4 Beurteilende Statistik oder Inferenz

4.1 Aufgaben der beurteilenden Statistik

Wenn die Verteilungsfunktion einer Zufallsvariablen mit allen ihren Parametern, wie Mittelwert, Streuung, weitere Momente usw. bekannt ist, lassen sich alle Fragen über die Wahrscheinlichkeiten von zufälligen Ereignissen, die die zu untersuchende Zufallsvariable betreffen, beantworten. In der Praxis ist das leider selten der Fall. Meist ist die Verteilungsfunktion der betreffenden Zufallsvariablen nicht vollständig bekannt. Wenn bisher in den Modellrechnungen bei normalverteilten Zufallsvariablen stets Mittelwert und Varianz als bekannt vorausgesetzt wurden, so ist der Informationsstand in der Praxis nicht immer so umfassend.

Mittelwert und Varianz sind in der Regel unbekannt. Selbst die Art der Verteilungsfunktion ist meist nur eine grobe Modellannahme, mit der man zunächst operiert und die man korrekterweise erst wieder überprüfen müßte. Der Statistiker gewinnt nun seine Information über die unbekannte Verteilungsfunktion aus den Beobachtungen der Zufallsvariablen. Er könnte sich rein theoretisch über die Verteilungsfunktion Klarheit verschaffen, indem er sich unendlich viele Realisationen der Zufallsvariablen anschaut, oder anders ausgedrückt, indem er die Grundgesamtheit betrachtet. Das ist jedoch aus vielen Gründen nicht möglich. In der Praxis wird man sich immer mit einigen wenigen Realisationen der Zufallsvariablen, eben mit einer Stichprobe aus der Grundgesamtheit aller hypothetisch möglichen Realisationen zufrieden geben müssen und versuchen, aus dieser Stichprobe Rückschlüsse auf die Grundgesamtheit oder die Verteilungsfunktion der Zufallsvariablen zu ziehen. Dazu muß man aber wissen, wie sich die interessierenden Eigenschaften (z.B. der Mittelwert) aller überhaupt möglichen Stichproben vom Umfang n verteilen. Die Wahrscheinlichkeitstheorie liefert das Rüstzeug, um Methoden und Verfahren zur statistischen Analyse von Beobachtungswerten und Zufallsvariablen anzugeben. Damit kommt man zum eigentlichen Kernpunkt der Statistik, zur **beurteilenden Statistik** oder **Inferenz**. Der große Statistiker R.A. Fisher (1890–1962) hat in den 50er Jahren herauskristallisiert, was man unter beurteilender Statistik oder Inferenz versteht. Er meint, die Aufgabe des Statistikers ist, aufgrund von empirischen Beobachtungen einen objektiven Induktionsschluß durchzuführen. Die empirischen Beobachtungen sind als Stichprobenwerte aufzufassen. Mit Hilfe dieses Induktionsschlusses soll eine Aussage über die Grundgesamtheit oder über die entsprechende Verteilung gewonnen werden, um aus der Situation der Ungewißheit oder der unvollständigen Information bzgl. der Grundgesamtheit soweit als möglich herauszukommen. Wenn man das zusammenfaßt, kommt man in etwa wieder zu der Definition der Statistik, die bereits in der Einleitung gegeben wurde: Beurteilende Statistik oder Inferenz ist die Überwindung der Ungewißheit durch induktive Schlüsse aufgrund von empirischen Beobachtungen. In diesem Zusammenhang sei noch ein Wort zur groben Unterscheidung zwischen Statistik und Wahrscheinlichkeitstheorie erwähnt. Bereits früher wurde grob unterschieden: Die Wahrscheinlichkeitsrechnung lehrt, wie man mit Wahrscheinlichkeiten rechnet, die Statistik zeigt, wie man Wahrscheinlichkeiten numerisch bestimmt. Die Wahrscheinlichkeitstheorie liefert das mathematische Rüstzeug für die Statistik. Insofern gehört die Wahrscheinlichkeitstheorie zur Mathematik, denn sie ist eine axiomatische Wissenschaft, die aufbauend aus postulierten, unbewiesenen oder nicht beweisbaren, weil evidenten Axiomen Lehrsätze folgert, die also alle richtig sind, solange die Axiome aufrechterhalten werden. Die Statistik hat dagegen ein ganz bestimmtes praktisches Ziel im Auge, nämlich mehr Information über eine unbekannte Grundgesamtheit aufgrund einer Stichprobe zu erfahren oder die Ungewißheit zu überwinden.

Methodisch besteht zwischen der Statistik und der Wahrscheinlichkeitstheorie folgender Unterschied: (vgl. Bild 4.1): In der Wahrscheinlichkeitstheorie wird deduziert, und zwar von bekannten bzw. als bekannt vorausgesetzten Verteilungen oder Modellen auf Realisationen. Wenn man z.B. eine bestimmte Verlustwahrscheinlichkeit in der Kälbermast kennt oder unterstellt, dann kann man das bekannte Bernoulli-Modell heranziehen und die Wahrscheinlichkeit berechnen, daß von beispielsweise 50 Tieren genau $0, 1, 2, \ldots$ usw. Tiere sterben. In der Statistik geht man in der Regel umgekehrt vor. Man schließt von gegebenen Beobachtungen oder Realisationen auf unbekannte Verteilungen und ihre Parameter. In dem eben zitierten Beispiel würde der Statistiker einen Kälbermastversuch durchführen und aus der beobachteten Verlustquote auf die unbekannte Verlustwahrscheinlichkeit schließen. Die beurteilende Statistik hat dabei im wesentlichen zwei Aufgaben:

- Die Schätzung unbekannter Parameter der Grundgesamtheit
- Die Prüfung von Hypothesen über diese Parameter

Es ist jedoch nicht so, daß in der Statistik überhaupt keine Deduktionen vorkommen. Die Betonung liegt im folgenden allerdings auf der induktiven Methode, d.h. auf Schlußweisen von Beobachtungen auf Grundgesamtheiten.

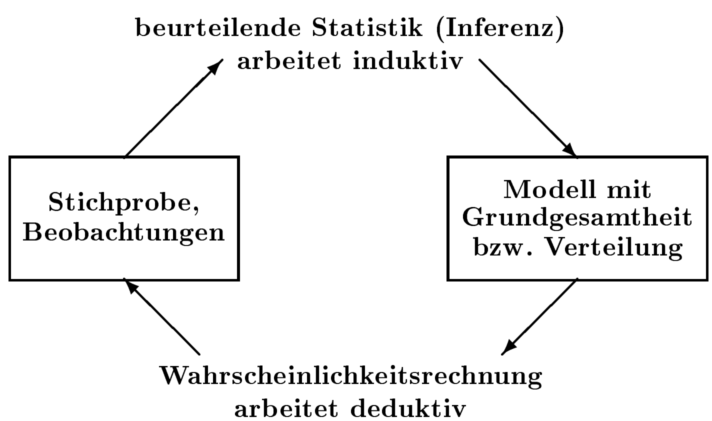

Bild 4.1: Abgrenzung Wahrscheinlichkeitstheorie und Statistik

4.2 Der Begriff der Stichprobe

Der Begriff der Stichprobe soll noch einmal aufgegriffen werden. Bisher wurde darunter eine endliche Teilmenge von Dingen oder ihre entsprechenden Merkmalswerte aus einer gegebenen Menge oder Grundgesamtheit verstanden. In diesem Zusammenhang muß als Grundgesamtheit auch eine unendlich große Menge zugelassen werden, z.B. die in Gedanken unendlich oft durchführbare Realisation eines Zufallsexperiments. Damit werden also auch hypothetische Grundgesamtheiten zugelassen.

