Bachelorseminar

Information und statistische Inferenz - ein Einblick

Induktion und Inferenz - Grundlagen

Ludwig- Maximilians- Universität München

Institut für Statistik

Autorin: Natalie Grimm

Betreuerin: Frau PD Dr. Dr. Schneider

23.01.2015

Inhaltsverzeichnis

1 Inferenz und Induktion

1.1 Grundannahmen

Die nachfolgende Arbeit beschreibt eine Zusammenfassung von Dr. Bernhard Rügers Lehrbuch Test- und Schätztheorie, erster Band: Grundlagen, von Seite 111 bis 131.
Um im Folgenden die Konzepte der Inferenz sowie die statistische Induktion näher behandeln zu können, ist es von Nöten zu Beginn die Grundannahmen der induktiven Statistik sowie das sogenannte Grundproblem statistischer Inferenz darzustellen.
Der kommende Abschnitt bezieht sich auf die Seiten 111 bis 112.
Die Definition eines statistischen Schlusses besagt: „Ein statistischer Schluss ist ein Schluss von der Beobachtung x einer Stichprobe X auf die Verteilung (...) der Zufallsgröße X (...).“ (Rüger, 1999, Seite 111) Kurz gesagt, ein „Schluss von einer Stichprobe auf die Grundgesamtheit“ (Rüger, 1999, Seite 111), gestützt auf die zugrunde gelegte Verteilungsannahme über X, als eine Familie oft parametrisierter Wahrscheinlichkeitsverteilungen. X ist gemäß eine dieser Verteilungen. Diese Definition ist vor allem in der klassischen Test- und Schätztheorie zu finden.
In der Theorie der statistischen Behandlung von „vorhandene[n] Unschärfen in den Beobachtungen oder Unsicherheiten im Vorwissen“ (Rüger, 1999, Seite 111), gibt es hingegen eine gelockerte Auffassung dieser Formulierung. Dabei werden die Beobachtungen als Realisationen von Zufallsgrößen gesehen, sie werden als Wahrscheinlichkeitsverteilungen bezeichnet. Um sich ebenfalls mit der Problematik statistischer Schlüsse genauer befassen zu können, ist eine Ausweitung der Definition notwendig. Verschiedene Formulierungsvorschläge werden hierbei gegeben, wobei jener exemplarisch herausgenommen wurde: „Ein statistischer Schluss (...) ist ein Verfahren, mit dem unter Berücksichtigung von Unsicherheiten aus empirisch gewonnenen Daten Aussagen oder Entscheidungen über mögliche Bedingungen oder Gesetze, unter denen die Daten stehen, getroffen werden.“ (Rüger, 1999, Seite 112)
Statistische Schlüsse sind auch induktive Schlüsse, aufgrund dessen wird die gesamte schließende Statistik induktive Statistik genannt.

Im Allgemeinen formuliert und dargestellt bedeutet dies:

Induktiver Schluss:	Beobachtung $\rightarrow$	Gesetz bzw. Hypothese
	Einzelfall	allgemeine Aussage
Deduktiver Schluss:	Beobachtung $\leftarrow$	Gesetz bzw. Hypothese

1.2 Problematik

Der folgende Abschnitt mit Bezug auf die Problematik eines induktiven Schlusses fasst die Seiten 112 bis 113 zusammen.
Dazu lohnt es sich zu Beginn einen genaueren Blick auf die Begriffe der Induktion, Deduktion, Hypothese, Beobachtung und ihre Rollen in der statistischen Inferenz zu werfen.
Der Begriff Induktion geht auf Aristoteles (384-322) zurück. M. T. Cicero (106-43) übersetzte Induktion mit Herbeiführung. Dieser Ausdruck hat im ursprünglichen Sinne drei verschiedene Bedeutungen.
Die erste beschreibt die oben genannte Aussage des Schlusses von einem einzelnen Fall auf ein allgemeingültiges Gesetz.
Im nächsten Fall, ist die Herbeiführung von Axiomen, die als gegeben vorausgesetzt werden können, gemeint.
Die dritte Bedeutung bezeichnet die rhetorische Variante als ein Verfahren bei dem ein Gesprächspartner zur Akzeptanz von gewissen Prämissen gebracht werden soll, auf denen wiederum deduktive Schlüsse aufbauen.
Das sogenannte Induktionsproblem forscht nach der Begründbarkeit oder Rechtfertigung eines induktiven Schlusses. Dabei ist ebenso wichtig, ob jener als korrekt nachgewiesen werden kann. Die Frage stellt sich, ob es eine allgemeingültige Theorie, eine Art induktive Logik, gibt, mit der ein induktiver Schluss als richtig bewiesen werden kann. D. Hume (1711-1776) verneint dies. Er argumentiert etwa in der Weise: Die Natur soll sich uniform verhalten, dass heißt in Raum und Zeit vergleichbar. Diese Uniformitätsannahme über die Natur ist jedoch nicht deduktiv oder induktiv begründbar und führt zur sogenannten Unlösbarkeit des Induktionsproblems.

