und das hat die Form (1.2) bzw. (1.3) mit $h(Y) = \prod_{i=1}^{n} h^*(y_i)$ und $\theta = \eta$.

Definition 1.5
Zwei Likelihood-Funktionen $L_1(Y_1, \theta)$ und $L_2(Y_2, \theta)$ heißen äquivalent, $L_1 \sim L_2$, wenn

$$L_1(Y_1, \theta) = a(Y_1, Y_2) L_2(Y_2, \theta) \tag{1.7}$$

mit einer von θ unabhängigen Funktion $a(Y_1, Y_2)$ ist.

Dann folgt aus Satz 1.1

Korollar 1.2
$M(\boldsymbol{Y})$ ist genau dann suffizient bezüglich θ, wenn die Likelihood-Funktion $L_M(M, \theta)$ von $\boldsymbol{M} = M(\boldsymbol{Y})$ äquivalent zur Likelihood-Funktion einer Zufallsstichprobe $\boldsymbol{Y}$ ist.

Beweis: Ist $M(\boldsymbol{Y})$ suffizient, so hat mit $L(Y, \eta)$ wegen

$$L_M(M, \theta) = a(Y) L(Y, \theta), a(Y) > 0 \tag{1.8}$$

auch $L_M(M, \theta)$ die Form (1.1). Gilt andererseits (1.8), so folgt, dass die bedingte Verteilung einer Zufallsstichprobe $\boldsymbol{Y}$ bei gegebenem $M(\boldsymbol{Y}) = M$ von θ unabhängig ist.

Beispiel 1.4
Die Komponenten $\boldsymbol{y}_i$ einer Zufallsstichprobe $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$ seien nach $N(\mu, 1)$ verteilt. Es gilt:

$$L(Y, \mu) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \mathrm{e}^{-\frac{1}{2}(Y-\mu e_n)^{\mathrm{T}}(Y-\mu e_N)} = \frac{1}{\left(\sqrt{2\pi}\right)^n} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\bar{y})^2} \mathrm{e}^{-\frac{n}{2}(\bar{y}-\mu)^2} \tag{1.9}$$

Da $M(\boldsymbol{Y}) = \bar{\boldsymbol{y}}$ nach $N(\mu, \frac{1}{n})$ verteilt ist, ist

$$L_M(\bar{y}, \mu) = \frac{\sqrt{n}}{\sqrt{2\pi}} \mathrm{e}^{-\frac{n}{2}(\bar{y}-\mu)^2} \tag{1.10}$$

und damit gilt $L_M(\bar{y}, \mu) \sim L(Y, \mu)$ und $\bar{\boldsymbol{y}}$ ist suffizient bezüglich μ.

Allgemein folgt unmittelbar aus Definition 1.4

Korollar 1.3
Ist $c > 0$ eine von θ unabhängig gewählte reelle Zahl und $M(\boldsymbol{Y})$ suffizient bezüglich θ, so ist auch $cM(\boldsymbol{Y})$ suffizient bezüglich θ.

So ist also z. B. mit $\boldsymbol{M} = \sum \boldsymbol{y}_i$ und $c = \frac{1}{n}$ auch $\frac{1}{n}\sum \boldsymbol{y}_i = \bar{\boldsymbol{y}}$ suffizient.

Man kann nun die Frage stellen, ob es unter den suffizienten Maßzahlen bezüglich einer Verteilungsfamilie $P^*(\theta), \theta \in \Omega$ solche Maßzahlen gibt, die in einem noch zu definierenden Sinne minimal sind, also möglichst wenige Komponenten enthalten. Wie das folgende Beispiel zeigt, ist diese Frage nicht abwegig.

Beispiel 1.5
Es sei $P^*(\theta), \theta \in \Omega$ die Familie der $N(\mu, \sigma^2)$-Normalverteilungen ($\sigma > 0$). Wir betrachten die Maßzahlen einer Zufallsstichprobe $\boldsymbol{Y}$ vom Umfang n:

$$\begin{aligned}
M_1(\boldsymbol{Y}) &= \boldsymbol{Y} \\
M_2(\boldsymbol{Y}) &= \left(\boldsymbol{y}_1^2, \dots, \boldsymbol{y}_n^2\right)^{\mathrm{T}} \\
M_3(\boldsymbol{Y}) &= \left(\sum_{i=1}^{r} \boldsymbol{y}_i^2, \sum_{i=r+1}^{n} \boldsymbol{y}_i^2,\right)^{\mathrm{T}}, \quad r = 1, \dots, n-1 \\
M_4(\boldsymbol{Y}) &= \left(\sum_{i=1}^{n} \boldsymbol{y}_i^2\right)
\end{aligned}$$

die alle bezüglich σ^2 suffizient sind. Das zeigt man sehr einfach mithilfe von Korollar 1.1 zum Zerlegungssatz. Die Likelihood-Funktion von $M_1(\boldsymbol{Y})$ und $\boldsymbol{Y}$ sind identisch (und damit äquivalent). Da mit den $\boldsymbol{y}_i$ auch die $\boldsymbol{y}_i^2$ unabhängig sind und

die $\frac{y_i^2}{\sigma^2} = \chi_i^2$ nach $CQ(1)$ χ^2-verteilt sind, folgt nach der Transformation $y_i^2 = \sigma^2 \chi_i^2$

$$L_M(M_2(\boldsymbol{Y}), \sigma^2) \sim L(Y, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \mathrm{e}^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2} \tag{1.11}$$

Analog verfährt man mit $M_3(\boldsymbol{Y})$ und $\boldsymbol{M}_4(\boldsymbol{Y})$.

Sicher stellt $M_4(\boldsymbol{Y})$ die weitestgehende Zusammenfassung der Komponenten einer Zufallsstichprobe $\boldsymbol{Y}$ dar und ist gegenüber den anderen Maßzahlen vorzuziehen.

Definition 1.6
Eine bezüglich θ suffiziente Maßzahl $M^*(\boldsymbol{Y})$ heißt minimal suffizient bezüglich θ, wenn sie sich als eine Funktion jeder anderen suffizienten Maßzahl $M(\boldsymbol{Y})$ darstellen lässt.

Betrachten wir Beispiel 1.5, so ist

$$M_4(\boldsymbol{Y}) = M_1^{\mathrm{T}}(\boldsymbol{Y}) M_{\mathbf{1}}(\boldsymbol{Y}) = \mathrm{e}_n^{\mathrm{T}} M_{\mathbf{2}} = \begin{pmatrix} 1 & 1 \end{pmatrix} M_3 \,, \quad r = 1, \ldots, n-1$$

Damit kann $M_4(\boldsymbol{Y})$ als Funktion aller suffizienten Maßzahlen des Beispiels geschrieben werden. Für $M_1(\boldsymbol{Y})$, $M_2(\boldsymbol{Y})$ und $M_3(\boldsymbol{Y})$ gilt das nicht, sie sind keine Funktionen von $M_4(\boldsymbol{Y})$. $M_4(\boldsymbol{Y})$ ist die einzige Maßzahl des Beispiels 1.5, die minimal suffizient bezüglich σ^2 sein könnte. Wir werden sehen, dass sie tatsächlich diese Eigenschaft besitzt. Wie kann man nun aber die Minimalsuffizienz feststellen? Wir überlegen uns, dass man mithilfe einer Maßzahl $M(\boldsymbol{Y})$ den Stichprobenraum in elementefremde Teilmengen derart zerlegen kann, dass alle Y, für die $M(Y)$ den gleichen Wert M ergibt, derselben Teilmenge angehören. Umgekehrt ist durch eine gegebene Zerlegung auch die Maßzahl definiert. Wir definieren nun eine Zerlegung, von der wir zeigen werden, dass durch sie eine minimal suffiziente Maßzahl gegeben ist.

Definition 1.7
Es sei $Y_0 \in \{Y\}$ ein fester Punkt im Stichprobenraum (ein bestimmter Wert einer realisierten Zufallsstichprobe), der die Realisationen einer Zufallsstichprobe $\boldsymbol{Y}$ mit Komponenten aus einer Familie $(P^*(\theta), \theta \in \Omega)$ von Wahrscheinlichkeitsverteilungen enthält. Über die Likelihood-Funktion $L(Y, \theta)$ wird durch

$$M(Y_0) = \{Y : L(Y, \theta) \sim L(Y_0, \theta)\} \tag{1.12}$$

eine Teilmenge in $\{Y\}$ definiert. Lassen wir Y_0 den ganzen Stichprobenraum $\{Y\}$ durchlaufen, so wird eine Zerlegung erzeugt. Diese Zerlegung heißt Likelihood-Zerlegung, die ihr entsprechende Maßzahl $M_{\mathrm{L}}(\boldsymbol{Y})$ für die $M_{\mathrm{L}}(\boldsymbol{Y}) = \text{konst.}$ für alle $Y \in M(\boldsymbol{Y}_0)$ und für jedes $\boldsymbol{Y}_0$ gilt, heißt Likelihood-Maßzahl.

Bevor wir mit dieser Methode minimal suffiziente Maßzahlen für einige Beispiele konstruieren, formulieren wir den

Satz 1.2
Die Likelihood-Maßzahl $M_{\mathrm{L}}(\boldsymbol{Y})$ ist minimal suffizient bezüglich θ.

Beweis: Für die Likelihood-Maßzahl $M_{\mathrm{L}}(\boldsymbol{Y})$ gilt mit $Y_1, Y_2 \in \{Y\}$

$$M_{\mathrm{L}}(\boldsymbol{Y}_1) = M_{\mathrm{L}}(\boldsymbol{Y}_2)$$

genau dann, wenn $L(Y_1, \theta) \sim L(Y_2, \theta)$ ist. Damit ist $L(Y, \theta)$ eine Funktion von $M_{\mathrm{L}}(\boldsymbol{Y})$ der Form

$$L(Y, \theta) = a(Y) g^*(M_{\mathrm{L}}(\boldsymbol{Y}), \theta) \tag{1.13}$$

und nach dem Zerlegungssatz ist $M_{\mathrm{L}}(\boldsymbol{Y})$ suffizient bezüglich θ. Ist $M(\boldsymbol{Y})$ eine beliebige andere bezüglich θ suffiziente Maßzahl und gilt für zwei Punkte $Y_1, Y_2 \in \{Y\}$ die Beziehung $M(\boldsymbol{Y}_1) = M(\boldsymbol{Y}_2)$ sowie $L(Y_i, \theta) > 0$ mit $i = 1, 2$, so folgt ebenfalls aus dem Zerlegungssatz

$$L(Y_1, \theta) = h(Y_1) g(M(\boldsymbol{Y}_1), \theta) = h(Y_2) g(M(\boldsymbol{Y}_2), \theta)$$

wegen $M(\boldsymbol{Y}_1) = M(\boldsymbol{Y}_2)$ und $L(Y_2, \theta) = h(Y_2) g(M(\boldsymbol{Y}_2), \theta)$ bzw. $g(M(\boldsymbol{Y}_2), \theta) = \frac{L(Y_2, \theta)}{h(Y_2)}$.

