Komponenten in der Statistik-Ausbildung[1]

Günther Sawitzki

StatLab Heidelberg
Im Neuenheimer Feld 294
D 69120 Heidelberg
&
VIROR
- Virtuelle Universität Oberrhein -

VIROR, die “Virtuelle Hochschule Oberrhein” ist ein regionales Projekt zur Nutzbarmachung von Medien- und Netzwerktechnologien für die Ausbildung. Als Teil des VIROR-Projekts werden die Möglichkeiten von Komponenten-Technologien für den Unterricht erprobt. Ein Schwerpunkt ist dabei die Entwicklung von “wiederverwendbaren” Basiskomponenten für die Statistikausbildung. Wir diskutieren an einigen Beispielen Erfahrungen mit diesem Ansatz.

1. VIROR

1.1 Das VIROR-Projekt

VIROR (Virtuelle Hochschule Oberrhein) ist ein gemeinsames Projekt von vier oberrheinischen Universitäten (Freiburg, Heidelberg, Karlsruhe, Mannheim). Das Projekt wird finanziert durch das Förderprogramm "Virtuelle Hochschule" des Ministeriums für Wissenschaft, Forschung und Kunst, Baden-Württemberg, im Rahmen der "medi@"-Initiative des Landes. Im Projekt VIROR soll das Potential multimedialer Techniken zur Verbesserung und Effektivierung der Lehre und Studiermöglichkeiten genutzt, weiterentwickelt und evaluiert werden. Insbesondere sollen multimediale Lehr- und Lernsysteme entwickelt werden.
Das Programm verfolgt vor allem folgende Ziele:

- Gemeinsame Nutzung verteilter Ressourcen zur Bereicherung des Lehr- und Studienangebots einzelner Hochschulen
- Verminderung der Zeit- und Ortsabhängigkeit des Studiums
- Individualisierung des Lerntempos
- Förderung selbstgesteuerten Lernens
- Bessere Erschließung komplexer Inhalte durch multimediale Elemente

Weitere Informationen zu VIROR sind im Internet unter <http://www.viror.de/> zugänglich.

[1] Berlin SFB 373, 20.Sep. 99. Dieses Diskussionspapier entstand während eines Gastaufenthalts am Sonderforschungsbereich 373 “Quantifikation und Simulation ökonomischer Prozesse” der Humboldt-Universität zu Berlin, unterstützt von der Deutschen Forschungsgemeinschaft. Dank an W. Härdle und an alle Mitgliedern seiner Arbeitsgruppe für die Gastfreundschaft und angeregte Diskussionen.

1.2 Themenbereich Statistik im VIROR

In vielen Bereichen des VIROR-Projekts ist die Statistik ein wesentlicher Ausbildungsinhalt, insbesondere in den Bereichen Medizin, Psychologie und Ökonomie. Ein Teil-Projekt, formell als Teilprojekt T1.2 notiert, befaßt sich schwerpunktmäßig mit Statistik. Die wesentlichen Beiträge in diesem Teilprojekt kommen aus dem Grundlagenbereich mathematische Statistik sowie aus den Anwendungsbereichen statistische Physik und Biometrie.

Teil der Arbeit in T1.2 ist die Entwicklung und Evaluierung von multimedialem Material im üblichen Sinne: Lehrbuch- und Kursmaterial unter Verwendung aktueller Technologien, insbesondere im Bereich elektronischer Medien. Während dieses Material auf der traditionellen Struktur von Kursen aufbaut (und dabei auch oft traditionelles Kursmaterial wiederverwenden kann), konzentriert sich eine zweite Entwicklungslinie auf Material, das eine flexiblere Organisation des Unterrichts und des Lernens ermöglicht. Während die erste Entwicklungslinie das "elektronische Buch" als eine typische Metapher hat, konzentriert sich die zweite Linie auf kleine, in sich abgerundete Komponenten. Beide Entwicklungslinien sind nicht getrennt, sondern haben eine enge Verbindung. Material aus in sich abgeschlossenen Kursen und "elektronischen Büchern" kann isoliert und zu Komponenten umgeformt werden; Komponenten können als Bausteine für Kurse und Buchmaterial benutzt werden.

Beispiele aktueller Entwicklungen für Komponenten im Rahmen des VIROR-Projekts werden im folgenden gegeben, einige offene Fragen diskutiert und einige Verweise auf verwandte Entwicklungen gegeben.

1.3 Komponenten

Unter Komponenten werden hier kleine, im wesentlichen in sich abgeschlossene Lehr- oder Lerneinheiten verstanden. Eine Komponente bezieht sich typischerweise nur auf einen Begriff oder einen Effekt. In klassischer textorientierter Darstellung (Skript, Buch) entspricht sie etwa einem Abschnitt. Im Unterricht erfordert sie eine Zeitspanne, die deutlich innerhalb der Aufmerksamkeitsspanne der Rezipienten liegt.

Der Organisation des Inhalts in kleine, im wesentlichen in sich abgeschlossene Komponenten entspricht eine feine Granularität der unterstützenden Software für interaktive Teile des Lehrmaterials. Anstatt große monolithische Software einzusetzen, hat die Entwicklung und der Einsatz von wiederverwendbaren Softwarekomponenten den Vorrang.

Der Komponenten-Ansatz eröffnet Möglichkeiten. Anstatt den Lehrenden als jemanden zu sehen, der Lehrmaterial nur empfängt und weitergibt, sehen wir ihn als aktiven Autor, der seinen Kurs aus eigenen Ideen sowie aus verschiedenen anderen Quellen komponiert. Eine

Aufteilung in Komponenten hilft, das Material an unterschiedliche Kursverläufe anzupassen.

Die Eröffnung von Möglichkeiten gilt auch für Studierende. Komponenten laden dazu ein, das Material neu zu organisieren, auch über Kurs-Grenzen hinweg, um eigenen Bedürfnissen oder Präferenzen zu folgen.

Im Hintergrund des Komponenten-Ansatzes stehen Erfahrungen mit klassischem Lehrmaterial. Obwohl Textbücher reichlich vorhanden sind, scheint ihr Einsatz im Unterricht in der Statistik sehr beschränkt zu sein. Es ist durchaus üblich, Material abschnittsweise zu übernehmen. Die Bereitschaft, einen vollständigen Kurs oder größere Kurssegmente zu übernehmen, scheint jedoch gering zu sein. Noch geringer ist die Bereitschaft, Kursmaterial über Kursgrenzen hinweg zu übernehmen.Während im Hinblick auf den statistischen Inhalt große Teile von Kursen in verschiedenen Anwendungsgebieten übereinstimmen, erlaubt es die übliche Organisation des Kursmaterials kaum, etwa Komponenten eines Kurses für Physiker in einem Biometrie-Kurs für Mediziner zu nutzen. Dadurch, daß Komponenten stärker modularisiert sind als bei klassischem Material, kann der Austausch über Kursgrenzen hinweg erleichtert werden.

2. Beispiele aus der laufenden Entwicklung

Unterrichtsbeispiele sollten immer zumindest aus zwei Perspektiven betrachtet werden: aus didaktischer Perspektive (mit der Frage, *was* vermittelt wird) und aus methodischer Perspektive (mit der Frage, *wie* es vermittelt wird). Die Beispiele sind hier in der Reihenfolge aufgeführt, wie sie in einem üblichen Curriculum aufeinander folgen. In dieser Anordnung steht das methodisch interessanteste Beispiel am Anfang; die Reihe geht über zu mehr konventionellen Beispielen.

Die statistischen Softwarekomponenten benutzen eine Implementierung des Oberon-basierten Voyager-Systems (Sawitzki 1996). Als Betriebssystem steht Oberon in mehreren Varianten zur Verfügung. Die hier gezeigten Beispiele (Voyager II) sind einer Implementierung unter BlackBox (Oberon Microsystems, <http://www.oberon.ch/> entnommen. Die in der Forschung benutzte Entwicklungsumgebung (Voyager I) basiert auf dem ETH Oberon (System 3) <http://statlab.uni-heidelberg.de/projects/oberon/>.

Beispiel 2.1 : Aus der Einführung zur Statistik

Das erste Beispiel stammt aus dem Ergänzungsmaterial zu einer Einführung in die Statistik. Der Kurs, für den dieses Material entwickelt wurde, wurde von R. Dahlhaus in Heidelberg 1998 gehalten, die Übungen dazu von M. Eichler. Das Ergänzungsmaterial stammt von M. Sahm und G. Sawitzki 1998. Der Kurs ist eine Grundveranstaltung im Mathematik-Studium, die Teilnehmer sind vorrangig Mathematik-Studenten im Grundstudium mit einer Grundausbildung in Analysis und Vorkenntnissen in elementarer Wahrscheinlichkeitstheorie. Das

Ergänzungsmaterial ist zum begleitenden Selbststudium geeignet.

Das Ergänzungsmaterial konzentriert sich auf Grundkonzepte der Statistik. Neben den üblichen Hyperlink-Möglichkeiten werden zwei Interaktionen benutzt: freie graphische Eingabe von Funktionen und “drag&drop”-Dateneingabe. Diese unkonventionellen Möglichkeiten werden in einer Einführung illustriert. Im Anhang 1 ist ein kommentierter Ausdruck dieser Einleitung und der ersten Lektionen wiedergegeben. Lektion 1 führt die empirische Verteilung ein; Lektion 2 die empirische Verteilungsfunktion; Lektion 3 gibt eine Einführung in statistische Tests. Die statische Version im Anhang sollte nur zur Orientierung dienen. Dem Leser ist die interaktive Version empfohlen, zugänglich über

</http://www.statlab.uni-heidelberg.de/projects/viror/statgdp/>.

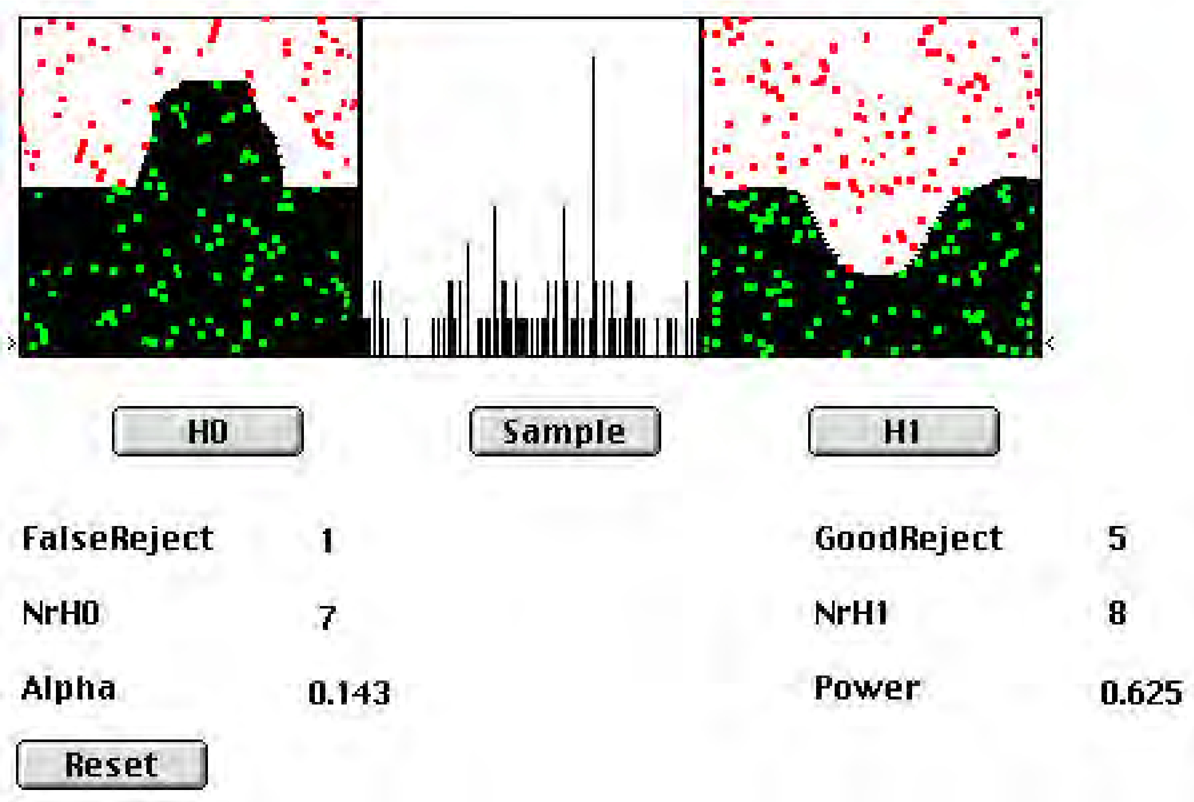


Abb 1. Aus dem Grundkurs für Statistik:
Anpassungstest gegen einfache Alternative, Details: siehe online-Version.