Wenn nun eine Stichprobe eine gewisse Information über die Grundgesamtheit liefern soll, dann muß die Stichprobe erstens eine Zufallsauswahl darstellen, d.h. jedes Element der Grundgesamtheit muß die gleiche Chance oder Wahrscheinlichkeit haben, in die Stichprobe zu kommen. Man kann z.B. nicht nur die besten Ergebnisse eines mehrmals durchgeführten Versuchs zur Auswertung heranziehen. Man nehme an, alle landwirtschaftlichen Betriebe in Bayern bilden eine Grundgesamtheit. Eine repräsentative Stichprobe kann dann nicht ausschließlich aus Betrieben des Gäubodens bestehen. In einem solchen Fall würde man von einer willkürlichen Auswahl sprechen. Vor der Beantwortung der Frage, wie man eine Zufallsauswahl bewerkstelligt, soll kurz die Frage behandelt werden, warum man überhaupt mit Stichproben arbeiten muß. Die wichtigsten Gründe für die Anwendung von Stichprobenverfahren sind ökonomischer Natur. Um Kosten zu sparen und Zeit zu gewinnen, macht man nur einige wenige Versuche, greift nur einige Tiere aus einer Population heraus. Interessiert eine im Meer lebende Fischart, kann man nicht alle Exemplare dieser Gattung fangen. Bei Unterstellung einer unendlichen Grundgesamtheit (unbegrenzte Durchführung eines Zufallsexperiments) kann man aus theoretischen Gründen nicht alle Elemente der Grundgesamtheit zur Untersuchung heranziehen. Aber auch bei einer endlichen Grundgesamtheit wird man prinzipielle Einwände gegen eine Untersuchung der ganzen Grundgesamtheit geltend machen, wenn z.B. Lebensdaueruntersuchungen durchgeführt werden.

Zweitens wird vorausgesetzt, daß die n Ausführungen des Experiments, die n Ziehungen, welche die Stichprobenergebnisse liefern, voneinander unabhängig sind. Diese Voraussetzung ist notwendig, weil jeder Ziehung eine Stichprobenvariable als zufällige Variable zugeordnet wird und bei der Ableitung von Verteilungsgesetzen stets n unabhängige Zufallsvariablen vorausgesetzt werden. Bei einer unendlichen Grundgesamtheit, insbesondere bei einer hypothetischen Grundgesamtheit ist die Unabhängigkeit der einzelnen Ziehungen unmittelbar einsichtig. Eine unendliche Grundgesamtheit wird in ihrer Zusammensetzung nicht verändert, wenn man endlich viele Elemente herausgreift. Etwas kritischer ist diese Frage zu beantworten, wenn eine endliche Grundgesamtheit vorliegt. Prinzipiell sind die Stichprobenvariablen nicht mehr unabhängig, wenn die Grundgesamtheit endlich ist und die Stichprobenelemente nicht mehr zurückgelegt werden. Diese Situation findet man in der Praxis vor, wenn man aus irgendwelchen Karteien, Listen oder Produktionen n Elemente auswählt, um einen Schluß auf die Gesamtheit zu ziehen. Für solche Fälle gibt es nun eine ganze Reihe von Stichprobenauswahlverfahren (vgl. Heinhold/Gaede), z.B. reine oder geschichtete Zufallsauswahl, Klumpenauswahlverfahren usw. Hier soll nicht näher auf diese Verfahren eingegangen werden, sondern die Unabhängigkeit einer Stichprobe aus einer endlichen Grundgesamtheit angenommen werden, wenn die Grundgesamtheit gegenüber der Stichprobe sehr groß ist oder anders formuliert, wenn es aus der Sachlage heraus klar ist, daß durch die Entnahme von n Elementen die Zusammensetzung der Grundgesamtheit (z.B. bzgl. der Eigenschaft "gut", "schlecht") nicht wesentlich verändert wird.

Die beiden Stichprobeneigenschaften Zufälligkeit und Unabhängigkeit sollen von nun an stets als gegeben vorausgesetzt werden. Üblicherweise spricht man dann von **Zufallsstichproben**. Führt man ein Experiment, z.B. einen Mastversuch oder einen Sortenversuch, n-mal unabhängig voneinander unter gleichen Bedingungen durch, wobei jeweils eine Ergebnisgröße X notiert wird, dann sind die Ergebnisse $x_1, x_2, \ldots, x_n$ als Zufallsstichprobe anzusehen. Die Unabhängigkeit ist durch die Versuchsbedingungen festgelegt. Die Zufälligkeit ist prinzipiell dadurch gewährleistet, daß der Ausgang jedes Versuchs vom Zufall beeinflußt wird, und daß man n aufeinanderfolgende Resultate nimmt und nicht willkürlich Resultate auswählt. Die dazugehörige Grundgesamtheit ist hypothetisch und unendlich groß.

Wie bewerkstelligt man nun eine Zufallsauswahl vom Umfang n aus einer vorliegenden endlichen Grundgesamtheit z.B. N Personen, Tiere oder Betriebe? Hier kann man alle Elemente der Grundgesamtheit numerieren und dann versuchen n Zahlen zufällig auszuwählen. Das kann durch Lose geschehen, die man in eine Urne steckt, um dann eben n Lose zu ziehen. Man könnte aber auch einen zehnseitigen Würfel verwenden und die gewürfelten Zahlen als Stichprobe aus der numerierten Grundgesamtheit heranziehen. Auf diese Weise kann man sich sog. **Zufallszahlen** erzeugen. Es existieren Tabellen von solchen Zufallszahlen, die einem das Würfeln mit einem zehnseitigen Würfel ersparen (vgl. Tab. A.8 im Anhang).

Hat man die Aufgabe, aus einer Grundgesamtheit von N Elementen eine zufällige Stichprobe vom Umfang n auszuwählen, so kann man folgendermaßen vorgehen: Man ordnet den N Elementen die Zahlen 1 bis N zu. Dann wählt man eine beliebige Ziffer der Tabelle als Anfangspunkt und liest ab dort fortlaufend jeweils Gruppen von s Ziffern ab, wenn N eine s-stellige Zahl ist. Ist die in der Tabelle der Zufallszahlen abgelesene Zahl kleiner oder gleich N, dann wird das entsprechende Element der Grundgesamtheit in die Zufallsstichprobe aufgenommen, andernfalls nicht. Man fährt solange mit dieser Methode fort, bis man n Elemente ausgewählt hat.

Beispiel:

Aus einer Grundgesamtheit von $N = 7520$ (numerierten) Elementen ist eine Stichprobe vom Umfang $n = 25$ auszuwählen. Als Startziffer sei die Ziffer 5 in der 19. Zeile in der 4. Dreiergruppe der Tab. A.8 im Anhang ausgewählt. Die erste Vierergruppe ist die Zahl 5413. Man liest zeilenweise weiter und notiert alle Vierergruppen, die kleiner oder gleich 7520 sind:

5413	4922	2976	2774	4666	3370	32	4083	3680	648
284	3543	6292	7299	1616	1105	2451	5441	2508	2106
3044	3707	4644	3181	4355					

Zufallszahlen können in MINITAB durch das Kommando `random` mit der Option `integer` erzeugt werden. Die allgemeine Syntax lautet:

`RANDOM` je `K` Werte in die Spalten `C,...,C`;
  `INTEGER` $a =$ `K` $b =$ `K`.