Damit wird $L(Y_1, \theta)$ zu

$$L(Y_1, \theta) = \frac{h(Y_1)}{h(Y_2)} L(Y_2, \theta), h(Y_2) > 0$$

sodass $L(Y_1, \theta) \sim L(Y_2, \theta)$ ist. Das ist aber gerade die Bedingung dafür, dass $M(\boldsymbol{Y}_1) = M(\boldsymbol{Y}_2)$ ist. Folglich ist $M_{\mathrm{L}}(\boldsymbol{Y})$ eine Funktion von $M(\boldsymbol{Y})$, wie $M(\boldsymbol{Y})$ auch gewählt wird und damit minimal suffizient.

Wir demonstrieren das Verfahren an zwei Beispielen.

Beispiel 1.6
Die Komponenten $\boldsymbol{y}_i$ einer Zufallsstichprobe $\boldsymbol{Y}$ seien nach $B(N, p)$, N fest, $0 < p < 1$ binomialverteilt. Es ist eine bezüglich p minimal suffiziente Maßzahl gesucht. Die Likelihood-Funktion ist

$$L(Y, p) = \prod_{i=1}^{n} \binom{N}{y_i} p^{y_i} (1-p)^{(N-y_i)}, \quad y_i = 0, 1, \ldots, N$$

Für alle $Y_0 = (y_{01}, \ldots, y_{0N})^{\mathrm{T}} \in \{Y\}$ mit $L(Y_0, p) > 0$ ist

$$\frac{L(Y, p)}{L(Y_0, p)} = \frac{\prod_{i=1}^{n} \binom{N}{y_i}}{\prod_{i=1}^{n} \binom{N}{y_{0i}}} \left(\frac{p}{1-p}\right)^{\sum_{i=1}^{n} (y_i - y_{0i})}$$

Damit ist $M(Y_0)$ auch durch $M(Y_0) = \{Y : \sum_{i=1}^n y_i = \sum_{i=1}^n y_{0i}\}$ definiert, da gerade dort $L(Y, p) \sim L(Y_0, p)$ gilt. Folglich ist $M(\boldsymbol{Y}) = \sum_{i=1}^n \boldsymbol{y}_i$ eine minimal suffiziente Maßzahl.

Beispiel 1.7
Die Komponenten $\boldsymbol{y}_i$ einer Zufallsstichprobe $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^{\mathrm{T}}$ seien gammaverteilt. Dann ist für $y_i > 0$

$$L(Y, a, k) = \frac{a^{nk}}{[\Gamma(k)]^n} \mathrm{e}^{-a\sum_{i=1}^n y_i} \prod_{i=1}^n y_i^{k-1}$$

Für alle $Y_0 = (y_{01}, \ldots, y_{0N})^{\mathrm{T}} \in \{Y\}$ mit $L(Y_0, a, k) > 0$ ist

$$\frac{L(Y, a, k)}{L(Y_0, a, k)} = \mathrm{e}^{-a(\sum_{i=1}^n y_i - \sum_{i=1}^n y_{0i})} \frac{\prod_{i=1}^n y_i^{k-1}}{\prod_{i=1}^n y_{0i}^{k-1}}$$

Ist a vorgegeben, so ist $\prod_{i=1}^n \boldsymbol{y}_i$ minimal suffizient bezüglich k. Ist k bekannt, so ist $\sum \boldsymbol{y_i}$ minimal suffizient bezüglich a. Sind a und k unbekannte Parameter, so ist $(\prod_{i=1}^n \boldsymbol{y}_i, \sum_{i=1}^{\boldsymbol{n}} \boldsymbol{y}_i)$ minimal suffizient bezüglich (a, k).

Allgemein gilt:

Satz 1.3
Ist $(P^*(\theta), \theta \in \Omega)$ eine k-parametrische Exponentialfamilie mit der Likelihood-Funktion in kanonischer Form

$$L(y, \theta) = \mathrm{e}^{\sum_{i=0}^k \eta_i M_i(y) - A(\eta)} h(y)$$

wobei die Dimension des Parameterraumes gleich k ist (d. h., die $\eta_1, \ldots, \eta_k$ linear unabhängig sind), dann ist

$$M(\boldsymbol{Y}) = \left(\sum_{i=1}^n M_1(\boldsymbol{y_i}), \ldots, \sum_{i=1}^n M_k(\boldsymbol{y_i}) \right)^{\mathrm{T}}$$

minimal suffizient bezüglich $(P^*(\theta), \theta \in \Omega)$.

Beweis: Die Suffizienz von $M(\boldsymbol{Y})$ folgt aus Korollar 1.1 zum Zerlegungssatz, und die Minimalsuffizienz folgt aus der Tatsache, dass $M(\boldsymbol{Y})$ die Likelihood-Maßzahl ist, denn es ist genau dann $L(Y, \theta) \sim L(Y_0, \theta)$, wenn

$$\sum_{j=1}^k \eta_j \sum_{i=1}^n [M_j(\boldsymbol{y}_i) - M_j(\boldsymbol{y}_{0i})] = 0$$

gilt, und wegen der linearen Unabhängigkeit der η_i ist das nur dann der Fall, wenn $M(\boldsymbol{Y}) = M(\boldsymbol{Y_0})$ gilt.

Beispiel 1.8
Es sei $(P^*(\theta), \theta \in \Omega)$ die Familie der zweidimensionalen Normalverteilungen mit der Zufallsvariablen $\binom{x}{y}$, dem Erwartungswertvektor $\mu = \binom{\mu_x}{\mu_y}$ und der Kovarianzmatrix $\Sigma = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$. Das ist eine vierparametrische Exponentialfamilie mit den natürlichen Parametern

$$\eta_1 = \frac{\mu_x}{\sigma_x^2}\,, \quad \eta_2 = \frac{\mu_y}{\sigma_y^2}\,, \quad \eta_3 = -\frac{1}{2\sigma_x^2}\,, \quad \eta_4 = -\frac{1}{2\sigma_y^2}$$

und den Faktoren

$$M_1\left[\binom{x}{y}\right] = x\,, \quad M_2\left[\binom{x}{y}\right] = y\,,$$
$$M_3\left[\binom{x}{y}\right] = x^2\,, \quad M_4\left[\binom{x}{y}\right] = y^2\,,$$
$$A(\eta) = \frac{1}{2}\left(\frac{\mu_x^2}{\sigma_x^2} + \frac{\mu_y^2}{\sigma_y^2}\right)$$

Ist $\dim(\Omega) = 4$, so ist

$$\boldsymbol{M} = \left(\sum_{i=1}^{n} \boldsymbol{M}_{1i}\,, \quad \sum_{i=1}^{n} \boldsymbol{M}_{2i}\,, \quad \sum_{i=1}^{n} \boldsymbol{M}_{3i}\,, \quad \sum_{i=1}^{n} \boldsymbol{M}_{4i}\right)^{\mathrm{T}}$$

minimal suffizient bezüglich $(P^*(\theta), \theta \in \Omega)$. Nehmen wir an, $(\check{P}^*(\theta), \theta \in \Omega) \subseteq (P^*(\theta), \theta \in \Omega)$ sei die Teilfamilie von $(P^*(\theta), \theta \in \Omega)$, für die $\sigma_x^2 = \sigma_y^2 = \sigma^2$ gilt, dann ist $\dim(\Omega) = 3$, und $\boldsymbol{M}$ ist nicht minimal suffizient bezüglich $\check{P}^*(\theta)$, $\theta \in \Omega$.

Die natürlichen Parameter von $\check{P}^*(\theta)$, $\theta \in \Omega$ sind

$$\eta_1 = \frac{\mu_x}{\sigma^2}\,, \quad \eta_2 = \frac{\mu_y}{\sigma^2}\,, \quad \eta_3 = -\frac{1}{2\sigma^2}$$

ferner ist $A(\eta) = \frac{1}{2\sigma^2}(\mu_x^2 + \mu_y^2)$ und die Faktoren der η_i sind

$$\check{M}_1\left[\binom{x}{y}\right] = x\,, \quad \check{M}_2\left[\binom{x}{y}\right] = y\,, \quad \check{M}_3\left[\binom{x}{y}\right] = x^2 + y^2$$

Bezüglich $\check{P}^*(\theta)$, $\theta \in \Omega$ ist

$$\check{\boldsymbol{M}} = \left(\sum_{i=1}^{n} \check{\boldsymbol{M}}_{1i}\,, \quad \sum_{i=1}^{n} \check{\boldsymbol{M}}_{2i}\,, \quad \sum_{i=1}^{n} \check{\boldsymbol{M}}_{3i}\right)^{\mathrm{T}}$$

minimal suffizient.

Wie in Kapitel 6 am Beispiel des Modells II der Varianzanalyse gezeigt wird, ist das Ergebnis von Satz 1.3 auch in komplizierteren Modellen geeignet, minimal suffiziente Maßzahlen zu finden.

Eine für die Schätztheorie weitere wichtige Eigenschaft ist die Vollständigkeit bzw. beschränkte Vollständigkeit, die wir gemeinsam durch folgende Definition einführen.

Definition 1.8
Eine Verteilungsfamilie $P = (P_\theta, \theta \in \Omega)$ mit der Verteilungsfunktion $F(y, \theta)$, $\theta \in \Omega$ heißt vollständig, wenn für jede P-integrierbare Funktion $h(\boldsymbol{y})$ der Zufallsvariablen $\boldsymbol{y}$ aus

$$E[h(\boldsymbol{y})] = \int h(y)\,\mathrm{d}F(y) = 0 \quad \text{für alle} \quad \theta \in \Omega \tag{1.14}$$

die Beziehung

$$P_\theta[h(\boldsymbol{y}) = 0] = 1 \quad \text{für alle} \quad \theta \in \Omega \tag{1.15}$$

folgt. Folgt (1.15) aus (1.14) nur für beschränkte Funktionen $h(y)$, so heißt $P = (P_\theta, \theta \in \Omega)$ beschränkt vollständig.