Beispiel 2.2: Lineare Regression

Im üblichen Mathematik-Curriculum hat die einfache lineare Regression an zwei Stellen ihren Platz. Sie kann im Rahmen der Einführung in die Statistik behandelt werden, oder als einführender Spezialfall im Rahmen der Statistik linearer Modelle. In der angewandten Statistik ist die einfache lineare Regression ein Standard-Bestandteil.

Das hier vorgestellte Material wurde für eine Vorlesung über Statistik linearer Modelle entwickelt; der Kurs wurde gehalten von E. Mammen, Heidelberg 1998/1999. Die Teilnehmer waren Mathematik-Studenten im Hauptstudium mit Grundausbildung in Statistik.

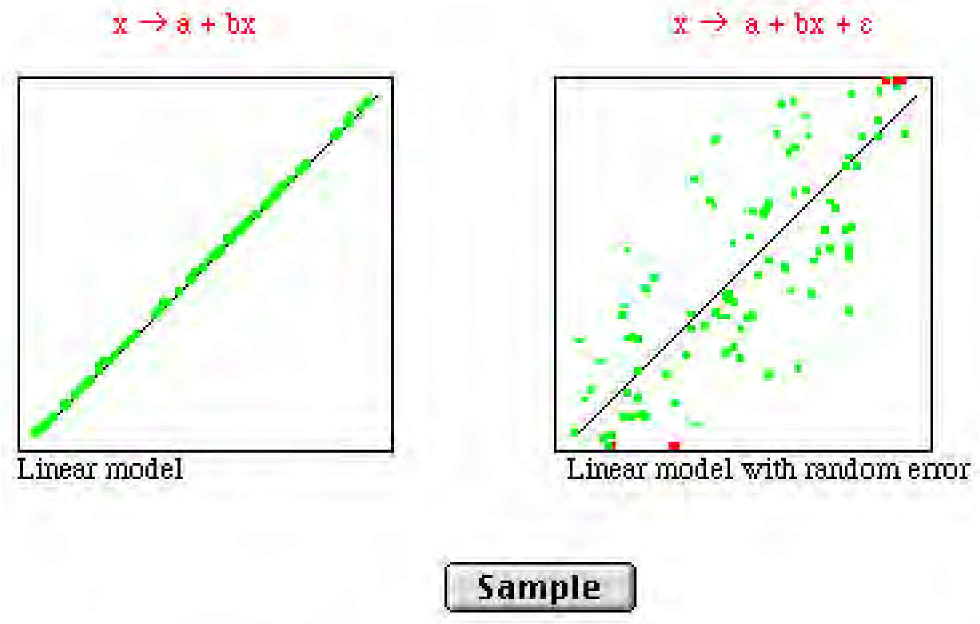


Abb. 2: Aus dem Kurs über lineare Regression.
Simulation eines einfachen linearen Modells

Der Kurs deckt das klassische Material über einfache lineare Regression ab. Dazu gehört auch eine Diskussion über das Verhalten bei Abweichungen von den Grundannahmen. Noch nicht immer üblich, aber von großer Bedeutung für die Praxis ist die Modelldiagnostik und Residuenanalyse. Diese Themen bedürfen einer eingehenderen Diskussion und praktischen Trainings. Im Rahmen des Kurses sind sie nur so weit andiskutiert, wie dies mit elementaren Mitteln möglich ist. Das Material kann als Kapitel in eine Vorlesung integriert werden oder eigenständig zum Selbststudium benutzt werden.

Eine online-Version ist über

</http://www.statlab.uni-heidelberg.de/projects/viror/>

zugänglich.

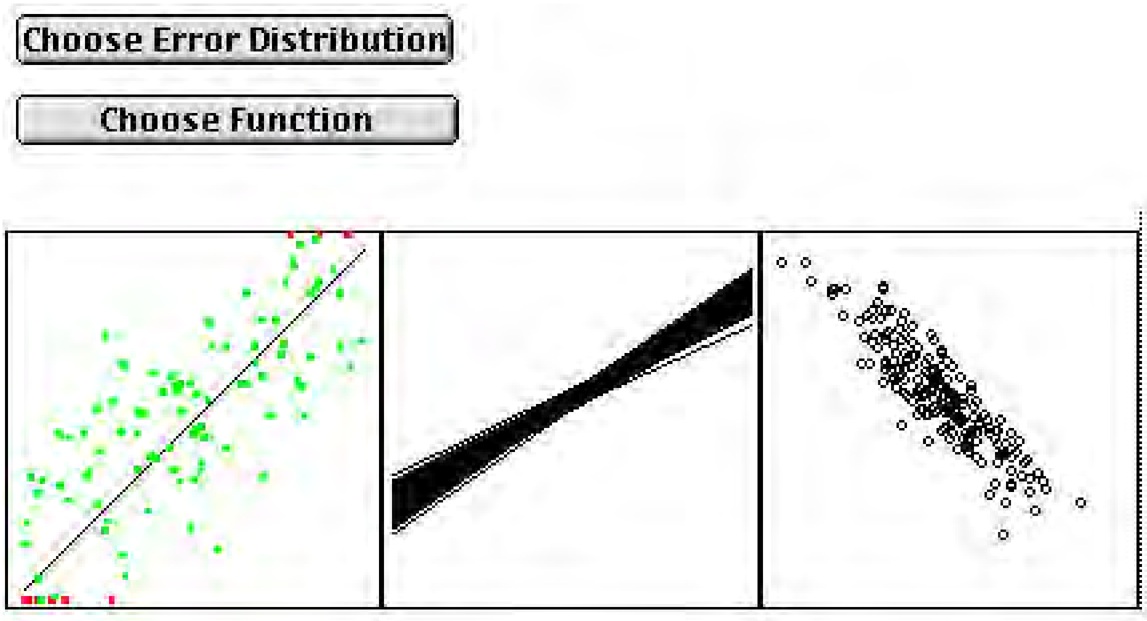


Abb.3:Aus dem Kurs über lineare Regression.
Monte-Carlo-Verteilung von geschätzten Regressionsgeraden und Parameterschätzern im einfachen linearen Modell

Beispiel 2.3: Kurvenschätzung

Glättung und Kurvenschätzung ist eine Spezialvorlesung im Rahmen der mathematischen Statistik. In der angewandten Statistik kann das Material am ehesten im Rahmen einer Weiterbildungsveranstaltung oder einer Graduiertenveranstaltung genutzt werden.

Das Material wurde entwickelt für eine Vorlesung über Kurvenschätzung, gehalten von E. Mammen und G. Sawitzki, Heidelberg 1999. Teilnehmer waren fortgeschrittene Mathematik-studenten oder Graduierte anderer Fachrichtungen. Das Kursmaterial hat den Charakter eines Vorlesungsskripts.

Schwerpunkt des Kurses ist die mathematische Untersuchung von Methoden der Kurven-schätzung. Der interaktive Teil des Kursmaterials hat dabei weitgehend die Aufgabe, interne Zusammenhänge der Methoden zu illustrieren.

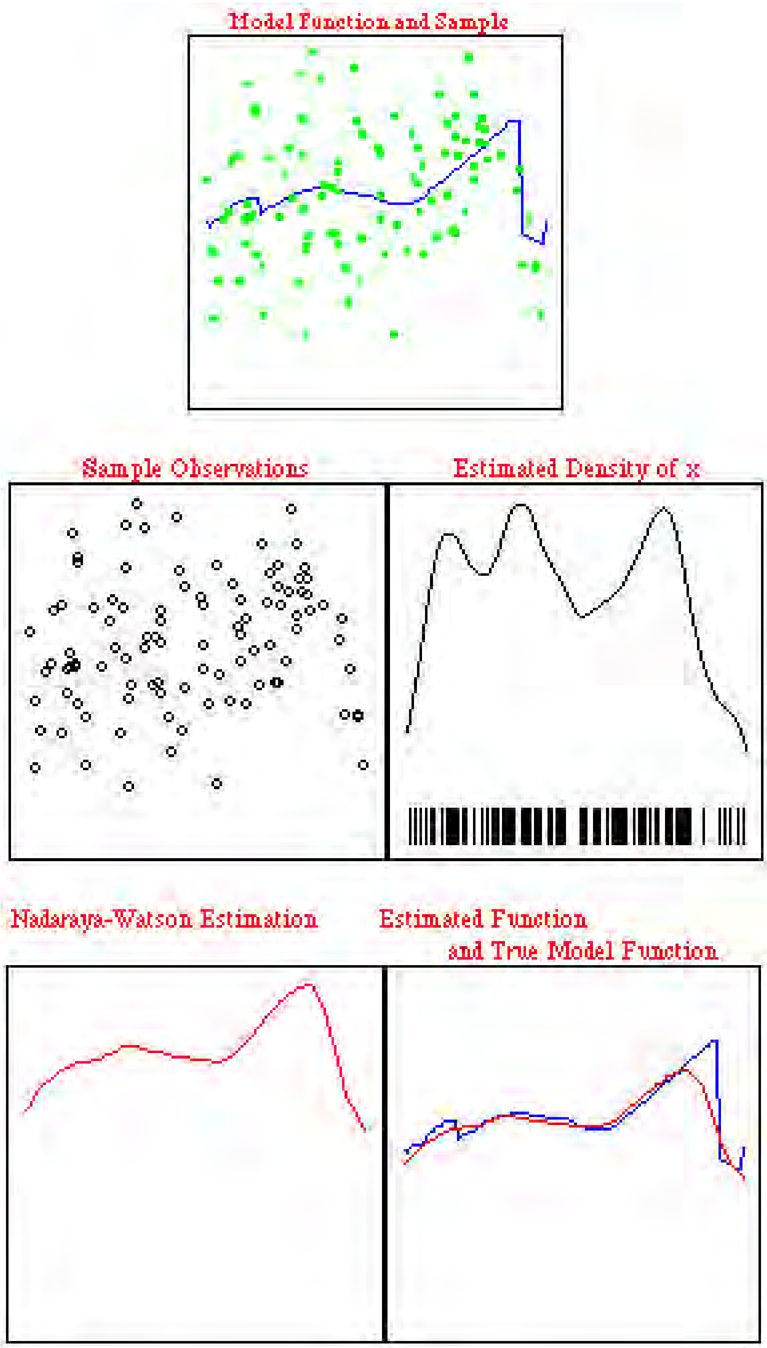


Abb.4:Aus dem Kurs über Kurvenschätzung: Nadaraya-Watson-Schätzung.

3. Diskussion

Komponenten bieten Möglichkeiten, die noch der kritischen Diskussion bedürfen. Diese Diskussion ist noch nicht geführt. Deshalb können hier nur - noch unverbundene und nicht zu Ende diskutiert - einzelne Aspekte beleuchtet werden.

3.1 Komponenten im Vergleich zu geschlossenen Kursen

Wie bereits erwähnt stehen Komponenten und geschlossene Kurse nicht im Gegensatz. Beide haben ihre Funktion, und wechselseitiger Austausch ist möglich. Den angemessenen Platz zu bestimmen ist eine offene Aufgabe.