In die Spalten C,...,C werden gleichverteilte diskrete Zufallszahlen geschrieben. Die Wahrscheinlichkeit ist für die ganzen Zahlen $a, a+1, \ldots b$ gleich groß und für alle anderen Zahlen gleich 0.

Damit kann die Stichprobe des vorangehenden Beispiels auf einfache Weise mit MINITAB gezogen werden, wenn man sich gleichverteilte Zufallszahlen zwischen 1 und dem Umfang der Grundgesamtheit $N = 7520$ erzeugt.

```
MTB > random 25 c1;
SUBC> integer 1 7520.
MTB > print c1

C1
  4477   3127   7313   4135   5397   5270   4421   3650   5002
  1056   4158    814   3930   2340   6727   6096   2441   4274
   257   1928    290   6060   5455   5021   3379
```

4.3 Der Hauptsatz der Statistik und der Begriff der Stichprobenvariablen

In vielen Fällen wird man sich eine hypothetische, unendlich große Menge von möglichen Merkmalswerten als Grundgesamtheit vorstellen, aus der man eine Stichprobe vom Umfang n gezogen hat. Die Realisationen der Stichprobe seien die Werte $x_1, x_2, \ldots, x_n$. Man kann nun folgende Erfahrungstatsache beobachten. Die relative Summenhäufigkeitsfunktion $\widetilde{F}_n(x)$, die man aus diesen n Werten $x_1, x_2, \ldots, x_n$ ermitteln kann, also die Funktion $\widetilde{F}_n(x) = \dfrac{m_n(x)}{n}$, wenn $m_n(x)$ die Anzahl der x_i ist mit $x_i \leq x$, nähert sich mit wachsendem n einer Verteilungsfunktion $F(x)$ an: $\widetilde{F}_n(x) \approx F(x)$. Diese Verteilungsfunktion $F(x)$ sieht man als kennzeichnendes Charakteristikum der Grundgesamtheit an, aus der die Stichprobe entnommen wurde. Man sagt auch: Die Grundgesamtheit ist nach $F(x)$ verteilt. In der Regel ist $F(x)$ unbekannt und man wird mit Hilfe der Stichprobe versuchen, die Ungewißheit über $F(x)$ in irgendeiner Weise zu überwinden. Das ist der Inhalt des sog. **Hauptsatzes der Statistik**.

Man interpretiert dazu die beobachteten Werte x_i $(i = 1, 2, \ldots, n)$ als Realisationen von n stochastisch unabhängigen Zufallsgrößen X_i, die alle dieselbe Verteilungsfunktion $F(x)$ besitzen. Man sagt auch, daß die Zufallsgrößen $X_1, X_2, \ldots, X_n$ identisch verteilt sind mit der Verteilungsfunktion $F(x)$. Die Stichprobenvariable X_i ist als die Zufallsvariable aufzufassen, die man der i-ten Ziehung oder der i-ten Wiederholung eines Experiments zuordnen kann.

4.4 Testverteilungen

Im folgenden Abschnitt werden einige Verteilungen eingeführt, die im Zusammenhang von statistischen Rückschlüssen auf die Grundgesamtheit aufgrund von Stichproben eine wichtige Rolle spielen. Es sind Verteilungen von Zufallsgrößen, deren Realisationen nicht unmittelbar als Ergebnisse von Zufallsexperimenten beobachtet werden, sondern es handelt sich um die Verteilung abgeleiteter Größen, z.B. der Verteilung der Quadratsumme von n normalverteilten Zufallsgrößen. Solche und ähnliche Verteilungen heißen auch **Testverteilungen** oder **Prüfverteilungen**, weil man sie zur Überprüfung von Hypothesen verwendet. In diesem Zusammenhang sollen einige Tatsachen und Sätze zitiert werden, auf die entsprechenden Beweise wird aber verzichtet. Der an Beweisen interessierte Leser möge diese z.B. bei Heinhold/Gaede nachvollziehen.

4.4.1 χ^2-Verteilung

Es seien n unabhängige normalverteilte Zufallsvariablen $X_1, X_2, \ldots, X_n$ gegeben mit dem Mittelwert $\mu = 0$ und der Varianz $\sigma^2 = 1$. Die Summe der Quadrate dieser Zufallsvariablen wird mit χ^2 (sprich: Chi-Quadrat) bezeichnet:

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2 \tag{4.1}$$

Die Funktion χ^2 stellt selbst wieder eine Zufallsvariable dar, die eine bestimmte Dichte- (Bild 4.2) und Verteilungsfunktion besitzt.

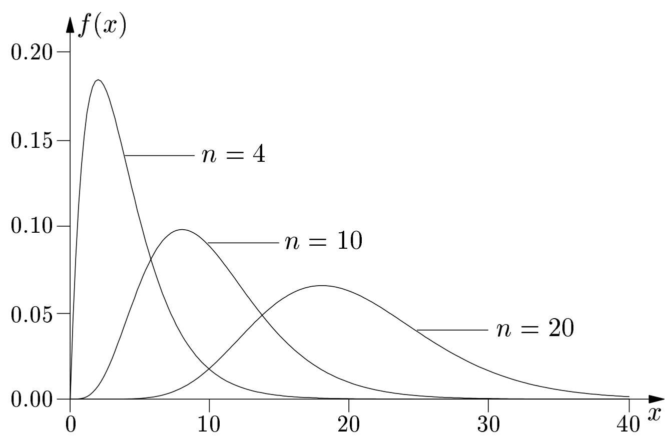

Bild 4.2: Dichtefunktionen der χ^2-Verteilung für verschiedene n

Es wird hier darauf verzichtet, diese Funktionen durch Formeln genauer zu beschreiben. Allgemein heißt die Verteilung einer Zufallsvariablen, die als Summe von n quadrierten normalverteilten Zufallsvariablen gebildet wird, **Chi-Quadrat-Verteilung**. Hierbei ist n eine positive ganze Zahl und gibt an, wieviele unabhängige $(0, 1^2)$-normalverteilte Zufallsvariablen $X_1, X_2, \ldots, X_n$ in die Größe χ^2 eingehen. n heißt auch die **Anzahl der Freiheitsgrade** der Verteilung. In Bild 4.2 ist jeweils die Dichtefunktion für einige Freiheitsgrade n skizziert. Es ist klar, daß für negative x die Dichtefunktion jeweils den

Wert Null hat, da die Größe χ^2 nur Werte größer oder gleich Null annehmen kann. Wichtige Fraktilenwerte sind für verschiedene Freiheitsgrade in Tab. A.3 des Anhangs festgehalten. Die χ^2-Verteilung der Zufallsgröße χ^2 hat den Erwartungswert $\mu = n$ und die Varianz $\sigma^2 = 2n$. Eine wichtige Eigenschaft der χ^2-Verteilung ist, daß man sie für großes n durch die Normalverteilung approximieren kann. Man kann zeigen (vgl. Heinhold/Gaede):