Wir wollen ein Beispiel für eine vollständige Verteilungsfamilie betrachten.

Beispiel 1.9
Es sei P die Familie $\{P_p\}$, $p \in (0, 1)$ der Binomialverteilungen mit der Wahrscheinlichkeitsfunktion

$$p(y, p) = \binom{n}{y} p^y (1-p)^{n-y} = \binom{n}{y} \nu^y (1-p)^n\,, \quad 0 < p < 1$$
$$y = 0, 1, \ldots, n\,, \quad \nu = \frac{p}{1-p}$$

Integrierbarkeit von $h(y)$ bedeutet Endlichkeit von $(1-p)^n \sum_{y=0}^{n} h(y)\binom{n}{y}\nu^y$ und aus (1.14) folgt

$$\sum_{y=0}^{n} h(y)\binom{n}{y}\nu^y = 0 \quad \text{für alle} \quad p \in (0, 1)$$

Das ist ein Polynom n-ten Grades in ν, das höchstens n reelle Nullstellen besitzt. Damit diese Gleichung für alle $\nu \in R^+$ erfüllt ist, muss $\binom{n}{y}h(y)$ für $y = 0, 1, \ldots, n$ verschwinden, und da alle $\binom{n}{y} > 0$ sind, impliziert das $P_\theta[h(\boldsymbol{y}) = 0] = 1$ für alle $p \in (0, 1)$.

Satz 1.4
Eine k-parametrische Exponentialfamilie der Verteilung der suffizienten Maßzahl ist unter den Voraussetzungen von Satz 1.3 ($\dim(\Omega)) = k$) vollständig.

Den Beweis findet man bei Lehmann (1959, S. 132).

Definition 1.9
Gegeben sei eine Zufallsstichprobe $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$, deren Komponenten einer Verteilung aus der Familie

$$P^* = (P_\theta, \theta \in \Omega)$$

folgen. Eine Maßzahl $M(\boldsymbol{Y})$, deren Verteilung von θ unabhängig ist, heißt Hilfsmaßzahl. Ist P die Familie der durch die Maßzahl $M(\boldsymbol{Y})$ aus P^* induzierten Verteilungen und ist P vollständig und $M(\boldsymbol{Y})$ suffizient bezüglich P^*, so heißt $M(\boldsymbol{Y})$ vollständig suffizient.

Beispiel 1.10
Es sei P^* die Familie der Normalverteilungen $N(\mu, 1)$ mit Erwartungswert $\mu = \theta$ und Varianz 1, d. h., es gilt $\Omega = R^1$. Das ist eine einparametrische Exponentialfamilie mit $\dim(\Omega) = 1$, die nach Satz 1.4 vollständig ist. Ist $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$ eine Zufallsstichprobe mit Komponenten aus P^*, so ist $M_1(\boldsymbol{Y}) = \bar{\boldsymbol{y}}$ nach $N(\mu, \frac{1}{n})$ verteilt. Die Familie der Verteilungen von P^* ist folglich auch vollständig. Wegen Satz 1.3 ist $\bar{\boldsymbol{y}}$ minimal suffizient und damit vollständig suffizient. Die durch $(n-1)M_2(\boldsymbol{Y}) = \sum \boldsymbol{y}_i^2 - n\bar{\boldsymbol{y}}^2$ induzierte Verteilungsfamilie der $CQ(n-1)$-Verteilungen (χ^2-Verteilungen mit $n-1$ Freiheitsgraden) ist von μ unabhängig. Folglich ist $\boldsymbol{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^2$ bezüglich $\mu = \theta$ eine Hilfsmaßzahl.

Wir schließen diesen Abschnitt ab mit

Satz 1.5
Es sei $\boldsymbol{Y}$ eine Zufallsstichprobe mit Komponenten aus $P = (P_\theta, \theta \in \Omega)$ und $M_1(\boldsymbol{Y})$ beschränkt vollständig suffizient bezüglich P. Ist ferner $M_2(\boldsymbol{Y})$ eine Maßzahl mit einer von θ unabhängigen Verteilung, so sind $M_1(\boldsymbol{Y})$ und $M_2(\boldsymbol{Y})$ unabhängig.

Beweis: Es sei $\{Y_0\} \subset \{Y\}$ eine Teilmenge des Stichprobenraumes $\{Y\}$. $M_2(\boldsymbol{Y})$ bildet $\{Y\}$ auf $\{M\}$ und $\{Y_0\}$ auf $\{M_0\}$ ab. Da die Verteilung von $M_2(\boldsymbol{Y})$ von θ unabhängig ist, ist $P[M_2(\boldsymbol{Y}) \in \{M_0\}]$ von θ unabhängig. Darüber hinaus ist wegen der Suffizienz von $M_1(\boldsymbol{Y})$ bezüglich θ auch $P[M_2(\boldsymbol{Y}) \in \{M_0\}|M_1(\boldsymbol{Y})]$ von θ unabhängig. Wir betrachten die Maßzahl

$$h(M_1(\boldsymbol{Y})) = P[M_2(\boldsymbol{Y}) \in \{M_0\}|M_1(\boldsymbol{Y})] - P[M_2(\boldsymbol{Y}) \in \{M_0\}]$$

die von $M_1(\boldsymbol{Y})$ abhängt, sodass analog zu (1.14)

$$E_\theta[h(M_1(\boldsymbol{Y}))] = E_\theta[P[M_2(\boldsymbol{Y}) \in \{M_0\}]M_1(\boldsymbol{Y}) - P[M_2(\boldsymbol{Y}) \in \{M_0\}]] = 0$$

für alle $\theta \in \Omega$ folgt. Da $M_1(\boldsymbol{Y})$ beschränkt vollständig ist, gilt für alle $\theta \in \Omega$ mit Wahrscheinlichkeit 1 analog zu (1.15) $P[M_2(\boldsymbol{Y}) \in \{M_0\}|M_1(\boldsymbol{Y})] - P[M_2(\boldsymbol{Y}) \in \{M_0\}] = 0$ und das bedeutet, dass $M_1(\boldsymbol{Y})$ und $M_2(\boldsymbol{Y})$ unabhängig sind.

1.4 Der Informationsbegriff in der Statistik

Bei der heuristischen Einführung suffizienter Maßzahlen in Abschn. 1.2 war davon die Rede, dass eine Maßzahl die Information einer Stichprobe weitgehend ausschöpfen sollte. Suffizient wird daher auch mit erschöpfend übersetzt. Was soll aber unter der Information einer Stichprobe eigentlich verstanden werden? Der Informationsbegriff wurde von R.A. Fisher in die Statistik eingeführt, und seine Definition ist auch heute noch von großer Bedeutung. Wir sprechen in diesem Zusammenhang von der Fisher-Information. Ein weiterer Informationsbegriff stammt von Kullback und Leibler (1951), wir wollen aber hier nicht weiter auf diese Definition eingehen. Wir beschränken uns in diesem Abschnitt zunächst auf Verteilungsfamilien

$$P = (P_\theta, \theta \in \Omega), \Omega \subset R^1$$

mit reellen Parametern θ. Mit $L(y, \theta)$ wird die Likelihood-Funktion $(\boldsymbol{Y} = \boldsymbol{y})$ von P bezeichnet.

Definition 1.10
Es sei $\boldsymbol{y}$ nach $P = (P_\theta, \theta \in \Omega)$, $\Omega \subset R^1$ verteilt. Weiter sei folgende Voraussetzung V1 erfüllt:

1. Ω ist ein offenes Intervall.
2. Für jedes $y \in \{Y\}$ und für jedes $\theta \in \Omega$ existiert $\frac{\partial}{\partial\theta} L(y, \theta)$ und ist endlich. Die Menge der Punkte, in denen $L(y, \theta) = 0$ ist, hängt nicht von θ ab.
3. Für jedes $\theta \in \Omega$ existiert ein $\varepsilon > 0$ und eine positive P_θ-integrierbare Funktion $k(y, \theta)$ derart, dass für alle θ_0 in einer ε-Umgebung von θ
$$\left| \frac{L(y, \theta) - L(y, \theta_0)}{\theta - \theta_0} \right| \leq k(y, \theta_0)$$
gilt.
4. $\frac{\partial}{\partial\theta} L(y, \theta)$ ist quadratisch P_θ-integrierbar und es ist für alle $\theta \in \Omega$
$$0 < E\left\{ \left[\frac{\partial}{\partial\theta} \ln L(\boldsymbol{y}, \theta) \right]^2 \right\}$$

Dann heißt

$$I(\theta) = E\left\{ \left[\frac{\partial}{\partial\theta} \ln L(\boldsymbol{y}, \theta) \right]^2 \right\} \tag{1.16}$$

die Fisher-Information der Verteilung P_θ bzw. von $\boldsymbol{y}$.