Beispiel 1 (Grundkurs Statistik) bezieht sich auf einen Bereich, der konsolidiert ist: die wesentlichen Begriffe sind klar gefasst; kritische Eigenschaften sind bekannt. Dieser Bereich kann vollständig in einem Komponenten-Modell abgedeckt werden. Die Aufteilung der Komponenten bedarf noch der Diskussion.Traditionelle Darstellungen sind im wesentlichen an parametrischen Familien von Verteilungen orientiert; in den Anwendungen ist sogar die Folge Normalverteilung → parametrische Modelle → Grundkonzepte weit verbreitet (und bisweilen ist die Normalverteilung das einzige parametrische Modell, das diskutiert wird). Dies entspricht dem Stand der praktischen Möglichkeiten der 50er Jahre. Wesentliche Konzepte (wie Fluktuation, Güte etc.) können einfacher und mit weniger Irrtumsquellen diskutiert werden, wenn sie abstrakt gefasst werden. Wie Beispiel 1 zeigt, können diese Konzepte auch algorithmisch angemessen repräsentiert werden. Damit ist es möglich, für die Praxis bedeutsame Konzepte (wie Fluktuation, Güte etc.) von Hilfskonstrukten (wie Varianz, Mittelwert, Score) zu trennen und die Aufmerksamkeit auf die zentralen Konzepte zu konzentrieren.

Beispiel 3 (Glättung) eröffnet einen anderen Bereich. Während für bestimmte Bereiche geschlossene Lösungen vorliegen, bedürfen andere Bereiche (Dimensionsproblem, Regularität etc.) noch der Diskussion. Diese kann am besten in einem geschlossenen Kontext erfolgen. Komponenten haben hier nur eine helfende Rolle als Bausteine. Beim offenen Stand der Diskussion scheint ein Dokumentenmodell hier eher eine angemessenere Repräsentation.

Die kritische Bedingung für ein Komponenten-Design wird durch die Referenzen gegeben. Domininieren übergreifende oder globale Referenzen, so spricht dies für ein Dokumentenmodell. Dominieren interne Referenzen, so ist ein Entwurf mit isolierten Komponenten sinnvoll.

Andere Wege sind gangbar, so etwa illustriert durch das Erklärungssystem der CADEMO-Software <http://www.biomathematics.com/eprod.htm>. Dieses System ist als Lexikon-System mit festen internen Links angelegt; von diesem “Gerüst” erfolgen Verweise zu einzelnen Moduln. Der am Modell eines Nachschlagewerks orientierte Aufbau führt hier natürlich zu einem System, das in Teilen eine Komponentenarchitektur hat.

Bei der Migration von bestehendem Material ist ein möglicher Zwischenschritt, dieses Material zunächst mit Links zu markieren und externe Hilfsmodule anzuschliessen. Dieser Weg wird z.B. in (Müller, 1998) beschrieben. Um von hier zu einer flexiblen Verwendbarkeit wie bei Dokumentenmaterial zu kommen, ist der nächste Schritt, die innere “Architektur” zu revidieren und in sich geschlossene Bestandteile zu isolieren, um sie dann im nächsten Schritt als “Bausteine” aufzubereiten.

3.1.a Software- und Erklärungskomponenten

Die Unterrichtskomponenten, die hier das Thema sind, bestehen in der Regel aus mindestens zwei Bestandteilen: einer Softwarekomponente und einer Erklärungskomponente. Idealerweise gilt dieselbe Flexibilität und Wiederverwendbarkeit wie für die Unterrichtskomponenten auch intern für deren Bestandteile, für Softwarekomponente und Erklärungskomponente. Dies ist nicht erforderlich, wenn Unterrichtskomponenten als ganzes übernommen und in einen anderen Kontext eingefügt werden sollen. Sollen sie jedoch integriert werden, d.h. in der Regel Bezeichnungen dem Kontext angepaßt, Konventionen aufeinander abgestimmt werden etc., so ist diese Trennbarkeit notwendig.

Diese Trennbarkeit ist zwangsläufig nahegelegt, wenn die Unterrichtskomponenten eine andere technische Basis benutzt als die Software, etwa HTML für die Erklärungskomponente und JAVA für die Software. Diesen Weg verfolgen etwa mm*stat oder JUMBO. Eine andere Frage ist, ob diese Trennung so weit an den potentiellen Coautor weitergereicht wird, daß sie effektiv genutzt werden kann. mm*stat benutzt hier extensiv Skripte und dynamisch geladene Seiten, die für eine Komponentennutzung praktisch Zugriff auf das dahinter stehende Skriptsystem und die Originalquellen voraussetzen. Im Vergleich dazu benutzt z.B. JUMBO im wesentlichen "konventionelle" HTML-Programmierung und lädt dazu ein, die präsentierten Seiten als Bausteine oder Fundquellen für Applets zu benutzen.

HTML-basierte Lösungen führen eine neue Komplexitätsschicht ein: auf die Seiten ist kein direkter Zugriff möglich, sondern nur eine mittelbarer durch die Veränderung des HTML-Codes, etwa mithilfe eines HTML-Editors.

Im Gegensatz dazu erlaubt es die in Voyager gewählte Implementierung, bei hinreichenden Berechtigungen durch cut&paste bzw. drag&drop direkt auf das Material zuzugreifen und es, falls gewünscht, interaktiv zu modifizieren.

3.1.b Softwarebasis und Transfer

Statistik, insbesondere Angewandte Statistik, ist in der Praxis weitgehend rechnergestützt. Als Inhalt multimedialer Lehr- und Lernmaterialien spielt sie dadurch eine besondere Rolle: während bei anderen Inhalten Lehr- und Anwendungsumgebung meist getrennt sind, benutzen multimediale Lernsituationen und Statistikanwendungen weitgehend dieselben Umgebungen. Dadurch ist es prinzipiell auch möglich, für die Lehrsituation und für die Anwendungssituation gemeinsame Software zu benutzen. Wird diese Möglichkeit nicht genutzt, so ist ein weiterer Lernschritt und ein Transfer nötig: neben den statistischen Methoden muß dann die Handhabung der statistischen Auswertungssoftware gelernt werden; gelerntes Wissen muß von der Lernumgebung in die Auswertungsumgebung transferiert werden.

Ein Blick auf verfügbare Einführungen und Benutzeranweisungen zeigt, daß das Training in der Anwendungssoftware einen vergleichbaren Umfang einnehmen kann wie die eigentliche Statistikausbildung. Von der statistischen Sache her marginale Unterschiede wie unterschiedliche Problemformulierungen, Syntax und Bezeichnungen oder Organisation des Materials können diesen Aufwand noch unnötig aufblähen, so daß hier die Gefahr besteht, wertvolle Ressourcen unnötig zu verschwenden.

Es ist möglich, diese Aufweitung zu vermeiden, wenn die Statistik-Ausbildung diejenige Software benutzt, die auch in der praktischen Auswertung eingesetzt werden kann. Je nach Themenbereich wird dies in unterschiedlichem Maße gelingen: wo es um grundlegende Konzepte und allgemeine Zusammenhänge geht, hat spezielle Demonstrationssoftware ihren Platz. Wo es um Methoden der Datenauswertung geht, sollte jedoch nach Möglichkeit auf Sondersoftware verzichtet werden und unnötige Transferschritte sollten vermieden werden.

Die Verbindung zwischen theoretischem und praktischem Unterrichtmaterial wird z.B. in (Härdle 1998, s. Müller, 1998) geschaffen. Die zugrundeliegende Software ist hier XploRe; Illustrationen und Beispiele verweisen auf XploRe-Makros oder XploRe-Aufrufe, die modifizierbar sind und in praktischer Arbeit als Schablonen zur Problemlösung genutzt werden können. (Härdle 1998) arbeitet hier zweistufig: das eigentliche Kursmaterial sind vorbereitete Screendumps und Outputs; der XploRe-Hintergrund ist durch Inspektion zugänglich.

Diese Verbindung wird in mm*stat nicht mehr aufrecht erhalten: das theoretische Material ist hier durch Beispiele ergänzt. Dem Studenten wird jedoch nicht die Möglichkeit geboten, zu inspizieren, wie diese Besipiele praktisch gerechnet werden. Modelle für den Einsatz in praktischer Auswertung werden nicht gegeben. Eine Möglichkeit der Inspektion wie im früheren (Härdle 1998)-Material wäre wünschenswert.

Velleman benutzt in ActivStat das DataDesk als Arbeitssoftware. Präsentation der Software wird im Kurs auf das Minimum reduziert, daß zur Durchführung des Kurses benötigt wird. Alle praktische Übungen benutzen jedoch DataDesk direkt, und die erworbenen Fähigkeiten sind für die praktische statistische Arbeit mit DataDesk unmittelbar einsetzbar.

In den Voyager-basierten Materialien wird ein (unmodifiziertes) Voyager-System benutzt. Da hier sowohl die Software als auch das Kursmaterial eine Komponentenarchitektur haben, können alle Elemente sowohl innerhalb des Kurses auf Benutzer-Daten angewandt werden, als auch autonomes Material übernommen werden.

Tendenziell bestünde diese Möglichkeit von der Architektur her auch bei JUMBO (Heineke & Köpcke 1998±). Dieses Material benutzt jedoch Java-Applets. Bei der Auswahl stand der Demonstrationscharakter im Vordergrund, so daß Ein- und Ausgabeprobleme, Performanceeinschränkungen und numerische Instabilitäten oft der praktischen Nutzung im

Wege stehen.

Eine Alternative ist es, die Statistik-Ausbildung mit der Einweisung in Auswertungssysteme zu verbinden. Dieser Weg wird z.B. in (Gentleman 1995), (Ripley 1996) gegangen. Der inhaltliche Schwerpunkt dieser Kurse verschiebt sich jedoch in Richtung auf die Einführung in Auswertungssysteme.

3.2 Kommunikation von Komponenten

Komponenten sind definitonsgemäß sehr spezifisch; Identifikation und Dokumentation dementsprechend komplex. Das kompliziert das Auffinden und die Wiederverwendung von Komponenten. Die naheliegende Lösung, Komponenten-Kataloge zu erstellen, wird damit schnell unpraktikabel.

Vorzuziehen ist es, Komponenten im Kontext darzustellen, etwa in Analogie zu den im Schulunterricht verbreiteten Unterrichtseinheiten. Dies sind größere, thematisch zusammenhängende Sequenzen, die als “Kurs-Block” eingesetz werden können. Die Einführung von Tests (Beispiel 1) oder einfacher linearer Regression (Beispiel 2) folgen diesem Modell. Jedes dieser Modelle fasst mehrere Komponenten zusammen und illustriert die Verwendung der Komponenten in einem Kontext. Im Idealfall ist dieser Kontext ein hinreichendes Beispiel, so daß keine weitere Dokumentation für die Komponenten benötigt wird.

Für den effektiven Einsatz ist eine zusätzliche Voraussetzung, daß die vorgestellten Komponenten in flexibler Weise (z.B. Cut&Paste) aus dem vorgestellten Kontext in den jeweiligen Anwendungskontext übernommen werden können.

3.3 Einbettung von Beispielen

Beispiele treten typischerweise in zwei Rollen auf: als motivierende Beispiele bei der Begriffsbildung und als illustrierende Beispiele im Anschluss an eine theoretische Darstellung.

Die Identifikation und Definition motivierender Beispiele bedarf in der Regel größeren Aufwands. Ist diese Arbeit geleistet, so sind motivierende Beispiele in der Regel so prägnant, daß ein Transfer zwischen verschiedenen Anwendungsbereichen möglich ist. Beispiele dieser Art gehören in die Grundmodule. Ein typischer Repräsentant dieser Art sind Beispiele für zensierte Daten. Die Problematik zensierter Daten ist zuerst umfassend im Bereich der Biometrie diskutiert worden (mit Rückgriff auf Methoden aus technischen Anwendungen). Überlebenszeit-Probleme sind nach wie vor am besten geeignet, die damit verbundenen statistischen Probleme zu demonstrieren.

Illustrierende Beispiele andererseits sind weitgehend austauschbar. Ihre Aufgabe ist es, theoretisch bereits definierte Konzepte oder Methoden im für den Rezipienten gewohnten Kontext zu illustrieren. Ein typischer Repräsentant dieser Art sind Beispiele für Regression. Gleichermassen relevante Beispiele finden sich in allen Anwendungsbereichen. Diese Beispiele haben nur ergänzenden Charakter. Eine kanonische Vorgabe ist hier nicht sinnvoll: in der Regel hat der jeweilige Dozent bessere, treffendere, dem Kursverlauf besser angepasste Beispiele.

Für die “Architektur” eines Kurses gibt dies die Empfehlung, technisch und strukturell die illustrierenden Beispiele vom Korpus des Kurses zu trennen. Im Korpus sollte eine funktionelle Beschreibung des Beispiels verankert sein; das eigentliche Beispiel kann in loser Kopplung kontext-abhängig ergänzt werden. Diese Trennung ist tendenziell in mm*stat gegeben, während andere Kurse wie AktivStat Beispiel und Kontext eng miteinander verknüpfen und dadurch nicht einfach übertragbar sind.