- Die Zufallsvariable χ^2 ist für $n \to \infty$ normalverteilt mit dem Erwartungswert $\mu = n$ und der Varianz $\sigma^2 = 2n$. Wenn man also mit $F(x)$ die Verteilungsfunktion der Zufallsvariablen χ^2 bezeichnet, so gilt für großes n:

$$F(x) \approx \phi\left(\frac{x \Leftrightarrow n}{\sqrt{2n}}\right) \tag{4.2}$$

- Die Zufallsvariable $\sqrt{2\chi^2}$ ist für $n \to \infty$ normalverteilt mit dem Erwartungswert $\mu = \sqrt{2n \Leftrightarrow 1}$ und der Varianz $\sigma^2 = 1$. Wenn man also mit $G(x)$ die Verteilungsfunktion der Zufallsvariablen $\sqrt{2\chi^2}$ bezeichnet, so gilt für großes n:

$$G(x) \approx \phi\left(x \Leftrightarrow \sqrt{2n \Leftrightarrow 1}\right) \tag{4.3}$$

Eine weitere wichtige Eigenschaft der χ^2-Verteilung ist ihre Additivität. Sie besagt, wenn zwei unabhängige Zufallsgrößen X und Y jeweils χ^2-verteilt sind mit n bzw. m Freiheitsgraden, so ist die Summe $X+Y$ ebenfalls χ^2-verteilt mit $n+m$ Freiheitsgraden.

Die Hauptanwendungsgebiete der χ^2-Verteilung sind erstens die Verteilung der Stichprobenvarianz S^2 (siehe unten) und zweitens die Analyse von Vier- und Mehrfeldertafeln in der sog. **Kontingenztafelanalyse**.

Die χ^2-Verteilung hängt sehr eng mit der Verteilung der empirischen Varianz $S^2 = \dfrac{1}{n \Leftrightarrow 1} \cdot \sum_{i=1}^{n}(X_i \Leftrightarrow \overline{X})^2$ einer Grundgesamtheit zusammen. Zunächst kann folgendes festgehalten werden: Zieht man aus einer beliebig (nicht unbedingt normal-) verteilten Grundgesamtheit mit der Varianz σ^2 eine Stichprobe $X_1, X_2, \ldots, X_n$ vom Umfang n, so ist S^2 eine erwartungstreue (vgl. 4.5.2) Schätzfunktion für σ^2, bzw. $s^2 = \dfrac{1}{n \Leftrightarrow 1} \cdot \sum_{i=1}^{n}(x_i \Leftrightarrow \overline{x})^2$ ist ein unverzerrter Schätzwert für das unbekannte σ^2. Dies gilt wohlgemerkt ohne besondere Voraussetzung über die Form der Verteilung in der Grundgesamtheit. Es sei noch festgehalten, daß man den unverzerrten Schätzwert s^2 für σ^2 einer Stichprobe berechnen kann, ohne irgendwelche Kenntnis über den Erwartungswert der Grundgesamtheit. Mit Hilfe der soeben eingeführten χ^2-Verteilung kann man die Verteilung von S^2 bestimmen. Man kann nämlich zeigen, daß die Zufallsgröße $\dfrac{(n \Leftrightarrow 1) \cdot S^2}{\sigma^2}$ eine χ^2-verteilte Zufallsgröße vom Freiheitsgrad $n \Leftrightarrow 1$ (kurz: $\boldsymbol{\chi^2}_{n-1}$) ist, unter der Voraussetzung, daß die X_i aus einer (μ, σ^2)-normalverteilten Grundgesamtheit kommen. Also:

$$\frac{(n \Leftrightarrow 1) \cdot S^2}{\sigma^2} = \boldsymbol{\chi^2}_{n-1} \quad \Leftrightarrow \quad S^2 = \frac{\sigma^2}{n \Leftrightarrow 1} \cdot \boldsymbol{\chi^2}_{n-1} \tag{4.4}$$

Außerdem kann man zeigen, daß S^2 und $\overline{X}$ unabhängige Zufallsgrößen sind. Diese besondere Eigenschaft wird sich noch im Zusammenhang mit der t-Verteilung als wichtig erweisen.

Die Tatsache, daß $\dfrac{(n \Leftrightarrow 1) \cdot S^2}{\sigma^2}$ $\boldsymbol{\chi^2}_{n-1}$-verteilt ist, liefert folgende Wahrscheinlichkeitsaussage, wenn man mit $\chi^2_{n-1;\alpha/2}$ und $\chi^2_{n-1;1-\alpha/2}$ die entsprechenden Fraktilen der χ^2-Verteilung bezeichnet:

$$
\begin{aligned}
&P\left(\chi^2_{n-1;\alpha/2} \leq \frac{(n \Leftrightarrow 1) \cdot S^2}{\sigma^2} \leq \chi^2_{n-1;1-\alpha/2}\right) = 1 \Leftrightarrow \alpha \qquad \text{bzw.} \\
&P\left(\frac{(n \Leftrightarrow 1) \cdot S^2}{\chi^2_{n-1;1-\alpha/2}} \leq \sigma^2 \leq \frac{(n \Leftrightarrow 1) \cdot S^2}{\chi^2_{n-1;\alpha/2}}\right) = 1 \Leftrightarrow \alpha
\end{aligned}
\tag{4.5}
$$

Mit Hilfe dieser Wahrscheinlichkeitsaussage kann später ein sog. Vertrauensintervall für die unbekannte Varianz σ^2 der Grundgesamtheit angegeben werden.

Fraktilen der χ^2-Verteilung sind in Tab. A.3 im Anhang für verschiedene Freiheitsgrade n tabelliert.

In MINITAB können die Fraktilen der χ^2-Verteilung mit dem Kommando `invcdf` und der Option `chisquare` berechnet werden. Die allgemeine Syntax ist:

`INVCDF` für Werte in `E` [nach `E`];
  `CHISQUARE` $n =$ `K`.

`E` kann eine Konstante oder eine Spalte sein. n ist die Anzahl der Freiheitsgrade.

```
MTB > invcdf 0.95;
SUBC> chisquare 23.
    0.9500   35.1724
```

Die 95%-Fraktile der χ^2-Verteilung mit 23 Freiheitsgraden ist also $\chi^2_{23;0.95} \approx 35.17$.

Eine Tabelle der χ^2-Fraktilen für verschiedene Wahrscheinlichkeiten $F(x)$ und Freiheitsgrade FG erzeugt folgende Befehlsfolge:

```
MTB > set c1
DATA> 0.05 0.1 0.5 0.9 0.95
DATA> end
MTB > invcdf c1 c2;
SUBC> chisquare 1.
MTB > invcdf c1 c3;
SUBC> chisquare 5.
MTB > invcdf c1 c4;
SUBC> chisquare 10.
MTB > name c1 'F(x)' c2 '1 FG' c3 '5 FG' c4 '10 FG'
MTB > print c1-c4
```

ROW	F(x)	1 FG	5 FG	10 FG
1	0.05	0.00393	1.1455	3.9403
2	0.10	0.01579	1.6103	4.8652
3	0.50	0.45494	4.3515	9.3418
4	0.90	2.70554	9.2364	15.9872
5	0.95	3.84146	11.0705	18.3070

4.4.2 *t*-Verteilung oder Student-Verteilung

Eine weitere Testverteilung, die bei einer Reihe wichtiger Tests gebraucht wird, ist von W.S. Gosset unter dem Pseudonym Student veröffentlicht worden. Sie ist folgendermaßen definiert. Es sei X eine $(0, 1^2)$-normalverteilte Zufallsvariable, Q sei eine von X unabhängige χ^2-verteilte Zufallsvariable vom Freiheitsgrad n. Dann nennt man die Verteilung der Zufallsgröße

$$\boldsymbol{T}_n = \frac{X}{\sqrt{Q/n}} \tag{4.6}$$

Student-Verteilung oder kurz ***t*-Verteilung** vom Freiheitsgrad n. Die Studentverteilung ist im Gegensatz zur χ^2-Verteilung eine symmetrische Verteilung, weil X eine zu $x = 0$ symmetrische Verteilung ist.