1.3 Lösungsvorschläge

Jener Abschnitt bezieht sich auf die Seiten 113 bis 116 und befasst sich mit verschiedenen Lösungsvorschlägen zu der Induktionsproblematik.
Eine mögliche Antwort auf dieses Problems wird von Hume in einer pragmatischen Art

und Weise selbst formuliert: Die Uniformitätsannahme über die Natur wird dem Bereich menschliches Verhalten anstatt dem Bereich menschlichen Erkennens zugeteilt, wobei die Gewohnheit der Menschen das Verhalten beeinflusst. Dies wird dadurch begründet, dass Naturgesetze zeitliche Kontinuität aufweisen. Fundiert durch die Erfahrung, dass Ereignisse aus der Vergangenheit unter bestimmten Bedingungen eingetreten sind und auch in Zukunft so eintreten werden. Diese „Macht der Gewohnheit“ als rationaler Instinkt liegt in der Notwendigkeit der menschlichen Existenz, die Gesetze der Natur erkennen zu wollen, um überlegt handeln zu können.

Ein anderer, nicht pragmatischer, Ansatz wird in der modernen Wissenschaftstheorie vorgelegt. Hierbei wird nicht mehr nach einer allgemeinen Rechtfertigung des Prinzips eines Schlusses geforscht, sondern nach Kriterien, die zur Begründung induktiver Schlussweisen herangezogen werden und an Annahmen gebunden sind. Es geht um die Beziehungsstruktur zwischen Beobachtung und Hypothese sowie zwischen Erfahrung und Theorie.

Neben naturwissenschaftlich orientierten Theorien, wie unter anderem bei Stegmüller, Popper und Lakatos, spielen auch geisteswissenschaftlich orientierte Richtungen wie bei den Konzepten des Pragmatismus, der Schlussweise der Abduktion, die die Induktion und Deduktion bedeutend ergänzt, und die Hermeneutik.

„Ein dritter Ansatz, das genannte Dilemma empirischen Forschens zu überwinden“ (Rüger, 1999, Seite 114) geht auf G. Galilei (1564-1642) und die Anfänge der neuzeitlichen empirischen Naturwissenschaften zurück. „Dieser Weg besteht in der Aufhebung des strikten Gegensatzes zwischen Erfahrung und Theorie, (...) [und] dass grundsätzlich jede Erfahrung theoriegeleitet und jede Theorie erfahrungsbedingt ist. (...) Theorie und Erfahrung bilden hierbei ein symmetrisches Paar, eine dialektische Zweisamkeit. Es gibt weder eine unerschütterliche Theorie noch eine unerschütterliche Erfahrung.“ (Rüger, 1999, Seite 114)

Die Statistik bietet eine Theorie für empirische Forschungsmethoden und setzt dabei zwei Ziele in Verbindung: das pragmatisch und anwendungsbezogene Methodenprogramm sowie die eigenständige statistische Begründung ihrer Methodologie, als eine Methodenlehre. Diese Entwicklung kam durch die Herauslösung aus den empirisch forschenden Einzelwissenschaften wie unter anderem Biologie, Psychologie und Soziologie. Die formale Theorie der praktischen Empirie berücksichtigt, dass die Unsicherheit und der Stützungsgrad einer Beobachtung aus einer Hypothese in die Konstruktion des Schlussverfahrens von Beginn an miteinbezogen wird.

2 Fragen an die statistische Inferenz

Während in Kapitel 1 Fragen und Probleme der Inferenz auf wissenschaftstheoretischer Ebene erläutert wurden, werden wir in diesem Kapitel diese und mögliche Lösungen auf dem Gebiet der Statistik betrachten und die Seiten 116 bis 124 zusammenfassen.
Die wichtigsten Inferenzkonzepte der Statistik sind beispielsweise die klassische Inferenz, Likelihood-Inferenz, Bayes-Inferenz, entscheidungstheoretische Inferenz, Fiduzial-Inferenz, Pivotal-Inferenz und die Struktural-Inferenz, wobei die letzten drei Inferenzkonzepte eher bruchstückhaft entwickelt sind.

2.1 Fragen

Drei Fragen nach der statistischen Inferenz stehen besonders im Vordergrund sowie die fundamentale Theorie der Unsicherheit und ebenso deren Umgang.
Sie sollen nicht nur die verwendeten unterschiedlichen Konzepte, sondern auch den Bereich ihrer Anwendung beleuchten. Dabei spielt auch eine Rolle von welcher Interpretation des Wahrscheinlichkeitsbegriffes ausgegangen wird. In Kapitel 2.3 wird darauf näher eingegangen.
Die erste der drei Fragen lautet: Dient ein Schluss aus einer Beobachtung einem kognitiven, das heißt einer Gewinnung von Erkenntnissen, oder einem dezisionistischen Zweck, der Herbeiführung von Entscheidungen?
Die zweite Frage analysiert welche Standpunkte objektiv oder subjektiv in die Argumentation des Schlusses aus einer Beobachtung eingehen.
Die dritte Frage besagt: Wie wird die Güte eines Schlusses beurteilt? Frequentistisch oder nichtfrequentistisch? Eine frequentistische Beurteilung der Güte bedeutet eine Antwort auf die Fragen: Wie sicher ist ein Schluss und wie häufig führt er zu einer richtigen Aussage? Die nichtfrequentistische Beurteilung dagegen: Wie plausibel ist Schluss bei vorliegender Beobachtung?