3.4 Interaktion

Die Rolle von interaktiven Elementen bedarf noch einer kritischen Diskussion. Ähnlich dem Wort “multimedial” hat “interaktiv” eine Inflation erfahren. In diesem Kontext soll interaktiv im engen Sinne eines möglichen Eingriffs in eine Aktion verstanden werden. Eine Aktion, ausgelöst durch einen Button oder ein Kommando, ist so wenig interaktiv wie der Druck auf einen Lichtschalter. In diesem Sinne wird hier zwischen Kommando-aktionen und interaktiven Aktionen unterschieden.

Die Beispiele aus der Grundvorlesung (Abschnitt 2.1) sind in diesem Sinne zunehmend echte Interaktionen. Dies geht jedoch mit zunehmender Komplexität einher. Die Lektionen beginnen mit dem Stichprobenziehen. Für dies ist Interaktion verzichtbar: die gleiche Funktion kann durch eine wohlausgewählte Reihe von statisch vorbereiteten Plots erreicht werden. Eingabe von Umfang der Grundgesamtheit und der Stichprobe sind Aktionen, die typischerweise auch in Kommando-Umgebungen möglich sind. Die interaktive Umgebung hat hier nur die Rolle, eine bequemere Umgebung zu bieten. Echte Interaktion kommt ins Spiel bei der Möglichkeit, Verteilungen frei zu definieren.

Erfahrungsgemäß bereitet die sinnvolle Nutzung hier die größten Schwierigkeiten. Die herausfordernde Frage ist hier: welche Eigenschaften der Verteilung sind kritisch für die Fluktuation? In der Regel ist die erwartete Antwort, daß prägnante Eigenschaften kritisch sind. Die in der Simulation gewinnbare Erfahrung ist jedoch, daß die flachen Abschnitte der Verteilung im statistischen Verhalten eher kritisch sind. Selbst in einer interaktiven Umgebung scheint hier eine Anleitung notwendig.

Die aus dem Kapitel über lineare Regression (Abschnitt 2.2) entnommenen Abbildungen geben Beispiele für weniger hilfreiche Simulationen. Die Interaktion ist hier auf einen Knopfdruck beschränkt - eine Handlungsmöglichkeit, die nicht den Namen “interaktiv” verdient. Das Beispiel (Abb. 2) ist zudem irreführend. Abb.2 gibt vor, typische Stichproben aus einfachen linearen Modellen zu simulieren. In der Simulation werden jedoch nur Stichproben mit uniform verteilten Regressoren gezogen. Dies ist ein Spezialfall, der gerade für den Anfänger nicht als solcher zu erkennen ist. Nicht homogen verteilte Regressoren oder auch systematische Designs gehören ebenso zu den typischen Fällen für einfache lineare Modelle wie der hier gezeigte Spezialfall. Die Simulation kann in diesem Beispiel besser durch eine wohl ausgewählte Serie von statischen Abbildungen ersetzt werden.

Im Beispiel (Abb. 3) gibt die Dynamik einer laufenden Simulation einen deutlicheren Eindruck von den Verteilungseigenschaften der Regressionsgeraden und dem geschätzten Parameter als ein statisches Bild - insbesondere auch beim Vergleich unterschiedlicher Fehlerverteilungen. Die Interaktion ist hier jedoch im wesentlichen auf die Auswahl aus einer Liste von Mittelwertsfunktionen und Fehlerverteilungen beschränkt. Ein gleichwertiger Effekt wird erreicht, wenn anstelle einer online-Simulation eine vorbereitete Sequenz von Bildern abgespielt wird.

Von unterschiedlicher Qualität ist die Simulation in Abb. 4 aus der Vorlesung über Kurven-schätzung. Ein komplexer Zusammenshang, die Nadaraya-Watson-Schätzung, wird hier in einer Serie von Teilschritten aufgelöst. Wie in Abb. 2 und 3 ist die erste Anwendung eine bloße Illustration, die auch mit einer Bildsequenz erreicht werden könnte. Darüber hinausgehend schließt sich jedoch natürlich die Frage an, wie flexibel die Nadaraya-Watson-Schätzung unter-schiedliche Mittelwertsfunktionen nachbilden kann. Die Interaktion bietet die Möglichkeit, die Modellfunktion frei graphisch zu definieren, und gewinnt dadurch gegenüber einer vordefinierten Bildsequenz.

3.5 Installationsvoraussetzungen

Im Gegensatz zu klassischen Medien wie Büchern gibt es bei neueren Medien zahlreiche Vor-aussetzungen, die erfüllt sein müssen, um die Medien zu nutzen. Die Verfügbarkeit ändert sich rasch. Deshalb scheint es angebracht, neben den Voraussetzung auch einen Zeit-Horizont zu spezifizieren, für den Produkte ausgelegt sind. Dies betrifft zum einen die notwendige Hard-ware und Systemsoftware. In zunehmendem Maße kommt vermittelnde Software wie Browser hinzu.

Vom technischer Seite sind in längerer Perspektive modulare Lösungen vorzuziehen, die sich auf wohl-definierte Funktionalität beschränken und auf allgemeinere Aufgaben auf System-leistungen zurückgreifen. Der gegenwärtige Stand der Systemsoftware erlaubt diese im Prinzip wünschenswerte Lösung nur in Ansätzen. Eine derzeit weit verbreitete Lösung ist der Rück-griff auf Browser. Damit wird allerdings eine zweite Abhängigkeit eingeführt; Nutzungs-voraussetzung sind nun System und Browser.

Browser erfüllen jedoch eine zweite für die Akzeptanz wichtige Funktion. Sie suggerieren für den Benutzer eine stabile, vertraute Umgebung. Dieser psychologische Effekt ist nicht zu unterschätzen und kann den Zugang erleichtern.

Vom technischen Standpunkt aus ist diese Stabilität oft nur an der Oberfläche gegeben. Damit Browser wie Netscape oder Micorsoft Explorer als stabil und homogen erscheinen, werden in der Praxis oft weitere Software-Module in Form von Plug-Ins oder Erweiterungen benötigt, die dazu dienen, weitere Funktionalität bereitzustellen oder Performance-Probeleme zu kompensieren. Dies schafft zum einen evtl. undokumentierte Installationsvoraussetzungen. Zum anderen wirft es massive Sicherheitsprobleme auf. Während die gängigen Browser hinreichend gut bekannt und ihre Sicherheitsprobleme weit diskutiert sind, öffnen Plug-Ins tendenziell weitere Hintertüren.

3.5.1 Software- und Hardwarevoraussetzungen; Zugänglichkeit

Eine kritische Frage ist dabei, inwieweit Multimedia-Produkte von anderer Software abhängig sind. Die hier vorgestellten Beispiele basieren softwareseitig auf dem Oberon-Betriebssystem. Oberon ist als natives Betriebssystem oder als Emulation unter gängigen Betriebssystemen auf allen heute üblichen Plattformen verfügbar. Dadurch ist Hardware- und Softwareunabhägigkeit gewährleistet.

Genauer: in der vorliegenden Form kann die Software unmittelbar auf jedem Rechner mit 32Bit-Windows (Windows 3.1 mit 32Bit-Extension, Windows 9x, Windows NT) oder MacOS (System 7 oder besser) eingesetzt werden. Keine weitere Software wird benötigt. Voraussetzungen für eine Speicherung sind 2MB Festplattenplatz/Kurs; zur Laufzeit sind mindestens 2MB Kernspeicher empfohlen. Damit liegt der Hardware- und Softwarebedarf an der unteren Grenze des heute zu Erwartenden.

Die Ökonomie des Oberon-Systems erlaubt es dabei, Laufzeitsystem, Anwendungsprogramm und Dokumente so zu bündeln, daß ein vollständiges System (incl. Dokumente) auf einer Diskette Platz findet.

Mit dem vollständigen Material im Umfang von einer Diskette kann das Material auch über Netze bereitgestellt werden und selbst unter ungünstigen Bedingungen mit vertretbarem Aufwand heruntergeladen werden.

3.5.2 Sicherstellung der Nutzungsvoraussetzungen

Das Voyager II -Material ist mit allem begleitenden Material gebündelt und hat einen Umfang, der eine Verbreitung auf Disketten oder über das Netz erlaubt. Die Bündelung des Materials stellt eine gewisse Redundanz zur Verfügung - Laufzeitsystem und unterstützende Komponenten sind bei jedem Kursmaterial in Kopie enthalten. Der modulare Aufbau und die Komponentenarchitektur ermöglichen auch eine Verteilung in Komponenten. Damit reduziert sich das kursspezifische Material auf deutlich unter 0.5 MB/Kurs.

3.5.3 Portabilität und Sicherung der Verfügbarkeit

Die aktuelle Voyager II-Software setzt ein entsprechendes Oberon-Laufzeitsystem voraus. Das verwendete System ist derzeit nur für Windows und Macintosh verfügbar und wird von einem kommerziellen Hersteller bereitgestellt.

Die eigentliche Entwicklungsumgebung für die interne Entwicklung (Voyager I) benutzt das Oberon-Laufzeitsystem der ETH Zürich. Dies ist auch für andere Plattformen wie Linux verfügbar. Als “second sources” dienen neben der ETH Zürch die UC Irvine und die Universität Linz; das System selbst ist im Quellcode verfügbar. Damit ist auch längerfristig gesichert, sodaß durch das System keine kritische Abhängigkeit geschaffen wird.

Voyager II hat derzeit die bessere Integration in unterschiedliche Betriebssystemumgebungen, bedarf aber einer jeweiligen Anpassung. Voyager I benutzt ein portables Objektformat und ist damit ohne Konversion und ohne jede Neukompilierung plattformunabhängig.

Die Frage der Dauer der sinnvollen Einsetzbarkeit läßt sich nur vage beantworten und hängt von der allgemeinen technischen Entwicklung ab. Die interne Planung bei Beginn des Voyager-Projekts ging für die aktive Software-Entwicklung von einen Zeitraum von ca. 7 Jahren aus. Die Prognosen, die dieser Planung zugrunde lagen, haben sich im wesentlichen bestätigt. Damit scheint eine weitere aktive Entwicklungsphase für ca. 4 Jahre sinnvoll. Frühere Projekte in ähnlicher Richtung lassen eine Nutzungsdauer von weiteren 10 Jahren erwarten, für die in vertretbarem Umfang eine Unterstützung zu sichern ist.

Im Sinne eines Software-Generationenmodells stellt sich die Frage der Code-Migration in andere Umgebungen. Für die aktuelle Voyager-Entwicklung ist Oberon als Programmiersprache eine kritische Grundentscheidung. Von Oberon sind automatische Konversionen nach Java und C verfügbar und erprobt. Damit ist auch von der Sprachen-Seite eine langfristige Erhaltung möglich.

3.6 Flexibilität

Das hier vorgestellte Material kann zum Selbststudium benutzt werden. In erster Linie wird es jedoch im Zusammenhang mit Kursen gesehen, sei es als Kursmaterial selbst, sei es als ergänzendes Material. Der Lehrende, der dieses Material benutzt, wird dabei nicht als Konsument, sondern als aktiver Autor gesehen. Damit stellt sich die Frage, welche Instrumente diesem Autor an die Hand gegeben werden, um aus den Komponenten und anderen Quellen einen eigenen Kurs zu gestalten.

In der hier verwendeten Umgebung sind alle Komponenten frei editierbar. Genauer: Dokumente können in unterschiedlichen Modi geöffnet werden. Im “Browser”-Modus können einzelne Komponenten nur benutzt, aber nicht verändert werden. Im “Edit”-Modus ist freies editieren möglich.

Bislang fehlt noch eine automatische Unterstützung für Hyperlink-Funktionen. Alle Links müssen derzeit explizit gesetzt werden und sind statisch. Dies widerspricht der Idee eines modifizierbaren Dokuments.

Wünschenswert ist eine dynamische Generierung von Links, mit zugehöriger Indizierung. Diese Möglichkeit bekommt besondere Bedeutung, wenn eine individualisierte Nutzung der Dokumente ermöglicht werden soll (Bookmarks).