Eine studentverteilte Zufallsgröße hat für $n = 2, 3, \ldots$ aus Symmetriegründen den Mittelwert 0. Für $n = 1$ existiert kein Mittelwert. Für $n = 1$ und $n = 2$ hat eine studentverteilte Zufallsgröße keine Varianz. Für $n = 3, 4, \ldots$ ist die Varianz jeweils $\sigma^2 = \dfrac{n}{n \Leftrightarrow 2}$. Es gilt außerdem, daß für $n \to \infty$ die Studentverteilung in eine $(0, 1^2)$-Normalverteilung übergeht. In Bild 4.3 sind die Dichtefunktionen für einige Freiheitsgrade skizziert.

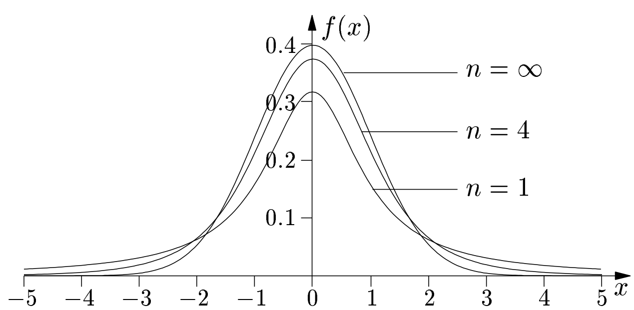

Bild 4.3: Dichtefunktionen der t-Verteilung für verschiedene n

In Tab. A.4 des Anhangs sind wichtige Fraktilenwerte für verschiedene Freiheitsgrade tabelliert.

Mit MINITAB können die Fraktilen der t-Verteilung mit dem Kommando `invcdf` und der Option `t` berechnet werden. Die allgemeine Syntax ist:

`INVCDF` für Werte in `E` [nach `E`];
`T` n = `K`.

`E` kann eine Konstante oder eine Spalte sein. n ist die Anzahl der Freiheitsgrade.

```
MTB > invcdf 0.99;
SUBC> t 7.
    0.9900    2.9980
```

Die 99%-Fraktile der t-Verteilung mit 7 Freiheitsgraden ist also $t_{7;0.99} \approx 3$.

Eine Tabelle der t-Fraktilen für verschiedene Wahrscheinlichkeiten $F(x)$ und Freiheitsgrade FG erzeugt folgende Befehlsfolge:

```
MTB > set c1
DATA> 0.9 0.95 0.975 0.99 0.995 0.999
DATA> end
MTB > name c1 'F(x)'
MTB > invcdf 'F(x)' c2;
SUBC> t 1.
MTB > invcdf 'F(x)' c3;
SUBC> t 10.
MTB > invcdf 'F(x)' c4;
SUBC> t 100.
MTB > name c2 '1 FG' c3 '10 FG' c4 '100 FG'
MTB > print c1-c4

 ROW    F(x)      1 FG     10 FG    100 FG

   1   0.900     3.078   1.37220   1.29007
   2   0.950     6.314   1.81246   1.66024
   3   0.975    12.706   2.22814   1.98398
   4   0.990    31.821   2.76378   2.36423
   5   0.995    63.657   3.16928   2.62592
   6   0.999   318.317   4.14371   3.17379
```

4.4.3 $\boldsymbol{F}$-Verteilung

X^2 sei eine χ^2-verteilte Zufallsgröße mit dem Freiheitsgrad m. Y^2 sei eine von X^2 unabhängige, ebenfalls χ^2-verteilte Zufallsgröße mit dem Freiheitsgrad n. Dann heißt die Verteilung der Zufallsgröße

$$\boldsymbol{F}_{m,n} = \frac{X^2/m}{Y^2/n} \tag{4.7}$$

$\boldsymbol{F}$-Verteilung mit den Freiheitsgraden m, n. In Bild 4.4 sind Dichtefunktionen F-verteilter Grundgesamtheiten für einige Freiheitsgrade gezeichnet.

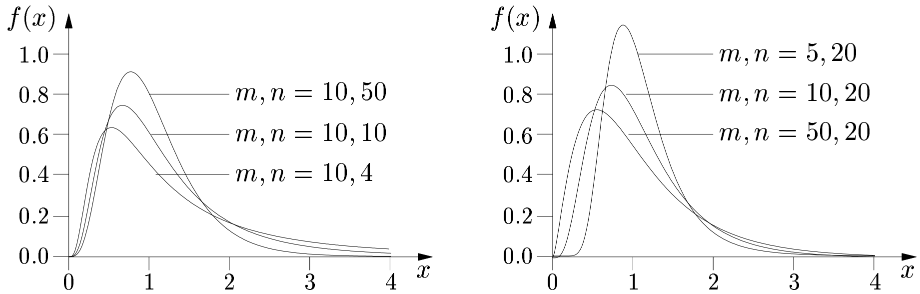

Bild 4.4: Dichtefunktionen der F-Verteilung für verschiedene Freiheitsgrade

In den Tabellen A.5 – A.7 des Anhangs sind die wichtigsten Fraktilen der F-Verteilung angegeben. Die Fraktilen zum Niveau $1 \Leftrightarrow \alpha$ einer F-Verteilung mit m Zähler- und n Nennerfreiheitsgraden, kurz einer $F_{m,n}$-Verteilung, sollen mit $F_{m,n;1-\alpha}$ bezeichnet werden. Es gilt:

$$P(\boldsymbol{F}_{m,n} \le F_{m,n;1-\alpha}) = 1 \Leftrightarrow \alpha \quad \text{und} \quad P(\boldsymbol{F}_{n,m} \ge F_{n,m;\alpha}) = 1 \Leftrightarrow \alpha \tag{4.8}$$

Es gilt zwischen $\boldsymbol{F}_{m,n}$ und $\boldsymbol{F}_{n,m}$ folgende Beziehung:

$$\frac{1}{\boldsymbol{F}_{m,n}} = \boldsymbol{F}_{n,m}. \tag{4.9}$$

Damit ergibt sich:

$$P\left(\frac{1}{\boldsymbol{F}_{m,n}} \ge F_{n,m;\alpha}\right) = 1 \Leftrightarrow \alpha \tag{4.10}$$

oder

$$P\left(\boldsymbol{F}_{m,n} \le \frac{1}{F_{n,m;\alpha}}\right) = 1 \Leftrightarrow \alpha \tag{4.11}$$

Vergleicht man (4.11) mit (4.8), so erhält man:

$$F_{m,n;1-\alpha} = \frac{1}{F_{n,m;\alpha}} \tag{4.12}$$

Mit dieser Gleichung kann man aus den $(1 \Leftrightarrow \alpha)$-Fraktilen die dazu komplementären α-Fraktilen ausrechnen. Man beachte allerdings, daß sich Zähler- und Nennerfreiheitsgrade vertauschen.