Aus der dritten Bedingung von V1 folgt, dass P_θ-Integration und Differentiation nach θ für $L(y, \theta)$ vertauscht werden können, und wegen

$$\frac{\partial}{\partial\theta} \ln L(y, \theta) = \frac{\frac{\partial}{\partial\theta} L(y, \theta)}{L(y, \theta)}$$

ist

$$E_\theta\left[\frac{\partial}{\partial\theta}\ln L(\boldsymbol{y},\theta)\right]=\int\limits_Y\frac{\partial}{\partial\theta}\ln L(y,\theta)L(y,\theta)\,\mathrm{d}y=\frac{\partial}{\partial\theta}\int\limits_Y L(y,\theta)\,\mathrm{d}y=\frac{\partial}{\partial\theta}1=0$$

für alle $\theta\in\Omega$. Damit ist

$$I(\theta)=\operatorname{var}\left\{\frac{\partial}{\partial\theta}\ln L(\boldsymbol{y},\theta)\right\} \tag{1.17}$$

Es existiere nun auch die zweite Ableitung von $\ln L(y,\theta)$ nach θ für alle y und θ, und es sei $\int_Y L(y,\theta)\,\mathrm{d}P_\theta$ zweifach differenzierbar, wobei man Integration und zweifache Ableitung vertauschen kann. Dann gilt wegen

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\ln L(y,\theta)&=\frac{L(y,\theta)\frac{\partial^2}{\partial\theta^2}L(y,\theta)-\left(\frac{\partial}{\partial\theta}L(\boldsymbol{y},\theta)\right)^2}{(L(y,\theta))^2}\\&=\frac{\frac{\partial^2}{\partial\theta^2}L(y,\theta)}{L(y,\theta)}-\left[\frac{\frac{\partial}{\partial\theta}L(\boldsymbol{y},\theta)}{L(y,\theta)}\right]^2\end{aligned}$$

und

$$0=\frac{\partial^2}{\partial\theta^2}\int\ln L(y,\theta)\,\mathrm{d}P_\theta=\int\frac{\partial^2}{\partial\theta^2}\ln L(y,\theta)\,\mathrm{d}P_\theta$$

die Beziehung

$$E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln L(\boldsymbol{y},\theta)\right]=-E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln L(\boldsymbol{y},\theta)\right]^2\right\}=-I(\theta)$$

und damit

$$I(\theta)=-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln L(\boldsymbol{y},\theta)\right] \tag{1.18}$$

Wir geben je ein Beispiel für eine diskrete und eine kontinuierliche Verteilung.

Beispiel 1.11
Es sei P die Familie der Binomialverteilungen mit gegebenem n und $\Omega=(0,1)$. Die Likelihood-Funktion ist

$$L(y,p)=\binom{n}{y}p^y(1-p)^{n-y}$$

Die Voraussetzung V1 ist erfüllt, denn das Quadrat von $\frac{\partial}{\partial p}\ln L(y,p)=\frac{y}{p}-\frac{n-y}{1-p}$ besitzt nach Übergang zu zufälligem $\boldsymbol{y}$ den endlichen Erwartungswert

$$I(p)=E_p\left\{\left[\frac{\partial}{\partial p}\ln L(\boldsymbol{y},p)\right]^2\right\}=\sum_{y=0}^{n}\left(\frac{y}{p}-\frac{n-y}{1-p}\right)^2\binom{n}{y}p^y(1-p)^{n-y}$$

und daraus folgt

$$I(p) = \frac{n}{p(1-p)}\,, \quad 0 < p < 1$$

Beispiel 1.12
Es sei P die Familie der $N(\mu, \sigma^2)$-Verteilungen mit bekanntem σ^2. Wir haben $\Omega = R^1$, und die Likelihood-Funktion hat die Form $L(y, \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$. Auch für diese Verteilung ist V1 erfüllt. Wir erhalten $\frac{\partial}{\partial\mu} L(y, \mu) = \frac{1}{\sigma^2}(y - \mu)$ und $I(\mu) = \frac{1}{\sigma^4} E[(\boldsymbol{y} - \mu)^2] = \frac{1}{\sigma^4} \operatorname{var}(\boldsymbol{y}) = \frac{1}{\sigma^2}$.

Wir beweisen nun die Additivitätseigenschaft der Fisher-Information.

Satz 1.6
Existiert für eine Familie P von Wahrscheinlichkeitsverteilungen mit $\Omega = R^1$ die Fisher-Information $I(\theta) = I_1(\theta)$ und ist $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^{\mathrm{T}}$ eine Zufallsstichprobe mit Komponenten $\boldsymbol{y}_i$ $(i = 1, \ldots, n)$, die nach $P_i \in P$ verteilt sind, so ist die Fisher-Information $I_n(\theta)$ der Verteilung von $\boldsymbol{Y}$ durch

$$I_n(\theta) = nI_1(\theta) \tag{1.19}$$

gegeben.

Beweis: Aus Definition 1.2 folgt, dass die Likelihood-Funktion $L_n(Y, \theta)$ einer Zufallsstichprobe $\boldsymbol{Y}$ gleich

$$L_n(Y, \theta) = \prod_{i=1}^{n} L(y_i, \theta)$$

ist. Damit ist

$$\ln L_n(Y, \theta) = \sum_{i=1}^{n} \ln L(y_i, \theta)$$

und

$$\frac{\partial}{\partial\theta} \ln L_n(Y, \theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln L(y_i, \theta)$$

Folglich wird wegen (1.17)

$$I_n(\theta) = \operatorname{var}\left\{ \frac{\partial}{\partial\theta} \ln L_n(\boldsymbol{Y}, \theta) \right\} = \sum_{i=1}^{n} \operatorname{var}\left\{ \frac{\partial}{\partial\theta} \ln L(\boldsymbol{y}_i, \theta) \right\} = nI_1(\theta)$$

Satz 1.7
Es sei $M(\boldsymbol{Y})$ eine suffiziente Maßzahl bezüglich der Verteilung $P_\theta \in P$, $\Omega \subseteq R^1$ der Komponenten der Zufallsstichprobe $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$. Die Verteilung P_θ erfülle die Bedingung V1 von Definition 1.10. Dann existiert die Fisher-Information

$$I_M(\theta) = \left\{ \left[\frac{\partial}{\partial \theta} \ln L_M(\boldsymbol{M}, \theta) \right]^2 \right\} \tag{1.20}$$

von $\boldsymbol{M} = M(\boldsymbol{Y})$, wobei $L_M(M, \theta)$ die Likelihood-Funktion von $\boldsymbol{M}$ ist, und es gilt

$$I_n(\theta) = I_M(\theta) \tag{1.21}$$

Beweis: Nach (1.2) bzw. (1.3) ist

$$L(Y, \theta) = h(Y) g(M(Y), \theta)$$

und damit

$$\frac{\partial}{\partial \theta} \ln L(Y, \theta) = \frac{\partial}{\partial \theta} \ln g(M(Y), \theta)$$

da $h(Y)$ nach Voraussetzung von θ unabhängig ist. Wegen Korollar 1.1 von Satz 1.1 gilt für die Likelihood-Funktion $L_M(M, \theta)$ von $\boldsymbol{M}$ auch Bedingung V1 von Definition 1.10, und damit existiert $I_M(\theta)$ von (1.20). Wegen der Äquivalenz von $L_M(M, \theta)$ und $L(Y, \theta)$ folgt weiterhin die Behauptung wegen $\frac{\partial}{\partial \theta} \ln L_M(M, \theta) = \frac{\partial}{\partial \theta} \ln\, g(M, \theta)$.

Folglich ist die Fisher-Information einer suffizienten Maßzahl gleich der der Zufallsstichprobe.

Ist $\theta \in \Omega \subseteq R^p$, so geben wir folgende

Definition 1.11
Es sei $\boldsymbol{y}$ nach $P_\theta \in P$, $\Omega \subseteq R^p$, $\theta = (\theta_1, \dots, \theta_p)^{\mathrm{T}}$ verteilt, und für jede Komponente θ_i $(i = 1, \dots, p)$ mögen die Bedingungen 2 bis 4 von V1 in Definition 1.10 erfüllt sein. Ω sei ein offenes Intervall im R^p. Ferner existiere der Erwartungswert von $\frac{\partial}{\partial \theta_i} \ln L(\boldsymbol{y}, \theta) \frac{\partial}{\partial \theta_j} \ln L(\boldsymbol{y}, \theta)$ für alle θ und alle $i, j = 1, \dots, p$. Dann heißt die quadratische Matrix der Ordnung p

$$I(\theta) = (I_{i,j}(\theta)) \,, \quad i, j = 1, \dots, p$$

mit

$$I(\theta) = E \left\{ \frac{\partial}{\partial \theta_i} \ln L(\boldsymbol{y}, \theta) \frac{\partial}{\partial \theta_j} \ln L(\boldsymbol{y}, \theta) \right\}$$

die (Fishersche) Informationsmatrix bezüglich P_θ.

Beispiel 1.13
Die Zufallsvariable $\boldsymbol{y}$ sei nach $N(\mu, \sigma^2)$ verteilt mit $\theta = (\mu, \sigma^2)^{\mathrm{T}} \in R^1 \times R^+ = \Omega$. Dann ist

$$\ln L(y, \theta) = -\ln \sqrt{2\pi} - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}(y - \mu)^2$$

und die Voraussetzung von Definition 1.11 ist mit $\theta_1 = \mu$ und $\theta_2 = \sigma^2$ erfüllt. Es ist

$$\frac{\partial}{\partial \mu} \ln L(y, \theta) = \frac{y - \mu}{\sigma^2} \quad \text{und} \quad \frac{\partial}{\partial \sigma^2} \ln L(y, \theta) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y - \mu)^2$$

Wegen $E[(\boldsymbol{y} - \mu)^2] = \mathrm{var}(\boldsymbol{y}) = \sigma^2$ ist $I_{11}(\theta) = \frac{1}{\sigma^2}$, und weil die Schiefe $\gamma_3 = 0$ ist und wegen $E(\boldsymbol{y} - \mu) = 0$ ist $I_{12}(\theta) = I_{21}(\theta) = 0$. Ferner gilt:

$$\left[\frac{\partial}{\partial \sigma^2} \ln L(y, \theta)\right]^2 = \frac{1}{4\sigma^4} - \frac{2}{4\sigma^6}(y - \mu)^2 + \frac{1}{4\sigma^8}(y - \mu)^4$$

und daher ist wegen $E[(\boldsymbol{y} - \mu)^4] = 3\sigma^4$ (denn es ist $\gamma_2 = 0$)

$$I_{22} = E\left\{\left[\frac{\partial}{\partial \sigma^2} \ln L(\boldsymbol{y}, \theta)\right]^2\right\} = \frac{1}{\sigma^4}\left[\frac{1}{4} - \frac{1}{2} + \frac{3}{4}\right] = \frac{1}{2\sigma^4}$$

und wir erhalten

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Definieren wir dagegen $\theta_1 = \mu$ und $\theta_2 = \sigma$, so ist $\frac{\partial}{\partial \sigma} \ln L(y, \theta) = -\frac{1}{\sigma} + \frac{1}{\sigma^2}(y - \mu)^2$. Während I_{11}, I_{12} und I_{21} unverändert bleiben, wird

$$I_{22} = E\left\{\left[\frac{\partial}{\partial \sigma} \ln L(\boldsymbol{y}, \theta)\right]^2\right\} = \frac{1}{\sigma^2}[1 - 2 + 3] = \frac{2}{\sigma^2}$$

und damit ist

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

Aus diesem Beispiel ersieht man, dass die Fisher-Information nicht invariant gegenüber Parametertransformationen ist. Aus der Kettenregel der Differenzialrechnung folgt allgemein der