3.6a Konsistenz

Flexibilität und Modifizierbarkeit werfen die Frage nach Konsistenz auf. Die damit verbundenen technischen Probleme sind an anderer Stelle ausführlicher diskutiert (Sawitzki 1999).

3.7 Komponenten als eigenständiges Lern-/Lehrmaterial

Ihrer “Feinkörnigkeit” wegen sind Komponenten als eigenständiges Lehrmaterial nur eingeschränkt sinnvoll. Sie konzentrieren sich auf einen Begriff oder einen Effekt, während eigenständiges Lernen einen größeren Kontext voraussetzt. Komponenten werden sinnvollerweise in einen Kontext eingebettet (wie in Abschnitt 2 illustriert) als Lern-/Lehrmaterial eingesetzt.

Sie können jedoch eine spezielle Bedeutung als “Teaser”, als Anreizmaterial bekommen. Damit sie diese Funktion erfüllen können, müssen sie einfach zugänglich sein. Aus dem Voyager-Material sind <http://statlab.uni–heidelberg.de/projects/random/> und <http://statlab.uni–heidelberg.de/projects/onedim/> Beispiele für den Komponenten-Einsatz in dieser Funktion. Die JUMBO-Sammlung von Applets <http://medweb.uni–muenster.de/institute/imib/lehre/skripte/biomathe/jumbo.html> können hier ebenfalls als Beispiele dienen.

3.8 Einsatz von Komponenten in Lehrveranstaltungen

Die Flexibilität von Komponenten und ihre Modifizierbarkeit unterstützt ihre Integration in Lehrverstaltungen. Der Einsatz kann dabei unterschiedliche Gestalt haben. Die Voyager II-Komponenten mit ihrem Begleitmaterial haben jeweils eine Form, die geeignet ist, als Modul in einen Kurs übernommen zu werden. Das Material ist soweit in sich abgerundet, daß es auch als begleitendes Kursmaterial zum Selbststudium übernommen werden kann. Die in Abschnitt 2 gegebenen Beispiele skizzierten unterschiedliche Reichweiten: während die angegebenen drei

Abschnitte (Beispiel 2.1) als Modul im Rahmen einer Statistik-Vorlesung miteinander verbunden oder in loser Folge eingestreut werden können, ist für die einfache lineare Regression eher eine geschlossene Übernahme als Kapitel sinnvoll. Beispiel 2.3, die Kurvenschätzung, hat einen Buch-Charakter mit wechselseitigen Verweisen und erfordert eine Anpassung, wenn der Text in einem anderen Ablauf benutzt wird.

Neben diesem durch die Beispiele vorgezeichneten Ablauf ermöglicht es die Komponenten-Architektur, einzelne Komponenten bis hinunter zu einzelnen Abbildungen zu übernehmen, unter voller Beibehaltung der Dynamik.

Für "klassische" Lehrveranstaltungen bietet sich insbesondere bei den Komponenten, für die die Interaktion nicht benötigt wird und die Dynamik nur marginal ist, eine Übernahme als Screen-Dumps oder Folien an, nachdem diese Beispiele der jeweiligen Lehrveranstaltung angepaßt und entsprechenden selektiert worden sind. Abschnitt 2.2 gibt Beispiele, die in dieser Form benutzt werden können.

3.9 Unterstützung

Weitergehende Unterstützung für Lehrende und Lernende ist eine naheliegende Ergänzung.

ActivStats bietet für registrierte professionelle Nutzer eine (zeitlich befristete) Web-basierte Unterstützung und statistische Beratung an. Daneben wird weiteres Material zur Unterrichtsunterstützung zur Verfügung gestellt.

Die Voyager-Software ist im alpha-Stadium. Für registrierte "Alpha-Nutzer" wird direkte Unterstützung angeboten.

Literatur:
(Bei Web-basiertem Material sind Veröffentlichungsjahre als ungefähre Angaben mit einer Kennzeichnung ± angegeben.)

Derby, N.; Härdle, W.; Rönz, B.: The Three Dimensions of Multimedia Teaching in Statistics. Preprint. Berlin 1999

Gentleman, R.: An Introduction to Statistical Computing Using R. Auckland 1995.

Härdle, W. et al: Applied Multivariate Statistical Analysis, Electronic Book (in preparation,1998)

Heinecke, A; Wolfgang Köpcke, W.: JUMBO: Java enhanced material for biometry. Münster 1998± <http://medweb.uni–muenster.de/institute/imib/lehre/skripte/biomathe/jumbo.html>

Müller, M.: Computer-assisted Statistics Teaching in Network Environments. COMPSTAT'98 Proceedings, Bristol, UK.

Sahm, M.; Sawitzki, G.: Grundkonzepte der Statistik. Heidelberg, 1998

<http://www.statlab.uni-heidelberg.de/projects/viror/statgdp/>
Sawitzki, G.: Extensible Statistical Software: On a Voyage to Oberon. Journal of Computational and Graphical Statistics Vol 5 (1996) 263 - 283. <http://www.statlab.uni–heidelberg.de/projects/voyager/>
Sawitzki, G.: Keeping Statistics Alive in Documents. Heidelberg 1999 <http://statlab.uni-heidelberg.de/users/gs/Components.pdf>
Velleman, P. et al.: ActivStats. Addison Wesley Longman 1996± <http://www.datadesk.com/ActivStats/>
Venables W. N. and Ripley B. D. (1997) Modern Applied Statistics with S-Plus. Second Edition. Springer, 462 pages, ISBN 0 387 98412 0.

Grundkonzepte der Statistik
- Kommentierter Auszug -

G Sawitzki
StatLab Heidelberg
und
VIROR - Virtuelle Universität Oberrhein
<http://www.statlab.uni-heidelberg.de/projects/viror/>

Noch in Vorbereitung. Stand: 4.10.99/Mac
Korrekturen/Kommentare bitte an <gs@statlab.uni-heidelberg.de>.

Dies ist ein kommentierter Auszug aus dem Kursmaterial zur Grundvorlesung Statistik. Eine aktuellere vollstaendige Fassung ist über <http://www.statlab.uni-heidelberg.de/projects/viror/statgdp/> zugänglich.

Idealerweise sollten die Aktionen/Interaktionen im Kurs "selbstverständlich" sein, d.h. Interaktions–modelle benutzen, die als bekannt vorausgesetzt werden können.

In diesem Kurs werden Aktionsfelder (Buttons) und Eingabefelder als bekannt vorausgesetzt. Zeichenfelder und "drag&drop" werden nicht als bekannt vorausgesetzt und hier durch ein Beispiel eingeführt.

Dichten oder Verteilungen in den Beispielen können mit der Maus gezeichnet werden, wenn das entsprechende Feld angeklickt ist.

Beispiel:

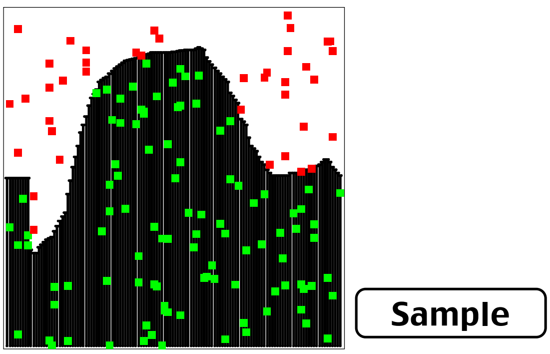


Abb 0.1Beispiel für Zeichenfelder:
Acceptance/Rejection Sampling

Datenlisten, Bilder von Verteilungen und Dichten dienen als Eingabe, wenn Sie auf Ausgabeplots verschoben werden. Benutzen Sie den Kopier-Modus Windows: rechte Maustaste; Mac: Optionstaste beim Verschieben gedrückt halten.

Beispiel: Selektieren Sie einige der Zahlen 1 4 8 12 14 2 19 2 17 33 22 11 18 5 1 12 2 19 2 24 und verschieben Sie diese auf den nachfolgenden Plot:

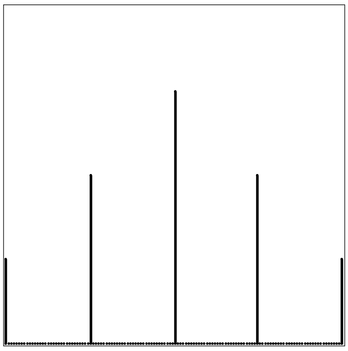


Abb 0.2 Beispiel für Drag&Drop:
empische Verteilung

Test-Kommandos, die noch im Text enthalten sind, habe ein Markierung, die aktiviert werden kann, wie z.B.
❗ "Dialog.Beep"

In "endgültigen" Versionen sollten diese Test-Kommandos entfernt werden.

Alle Komponenten dieses Kurses können im Kurs-Zusammenhang benutzt oder in andere Skripte integriert werden. Während der Skript-Erstellung sind Kommandos bisweilen einfacher zu handhaben als direkte Kopieraktionen.
Nach der Skript-Erstellung sollten sie jedoch entfernt werden - sie geben eine zusätzliche Interaktionsmöglichkeit, die nicht notwendig ist und auf die verzichtet werden sollte.

Für "Autoren" gibt es die Möglichkeit, die Zugriffsmöglichkeiten zu steuern. Ein Skript kann im "Edit Mode" sein. Dann sind Änderungen von Text oder Textkomponenten beliebig möglich. Im "Browser Mode" sind nur die üblichen Browser-Aktionen offen. Der Zustand eines Skripts kann dynamisch über ein Menu gesteuert werden.

Der Autor hat verschiedene Möglichkeiten, ein erstelltes Skript zu bündeln. Wird das Entwicklungssystem nicht mit verteilt, so hat der Empfänger nicht die Möglichkeit, den Zugriffsmodus zu ändern - es sei denn, er/sie beschafft sich ein Entwicklungssystem von anderer Seite.

Theoretische und empirische Wahrscheinlichkeiten

M. Sahm, G. Sawitzki
StatLab Heidelberg
und
VIROR - Virtuelle Universität Oberrhein
<http://www.statlab.uni-heidelberg.de/projects/viror/>

Dies ist ein kommentierter Auszug aus dem Kursmaterial zur Grundvorlesung Statistik. Eine aktuellere vollstaendige Fassung ist über <http://www.statlab.uni-heidelberg.de/projects/viror/statgdp/> zugänglich.

Noch in Vorbereitung. Stand: 10.12.98/04.10.99

Korrekturen/Kommentare bitte an <gs@statlab.uni-heidelberg.de>.

Dieses Kapitel soll die Grundfrage der Statistik vemitteln, die Frage nach dem Verhältnis zwischen Stichprobeninformation und Information über die Grundgesamtheit. Im Hinblick auf die weitere Diskussion wird dabei von Anfang an ein probabilistischer Zugang gewählt.

Theoretische Verteilungen

Auf einer endlichen Menge können wir Wahrscheinlichkeitsmaße betrachten,während wir von mathematischen Schwierigkeiten allgemeiner Maße noch verschont bleiben.

Unsere Menge Ω zählen wir ab. Der Einfachheit halber nehmen wir an $\Omega=\{1, 2,\ldots, N\}$ für ein festes N>0. Die Elementarereignisse, für die wir mit Wahrscheinlichkeiten rechnen wollen, ent– sprechen den Punkten dieser Menge. Wenn keine Unklarheit besteht, identifizieren wir die Punkte mit den einpunktigen Mengen, also i mit $\{i\}$ und schreiben $P(i)$ anstelle von $P(\{i\})$ für die Wahr– scheinlichkeit der einpunktigen Menge $\{i\}$ mit $i \in \{1, 2,\ldots, N\}$. Bisweilen kürzen wir weiter ab und schreiben p_i anstelle von $P(\{i\})$.

Auf einer endlichen Menge ist ein Wahrscheinlichkeitsmaß durch die Wahrscheinlichkeiten der einpunktigen Teilmengen bereits eindeutig bestimmt. Für eine Teilmenge $A \subset \Omega$ ist

$$P(A) = \Sigma_{i \in A} P(i) \qquad \text{für } A \subset \{1, 2,\ldots, N\}.$$

Wir können dann eine Wahrscheinlichkeit angeben, indem wir den Graphen von $i \to P(i)$ angeben.