2.2 Auswertung der Fragen

Die Beantwortung dieser drei Fragen führt zu einer jeweilig unterschiedlichen Sicht der Inferenzkonzepte. Diese Situation liegt in parametrisierten Verteilungsfamilien $\{P_\theta : \theta \in \Theta\}$ mit Θ als Parameterraum zugrunde.
Der Zweck einer statistischen Auswertung der Beobachtung ist die Erkenntnis einer Aussage, die auf der einen Seite möglichst präzise und auf der anderen Seite möglichst zuverlässige Informationen über einen unbekannten Parameter θ liefern soll. Eine Auswertung ist dabei ein statistisches Entscheidungsverfahren, bei dem das Ziel ist eine Entscheidung herbeizuführen. Es spielt auch der Gewinn oder Verlust, den eine Entscheidung mit sich bringt, eine große Rolle bei der Entscheidungsfindung.
In der zweiten Fragestellung wird nach den Vorkenntnissen über die Versuchsbedingungen, unter denen die Beobachtungen durchgeführt werden, gefragt.
Auf der einen Seite gibt es die Vertreter eines objektiven Standpunktes, auch Objektivisten genannt. Diese fordern, dass Unsicherheiten oder Lücken im Modell oder in den Vorkenntnissen nicht durch subjektive Annahmen der Behauptungen geschlossen werden dürfen. Bei einem Objektivisten können a-priori sowie a-posteriori Verteilungen verwendet werden, jedoch müssen sie aus objektiven Vorgängen entspringen. Auf diese wird in Kapitel 2.3 näher eingegangen. „Im Zusammenhang mit dem Satz von Bayes werden die Wahrscheinlichkeiten $P(A_i)$ auch als a-priori Wahrscheinlichkeiten und $P(A_i \mid B)$ als a-posteriori Wahrscheinlichkeiten bezeichnet, da $P(A_i)$ das Eintreten von A_i vor Kenntnis des Ereignisses B und $P(A_i \mid B)$ das Eintreten dieses Ereignisses nach Kenntnis von B bewertet.“ (Fahrmeir et al, 2011, Seite 213)
Auf der anderen Seite gibt es die Vertreter eines subjektiven Standpunktes, die Subjektivisten. Bei den Vertretern des subjektiven Standpunktes ist die Verteilung, die vor der Beobachtung zu Kenntnissen führt, durch ein a-priori Wissen über den unbekannten Parameter θ gegeben. Diese führt daraufhin zu einer a-posteriori Verteilung.
Beide Gruppen verwenden dabei insbesonders die Grundannahmen der Test- und Schätztheorie, die in Kapitel 1.1 erläutert wurden. Hierbei wird die Beobachtung und die Verteilungsannahme als objektiv angenommen. Erweiterungen dieses Modells sind entweder subjektiv oder objektiv.
Im Rahmen der dritten Frage, der Beurteilung einer Güte eines statistischen Schlusses, wird zwischen einer frequentistischen und einer nichtfrequentistischen Beurteilungsweise unterschieden.
Bei der frequentistischen Beurteilung wird der Schluss unter dem Gesichtspunkt des

Schlussverfahrens betrachtet unter dem er gewonnen wurde. Wie gut ein solches Verfahren ist, dass zu richtigen oder falschen Aussagen und zu guten oder schlechten Entscheidungen führen kann, wird mit dem Begriff des Sicherheitsgrades bezeichnet. Die Güte eines statistischen Verfahrens wird danach beurteilt, wie häufig oder auf lange Sicht gesehen es zu richtigen Aussagen oder guten Entscheidungen führt. Folglich ist auch wichtig, wie gut ein Schluss aus einer vorliegenden Beobachtung ist oder wie gut das Verfahren ist, aus dem der Schluss gewonnen wurde. Dadurch steht der Aspekt, dass von einem Einzelfall auf andere Fälle geschlossen werden soll, hinsichtlich einer Güte des Verfahrens, besonders im Vordergrund.
Bei einer nichtfrequentistischen Beurteilungsweise der Güte steht der Schluss im Einzelfall im Mittelpunkt. „Hier wird die Güte eines Schlusses nur an Hand der einen Beobachtung, aus der er gewonnen wurde, beurteilt, indem ein Maß oder Grad für die Stützung oder Plausibilität, die die Beobachtung dem Schluss verleiht, bestimmt wird. Entsprechend steht einem Sicherheitsgrad (...) ein Stützungsgrad der nichtfrequentistischen Beurteilungsweise gegenüber.“ (Rüger, 1999, Seite 119)