Satz 1.8
Es sei $\psi = h(\theta)$ eine in $\Omega \subseteq R^1$ monotone und bezüglich θ differenzierbare Funktion, h bilde Ω auf Π ab. Dann existiert die nach ψ differenzierbare Umkehrfunktion $\theta = g(\psi)$. Unter den Voraussetzungen von Definition 1.10 sei $I(\theta)$ die Fisher-Information der Verteilung $P_0 \in \Pi$. Die Fisher-Information $I^*(\psi)$ der Verteilung P_ψ (d. h., P_θ geschrieben mit dem transformierten Parameter) ist dann

$$I^*(\psi) = I(\theta)\left(\frac{\mathrm{d}}{\mathrm{d}\psi} g(\psi)\right)^2 \tag{1.22}$$

In Beispiel 1.13 war (für festes μ) $\theta = \sigma^2$, $\psi = \sqrt{\theta} = \sigma$ und $\frac{\mathrm{d}\theta}{\mathrm{d}\psi} = 2\psi = 2\sigma$. Nach Satz 1.8 ist also mit $I(\sigma^2) = \frac{1}{\sigma^2}$

$$I^*(\psi) = I(\sigma^2)4\sigma^2 = \frac{2}{\sigma^2}$$

In Kapitel 2 benötigen wir folgende Ungleichung:

Satz 1.9 Ungleichung von Rao und Cramér
Es gelte Bedingung VI von Definition 1.10 für die Komponenten der Zufallsstichprobe $\boldsymbol{Y}$, deren Likelihood-Funktion $L(Y, \theta)$ ist. Die Menge $\{Y_0\} = \{Y \in \{Y\} : L(Y, \theta) = 0\}$ der Punkte des Stichprobenraumes, für die $L(Y, \theta) = 0$ ist, hänge nicht von θ ab. Es sei $P_\theta \in P = (P_\theta, \theta \in \Omega)$, $\Omega \subseteq R^1$ die Verteilung der Komponenten, und $M(\boldsymbol{Y})$ sei eine Maßzahl mit Erwartungswert $E[M(\boldsymbol{Y})]$ und Varianz $\mathrm{var}[M(\boldsymbol{Y})]$, die den Stichprobenraum $\{Y\}$ in Ω abbildet. Dann gilt die Rao-Cramér-Ungleichung

$$\mathrm{var}[M(\boldsymbol{Y})] \geq \frac{\left(\frac{\mathrm{d}E[M(\boldsymbol{Y})]}{\mathrm{d}\theta}\right)^2}{nI(\theta)} \tag{1.23}$$

Beweis: Mit $M(\boldsymbol{Y}) = \boldsymbol{M}$ ist $E = E[\boldsymbol{M} - E(\boldsymbol{M})] = 0$, sodass

$$\frac{\mathrm{d}E}{\mathrm{d}\theta} = -\int\limits_{\{Y\}} \frac{\mathrm{d}E[M(\boldsymbol{Y})]}{\mathrm{d}\theta}\,\mathrm{d}P_\theta + \int\limits_{\{Y\}} (\boldsymbol{M} - \boldsymbol{E}(\boldsymbol{M}))\frac{\mathrm{d}}{\mathrm{d}\theta}L(Y, \theta)\,\mathrm{d}P_\theta = 0$$

bzw.

$$\frac{\mathrm{d}E}{\mathrm{d}\theta} = -\frac{\mathrm{d}E[M(\boldsymbol{Y})]}{\mathrm{d}\theta}\int\limits_{\{Y\}} \mathrm{d}P_\theta + \int\limits_{\{Y\}} (\boldsymbol{M} - \boldsymbol{E}(\boldsymbol{M}))\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(Y, \theta)\,\mathrm{d}Y = 0$$

gilt. Daraus folgt

$$E\left\{(\boldsymbol{M} - \boldsymbol{E}(\boldsymbol{M}))\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(Y, \theta)\right\} = \frac{\mathrm{d}E(\boldsymbol{M})}{\mathrm{d}\theta}$$

Wegen der Schwarzschen Ungleichung folgt weiter

$$\begin{aligned}\left\{\frac{\mathrm{d}E(\boldsymbol{M})}{\mathrm{d}\theta}\right\}^2 &= \left\{E[(\boldsymbol{M} - E(\boldsymbol{M}))]\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(Y, \theta)\right\}^2 \\ &\leq E\left[(\boldsymbol{M} - E(\boldsymbol{M}))^2 E\left\{\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\boldsymbol{Y}, \theta)\right]^2\right\}\right]\end{aligned}$$

und das ergibt wegen (1.16) den Beweis.

Ist speziell $E(\boldsymbol{M}) = \theta$, so hat die Rao-Cramér-Ungleichung die Form

$$\mathrm{var}[M(\boldsymbol{Y})] \geq \frac{1}{nI(\theta)} \tag{1.24}$$

Wir beweisen noch

Satz 1.10
Ist $\boldsymbol{y}$ nach einer einparametrischen Exponentialfamilie verteilt, und ist $g(\theta) = \eta = E(\boldsymbol{M})$, so gilt:

$$I^*(\eta) = \frac{1}{\operatorname{var}(\boldsymbol{M})} \tag{1.25}$$

Beweis: Da die Bedingung V1 von Definition 1.10 erfüllt ist, existiert $I(\eta)$, und wegen

$$\frac{\mathrm{d}}{\mathrm{d}\eta} \ln L(Y, \eta) = M(Y) - \frac{\mathrm{d}}{\mathrm{d}\eta} A(\eta)$$

und nach Satz 1.8 ist $\operatorname{var}(\boldsymbol{M}) = I^*(\eta) = I(\theta)[\operatorname{var}(\boldsymbol{M})]^2$, und daraus folgt die Behauptung.

Aus der Schwarzschen Ungleichung für zweite Momente für jede Maßzahl $M(\boldsymbol{Y})$ mit endlichem zweiten Moment und eine beliebige Funktion $h(\boldsymbol{Y}, \theta)$, deren zweites Moment ebenfalls existiert, folgt:

$$\operatorname{var}(\boldsymbol{M}) \geq \frac{\operatorname{cov}^2[\boldsymbol{M}, \boldsymbol{h}(\boldsymbol{Y}, \theta)]}{\operatorname{var}[h(\boldsymbol{Y}, \theta)]}$$

Satz 1.11
Es sei $M(\boldsymbol{Y})$ eine Maßzahl mit dem Erwartungswert $g(\theta)$ und existierendem zweiten Moment, und $h_j(\boldsymbol{Y}, \theta)$, $j = 1, \ldots, r$ seien Funktionen, deren zweite Momente existieren. Mit

$$c_j = \operatorname{cov}(M(\boldsymbol{Y}), \boldsymbol{h}_j)\,, \quad \sigma_{ij} = \operatorname{cov}(\boldsymbol{h}_i, \boldsymbol{h}_j)\,, \quad c^{\mathrm{T}} = (c_1, \ldots, c_r)$$

und $\Sigma = (\sigma_{ij})$, $|\Sigma| \neq 0$ gilt dann stets

$$\operatorname{var}(\boldsymbol{M}) \geq c^{\mathrm{T}} \Sigma^{-1} c \tag{1.26}$$

Beweis: Die Behauptung folgt aus $\frac{c^{\mathrm{T}} \Sigma^{-1} c}{\operatorname{var}(\boldsymbol{y})} \leq 1$.

Mithilfe von (1.26) lässt sich die Rao-Cramér-Ungleichung (1.24) auf den p-dimensionalen Fall verallgemeinern.

Satz 1.12
Die Komponenten einer Zufallsstichprobe $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^{\mathrm{T}}$ seien nach

$$P_\theta \in P = (P_\theta, \theta \in \Omega)\,, \quad \Omega \subseteq R^p\,, \quad \theta^{\mathrm{T}} = (\theta_1, \ldots, \theta_p)\,, \quad p > 1$$

verteilt. $L(Y, \theta)$ sei die Likelihood-Funktion von $\boldsymbol{Y}$. Ferner seien die Voraussetzungen von Definition 1.10 erfüllt, und die Menge der Punkte in $\{Y\}$, für die

$L(Y, \theta) = 0$ ist, hänge nicht von θ ab. Es sei $M(\boldsymbol{Y})$ eine Maßzahl, deren Erwartungswert $E[M(\boldsymbol{Y})] = w(\theta)$ existiert und nach den θ_i ableitbar ist. Dann gilt

$$\mathrm{var}[M(\boldsymbol{Y})] \geq a^{\mathrm{T}} I^{-1} a$$

wobei I^{-1} die Inverse von $I(\theta)$ und a der Vektor der Ableitungen von $w(\theta)$ nach den θ_i ist.

Beweis: Da $I(\theta)$ positiv definit ist und damit I^{-1} existiert, folgt die Behauptung mit $h_j = \frac{\mathrm{d}}{\mathrm{d}\theta_j}$ aus (1.26) und nach Definition 1.11.

1.5 Statistische Entscheidungstheorie

Wir formulieren zunächst das allgemeine statistische Entscheidungsproblem und gehen von einer Menge von Zufallsvariablen $\{\boldsymbol{y}_t\}$ mit $t \in R^1$ aus, deren Verteilung $P_\theta \in P = (P_\theta, \theta \in \Omega)$, $\dim\{\Omega\} = p$ zumindest teilweise unbekannt ist.

Wir beschränken uns hier auf den Fall, dass ausschließlich Aussagen über $\psi = g(\theta)$ zu machen sind, wobei Ω durch g auf Z abgebildet wird und $\dim(Z) = s$ ist. Z heißt Zustandsraum.

Für die Aussagen über ψ steht dem Statistiker eine Menge $\{E\}$ von Entscheidungen zur Verfügung, $\{E\}$ heißt Entscheidungsraum. Für jedes feste t_i sei $Y_{t_i} = (\boldsymbol{y}_{t_i 1}, \ldots, \boldsymbol{y}_{t_i n_i})^{\mathrm{T}}$ eine Zufallsstichprobe vom Umfang n_i.