❗ "StatgdpMouseData.Deposit; StdCmds.PasteView"

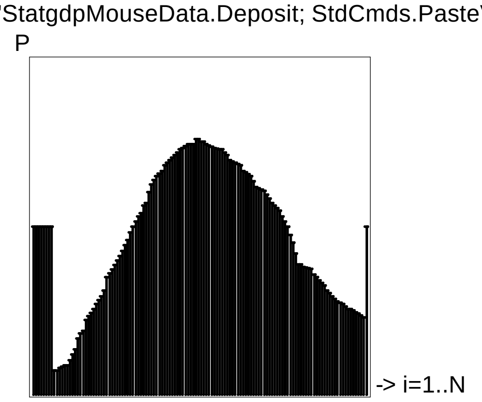


Abb 1.1: Verteilung auf einer diskreten Grundgesamtheit

Die Frage nach dem Skalenniveau wird hier umgangen. Das Modell beschreibt eine endliche Grundgesamtheit mit ordinalen Ausprägungen. Dies ist eine Modellvorstellung, die für "Anfänger" akzeptierbar scheint.

Bei den bisherigen Erfahrungen scheinen ausge–bildete Statistiker damit mehr Schwierigkeiten zu haben. Obwohl die Grundgesamtheit hier mit {1, 2,..., N} angegeben ist, ergeben sich mit ausgebildeten Statistikern regelmässig Dis–kussionen, ob man nicht ein kontinuierliches Modell benutzen sollten. Erfahrungsgemäß treten diese Diskussionen mit Studenten nicht auf.

Der Skalenfaktor der Abszisse kann dabei ignoriert werden: durch die Normierungsbedingung

$$P(\Omega) = 1$$

definiert jede Wahl von $i \to P'(i)$ mit $P'(i) \geq 0$, $P'(\Omega) = \Sigma_{i \in \Omega} P(i) > 0$ eine Wahrscheinlichkeit P, indem wir setzen

$$P(i) = P'(i)/P'(\Omega).$$

Damit wir uns um diese triviale Normierung nicht mehr kümmern müssen, führen wir die Bezeichnung

$$P \propto P'$$

ein, wenn P bis auf Normierung proportional zu P' ist.

Die Standardnormierung $0 \leq P(A) \leq 1$ hat keine besondere mathematische Rechtfertigung und ist erfahrungsgemäß eine Schwierigkeit für Anfänger. Deshalb wird hier diese Normierung deutlich als bloße Konvention eingeführt. Odds oder andere gängige Normierungen sind geeignete Beispiele für Alternativen, die in der Diskussion erwähnt werden sollten.
Die Einführung der in der Bayes-Statistik üblichen $\propto$-Notation vermeidet Sonderbehandlungen der Normierung.

Empirische Verteilungen

In der Statistik wollen wir Aussagen aufgrund von Beobachtungen machen. Wir bleiben dabei nicht bei den einzelnen Beobachtungen, sondern fragen nach dem Maß, nach dem die Beobachtungen verteilt sein könnten. Durch diese Betrachtungsweise können wir auch Situationen modellieren, die so komplex sind, daß uns die Einzelbeobachtung nicht mehr interessiert (z.B. Vielteilchen–systeme, Thermodynamik), oder bei denen wir die interne Dynamik im Einzelfall nicht modellieren können oder wollen, oder die zufällige Anteile haben.

Die Grundgesamtheit Ω beschreibt dabei die möglichen Werte unserer Einzelbeobachtungen. Beobachtungen werden modelliert als Zufallsvariable X mit Werten in Ω. Typischerweise haben wir nicht nur eine Beobachtung X, sondern eine Reihe von Beobachtungen X_k, k=1,...,n, jede mit Werten in Ω. Diese Beobachtungen bilden die Stichprobe $X_1, ..., X_n$, von der wir ausgehen.

Einer Stichprobe können wir wieder ein Wahrscheinlichkeitsmaß $P_{X_1, ..., X_n}$ zuordnen. Wir schreiben kurz P_n für $P_{X_1, ..., X_n}$ und setzen

$P_n(i) \propto n_i$ wenn i in der Stichprobe n_i mal auftritt:

d.h. $n_i = \#\{k : X_k = i\}$

P_n heißt die *empirische Wahrscheinlichkeit* zur Stichprobe $X_1, \ldots, X_n$. Normiert mit der normierenden Konstante $1/\Sigma_{i \in \Omega} n_i = 1/n$ ist

$P_n(i) = n_i/n.$

❗ "StatgdpScatterDens.Deposit; StdCmds.PasteView"

Sample: 1 4 8 12 14 2 19 2 17 33 22 11 18 5 1 12 2 19 2 24

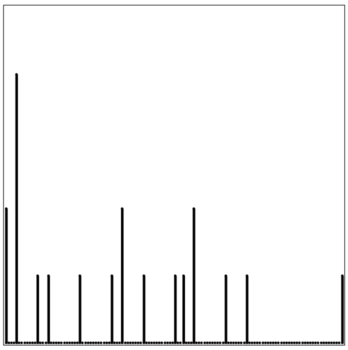


Abb 1.2: empirische Verteilung

X heißt *Zufallsvariable mit der Verteilung P*, wenn P die "Bildverteilung" von X ist, d.h. $Pr(X \in A) = P(A)$ für jedes Ereignis A. Die Beobachtungen $X_1, \ldots, X_n$ heißen *unabhängige Stichprobe mit der Verteilung P*, wenn P^n die "Bildverteilung" von $X_1, \ldots, X_n$ ist, d.h. $Pr(X_1, \ldots, X_n \in A_1 \times \ldots \times A_n) = P(A_1) \bullet \ldots \bullet P(A_n)$ für jedes Ereignis $A_1 \times \ldots \times A_n$. Zufallsvariablen können abhängig voneinander sein, und Beobachtungen mit unabhängigen Zufallsvariablen zu generieren kann beträchtlichen experimentellen und organisatorischen Aufwand bedeuten. Unabhängige Stichproben sind theoretisch einfacher zu behandeln und oft die Grundlage. Wir konzentrieren uns hier auf unabhängige Stichproben.

Statistische Fragestellungen

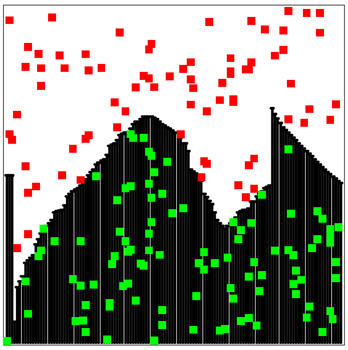
Am Beispiel der empirischen Verteilung können wir statistische Fragestellungen untersuchen. Wir können uns dem von zwei Seiten nähern.

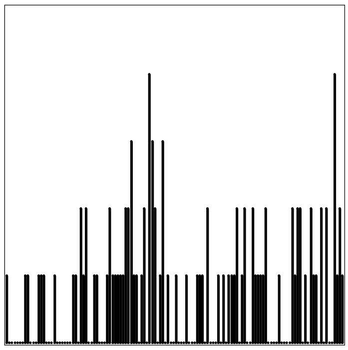
Angenommen, $X_1, \ldots, X_n$ ist eine unabhängige Stichprobe mit hypothetischer Verteilung P. Was wissen wir über die empirische Verteilung von $X_1, \ldots, X_n$?

Oder: Angenommen, $X_1, \ldots, X_n$ ist eine unabhängige Stichprobe mit unbekannter Verteilung P. Was sagt uns die empirische Verteilung von $X_1, \ldots, X_n$ über die Verteilung P?

Dies sind vorbereitende Fragestellungen, die dazu dienen, den Begriff der Fluktuation herauszuschälen.

"StatgdpMouseData.Deposit; StdCmds.PasteView" "StatgdpScatterDens.Deposit; StdCmds.PasteView"





Sample

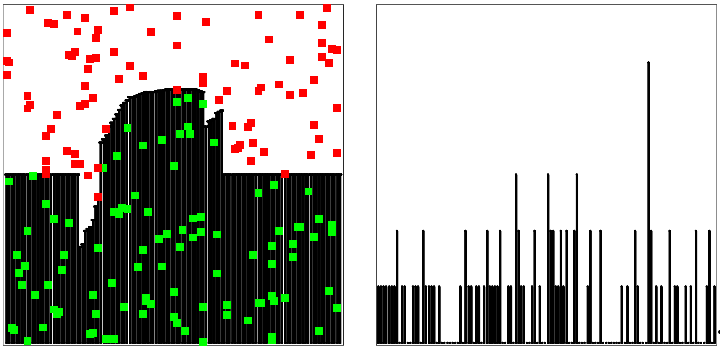
Abb 1.3 Theoretische und empirische Verteilung; Stichproben werden mit Acceptance/Rejectance generiert.

Zur Simulation wird acceptance/rejectance sampling benutzt. Für Mathematik-Studenten schliesst sich eine Übungsaufgabe an: Zeige, daß die durch acceptance/rejectance sampling generierte Stichprobe die Modellverteilung hat.

Die empirische Verteilung hat eine Fluktuation: unterschiedliche Stichproben aus derselben theoretischen Verteilung ergeben unterschiedliche empirische Verteilungen. Während das Verhalten im Einzelfall fluktuiert, kann das Verhalten über verschiedene Stichproben hinweg statistisch untersucht werden.

Das statistische Verhalten der empirischen Verteilung häng von der Mächtigkeit der Grundgesamt–heit und der Gestalt der zugrundeliegenden theoretischen Verteilung ab. Es wird aber auch vom Umfang der Stichprobe beeinflußt. Die theoretische Verteilung ist in der Regel unbekannt und das eigentliche Ziel der statistischen Untersuchung. Der Umfang der Stichprobe und die Regeln, nach denen die Stichprobe erhoben werden, sind jedoch oft unter der Kontrolle des Statistikers. Hier ist eine Wahl zu treffen, die eine möglichst gute Verläßlichkeit der erhobenen Beobachtungen erlaubt.

❶StatgdpController.ExplDistSample;
❶StatgdpController.ExplDistSampleDoSample
❶"StatgdpInstall.Form('StatgdpController.ExplDistSampleCtl');StdCmds.SetMaskMode"

>  <

Sample

PossibleValues 128

SampleSize 100 **Set**

Abb 1.4 Theoretische und empirische Verteilung; Umfang der Grundgesamtheit und Stichprobenumfang können variiert werden.

Statistische Fragestellung:

Bei welchem Stichprobenumfang n werden die groben Kontouren der Wahrscheinlichkeitsverteilung in der empirischen Verteilung erkennbar? Bei welchem Stichprobenumfang n ist die empirische Wahrscheinlichkeit im wesentlichen ähnlich zur theoretischen Verteilung? Verändern Sie Stichprobenumfang n und Umfang der Grundgesamtheit N!

Wie verändert sich die empirische Verteilung bei wachsendem Stichprobenumfang?

Wie unterscheiden sich empirische Verteilungen unterschiedlicher Stichproben aus derselben theoretischen Verteilung? Bei kleinem Stichprobenumfang? Bei großen Stichproben?

Hängt der notwendige Stichprobenumfang von der Gestalt der theoretischen Verteilung ab?

Die erste Fragegruppe (Umfang der Stich–probe/Grundgesamtheit) kann z.B. durch ein Dia–gramm (Stichprobenumfang vs. Umfang der Grundgesamtheit) angegangen werden, in dem "erfolgreiche" Kombinationen markiert werden. Die Grenzkurve zeigt asymptotisch eine √n–Charakteristik.
Die Frage nach der Gestalt der Verteilung macht er–fahrungsgemäß die größten Schwierigkeiten. In der

Regel werden Verteilungen mit prominenten Merkmalen (Moden) als besonders schwierig er–wartet. Im Gegenteil dazu sind flache Verteilungen besonders problematisch - Extremfall: uniforme Verteilung.

Statistische Fragestellungen können unter verschiedenen Rahmenbedingungen auftreten. Einige dieser Rahmenbedingungen haben zu Beispielen geführt, die heute in der theoretischen Analyse Standard sind. Eines dieser Beispiele ist die Entscheidung zwischen einem Paar von Alternativen. Die oben diskutierten Fragen übertragen sich auf diese Beispiel-Situation.