2.3 Auffassung von dem Begriff „Wahrscheinlichkeit“

Wie bereits in den vorherigen Kapiteln angedeutet wurde, gibt es zwei unterschiedliche Auffassungen des Wahrscheinkeitsbegriffes.
Der erste, der objektive Wahrscheinlichkeitsbegriff, lässt sich in drei jeweils verschiedene Begrifflichkeiten unterteilen: klassische, frequentistische und logische Wahrscheinlichkeit.
Die klassische Wahrscheinlichkeit basiert auf a-priori Kenntnissen über Symmetrieeigenschaften eines Zufallsexperiments. Bei einer endlichen Ergebnismenge wird jedes Ergebnis einzeln für gleich möglich gehalten. Dagegen müssen bei unendlichen Ergebnismengen Nebenbedingungen definiert werden.
Die frequentistische Wahrscheinlichkeit ist eine a-posteriori Wahrscheinlichkeit. Die Wahrscheinlichkeit eines Ergebnisses A aus den relativen Häufigkeiten, bei denen A aus einer Reihe von Versuchswiederholungen auftritt, wird damit bezeichnet. Sie stützt sich auf die prinzipielle Wiederholbarkeit des Experiments.
Die letzte Version des objektiven Wahrscheinlichkeitsbegriffes ist die logische Wahrscheinlichkeit. Sie ist jedoch nicht völlig ausgearbeitet. Dabei wird dem Ereignis A nicht mehr eine Wahrscheinlichkeit $P(A)$ zugeordnet, sondern eine zweistellige Beziehung, $P(B;A)$, aufgestellt. Zum einen kann sie als die Wahrscheinlichkeit von B unter der Annahme A, zum anderen als die Wahrscheinlichkeit das der Implikation aus A folgt B, die Korrektheit dieser Implikation folgt, verstanden werden. Der Begriff des Vertrauens-

grades beschreibt den Schluss von einer Beobachtung A auf die Aussage B.
Im Prinzip vom unzureichenden Grund oder auch Indifferenzprinzip genannt, finden alle drei Versionen des objektiven Wahrscheinlichkeitsbegriffs ihre Gemeinsamkeit.
Das Indifferenzprinzip besagt, dass im Bezug auf die Bayes-Laplace-Regel „nach dem Prinzip vom unzureichenden Grund die Situation eines a-priori Nichtwissens dadurch charakterisiert [wird], dass kein Grund vorhanden sei, irgendeinen Parameterwert vor einem anderen als glaubwürdiger auszuzeichnen, eine Charakterisierung, die mit der Indifferenz gegenüber den möglichen Parameterwerten zusammenfällt und zur Gleichverteilung nichtinformativer (besser: indifferenter) Verteilung führt.“ (Rüger, 1999, Seite 271)
Die subjektive Wahrscheinlichkeit bildet die zweite Auffassung des Begriffes und steht im Gegensatz zur objektiven Wahrscheinlichkeit. Die Wahrscheinlichkeit $P(A)$ teilt einem Ereignis A einen Grad zu, mit dem eine Person A an das Eintreten jenes Ereignisses glaubt, oder die Chance, die die Person dem Ereignis gibt, zu. Diese Wahrscheinlichkeit hängt von dieser Person A ab und kann von Person zu Person unterschiedlich sein. Beispielsweise bewerten Fans von Sportmannschaften ein Spiel ihrer eigenen Mannschaft hinsichtlich eines Sieges anders als Anhänger der gegnerischen Mannschaft. Solch eine gewisse Voreingenommenheit oder auch Vorkenntnis kann die Person A zu einer Entscheidung hinsichtlich ihres planerischen Verhaltens bringen. Dieser verhaltenstheoretische Aspekt spielt eine große Rolle bei der subjektiven Wahrscheinlichkeit. Insbesondere bei Wetten auf das Eintreten eines bestimmten Ereignisses A sowie dem Wettquotient für Einzahlung oder Auszahlung, lässt sich bemessen in wie weit diese Person in ihrer Entscheidung für die Wahrscheinlichkeit $P(A)$ gehen würde. Dabei muss die wettende Person natürlich eine Nutzenskala besitzen, bei der der Nutzen proportional zum Gewinn gewertet wird. Somit ist diese Person an gewisse Regeln gebunden. Diese Axiomensysteme, bei denen die Beziehung zwischen Wahrscheinlichkeit und Wettquotient dargestellt werden, sind bei der subjektiven Wahrscheinlichkeit von Bedeutung.
Die subjektive Wahrscheinlichkeit lässt sich ebenfalls weiter unterteilen.
Auf der einen Seite steht das subjektive Wissen im Vordergrund und wie sich subjektives Vorwissen durch Beobachtungen verändert. Dabei wird von Lernen durch Beobachtung gesprochen.
Auf der anderen Seite ist der Zusammenhang zwischen dem Wettverhalten der Person und ihren subjektiven Wahrscheinlichkeiten besonders wichtig. Es wird angenommen, dass sich eine Person hinsichtlich ungewissen Ereignissen so verhält, dass der Erwartungswert ihres „Nutzens“ am größten wird. Somit stehen dabei spiel- oder entscheidungstheoretische Aspekte im Vordergrund.
Bei beiden Fällen ist ein gemeinsamer Bezug zu Wetten und Wettquotienten gegeben.