Die Gesamtheit der Ergebnisse eines Versuches, anhand dessen eine Entscheidung zu fällen (d. h. aus $\{E\}$ auszuwählen) ist, sei mit

$$N = \sum_{i=1}^{k} n_i, A_k = (Y_{t_i}, \ldots, Y_{t_k}) \in \prod_{i=1}^{k} \{Y_{t_i}\} = \{Y_{k,N}\}$$

die Realisation einer Zufallsvariablen $\boldsymbol{A}_k = (\boldsymbol{Y}_{\boldsymbol{t}_i}, \ldots, \boldsymbol{Y}_{\boldsymbol{t}_k})$.

Nun sei $d \in D$ eine messbare Abbildung von $\{Y_{k,N}\}$ auf E, die jedem A_k eine Entscheidung $d(A_k)$ zuordnet, d heißt Entscheidungsfunktion, und D ist die Menge der zugelassenen Entscheidungsfunktionen. A_k wird von der Verteilung von $\boldsymbol{A}_k$ und den k-Tupeln $\mathfrak{S}_k = (t_1, \ldots, t_k)$ dem Spektrum des Versuchs, und $\mathfrak{N}_k = (n_1, \ldots, n_k)$, der Belegung des Spektrums, abhängen. Mit

$$\begin{pmatrix} \mathfrak{S}_k \\ \mathfrak{N}_k \end{pmatrix} = \begin{pmatrix} t_1, \ldots, t_k \\ n_1, \ldots, n_k \end{pmatrix} \in V_n$$

bezeichnen wir den konkreten Versuchsplan, der Element einer Menge V zugelassener Versuchspläne sei. Außerdem sei eine Verlustfunktion L als messbare Abbildung von $E \times Z \times R^1$ in den R^m gegeben (ihre Festlegung und damit die von m ist ein außermathematisches Problem), d. h., es ist

$$L = L[d(A_k), \psi, f(M)] \,, \quad d(A_k) \in E, \psi \in Z \tag{1.27}$$

mit der nichtnegativen reellen Funktion $f(M)$.

Die Funktion L gibt den Verlust an, der eintritt, wenn $d(A_k)$ gewählt wird und ψ der Wert im transformierten Parameterraum ist, $f(M)$ entspricht den Kosten für die Realisierung von $M = (d, \mathfrak{S}_k, \mathfrak{N}_k)$. Die Aufgabe der Statistik besteht in der Bereitstellung von Methoden zur Auswahl von Tripeln $M = (d, \mathfrak{S}_k, \mathfrak{N}_k)$, die ein Risikofunktion R genanntes Funktional $R(d, \mathfrak{S}_k, \mathfrak{N}_k, \psi, f(M))$ des zufälligen Verlustes minimieren. Wir werden mit d entweder eine Entscheidungsfunktion (bei festem n) oder eine Folge von Entscheidungsfunktionen gleicher Struktur, deren Elemente sich nur hinsichtlich des Stichprobenumfanges n unterscheiden, bezeichnen. Wir wollen annehmen, dass

$$R(d, \mathfrak{S}_k, \mathfrak{N}_k, \psi, f(M)) = F(d, \mathfrak{S}_k, \mathfrak{N}_k, \psi) + f(\mathfrak{S}_k, \mathfrak{N}_k) \tag{1.28}$$

gilt, wobei f nicht von d abhängt und d^* die Entscheidungsfunktion (Folge von Entscheidungsfunktionen) ist, für die

$$\min_{d \in D}(d, \mathfrak{S}_k, \mathfrak{N}_k, \psi) = F\left(d^*, \mathfrak{S}_k, \mathfrak{N}_k, \psi\right) \tag{1.29}$$

gilt. Dann kann R in zwei Schritten minimiert werden. Zunächst bestimmt man d^* so, dass (1.29) erfüllt ist; im zweiten Schritt bestimmt man $(\mathfrak{S}_k^*, \mathfrak{N}_k^*)$ so, dass

$$R\left(d^*, \mathfrak{S}_k^*, \mathfrak{N}_k^*, \psi, f(\mathfrak{S}_k^*, \mathfrak{N}_k^*)\right) = \min_{\binom{\mathfrak{S}_k}{\mathfrak{N}_k} \in V} R\left(d^*, \mathfrak{S}_k, \mathfrak{N}_k, \psi, f(\mathfrak{S}_k, \mathfrak{N}_k)\right)$$

gilt.

Definition 1.12
Ein Tripel $M^* \in V \times D$ heißt lokal R-optimal an der Stelle $\psi_0 \in Z$ bezüglich $V \times D$, wenn für alle $M \in V \times D$

$$R[M^*, \psi_0, f(M^*)] \leqslant R[M, \psi_0, f(M)]$$

gilt. Ist M^* für alle $\psi_0 \in Z$ lokal R-optimal, so heißt M^* global R-optimal.

Beispiel 1.14
Es sei $k = 1$ und $\boldsymbol{y}_{t_1} = \boldsymbol{y}$ nach $N(\mu, \sigma^2)$ verteilt. Dann ist $\theta = \binom{\mu}{\sigma^2} \in \Omega = R^1 \times R^+$ und $\boldsymbol{A}_1 = \boldsymbol{Y}$. Ferner sei $\psi = g(\theta) = \mu \in R^1$ und $d(\boldsymbol{Y}) = \hat{\boldsymbol{\mu}}$ eine statistische Maßzahl mit Realisationen im $R^1 = E$. Es sei D die Klasse der statistischen Maßzahlen mit endlichem zweiten Moment und Realisationen in R^1. Als Verlustfunktion wählen wir

$$L[\hat{\mu}, \mu, f(T)] = c_1(\hat{\mu} - \mu)^2 + c_2 nK\ , \quad c_1, c_2, K > 0$$

wobei K die Kosten einer Messung darstellen. Als Risiko R wird der erwartete zufällige Verlust

$$R(\hat{\mu}, n, \mu, Kn) = E[c_1(\mu - \hat{\mu})^2 + c_2 nK] = c_2 nK + c_1[\mathrm{var}(\hat{\boldsymbol{\mu}}) + B(\hat{\boldsymbol{\mu}})^2]$$

gewählt, in dem $B(\hat{\boldsymbol{\mu}}) = E(\hat{\boldsymbol{\mu}}) - \mu$ ist. In der Klasse D ist für die Entscheidung $\hat{\mu} = \psi_0$ zusammen mit $n = 0$ lokal R-optimal, für (ψ_0, n) ist R gleich 0. Um diesen unbefriedigenden trivialen Fall auszuschließen, kann man D einschränken. Mit $D_E \subset D$ bezeichnen wir die Teilmenge in D, für die $B(\hat{\boldsymbol{\mu}}) = 0$ ist. Dann wird

$$R(\hat{\mu}, n, \mu, Kn) = c_2 nK + c_1 \operatorname{var}(\hat{\boldsymbol{\mu}}) , \quad \hat{\mu} \in D_E$$

und hat die Form (1.28). Wir werden in Kapitel 2 sehen, dass $\operatorname{var}(\hat{\boldsymbol{\mu}})$ für $\hat{\boldsymbol{\mu}}^* = \bar{\boldsymbol{y}}$ zum Minimum wird.

Da $\operatorname{var}(\bar{\boldsymbol{y}}) = \frac{\sigma^2}{n}$ ist, gilt daher im Ergebnis des ersten Schrittes der Minimierung von R

$$\min_{d \in D_E} c_1 \operatorname{var}(\hat{\boldsymbol{\mu}}) = \frac{c_1}{n}\sigma^2$$

und es ist

$$R(\hat{\mu}, n, \mu, Kn) = c_2 Kn + \frac{c_1}{n}\sigma^2$$

Leiten wir die rechte Seite nach n ab und setzen die Ableitung gleich 0, so erhalten wir $n^* = \sigma\sqrt{\frac{c_1}{Kc_2}}$, und das hängt ebenso wie $\bar{\boldsymbol{y}}$ nicht von $\psi = \mu$ ab, und wegen der Konvexität der Ableitungsfunktion handelt es sich tatsächlich um ein Minimum. Daher ist die in Z global (jedoch in Ω wegen der Abhängigkeit von σ lokal) R-optimale Lösung des Entscheidungsproblems in $E \times Z$ gegeben durch

$$M^* = \left(\bar{y}, n^* = \sigma\sqrt{\frac{c_1}{Kc_2}}\right)$$

Wählen wir $\psi = g(\theta) = \sigma^2 > 0$, dann ist $E = R^+$, $k = 1$, $\boldsymbol{A}_1 = \boldsymbol{Y}$ und $N = n$. Die Verlustfunktion sei

$$L[d(Y), \sigma^2, f(M)] = c_1(\sigma^2 - d)^2 + c_2 nK , \quad c_i > 0 , \quad K > 0$$

Wählen wir als Risiko wieder

$$R(d(Y), n, \sigma^2, Kn) = R = E(\boldsymbol{L}) = c_1 E\left\{(\sigma^2 - d(\boldsymbol{Y}))^2\right\} + c_2 nK$$

so ist das wieder von der Form (1.28). Schränken wir uns aus zum vorigen Fall analogen Gründen auf die $d \in D_E$ ein, für die $E[d(\boldsymbol{Y})] = \sigma^2$ gilt, so ist, wie wir in Kapitel 2 sehen werden, der erste Summand von R für

$$d(\boldsymbol{Y}) = \boldsymbol{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^2$$

minimal.

Da $\frac{\boldsymbol{s}^2}{\sigma^2}(n-1)$ nach $CQ(n-1)$ verteilt ist und damit die Varianz $2(n-1)$ hat, ist

$$\operatorname{var}(\boldsymbol{s}^2) = \frac{2\sigma^4}{n-1}$$

und es ist nach dem ersten Schritt der Optimierung

$$R(\boldsymbol{s}^2, n, \sigma^2, Kn) = c_1 \frac{2\sigma^4}{n-1} + c_2 nK$$

Das R-optimale n ergibt sich zu

$$n^* = 1 + \sigma^2 \sqrt{\frac{2c_1}{Kc_2}}$$

und die lokale R-optimale Lösung des Entscheidungsproblems ist

$$M^* = \left(\boldsymbol{s}^2, n^* = 1 + \sigma^2 \sqrt{\frac{2c_1}{Kc_2}} \right)$$

Weiterführende theoretische Details und weitere Anwendungsfälle werden in den folgenden Kapiteln bei der Wahl des minimalen Stichprobenumfangs behandelt. Wir wollen davon ausgehen, dass d bei festem $\mathfrak{S}_k$ und $\mathfrak{N}_k$ bezüglich einer bestimmten Risikofunktion R-optimal zu wählen ist. Bezüglich der optimalen Wahl von $\binom{\mathfrak{S}_k}{\mathfrak{N}_k}$ verweisen wir auf die Kapitel 8 und 9 zur Regressionsanalyse. Wir schreiben daher

$$R(d, \psi) = E\{L[d(\boldsymbol{Y}), \psi]\} = r(d, \tau) \tag{1.30}$$

Um trivial lokal R-optimale Entscheidungsfunktionen d zu vermeiden, wurde in Beispiel 1.14 eine Einschränkung auf eine Teilklasse $D_E \subseteq D$ vorgenommen. Hier sollen zwei weitere allgemeine Vorgehensweisen zur Überwindung solcher Probleme vorgestellt werden.