! StatgdpController.ExplDensSample2;
! StatgdpController.ExplDensSampleDoSample2
! "StatgdpInstall.Form('StatgdpController.ExplDensSample2Ctl');StdCmds.SetMaskMode"

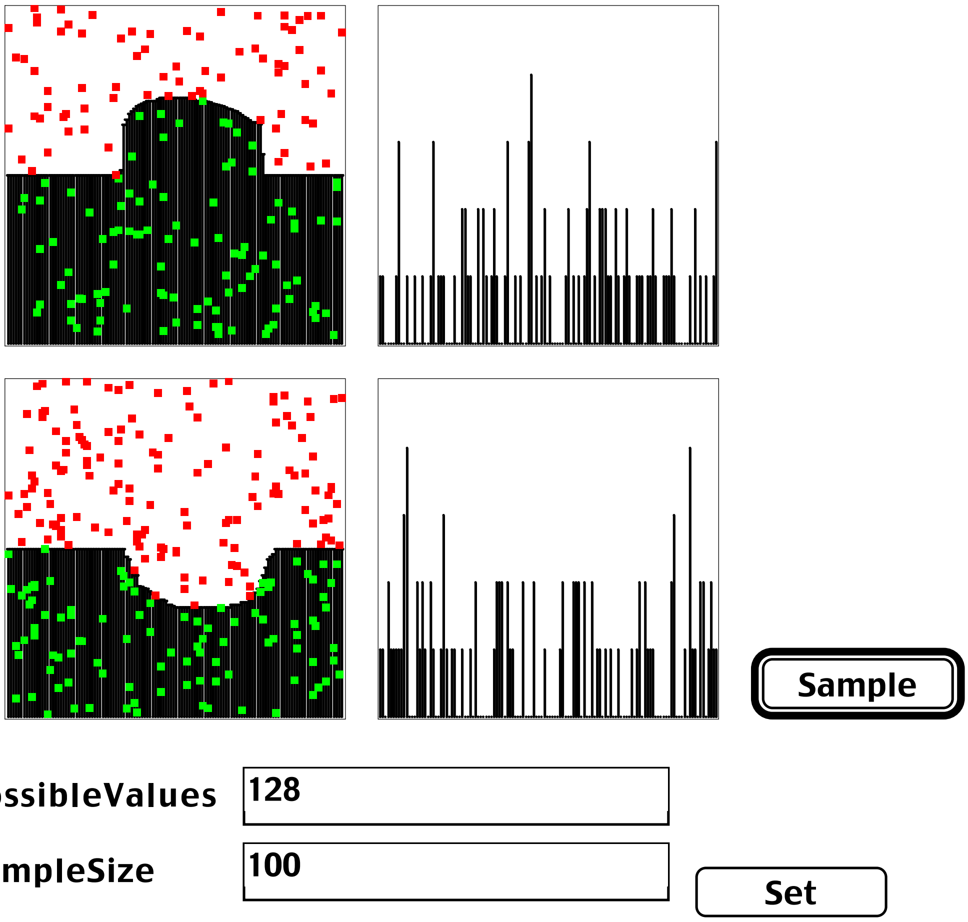


Abb 1.5 Theoretischer und empirischer Vergleich zweier Verteilungen

Statistische Fragestellung:

Bei welchem Stichprobenumfang werden unterschiedliche Verteilung anhand der empirischen Verteilung unterscheidbar?

Vorgehen: analog vorheriger Aufgabe.

Die empirische Wahrscheinlichkeit hat eine Fluktuation, die wir für unabhängige Stichproben mit einfachen Mitteln punktweise abschätzen.
Ist $p_i = P(\{i\})$ und $X_1, \ldots, X_n$ eine unabhängige Stichprobe mit Verteilung P, so ist $n_i = \#\{k: X_k = i\}$ eine (n, P_i)-binomialverteilte Zufallsvariable. Also ist

$$E(n_i) = n \bullet p_i \qquad Var(n_i) = n \bullet p_i \bullet (1 - p_i)$$

Für $p_{n,i} = P_n(\{i\}) = n_i / n$ ist dann

$$E(p_{n,i}) = p_i \qquad Var(p_{n,i}) = p_i \bullet (1 - p_i) / n$$

Im statistischen Mittel stimmt also an jedem Punkt die empirische Wahrscheinlichkeit mit der theoretischen überein. Die Varianz hängt von der Größe der theoretischen Wahrscheinlichkeit p_i ab. Einfach interpretierbar als die Varianz $Var(p_{n,i})$ ist der relative mittlere Fehler

$$\sqrt{E(p_{n,i} - p_i)^2} / p_i$$

Weil $E(p_{n,i}) = p_i$ ist $E(p_{n,i} - p_i)^2 = Var(p_{n,i})$ und der relative mittlere Fehler ist

$$\sqrt{(1 / n \bullet (1 - p_i) / p_i)}.$$

Mit wachsendem Stichprobenumfang nimmt der relative Fehler ab wie $\sqrt{(1/n)}$. Der relative Fehler hängt vom theoretischen Wert p_i ab und verhält sich wie $(1 - p_i) / p_i$.

Binomialverteilte Zufallsvariable, Erwartungswert und Varianz werden hier als bekannte Begriffe vorausgesetzt. Hier wird nur eine probabilistische Interpretation benötigt, wie sie z.B. schon in Kombinatorik-Lektionen vermittelt wird. Eine kritische statistische Diskussion dieser Begriffe folgt in einer späteren Lektion.

Programmier-Aufgabe:

Programmieren Sie einen "Taschenrechner", der aus zwei der drei Angaben: untere Grenze für p, Stichprobenumfang, relativer mittlerer Fehler jeweils den dritten Wert ermittelt. Interface-Problem: je nach Ziel hat der Benutzer implizit drei verschiedene Ziele. Wie wird das jeweilige Ziel selektiert?

Ausblick

Der mittlere relative Fehler wird drastisch schlecht, wenn die theoretische Wahrscheinlichkeit p_i klein ist. Er ist gut, wenn p_i groß ist. Das Verhalten für große p_i ist ein Schein-Erfolg: wenn p_i groß ist, so ist oft die praktische Frage, wie nahe p_i am Maximalwert 1 ist. Das relevante Er–folgskriterium ist dann nicht der relative mittlere Fehler von p_i, sondern von $(1 - p_i)$. Dieser ist $\sqrt{(1 / n \bullet (p_i) / (1 - p_i))}$ - symmetrisch zum Fall kleiner p_i.

Der mittlere relative Fehler wird ein katastrophales Problem, wenn wir an einen Limes-Über–gang $N \to \infty$ oder an einen kontinuierlichen Limes denken. Dann ist der typische Fall $P(\{i\}) \to 0$ für alle Einzelpunkte i, und der relative Fehler in jedem typischen Punkt explodiert. Hier müssen zu–sätzliche Überlegungen einsetzen.

Eine wesentliche Möglichkeit dazu bietet die folgende Überlegung: Die Rechnung gilt nicht nur für Punktmengen, sondern allgemein. Ist A eine Menge mit Wahrscheinlichkeitsmasse $P(A)$, so ist der mittlere relative Fehler der empirischen Verteilung $P_n(A)$

$$\sqrt{(1 / n \bullet (1 - P(A)) / P(A))}$$

Wir können also den mittleren relativen Fehler reduzieren, wenn wir die Genauigkeit unserer Beobachtungsauflösung reduzieren und uns anstelle auf Punktereignisse auf Teilmengen A mit hinreichend großer Wahrscheinlichkeitsmasse P(A) konzentrieren.

Diese Rechnungen gelten jedoch nur punktweise - egal ob der "Punkt" eine einpunktige oder eine komplexere Menge ist.. Die Fluktuation der empirischen Verteiung an verschiedenen Punkten ist stochastisch voneinander abhängig. Die gemeinsame Verteilung kann aus der Multinomialverteilung abgeleitet werden.

Hier wird die Verbindung zum nächsten Abschnitt (Verteilungsfunktion) geschaffen. Als spezielle Mengen werden dort halboffene Intervalle betrachtet.

Verteilungsfunktionen

M. Sahm, G. Sawitzki
StatLab Heidelberg
und
VIROR - Virtuelle Universität Oberrhein
<http://www.statlab.uni-heidelberg.de/projects/viror/>

Noch in Vorbereitung. Stand: 10.12.98/04.10.99
Korrekturen/Kommentare bitte an <gs@statlab.uni-heidelberg.de>.
Dies ist ein kommentierter Auszug aus dem Kursmaterial zur Grundvorlesung Statistik. Eine aktuellere vollstaendige Fassung ist über <http://www.statlab.uni-heidelberg.de/projects/viror/statgdp/> zugänglich.

Dieses Kapitel ist analog dem Kapitel über theoretische und empirische Verteilungen aufgebaut. Es benutzt implizit die Annahe, daß die betrachtete Zufallsvariable ordinal skaliert ist.

Testmengen

Auf einer endlichen Menge $\Omega=\{1, 2,\dots, N\}$ ist ein Wahrscheinlichkeitsmaß durch die Wahrschein–lichkeiten $P_i = P(\{i\})$ der einpunktigen Mengen bereits eindeutig bestimmt. Für eine Stichprobe $X_1, \dots,X_n$ können wir entsprechend die empirische Wahrscheinlichkeit $P_n = P_{X_1, \dots,X_n}$ bestimmen

$P_n(i)= n_i / n$ wenn i in der Stichprobe n_i mal auftritt:
d.h. $n_i = \#\{k:X_k=i\}$

Die empirische Wahrscheinlichkeit hat eine Fluktuation, die wir für unabhängige Stichproben punktweise abschätzen können. Die Fluktuation wird kritisch, wenn die Punktmassen klein werden. Der mittlere relative Fehler wird ein katastrophales Problem, wenn wir an eine Limes-Übergang $N\to\infty$ oder an einen kontinuierlichen Limes denken. Wir können den mittleren relativen Fehler reduzieren, wenn wir die Genauigkeit unserer Beobachtungsauflösung reduzieren und uns anstelle auf Punktereignisse auf Teilmengen A mit hinreichend großer Wahrscheinlichkeitsmasse P(A) konzentrieren.

Das System $\mathcal{A}$ der "Testmengen" $A\in\mathcal{A}$ muß hinreichend reichhaltig sein, so daß die Wahrscheinlichkeit P durch die Werte auf $A\in\mathcal{A}$ definiert ist. Andererseits sollten die Testmengen so ausgewählt werden, daß P(A) durch empirische Versionen $P_n(A)$ hinreichend gut geschätzt werden kann.

Der Einfachheit halber nehmen wir jetzt an, dass unsere Beobachtungen als Zahlen aufgefasst werden könnnen, d.h. wir unterstellen, dass die Codierung unserer Grundmenge Ω als Zahlen 1..N angemessen ist. Ein konventionelles System von Testmengen in dieser Situation sind die Mengen der Form $A=\{X\le a\}$. Dies können wir für endliche Mengen der Form $\Omega=\{1, 2,\dots, N\}$ anwenden. Diese Testmengen sind aber auch oft für andere Grundmengen geeignet, z. B. für die unendliche Grundgesamtheit $\Omega=\{1, 2,\dots\}$ oder für reellwertige Verteilungen.

Verteilungsfunktion

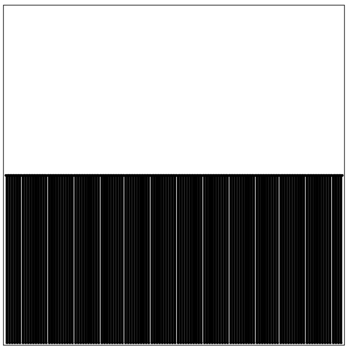
Für eine Wahrscheinlichkeitsverteilung P auf einer Menge $\Omega\subset\mathbf{R}$ heißt die Funktion F mit

$$F(x) = P(X \leq x)$$

die *Verteilungsfunktion* von P.

Für endliche Grundmengen $\Omega \subset \mathbf{R}$ - insbesondere für Mengen der Form $\Omega = \{1, 2, \ldots, N\}$ - ist die Wahrscheinlichkeitsverteilung durch die Verteilungsfunktion eindeutig bestimmt. Für $\Omega = \{1, 2, \ldots, N\}$ ist $P(\{i\}) = F(i) - F(i-1)$. In dieser Situation können wir Wahrscheinlichkeitsverteilung und Verteilungsfunktion direkt ineinander übersetzen.