Nun lassen sich die unterschiedlichen Auffassungen von Wahrscheinlichkeit auf die drei Fragen aus Kapitel 2.2 anwenden.
Zur besseren Übersicht erfolgt dies in einer Tabelle:

	Objektiver Wahrscheinlichkeitsbegriff	Subjektiver Wahrscheinlichkeitsbegriff
Zweck einer stat. Auswertung	Kognitivistisch & Dezisionistisch - offen	Kognitivistisch - subjektives Wissen Dezisionistisch - verfahrensorientiert
Standpunkt einer stat. Auswertung	Kriteriengeleitet (objektiv)	Posteriori- Verteilung (objektiv & subjektiv)
Beurteilung einer stat. Auswertung	Frequentistisch - frequentistisch Klassisch & Logisch - offen	Subjektiv - nichtfrequentistisch Verhaltensorientiert - frequentistisch & nicht frequentistisch

3 Schätzungen und Tests

3.1 Schätzverfahren

Bei der Auswertung von Beobachtungen stehen neben dem Testverfahren, die in Kapitel 3.2 näher erläutert werden, die Schätzverfahren im Vordergrund. Sie lassen sich in Punkt- und Intervallschätzung unterteilen. Bei beiden ist davon auszugehen, dass der Anwender unvoreingenommen gegenüber der unbekannten Größe θ ist. Die Information, die darin enthalten ist, soll bei der Methodik konzentriert und vollständig abgebildet werden.
Dieser Abschnitt bezieht sich auf die Seiten 125 bis 126.

3.1.1 Punktschätzung

Eine Punktschätzung für einen unbekannten Parameter θ ist eine messbare Abbildung U, von einem Stichprobenraum $\mathfrak{X}$ in $\mathbb{R}_r$, bei dem jedem $x \in \mathfrak{X}$ ein Schätzwert $\hat{\theta} = U(x)$ zugeordnet wird. $U = U(X)$ wird vor der Beobachtung als Zufallsgröße aufgefasst. Die Aussage $\theta = \hat{\theta}$ beschreibt dabei ein Ergebnis der Schätzung nach einer Beobachtung x.
Es soll bei einer Punktschätzung die Frage beantwortet werden, welcher Wert $\hat{\theta}$ anstatt von θ bei einer Beobachtung x angenommen werden soll. Dadurch können zwei unterschiedliche Ziele von einem Anwender in den Mittelpunkt gestellt werden.
Es kann entweder auf die Genauigkeit der Schätzung und Schätzfehler abgezielt oder auf die Bestimmung von $\hat{\theta}$, der durch die Beobachtung x gestützt wird, werden.
Als einfachstes Beispiel für eine Punktschätzung lässt sich folgendes formulieren:
Ein Merkmal X sei normalverteilt mit $X \sim N(\mu, \sigma^2)$. Aus Beobachtungen, die unabhängig sind, $(X_1, ..., X_n)$, wird der Erwartungswert $E(X)$ geschätzt. Für $E(X) = \mu$ und die Varianz $Var(X) = \sigma^2$ erhält man das arithmetischen Mittel $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ für alle $n = 1, ..., n$ als Schätzer.