Definition 1.13
Es sei $\boldsymbol{\theta}$ eine Zufallsvariable mit Realisationen $\theta \in \Omega$ mit der Wahrscheinlichkeitsverteilung P_τ, $\tau \in \mathfrak{T}$. Bezüglich P_τ möge der Erwartungswert von (1.30)

$$\int_{\Omega} R(d, \psi) \, \mathrm{d}\Pi_\tau = r(d, \tau) \tag{1.31}$$

der Bayessches Risiko bezüglich der a-priori-Verteilung P_τ genannt wird, existieren.

Eine Entscheidungsfunktion $d_0(\boldsymbol{Y})$, die

$$r(d_0, \tau) = \min_{d \in D} [r(d, \tau)]$$

erfüllt, heißt Bayessche Entscheidungsfunktion bezüglich der a-priori-Verteilung P_τ.

Definition 1.14
Eine Entscheidungsfunktion $d_0 \in D$ heißt Minimax-Entscheidungsfunktion, wenn

$$\max_{\theta\in\Omega} R(d_0, \psi) = \min_{d\in D} \max_{\theta\in\Omega} R(d, \psi) \tag{1.32}$$

gilt.

Definition 1.15
Es seien $d_1, d_2 \in D$ Entscheidungsfunktionen für ein bestimmtes Entscheidungsproblem mit der Risikofunktion $R(d, \psi)$ mit $\psi = g(\theta)$, $\theta \in \Omega$. Dann heißt d_1 nicht schlechter als d_2, wenn $R(d_1, \psi) \leqslant R(d_2, \psi)$ für alle $\theta \in \Omega$ gilt, d_1 heißt besser als d_2, wenn neben $R(d_1, \psi) \leqslant R(d_2, \psi)$ für alle $\theta \in \Omega$ für wenigstens ein $\theta^* \in \Omega$ die Ungleichung $R(d_1, \psi^*) < R(d_2, \psi^*)$ mit $\psi^* = g(\theta^*)$ gilt. Eine Entscheidungsfunktion d heißt zulässig in D, wenn es in D keine Entscheidungsfunktion gibt, die besser als d ist. Ist eine Entscheidungsfunktion nicht zulässig, so heißt sie unzulässig.

Eine weitere Darstellung der Entscheidungstheorie ist hier nicht erforderlich. Wir werden in Kapitel 2 die Theorie der Punktschätzungen behandeln, dort ist $d(\boldsymbol{Y}) = S(\boldsymbol{Y})$ eine Entscheidungsfunktion. In der Testtheorie in Kapitel 3 ist $d(\boldsymbol{Y})$ die Wahrscheinlichkeit für die Ablehnung einer Nullhypothese und in der Konfidenzschätzung ein Bereich in Ω, der den Wert θ der Verteilung P_θ mit einer vorgegebenen Wahrscheinlichkeit überdeckt. Auswahlregeln und multiple Vergleichsprozeduren sind andere Spezialfälle von Entscheidungsfunktionen.

1.6 Übungsaufgaben

Aufgabe 1.1
Um das Durchschnittseinkommen der Bewohner einer Großstadt zu schätzen, wird das Einkommen der Besitzer jedes 20. Privatanschlusses in einem Telefonbuch ermittelt.
Handelt es sich bei dieser Stichprobe um eine Zufallsstichprobe der Bevölkerung der Stadt?

Aufgabe 1.2
Aus einer Grundgesamtheit mit den Elementen 1, 2, 3 kann man mit Zurücklegen $3^4 = 81$ verschiedene Stichproben vom Umfang $n = 4$ auswählen. Man schreibe alle möglichen Stichproben auf, berechne $\bar{y}$ und s^2 und stelle die Häufigkeitsverteilung von $\bar{y}$ und s^2 als Streifendiagramm dar.

Aufgabe 1.3

Man beweise, dass die jeweilige Maßzahl $M(\boldsymbol{Y})$ suffizient bezüglich θ ist, wobei $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$ $n \geq 1$ eine Zufallsstichprobe aus einer Grundgesamtheit mit der Verteilung P_θ mit $\theta \in \Omega$ ist, indem man die bedingte Verteilung von $\boldsymbol{Y}$ bei gegebenem $M(\boldsymbol{Y})$ bildet.

a) $M(\boldsymbol{Y}) = \sum_{i=1}^{n} \boldsymbol{y}_i$ und P_θ ist die Poisson-Verteilung mit dem Parameter $\theta \in \Omega \subset R^+$.
b) $M(\boldsymbol{Y}) = (\boldsymbol{y}_{(1)}, \boldsymbol{y}_{(n)})^{\mathrm{T}}$ und P_θ ist die Gleichverteilung im Intervall $(\theta, \theta + 1)$ mit $\theta \in \Omega \subset R^1$.
c) $M(\boldsymbol{Y}) = \boldsymbol{y}_{(n)}$ und P_θ ist die Gleichverteilung im Intervall $(0, \theta)$ mit $\theta \in \Omega = R^+$.
d) $M(\boldsymbol{Y}) = \sum_{i=1}^{n} \boldsymbol{y}_i$ und P_θ ist die Exponentialverteilung mit dem Parameter $\theta \in \Omega = R^+$.

Aufgabe 1.4

Es sei $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$ $n \geq 1$ eine Zufallsstichprobe aus einer Grundgesamtheit mit der Verteilung P_θ, $\theta \in \Omega$. Man bestimme mithilfe des Korollars 1.1 zum Zerlegungssatz eine suffiziente Maßzahl bezüglich θ, wenn P_θ, $\theta \in \Omega$ die Dichtefunktion

a) $f(y, \theta) = \theta y^{\theta - 1}$, $0 < y < 1$; $\theta \in \Omega = R^+$
b) der Weibull-Verteilung

$$f(y, \theta) = \theta a (\theta y)^{a-1} \mathrm{e}^{-(\theta y)^a}, \quad y \geq 0, \quad \theta \in \Omega = R^+, a > 0 \text{ bekannt}$$

c) der Pareto-Verteilung

$$f(y, \theta) = \frac{\theta a^\theta}{y^{\theta+1}}, \quad y > a > 0, \quad \theta \in \Omega = R^+ \text{ bekannt}$$

besitzt.

Aufgabe 1.5

Man bestimme eine minimal suffiziente Maßzahl $M(\boldsymbol{Y})$ für den Parameter θ, wenn $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$ $n \geq 1$ eine Zufallsstichprobe aus einer Grundgesamtheit mit der folgenden Verteilung P_θ ist:

a) geometrische Verteilung mit der Wahrscheinlichkeitsfunktion

$$p(y, p) = p(1 - p)^{y-1}, \quad y = 1, 2, \dots, 0 < p < 1$$

b) hypergeometrische Verteilung mit der Wahrscheinlichkeitsfunktion

$$p(y, M, N, n) = \frac{\binom{M}{y}\binom{N-M}{n-y}}{\binom{N}{n}}, \quad n \in \{1, \dots, N\}$$

c) negative Binomialverteilung mit der Wahrscheinlichkeitsfunktion

$$p(y, p, r) = \binom{y-1}{r-1} p^r (1-p)^{y-r}\,, \quad 0 < p < 1,\ y \geq r \ \text{ ganz},$$
$$r \in \{0, 1, \dots\}$$

und i) $\theta = p$ und b bekannt; ii) $\theta^{\mathrm{T}} = (p, r)$.

d) Betaverteilung mit der Dichtefunktion

$$f(y, \theta) = \frac{1}{B(a, b)} y^{a-1}(1-y)^{b-1}\,, \quad 0 < y < 1\,, \quad 0 < a, b < \infty$$

und i) $\theta = a$ aber b bekannt; ii) $\theta = b$ aber a bekannt.

Aufgabe 1.6

Man beweise, dass die folgenden Verteilungsfamilien $\{P_\theta, \theta \in \Omega\}$ vollständig sind:

a) P_θ ist die Poisson-Verteilung mit dem Parameter $\theta \in \Omega = R^+$.
b) P_θ ist die Gleichverteilung im Intervall $(0, \theta)$, $\theta \in \Omega = R^+$.

Aufgabe 1.7

Es sei $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$ $n \geq 1$ eine Zufallsstichprobe, deren Komponenten im Intervall $(0, \theta)$, $\theta \in \Omega = R^+$ gleichverteilt sind. Man zeige, dass $M(\boldsymbol{Y}) = \boldsymbol{y}_{(n)}$ vollständig suffizient ist.

Aufgabe 1.8

Es besitze $\boldsymbol{y}$ die diskrete Verteilung P_θ mit der Wahrscheinlichkeitsfunktion

$$p(y, \theta) = P(\boldsymbol{y} = y) = \begin{cases} \theta & \text{für } y = -1 \\ (1-\theta)^2 \theta^y & \text{für } y = 0, 1, 2 \end{cases}$$

Man zeige, dass die entsprechende Verteilungsfamilie mit $\theta \in (0, 1)$ beschränkt vollständig, aber nicht vollständig ist.

Aufgabe 1.9

Gegeben sei eine einparametrische Exponentialfamilie mit der Dichte- oder Wahrscheinlichkeitsfunktion

$$f(y, \theta) = h(y) \mathrm{e}^{\{\eta(\theta) M(y) - B(\theta)\}}, \theta \in \Omega$$

a) Man drücke die Fisher-Information dieser Verteilung durch die Funktionen $\eta(\theta)$ und $B(\theta)$ aus.
b) Man benutze das Ergebnis von a) zur Berechnung von $I(\theta)$ für die
   (i) Binomialverteilung mit dem Parameter $\theta = p$,
   (ii) Poisson-Verteilung mit dem Parameter $\theta = \lambda$,
   (iii) Exponentialverteilung mit dem Parameter θ,
   (iv) Normalverteilung $N(\mu, \sigma^2)$ mit $\theta = \sigma, \mu$ fest.