Die F-Verteilung wird in der statistischen Praxis in den meisten Fällen als Verteilung des Verhältnisses zweier Varianzen verwendet. Seien $X_1, X_2, \ldots, X_m$ und $Y_1, Y_2, \ldots, Y_n$ zwei voneinander unabhängige Stichproben einer (μ_1, σ_1^2)- bzw. (μ_2, σ_2^2)-normalverteilten Grundgesamtheit. Dann sind folgende Summen $\sum_{i=1}^{m} \left(\frac{X_i \Leftrightarrow \overline{X}}{\sigma_1} \right)^2 = \frac{(m \Leftrightarrow 1)S_1^2}{\sigma_1^2}$ und $\sum_{j=1}^{n} \left(\frac{Y_j \Leftrightarrow \overline{Y}}{\sigma_2} \right)^2 = \frac{(n \Leftrightarrow 1)S_2^2}{\sigma_2^2}$ jeweils χ^2-verteilt mit $(m \Leftrightarrow 1)$ bzw. $(n \Leftrightarrow 1)$ Freiheitsgraden. Nach der Definition für die F-Verteilung ist daher die Größe $\boldsymbol{F}_{m-1,n-1} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ F-verteilt mit $(m \Leftrightarrow 1), (n \Leftrightarrow 1)$ Freiheitsgraden. Diese Tatsache wird später benutzt, um zwei unbekannte Streuungen σ_1^2 und σ_2^2 zweier normalverteilter Grundgesamtheiten miteinander zu vergleichen, insbesondere um die Frage zu entscheiden, ob die Streuungen gleich sind, oder anders ausgedrückt, um zu entscheiden, ob die beiden Grundgesamtheiten bzgl. ihrer Streuung als homogen anzusehen sind. In diesem Sinne spricht man manchmal auch von einem **Homogenitätstest**.

Fraktilen der F-Verteilung können in MINITAB mit dem Kommando `invcdf` und der Option `f` erzeugt werden. Die allgemeine Syntax lautet:

`INVCDF` für Werte in `E` [nach `E`];
  `F` m = `K` n = `K`.

`E` kann eine Konstante oder eine Spalte sein.

```
MTB > invcdf 0.975;
SUBC> f 7 13.
    0.9750    3.4827
```

Die 97.5%-Fraktile der F-Verteilung mit 7 Zähler- und 13 Nennerfreiheitsgraden ist also $F_{7,13;0.975} \approx 3.5$.

Eine Tabelle der F-Fraktilen für verschiedene Wahrscheinlichkeiten $F(x)$ sowie Zählerfreiheitsgrade m und Nennerfreiheitsgrade n erzeugt folgende Befehlsfolge:

```
MTB > set c1
DATA> 0.95 0.975 0.99
DATA> end
MTB > invcdf c1 c2;
SUBC> f 5 10.
MTB > invcdf c1 c3;
SUBC> f 20 8.
MTB > name c1 'F(x) c2 'm=5,n=10' c3 'm=20,n=8'
MTB > print c1-c3

 ROW    F(x)  m=5,n=10  m=20,n=8

   1   0.950   3.32584   3.15036
   2   0.975   4.23609   3.99946
   3   0.990   5.63633   5.35910
```

Die F-Verteilung mit einem Zählerfreiheitsgrad ($m = 1$) ist übrigens identisch mit dem Quadrat der t-Verteilung: $\boldsymbol{F}_{1,n} = (\boldsymbol{T}_n)^2$

Für die Fraktilen gilt:

$$F_{1,n;1-\alpha} = \left(t_{n;1-\alpha/2}\right)^2 \tag{4.13}$$

4.5 Schätzung von Parametern: Punktschätzungen

Unter einer **Punktschätzung** versteht man die Angabe eines Näherungswerts für einen unbekannten Parameter einer Grundgesamtheit. Angenommen eine Grundgesamtheit oder Population sei normalverteilt. Die beiden Parameter Mittelwert μ und Varianz σ^2 seien unbekannt. Eine Schätzung gibt nun aufgrund einer Stichprobe $x_1, x_2, \ldots, x_n$ **Schätzwerte** für diese unbekannten Parameter an. Das Verfahren heißt Punktschätzung, weil ein bestimmter Punkt im Parameterraum ausgewählt wird. Eine Punktschätzung stellt also eine einzelne Zahl dar, die aus den Stichprobenwerten ermittelt wird.

4.5.1 Die Momentenmethode

Es ist einleuchtend, daß man den Mittelwert $\overline{x}$ einer Stichprobe $x_1, x_2, \ldots, x_n$ als Punktschätzung für den unbekannten Erwartungswert μ der Grundgesamtheit heranzieht: $\mu \approx \overline{x}$.

Genauso kann man die empirische Varianz s^2 als Punktschätzung für die unbekannte Varianz σ^2 der Verteilung oder der Grundgesamtheit nehmen: $\sigma^2 \approx s^2$.

Wenn eine Stichprobe eine relative Häufigkeit h für das Auftreten einer bestimmten Eigenschaft A hat, so kann man diesen Wert als Schätzwert für die unbekannte Wahrscheinlichkeit p einer binomialverteilten Grundgesamtheit verwenden: $p \approx h$.

Damit wurden einige Beispiele der sog. **Momentenmethode** gebracht. Es wurde also sowohl in der beschreibenden Statistik der Begriff des Moments einer Stichprobe eingeführt, wie auch in der Wahrscheinlichkeitstheorie der Begriff des Moments einer Zufallsvariablen.

Es sei daran erinnert, daß bei Stichprobenwerten das Moment erster Ordnung in Bezug auf den Nullpunkt ($a = 0$) nichts anderes als das arithmetische Mittel $\overline{x}$ darstellt. Das zentrierte Moment zweiter Ordnung ($a = \overline{x}$) ist die empirische Varianz. Analog dazu ist das erste Moment einer Zufallsvariablen X gleich dem Erwartungswert μ, das zweite zentrierte Moment von X ist gleich der Varianz σ^2. Die Momentenmethode als Punktschätzung benutzt nun ein bestimmtes Moment, das aus einer Stichprobe berechnet wird, als Schätzwert für das analoge Moment der Verteilung oder der Zufallsvariablen. Anstelle von Moment spricht man auch gelegentlich von **Parameter**.

4.5.2 Kriterien für die Güte von Schätzungen

Das Stichprobenmoment wird im allgemeinen nicht mit dem unbekannten Moment oder Parameter der Verteilung zusammenfallen, da das Ergebnis der Stichprobe zufallsbedingt ist. Es ist jedoch verständlich, daß man verlangt, daß der Schätzwert ein "guter" Schätzwert für den unbekannten Parameter ist. Dazu muß der Begriff "gut" etwas näher erläutert werden. Die statistische Theorie verlangt für einen "guten" Schätzwert, daß die Schätzfunktion bestimmte Kriterien erfüllt, z.B. die **Erwartungstreue**, die **Konsistenz** und eine hohe **Effizienz**.

Eine Schätzfunktion bzw. ein Schätzwert $\widehat{\Theta}$ heißt **erwartungstreu** oder **unverzerrt** bzgl. eines Parameters Θ, wenn der Erwartungswert der Schätzfunktion gleich dem unbekannten Parameter Θ ist, also $\mathrm{E}(\widehat{\Theta}) = \Theta$ (Bild 4.5 links). Ein Schätzwert $\widehat{\Theta}$, der z.B. die Dichtefunktion in Bild 4.5 rechts besitzt, ist ein **verzerrter Schätzer**. Es hat auf lange Sicht eine **Verzerrung** oder **Bias**.