3.1.2 Intervallschätzung

Eine Intervallschätzung für θ kann als eine Abbildung C auf den Stichprobenraum $\mathfrak{X}$, das jedem x ein Element von $\mathfrak{X}$ auf dem Intervall $C(x)$ zuordnet, verstanden werden.
$C(x)$ ist dabei ein abgeschlossenes Intervall und hat als untere Grenze $C_1(x)$ sowie $C_2(x)$ als obere. Vor der Beobachtung wird $C(X) = [C_1(X); C_2(X)]$ als zufälliges Intervall beschrieben. Diese Grenzen sind r-dimensionale Stichprobenfunktionen und Zufallsgrößen. Nach der Beobachtung x ist $C(x)$ ein festes Intervall und das Ergebnis, einer Aussage der Intervallschätzung im Bezug auf den unbekannten Parameter, ist, dass $C_1(x) \leqq \theta \leqq C_2(x)$ und $\theta \in C(x)$ vorliegt. Im dem Spezialfall $C_1(x) = C_2(x)$ geht eine Intervallschätzung in eine Punktschätzung über.
Eine Intervallschätzung soll die Frage beantworten, in welchem Intervall das unbekannte θ bei Annahme einer Beobachtung x mit einer hohen Wahrscheinlichkeit bzw. Plausibilität liegt. Hierbei können zwei unterschiedliche Ziele vom Anwender in den Vordergrund gestellt werden.
Auf der einen Seite soll die Glaubwürdigkeit der Aussage $\theta \in C(x)$ überprüft werden, auf der anderen Seite soll die Präzision dieser Aussage bestimmt werden. Verallgemeinert gesprochen bedeutet dies: Je größer das Intervall $C(x)$ ist, desto glaubwürdiger jedoch aber auch unpräziser ist diese Aussage. Daraus lässt sich ebenfalls der Umkehrschluss ableiten.
Mithilfe eines Konfidenzintervalls lässt sich ein Kompromiss für beide Ziele finden. Ein Vertrauensgrad γ wird durch die Aussage $P_\theta(C_1(X) \leqq \theta \leqq C_2(X)) \geqq \gamma$ für alle θ gültig bekräftigt. Dazu wird das kleinste der Konfidenzintervalle mit dem festen Vertrauensgrad γ gebildet, was im Mittel gesehen, die präzisesten Aussagen liefern kann.
Als einfachstes Beispiel für eine Intervallschätzung lässt sich ein Fall bei bekannter Varianz σ^2 formulieren:
Ein Merkmal X sei normalverteilt mit $X \sim N(\mu, \sigma^2)$. Der unbekannte Erwartungswert beträgt $E(X) = \mu$ und als ein Schätzer wird das arithmetischen Mittel $\bar{X}$ verwendet. $\bar{X}$ lässt sich standardisieren. Somit gilt: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. „Für diese Statistik lässt sich (...) ein zweiseitig beschränkter Bereich angeben, in dem sie mit Wahrscheinlichkeit $1 - [\gamma]$ liegt.“ (Fahrmeir et al, 2011, Seite 388) Es gilt: $P(-z_{1-\gamma/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\gamma/2}) = 1 - \gamma$. Wobei $z_{1-\gamma/2}$ das $(1 - \gamma/2)$- Quantil der Standardnormalverteilung bezeichnet. Durch weiteres Umformen dieser Gleichung ergibt sich für das $(1 - \gamma)$ - Konfidenzintervall: $[\bar{X} - z_{1-\gamma/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\gamma/2}\frac{\sigma}{\sqrt{n}}]$.

3.2 Testverfahren

Statistische Tests sind Verfahren, in denen verschiedene Hypothesen über einen unbekannten Parameter θ durch eine Beobachtung x geprüft werden sollen. Dieser Abschnitt bezieht sich auf die Seiten 127 bis 128.
Bei statistischen Tests ist eine gewisse Voreingenommenheit des Betrachters durch die Aufstellung der Hypothesen über θ gegeben, wie sie im Gegensatz zu Intervall- und Punktschätzern nicht vorkommen.
Im einfachsten Fall werden zwei Hypothesen über θ formuliert, die sogenannte Nullhypothese H_0 und die Alternativhypothese H_1. Diese schließen sich einander aus und sind disjunkte und nichtleere Teilmengen von Θ. Die Teilmenge in Θ, die für H_0 zutrifft wird mit H_0, die Teilmenge von H_1 mit H_1 bezeichnet. Somit werden beide zu disjunkten Teilmengen in θ. Bei einem nichtrandomisierten Test von H_0 gegen H_1 wird der Stichprobenraum $\mathfrak{X}$ in zwei messbare Teilmengen K_0 und K_1 zerlegt.
Folgende Aussagen als Testergebnisse werden dabei offengelegt:

„Für $x \in K_0$ erfolgt die Aussage A_0 zugunsten von H_0." (Rüger, 1999, Seite 127)
Die Aussage A_0 kann dabei zum Beispiel wie folgt formuliert werden:
„H_0 wird aufgrund der Beobachtung nicht verworfen, H_0 kann beibehalten werden." (Rüger, 1999, Seite 127)

„Für $x \in K_1$ erfolgt die Aussage A_1 zugunsten von H_1." (Rüger, 1999, Seite 127)
Die Aussage A_1 kann dabei zum Beispiel wie folgt formuliert werden:
„H_1 ist aufgrund der Beobachtung signifikant." (Rüger, 1999, Seite 127)