Aufgabe 1.10

Es seien die Voraussetzungen aus Definition 1.11 erfüllt. Außerdem mögen die zweiten Ableitungen $\frac{\partial^2}{\partial\theta_i\partial\theta_j}L(y,\theta)$ für alle $i, j = 1, \dots, p$ und $y \in \{Y\}$ und ihre Erwartungswerte für zufälliges $\boldsymbol{y}$ existieren, und es sei $\int_{\{Y\}} L(y,\theta)\,\mathrm{d}y$ zweimal differenzierbar, wobei Integration und Differentiation vertauscht werden können. Man beweise, dass dann die Elemente der Informationsmatrix in Definition 1.11 die Gestalt

a) $I_{i,j}(\theta) = \operatorname{cov}\left[\frac{\partial}{\partial\theta_i}\ln L(\boldsymbol{y},\theta), \frac{\partial}{\partial\theta_j}\ln L(\boldsymbol{y},\theta)\right]$

b) $I_{i,j}(\theta) = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln L(\boldsymbol{y},\theta)\right]$

besitzen.

Aufgabe 1.11

Es sei $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)^{\mathrm{T}}$ eine Zufallsstichprobe aus einer Grundgesamtheit mit der Verteilung P_θ, $\theta \in \Omega$ und $M(\boldsymbol{Y})$ eine gegebene Maßzahl. Man berechne $E[M(\boldsymbol{Y})]$, $\operatorname{var}[M(\boldsymbol{Y})]$, die Fisher-Information $I(\theta)$ der Verteilung und die Rao-Cramér-Schranke für $\operatorname{var}[M(\boldsymbol{Y})]$.
Gilt in der Rao-Cramér-Ungleichung das Gleichheitszeichen, wenn

a) P_θ die Poisson-Verteilung mit dem Parameter $\theta \in R^+$ und

$$M(\boldsymbol{Y}) = \begin{cases} 1 & \text{für } y = 0 \\ 0 & \text{sonst} \end{cases}$$

ist (hier ist $n = 1$, d. h. $\boldsymbol{y} = \boldsymbol{Y}$);

b) P_θ die Poisson-Verteilung mit dem Parameter $\theta \in R^+$ und $M(\boldsymbol{Y}) = (1 - \frac{1}{n})^{n\bar{y}}$. (Verallgemeinerung von a) auf den Fall $n > 1$) ist;

c) $f(y,\theta) = \theta y^{\theta-1}$, $0 < y < 1$, $\theta \in R^+$ die Dichtefunktion von P_θ und $M(\boldsymbol{Y}) = -\frac{1}{n}\sum_{i=1}^{n}\ln \boldsymbol{y_i}$ ist?

Aufgabe 1.12

In einem Gebiet soll nach Öl gebohrt werden. Der Besitzer der Bohrrechte muss sich auf eine Strategie aus $E = \{E_1, E_2, E_3\}$ festlegen. Dabei bedeute: E_1 – Bohrung wird selbst durchgeführt, E_2 – die Bohrrechte werden verkauft, E_3 – ein Teil der Bohrrechte wird veräußert.
Es ist jedoch nicht bekannt, ob in dem Gebiet tatsächlich Öl vorkommt.
Es sei $\Omega = \{\theta_1, \theta_2\}$ wobei $\theta = \theta_1$ – Öl ist dort vorhanden, $\theta = \theta_2$ – Öl ist dort nicht vorhanden, bedeuten soll. Die Verlustfunktion $L(d,\theta)$ hat für die Entscheidungen $d = E_i$, $i = 1, 2, 3$ und $\theta = \theta_j$, $j = 1, 2$, die Form

	E_1	E_2	E_3
θ_1	0	10	5
θ_2	12	1	6

Die Entscheidung wird aufgrund von Gutachten über die geologischen Verhältnisse in dem Gebiet getroffen: Das Ergebnis der Gutachten sei durch $y \in \{0, 1\}$ gekennzeichnet.
Die Wahrscheinlichkeitsfunktion – in Abhängigkeit von θ – der Zufallsvariablen $\boldsymbol{y}$ sei $p_\theta(y)$ mit den Werten

	$y = 0$	$y = 1$
θ_1	0,3	0,7
θ_2	0,6	0,4

$\boldsymbol{y}$ gibt also die aus dem „Zufallsexperiment" der geologischen Gutachten erhaltene Information über das Vorhandensein ($y = 1$) oder Fehlen ($y = 0$) von Ölvorkommen in dem Gebiet an. Die Menge D der Entscheidungsfunktionen $d(\boldsymbol{y})$ enthalte alle nur möglichen 3^2 diskreten Funktionen:

	1	2	3	4	5	6	7	8	9
$d_i(0)$	E_1	E_1	E_1	E_2	E_2	E_2	E_3	E_3	E_3
$d_i(1)$	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3

a) Man bestimme das Risiko $R(d(\boldsymbol{y}), \theta) = E_\theta[L\{d(\boldsymbol{y}), \theta\}]$ für alle obigen 18 Fälle.
b) Man ermittle die Minimax-Entscheidungsfunktion.
c) Nach Meinung von Experten der Bohrtechnik ist die Wahrscheinlichkeit, bei der Niederbringung einer Bohrung in diesem Gebiet auf Öl zu stoßen, gleich 0,2. Dann kann $\boldsymbol{\theta}$ als Zufallsvariable mit der Wahrscheinlichkeitsfunktion

$\boldsymbol{\theta}$	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$
$\pi(\theta)$	0,2	0,8

betrachtet werden. Man bestimme für jede Entscheidungsfunktion das Bayessche Risiko $r(d_i, \pi)$ und anschließend die Bayessche Entscheidungsfunktion.

Aufgabe 1.13
Es sollen die Behandlungsstrategien beim Einsatz zweier Medikamente M_1 und M_2 beurteilt werden. Drei derartige Strategien stehen zur Verfügung: E_1 – Behandlung mit dem blutdruckerhöhenden Medikament M_1; E_2 – Behandlung ohne Medikamente; E_3 – Behandlung mit dem blutdrucksenkenden Medikament M_2; θ charakterisiert den (geeignet transformierten) Blutdruck eines Patienten: $\theta < 0$ zu niedriger Blutdruck, $\theta = 0$ Blutdruck normal, $\theta > 0$ zu hoher Blutdruck.

Die Verlustfunktion ist folgendermaßen definiert:

	E_1	E_2	E_3
$\theta < 0$	0	c	$b + c$
$\theta = 0$	b	0	b
$\theta > 0$	$b + c$	c	0

Bei einem Patienten wird der Blutdruck gemessen. Die Messung $\boldsymbol{y}$ sei nach $N(\theta, 1)$ verteilt und wird n-mal unabhängig voneinander durchgeführt: $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^{\mathrm{T}}$, aufgrund dieser Stichprobe wird die Entscheidungsfunktion

$$d_{r,s} = \begin{cases} E_1\,, & \text{falls } \bar{y} < r \\ E_2\,, & \text{falls } r \le \bar{y} \le s \\ E_3\,, & \text{falls } \bar{y} > s \end{cases}$$

definiert.

a) Man bestimme die Risikofunktion $R(d_{r,s}(\bar{\boldsymbol{y}}), \theta) = E\{L[d_{r.s}(\bar{\boldsymbol{y}}), \theta]\}$.
b) Man skizziere die Risikofunktion im Fall $b = c = 1$, $n = 1$ für i) $r = -s = -1$; ii) $r = -\frac{1}{2}s = -1$.

Für welche Werte von θ ist die Entscheidungsfunktion $d_{-1,1}(y)$ der Funktion $d_{-1,2}(y)$ vorzuziehen?

Literatur

Bahadur, R.R. (1955) Statistics and subfields. *Ann. Math. Stat.*, **26**, 490–497.

Blackwell, D. (1947) Conditional expectations and unbiased sequential estimation. *Ann. Math. Stat.*, **18**, 105–110.

Cochran, W.G. und Boing, W. (1972) *Stichprobenverfahren*, De Gruyter, Berlin, New York.

Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh.

Halmos, P.R. und Savage, L.J. (1949) Application of the Radon-Nykodin theorem to the theory of sufficient statistics. *Ann. Math. Stat.*, **20**, 225–241.

Kauermann, G. und Küchenhoff, H. (2011) *Stichproben: Methoden und praktische Umsetzung mit R*, Springer, Heidelberg.

Kullback, S. und Leibler, R.A. (1951) On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86.

Lehmann, E.L. und Romano, J.P. (2008) *Testing Statistical Hypothesis*, Springer, Heidelberg.

Lehmann, E.L. und Scheffé, H. (1950) Completeness, similar regions and unbiased estimation. *Sankhya*, **10**, 305–340.

Quatember, A. (2014) *Datenqualität in Stichprobenerhebungen*, Springer, Berlin.

Rao, C. R. (1945) Information and accuracy attainable in estimation of statistical parameters. *Bull. Calc. Math. Soc.*, **37** (3), 81–91.

Rasch, D. (1995) *Mathematische Statistik*, Joh. Ambrosius Barth, Berlin, Heidelberg

Rasch, D., Tiku, M.L. und Sumpf, D. (Hrsg.) (1994) *Elsevier's Dictionary of Biometry*, Elsevier, Amsterdam, London, New York.

Rasch, D., Herrendörfer, G., Bock, J., Victor, N. und Guiard, V. (Hrsg.) (2008) *Verfahrensbibliothek Versuchsplanung und -aus-*

wertung, 2. verbesserte Auflage in einem Band mit CD, R. Oldenbourg, München, Wien
(frühere Auflagen mit den Herausgebern Rasch, Herrendörfer, Bock, Busch (1978, 1981), Deutscher Landwirtschaftsverlag Berlin und (1995, 1996) Oldenbourg, München Wien).

Stigler, S.M. (1986, 1990) *The History of Statistics: The Measurement of Uncertainty Before 1900*, Harvard University Press, Cambridge.