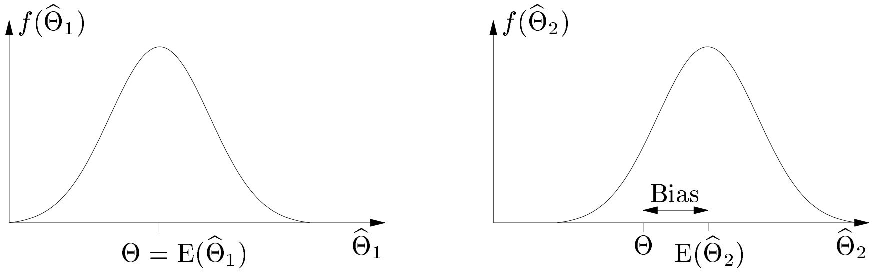

Bild 4.5: Dichtefunktion eines unverzerrten Schätzers $\widehat{\Theta}_1$ und eines nach oben verzerrten Schätzers $\widehat{\Theta}_2$

Beispiele:

1. *Das arithmetische Mittel $\overline{X}$ ist eine erwartungstreue Schätzfunktion für den Erwartungswert μ einer Grundgesamtheit. Es ist nämlich:*

$$\begin{aligned} \mathrm{E}(\overline{X}) &= \mathrm{E}\left(\frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)\right) = \\ &= \frac{1}{n} \cdot \mathrm{E}(X_1 + X_2 + \ldots + X_n) = \\ &= \frac{1}{n} \cdot (\mathrm{E}(X_1) + \mathrm{E}(X_2) + \ldots + \mathrm{E}(X_n)) \end{aligned}$$

Da die X_i alle die gleiche Verteilung haben, gilt:

$\mathrm{E}(X_i) = \mu$ *für* $i = 1, 2, \ldots, n$, *also:* $\mathrm{E}(\overline{X}) = \dfrac{1}{n} \cdot n \cdot \mu = \mu$

2. *Die empirische Varianz S^2 ist ebenfalls eine erwartungstreue Schätzfunktion für die Varianz σ^2 einer Grundgesamtheit, denn es ist:*

$$\mathrm{E}(S^2) = \mathrm{E}\left(\frac{1}{n \Leftrightarrow 1} \cdot \sum_{i=1}^{n} (X_i \Leftrightarrow \overline{X})^2\right)$$

Zunächst wird $\displaystyle\sum_{i=1}^{n} (X_i \Leftrightarrow \overline{X})^2$ *umgeformt:*

$$\begin{aligned} \sum (X_i \Leftrightarrow \overline{X})^2 &= \sum \big((X_i \Leftrightarrow \mu) + (\mu \Leftrightarrow \overline{X})\big)^2 = \\ &= \sum (X_i \Leftrightarrow \mu)^2 + 2(\mu \Leftrightarrow \overline{X}) \cdot \sum (X_i \Leftrightarrow \mu) + n(\mu \Leftrightarrow \overline{X})^2 = \\ &= \sum (X_i \Leftrightarrow \mu)^2 + 2(\mu \Leftrightarrow \overline{X}) \cdot n(\overline{X} \Leftrightarrow \mu) + n(\overline{X} \Leftrightarrow \mu)^2 = \\ &= \sum (X_i \Leftrightarrow \mu)^2 \Leftrightarrow 2n(\overline{X} \Leftrightarrow \mu)^2 + n(\overline{X} \Leftrightarrow \mu)^2 = \\ &= \sum (X_i \Leftrightarrow \mu)^2 \Leftrightarrow n(\overline{X} \Leftrightarrow \mu)^2 \end{aligned}$$

Daraus folgt:

$$\mathrm{E}(S^2) = \frac{\sum \mathrm{E}(X_i \Leftrightarrow \mu)^2 \Leftrightarrow n \cdot \mathrm{E}(\overline{X} \Leftrightarrow \mu)^2}{n \Leftrightarrow 1} = \frac{n\sigma^2 \Leftrightarrow n\sigma^2/n}{n \Leftrightarrow 1} = \sigma^2$$

Damit wird auch ersichtlich, warum bei der Berechnung der empirischen Varianz nach Gleichung (1.7) die Abweichungsquadratsumme $\sum_{i=1}^{n}(x_i \Leftrightarrow \overline{x})^2$ *durch* $n \Leftrightarrow 1$ *und nicht durch* n *dividiert wurde. Eine Division durch* n *würde die Erwartungstreue zerstören. Die mittlere quadratische Abweichung* $\frac{1}{n} \cdot \sum_{i=1}^{n}(x_i \Leftrightarrow \overline{x})^2$ *stellt also einen verzerrten Schätzwert für das unbekannte* σ^2 *dar, denn auf lange Sicht wird mit diesem Schätzwert die Streuung* σ^2 *etwas unterschätzt.*

Eine Schätzfunktion bzw. der entsprechende Schätzwert $\widehat{\Theta}$ heißt **konsistent**, wenn die Wahrscheinlichkeit dafür, daß sich der Schätzwert $\widehat{\Theta}$ um weniger als ε von dem wahren, zu schätzenden Parameter Θ unterscheidet, für jede noch so kleine, aber feste Zahl ε mit wachsendem Stichprobenumfang gegen 1 strebt:

$$P\left(\left|\widehat{\Theta} \Leftrightarrow \Theta\right| < \varepsilon\right) \to 1 \qquad \text{für } n \to \infty \tag{4.14}$$

Die Wahrscheinlichkeit, daß der Schätzfehler beliebig klein wird, geht also gegen 1 bzw. 100%.

Beispiel:

Das arithmetische Mittel $\overline{X}$ *ist eine konsistente Schätzfunktion für den Erwartungswert* μ. *Ebenso ist* S^2 *konsistent für* σ^2.

Um mehrere unverzerrte Schätzer miteinander vergleichen zu können, wird zunächst der Begriff der **relativen Effizienz** eingeführt. Die relative Effizienz von $\widehat{\Theta}_1$ verglichen mit $\widehat{\Theta}_2$ ist das Verhältnis der entsprechenden Varianzen der Schätzer, also $\frac{\mathrm{Var}(\widehat{\Theta}_2)}{\mathrm{Var}(\widehat{\Theta}_1)}$. Derjenige Schätzer ist also effizienter, der die kleinere Varianz hat (vgl. Bild 4.6). Ein Schätzer heißt schlechthin **effizient**, wenn es keine anderen Schätzer mit einer kleineren Varianz gibt.