Ein statistischer Test ϕ kann als Abbildung vom Stichprobenraum $\mathfrak{X}$ in einer zweielementigen Menge $\{A_0, A_1\}$ verstanden werden.
Formal definiert ist ein nichtrandomisierter Test ϕ als messbare Abbildung in $\mathfrak{X}$ mit $\{0, 1\}$ zu verstehen. Dabei ist für $\phi(x) = 1$ für $x \in K_1$ die Aussage A_1 zutreffend, für $\phi(x) = 0$ mit $x \in K_0$ die Aussage A_0.
Als Kriterien zur Bewertung der Qualität statistischer Tests wurden die Fehler 1. Art, $P(\text{Fehler 1. Art}) = P(H_0 \text{ ablehnen} \mid H_0 \text{ wahr})$, und Fehler 2. Art, $P(\text{Fehler 2. Art}) = P(H_0 \text{ beibehalten} \mid H_1 \text{ wahr})$, eingeführt. (Fahrmeir et al, 2011, vgl. Seite 421)
Als Beispiel lässt sich folgendes formulieren:
Bei „vorgegeben[en] Signifikanzniveau (...) [γ] und festen Stichprobenumfang n gibt eine

Gütefunktion g die Wahrscheinlichkeit für einen statistischen Test an, die Nullhypothese zu verwerfen. Speziell für den Gauß-Test ergibt sich die Gütefunktion $g(\mu)$ im Fall des Testproblems: $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ als $g(\mu) = \Phi(-z_{1-\gamma/2} + \frac{\mu-\mu_0}{\sigma}\sqrt{n}) + \Phi(-z_{1-\gamma/2} - \frac{\mu-\mu_0}{\sigma}\sqrt{n})$ (...) wobei Φ die Verteilungsfunktion der $N(0,1)$-Verteilung bezeichnet."(Fahrmeir et al, 2011, Seite 424)

4 Prinzipien für die statistische Inferenz

4.1 Einführung

Ein weiterer wichtiger Punkt bei der Einführung in die statistische Inferenz und Induktion sind die Prinzipien der statistischen Inferenz. Sie richten sich direkt an das Konzept, nach dem ein Schluss aus einer Beobachtung x aufzustellen ist. Dabei greifen sie in den Aufbau ein und spielen eine übergeordnete Rolle.
Nun werden im folgendem Abschnitt, einer Zusammenfassung der Seiten 128 bis 131, das Suffizienz- und Invarianz-Prinzip näher erläutert.

4.2 Suffizienz-Prinzip

Zu Beginn stellt sich die Frage: Wann gilt eine Schätzfunktion als gut? Was heißt dabei besonders gut?
„Ausgangspunkt der allgemeinen Definition des Suffizienzbegriffes ist eine Verteilungsannahme $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ über die Stichprobe X, also eine Familie von Verteilungen P_θ auf dem Stichprobenraum $\mathfrak{X}$, genauer auf dem messbaren Raum $(\mathfrak{X}, \mathfrak{B})$. Ist $T = T(x)$ eine Statistik, also eine messbare Abbildung von $(\mathfrak{X}, \mathfrak{B})$ in $(\mathfrak{T}, \mathfrak{C})$, so bilden für jedes feste $\theta \in \Theta$ und $B \in \mathfrak{B}$ die bedingten Wahrscheinlichkeiten $P_\theta(B \mid t) = P_\theta(X \in B \mid T = t)$ eine in t messbare Funktion ($\mathfrak{C}$-messbare Funktion).“(Rüger, 1999, Seite 38) „Eine Statistik T ist genau dann suffizient, wenn es eine Festlegung der bedingten Verteilung von X unter T gibt, die unabhängig von θ ist.“ (Rüger, 1999, Seite 38) Dabei ist T in den meisten Fällen einen k-dimensionaler Zufallsvektor und $(\mathfrak{T}, \mathfrak{C}) = (\mathbb{R}_k, \mathfrak{B}_k)$. $\mathfrak{T}$ ist ein messbarer Raum, indem sich der Stichprobenraum $\mathfrak{X}$ befindet. $\mathfrak{C}$ bezeichnet die geeignete σ-Algebra über $\mathfrak{T}$. (Rüger, 1999, vgl. Seite 34)
Das heißt: Aus $P_\theta(X \in B \mid T = t)$ folgt $P_{\not\theta}(X \in B \mid T = t)$.
Durch das Neymann-Kriterium, als eine notwendige Erweiterung der Definition der Suf-

fizienz, lässt sich eine suffiziente Statistik finden, die eine möglichst weitgehende Reduktion der Stichprobe erlaubt. Es besagt: „Liegt über die Stichprobe X eine dominierte Verteilungsannahme $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ vor, so ist eine Statistik $T = T(x)$ genau dann suffizient, wenn sich (für jedes $\theta \in \Theta$) die Dichten $f(x;\theta)$ von X (mit Ausnahme einer $\mathfrak{P}$-Nullmenge) faktorisieren lassen gemäß: $f(x;\theta) = h(x) \cdot g(T(x);\theta)$ wobei $h(x)$ eine von θ unabhängige (messbare) Funktion auf $\mathfrak{X}$ und $g(t;\theta)$ eine (messbare) Funktion auf $\mathfrak{T}$ ist.“ (Rüger, 1999, Seite 39)
Allgemein formuliert bedeutet es, dass sich eine Beobachtung x reduzieren lässt, wenn man x durch $T(x)$ als Stichprobenfunktion ersetzt und dabei den Schluss aus der beobachteten Stichprobe nur auf $T(x)$ stützt.
Bei dieser Reduktion von Beobachtung darf kein Verlust an Information im Schlussverfahren auftreten. Mittels dem Übergang von $(X, \mathfrak{P})$ auf $(T, \mathfrak{P}^T)$ lässt sich der Umfang einer Stichprobe reduzieren. Das Suffizienz-Prinzip weist nicht nur eine große Bedeutung für die Schätz- und Testtheorie auf.

4.3 Invarianz-Prinzip

Als weiteres Prinzip für die statistische Inferenz wird das Invarianz-Prinzip behandelt. Im Gegensatz zu dem Suffizienzprinzip ist es nur auf invariante Verteilungsannahmen $\mathfrak{P}$ anwendbar.
Die Definition von Invarianz lautet: $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ ist gegenüber einer Gruppe $\mathfrak{G}$ von Transformationen, als eine Familie von Abbildungen $g : \Psi \to \Psi$, eine invariante Verteilungsfamilie. Äquivariant einer Schätzung U für das unbekannte θ gegenüber $\mathfrak{G}$ bedeutet, wenn für alle x und $\tilde{x}$ in $\mathfrak{X}$ und alle g in $\mathfrak{G}$ gilt: aus $U(x) = U(\tilde{x})$ folgt die Transformation $U(gx) = U(g\tilde{x})$. Die beiden Beobachtungen führen zu dem selben Schätzwert, unabhängig von den ursprünglichen Beobachtungen und der mit g transformierten Skala.
Verallgemeinert gesprochen bedeutet dies, dass bei einer invarianten Verteilungsannahme äquivalente Schätzungen verwendet werden können. Solange die Verteilung bijektiv abbildbar ist, können die Schätzungen praktisch „umgetauft“ werden.
Beispielsweise kann bei einer Annahme der Normalverteilung, $N(\mu, 1)$, $\bar{x}$ durch μ geschätzt werden. Das „Umtaufen“ von $\bar{x}$ zu $(\bar{x} + b)$ führt zu μ_b und zum selben Schätzwert.

5 Zusammenfassung

Die vorherig beschriebenen Kapitel sollen eine allgemeine und grundlegende Einführung in die statistische Inferenz und Induktion darstellen und als Basis für weitere Vertiefungen in die Materie dienen, so wie Rüger sie in Test- und Schätztheorie auf den Seiten 111 bis 131 zusammengefasst hat.
Diese hier vorgestellten Inferenzkonzepte sind nicht nur für die Bereiche der induktiven Statistik bedeutungsvoll, sondern haben ebenfalls Auswirkungen auf alle Teilbereiche der Statistik.
Zusammenfassend ist zu sagen, dass der Umgang mit der Inferenz (vgl. Kapitel 1 und 2) gewisse Problematiken hervorrufen kann, da Unsicherheiten in den Daten oder im Vorwissen des Anwenders nicht ausgeschlossen werden können und ein Umgang mit jenen gefunden werden muss. Das sogenannte Induktionsproblem beschreibt die Problematik der Begründbarkeit oder Rechtfertigung eines induktiven Schlusses. (vgl. Kapitel 1 und 2)
Eine Unterscheidung des Begriffes der Wahrscheinlichkeit in subjektive und objektive Lager lässt sich hinsichtlich dem Zweck, Standpunkt und der Beurteilung einer statistischen Auswertung finden. (vgl. Kapitel 2.3)
Auswertungen erfolgen bei unvoreingenommenen Anwendern in einem Schätzverfahren, welche sich in Intervall- und Punktschätzung untergliedern lassen. (vgl. Kapitel 3.1)
In Testverfahren werden dagegen Hypothesen gegenüber einem unbekannten Parameter θ überprüft. (vgl. Kapitel 3.2)
Das Suffizienzprinzip ist das wichtigste der vorgestellten Inferenzprinzipien und wird in Kapitel 4.2 näher erläutert. Damit lässt sich eine Stichprobe hinsichtlich ihres Umfanges reduzieren ohne dabei einen Informationsverlust zu riskieren. „Ist T = T(x) eine Statistik,(...) so bilden für jedes feste $\theta \in \Theta$ und $B \in \mathfrak{B}$ die bedingten Wahrscheinlichkeiten $P_\theta(B \mid t) = P_\theta(X \in B \mid T = t)$ eine in t messbare Funktion (...).“(Rüger, 1999, Seite 38)

Literaturverzeichnis

Rüger, Dr. B. (1999). Test- und Schätztheorie, Vol. Band 1, R. Oldenbourg Verlag München Wien.

Fahrmeir, L. et al. (2011). Statistik - Der Weg zur Datenanalyse, Vol. 7, Springer-Verlag Berlin Heidelberg.