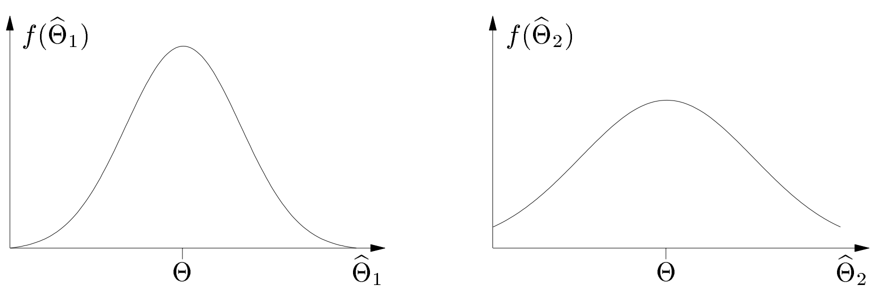

Bild 4.6: Relative Effizienz zweier Schätzer $\widehat{\Theta}_1$ und $\widehat{\Theta}_2$

Beispiel:

Das arithmetische Mittel $\overline{X}$ *und der Median* $\widetilde{X}$ *sind Schätzer des unbekannten Erwartungswerts* μ *einer normalverteilten Grundgesamtheit. Die Varianz* $\mathrm{Var}(\overline{X})$ *von* $\overline{X}$ *ist* $\dfrac{\sigma^2}{n}$ *(vgl. 3.1.3). Auch der Median* $\widetilde{X}$ *einer Stichprobe stellt einen unverzerrten Schätzer für* μ *dar. Seine Varianz* $\mathrm{Var}(\widetilde{X})$ *beträgt aber* $\dfrac{\sigma^2}{n} \cdot \dfrac{\pi}{2}$*, ist also rund 57% größer als die Varianz von* $\overline{X}$*. Anders ausgedrückt kann man sagen: Die relative Effizienz des arithmetischen Mittels* $\overline{X}$ *verglichen mit dem Median* $\widetilde{X}$ *ist* $\dfrac{\pi}{2} \approx 1.57$ *bzw. 157%. Vergrößert man den Umfang* n *der Stichprobe, so kann man jeweils die Varianzen der beiden Schätzer, arithmetisches Mittel und Median, verkleinern. Man könnte also die schlechtere Effizienz des Medians dadurch ausgleichen, daß man eine um 57% größere Stichprobe zum Schätzen für den Parameter* μ *heranzöge, was natürlich einen beträchtlichen Mehraufwand hinsichtlich der Gewinnung von Stichprobenwerten bedeuten würde.*

Es wurde deutlich gemacht, daß bei einem Vergleich zwischen unverzerrten Schätzern derjenige mit der kleineren Varianz den Vorzug hat. Vergleicht man aber beliebige Schätzer (also verzerrte und unverzerrte), so ist es verständlich, daß die Varianz der Schätzer nicht mehr allein entscheidend sein kann.

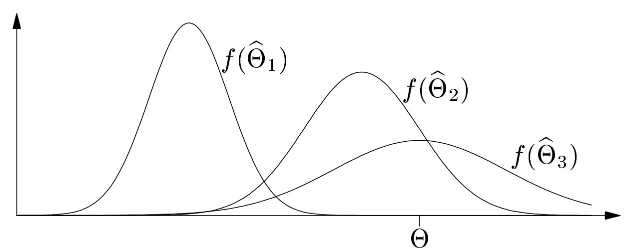

Bild 4.7: Vergleich von beliebigen Schätzern

Aus Bild 4.7 wird unmittelbar ersichtlich, daß der Schätzer $\widehat{\Theta}_2$ den besten Kompromiß hinsichtlich der Kleinheit von Verzerrung und Varianz bietet. Der Schätzer $\widehat{\Theta}_1$ hat zwar die kleinste Varianz, aber er hat sicherlich eine nicht akzeptable Verzerrung. Beim Schätzer $\widehat{\Theta}_3$ ist es gerade umgekehrt. Um einen Vergleich anstellen zu können, soll daher ein Kriterium vorgeschlagen werden, welches sowohl die Verzerrung als auch die Varianz berücksichtigt. Man kann sich leicht überzeugen, daß der mittlere quadratische Fehler MSE (von engl. *mean square error*) ein solches geeignetes Kriterium darstellt:

$$\mathrm{MSE} = \mathrm{E}\big(\widehat{\Theta} \Leftrightarrow \Theta\big)^2 \tag{4.15}$$

Der Unterschied des MSE zur Varianz $\mathrm{Var}(\widehat{\Theta})$ des Schätzers liegt darin, daß die Varianz bekanntlich als $\mathrm{E}\big(\widehat{\Theta} \Leftrightarrow \mathrm{E}(\widehat{\Theta})\big)^2$ definiert ist. Man kann leicht zeigen, daß folgende Beziehung gilt:

$$\mathrm{MSE} = \mathrm{Var}(\widehat{\Theta}) + \big(\mathrm{E}(\widehat{\Theta}) \Leftrightarrow \Theta\big)^2 = \text{Varianz} + \text{Verzerrung}^2 \tag{4.16}$$

Das MSE-Kriterium stellt also im allgemeinen Fall ein geeignetes Vergleichskriterium für Schätzwerte dar.

Die Eigenschaft der Effizienz für den arithmetischen Mittelwert setzt jeweils Stichproben aus normalverteilten Grundgesamtheiten voraus. Wenngleich diese Annahme der Normalverteilung sehr häufig gemacht wird, vor allem, weil dann gängige statistische Verfahren zur Verfügung stehen, so weiß man im Einzelfall nicht immer, ob diese Annahme zurecht besteht. Es kann manchmal durchaus vernünftig sein, andere Verteilungen für die Grundgesamtheit anzunehmen, z.B. eine Laplace-Verteilung oder eine Cauchy-Verteilung (Bild 4.8). Diese Verteilungen unterscheiden sich von der Normalverteilung dadurch, daß die Enden ihrer Dichtefunktionen nicht so schnell gegen Null gehen (also "dickschwänziger" sind), und damit die Wahrscheinlichkeit für extreme Werte größer ist als bei der Normalverteilung. In diesen Fällen zeigt sich, daß der Median ein effizienterer Schätzwert für den unbekannten Erwartungswert ist als das arithmetische Mittel.

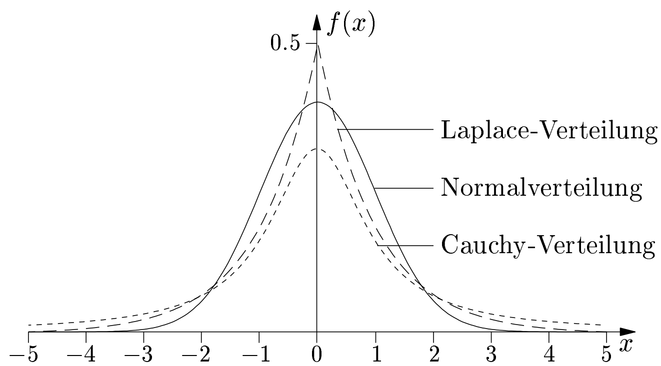

Bild 4.8: Wahrscheinlichkeitsdichten der Laplace-, Normal- und Cauchy-Verteilung

Daraus kann man ersehen, daß die Eigenschaft der Effizienz von der Verteilung der zugrundeliegenden Population abhängt. Wenn aber z.B. nichts oder wenig über diese Verteilung bekannt ist, weiß man auch nicht, welcher der beste bzw. effizienteste Schätzwert ist. Darum hat man in den letzten Jahren versucht, sog. **robuste Schätzer** zu finden, welche sich auch dann noch möglichst effizient verhalten, wenn Abweichungen von den unterstellten Voraussetzungen hinsichtlich der Verteilung der Grundgesamtheit eintreten. So zeigte man z.B. daß das arithmetische Mittel nicht robust gegenüber Veränderungen der Verteilungen in der Grundgesamtheit ist. Man hat robustere Schätzwerte vorgeschlagen, auf die an dieser Stelle jedoch nicht im einzelnen eingegangen werden kann.