"StatgdpMouseData.Deposit; StdCmds.PasteView" "StatgdpMouseDistr.Deposit; StdCmds.PasteView"



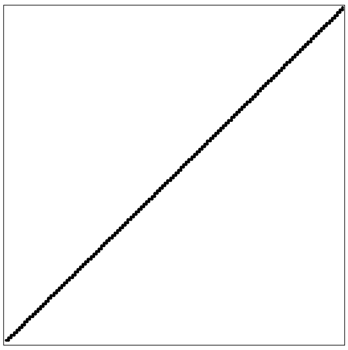


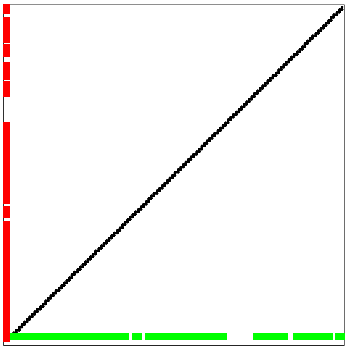
Abb 2.1 Theoretische Verteilung und Verteilungs–funktion. Die Plots sind Zeichenflächen, die inter–aktiv verändert werden können; beide Plots sind dynamisch miteinander verbunden.

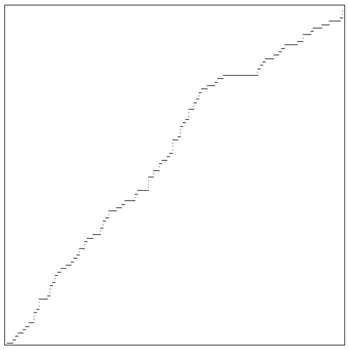
Wie bei der Wahrscheinlichkeitsverteilung haben wir auch eine empirische Version der Verteilungsfunktion. Für die Stichprobe $X_1, \ldots, X_n$ ist

$$F_n(i) = \#\{k : X_k \leq i\} / n$$

Für endliche Mengen $\Omega \subset \mathbf{R}$ ist F_n die Verteilungsfunktion der empirischen Verteilung P_n. Die Definition ist aber auch brauchbar für beliebige Mengen $\Omega \subset \mathbf{R}$, insbesondere auch für reelle Zufallsvariable, d.h. $\Omega = \mathbf{R}$.

"StatgdpMouseDistr.Deposit; StdCmds.PasteView" "StatgdpDistribution.Deposit; StdCmds.PasteView"





Sample

PossibleValues 128

SampleSize 100 **Set**

Abb 2.2 Theoretische und empirische Ver–teilungsfunktion, mit kontrollierbarem Unfang der Grundgesamtheit und der Stichprobe.

Zur Simulation wird inversion sampling benutzt. Für Mathematik-Studenten schliesst sich ein Übungsaufgabe an: Zeige, daß die durch inversion sampling generierte Stichprobe die Modellverteilung hat.

Statistische Fragestellung:

Wie verändert sich die empirische Verteilungsfunktion bei wachsendem Stichprobenumfang?

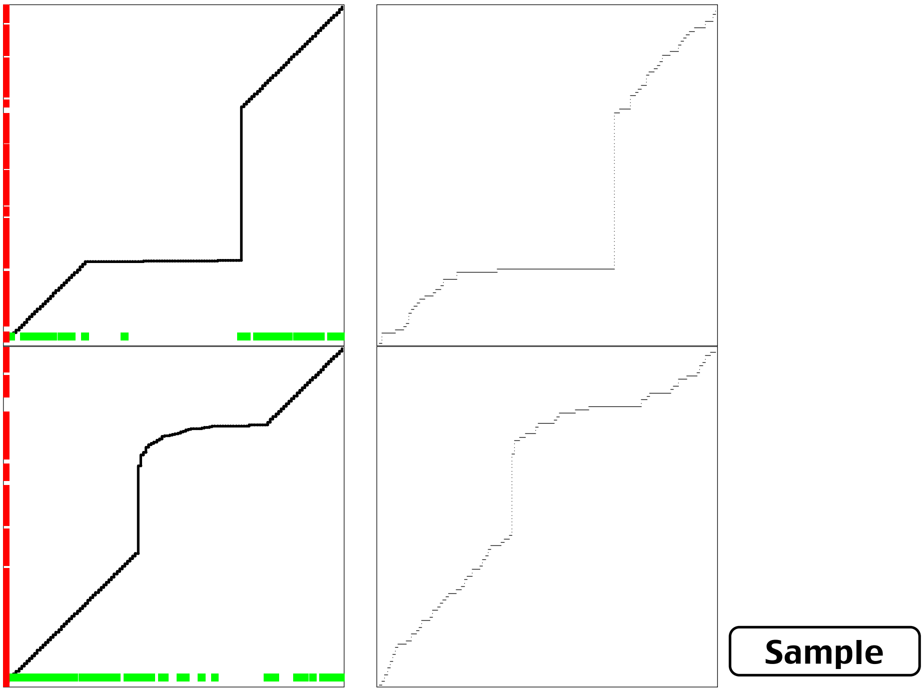
Wie unterscheiden sich empirische Verteilungsfunktionen unterschiedlicher Stichproben aus derselben theoretischen Verteilungsfunktion? Bei kleinem Stichprobenumfang? Bei großen Stichproben? Hinweis: Untersuchen Sie $\sup_x$ | Fn(x)-F(x) | in Abhängigkeit vom Stichprobenumfang N.

Bei welchem Stichprobenumfang n werden die groben Kontouren der Wahrscheinlichkeitsverteilung in der empirischen Verteilungsfunktion erkennbar? Bei welchem Stichprobenumfang n ist die empirische Wahrscheinlichkeit im wesentlichen ähnlich zur theoretischen Verteilungsfunktion? Verändern Sie Stichprobenumfang n und Umfang der Grundgesamtheit N!

Hängt der notwendige Stichprobenumfang von der Gestalt der theoretischen Verteilungsfunktion ab?

StatgdpController.ExplDistrSample2;
StatgdpController.ExplDistrSampleDoSample2

"StatgdpMouseDistr.Deposit; StdCmds.PasteView" "StatgdpDistribution.Deposit; StdCmds.PasteView"



PossibleValues 128

SampleSize 100 **Set**

Abb 2.3: Theoretischer und empirischer Vergleich zweier Verteilungsfunktionen

Statistische Fragestellung:

Bei welchem Stichprobenumfang werden unterschiedliche Verteilungsfunktionen anhand der empirischen Verteilungsfunktion unterscheidbar?

Verbesserte Darstellung: P-P-Plot

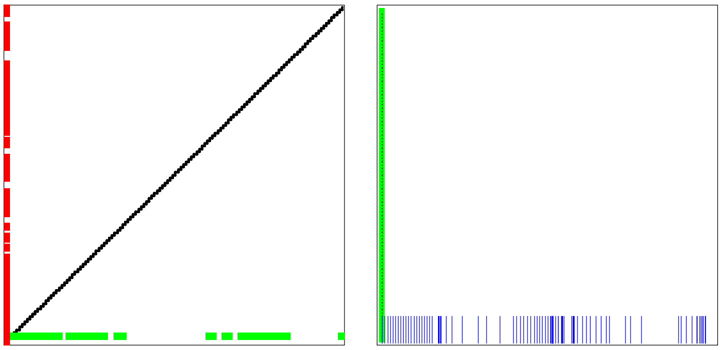
Der erste Eindruck von Verteilungsfunktionen wird durch ihre allgemeinen Eigenschaften bestimmt:

$$\lim_{x \to -\infty} F(x) = 0$$
$$\lim_{x \to +\infty} F(x) = 1$$
$$x<y \Rightarrow F(x) \leq F(y)$$

Um Verteilungsfunktionen F', F graphisch besser miteinander vergleichen zu können, betrachten wir den P-P-Plot, das ist der Graph von $x \to (F'(x), F(x))$. Dies ist eine diagonale Gerade, wenn die beiden Verteilungsfunktionen übereinstimmen.

Für unsere Zwecke, um die empirische Verteilungsfunktion mit der (einer) theoretischen zu vergleichen, betrachten wir also $(F_n(x), F(x))$.

"StatgdpPPlot.Deposit; StdCmds.PasteView" "StatgdpMouseDistr.Deposit; StdCmds.PasteView"



Sample

Abb 2.4 PP-Plot

Die Wahl der Mengen $\{X \leq x\}$ als Testmengen ist nur eine Konvention. Genau so gut hätte man z.B. $\{X<x\}$ wählen können und dann

$$F'_n(i) = \#\{k : X_k < i\}/n$$

als empirische Verteilungsfunktion definiert. Eine ganze Reihe anderer Kandidaten sind gleichermaßen möglich.

Wenn man die empirische Verteilungsfunktion als Versuch ansieht, eine (unbekannte) theoretische Verteilungsfunktion anhand der Daten zu schätzen, kann sich die Auswahl geeigneter Kandidaten nach Annahmen über die Struktur der theoretischen Verteilungsfunktion richten. Durch lokale Glättung können wir z.B. eine stetige Variante gewinnen. Wenden wir dies auf diskrete Variable an und rechnen dann zurück auf die Verteilung, so erhalten wir eine geglättete Variante der empirischen Verteilung.

❗ "StatgdpMouseDistr.Deposit; StdCmds.PasteView"

❗"StatgdpSmoothDistribution.Deposit; StdCmds.PasteView"

❗"StatgdpSmoothDensity.Deposit; StdCmds.PasteView"

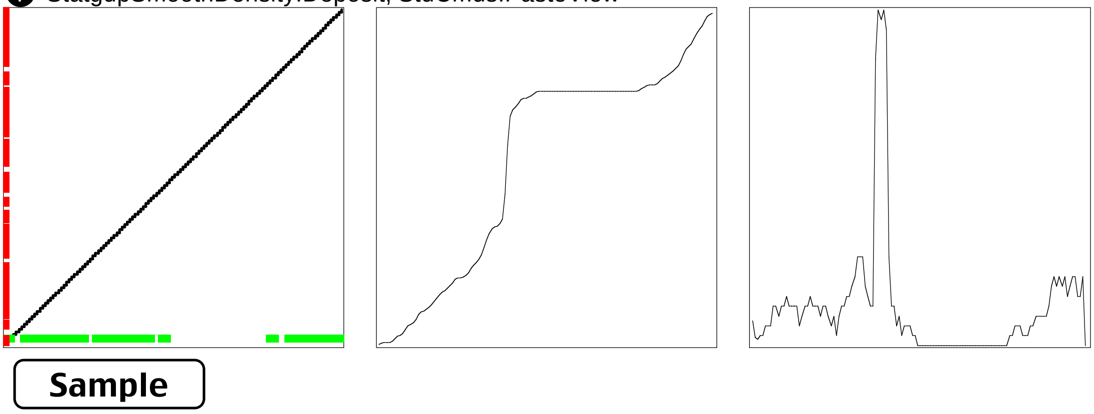


Abb 2.5 Glättung und Dichte

Statistische Fragestellung:

Führen Sie für die geglättete Verteilung die analogen Untersuchungen durch wie für die empirische Verteilung. Unterscheidet sich der Stichprobenumfang, der benötigt wird, um Strukturen zu erkennen? Diskutieren Sie die Detailfragen wie für die empirische Verteilung Schritt für Schritt.

Programmier-Aufgabe:

Entwerfen Sie Glättungsalgorithmen, die die empirische Verteilungsfunktion bzw. die empirische Verteilung anhand benachbarter Werte glätten. Können Sie ein Glättungsverfahren entwerfen, das die optisch wahrgenommenen Zacken im obigen Beispiel vermeidet?

Tests

M. Sahm, G. Sawitzki
StatLab Heidelberg
und
VIROR - Virtuelle Universität Oberrhein
<http://www.statlab.uni-heidelberg.de/projects/viror/>

Dies ist ein kommentierter Auszug aus dem Kursmaterial zur Grundvorlesung Statistik. Eine aktuellere vollstaendige Fassung ist über <http://www.statlab.uni-heidelberg.de/projects/viror/statgdp/> zugänglich.

Noch in Vorbereitung. Stand: 10.12.98/04.10.99
Korrekturen/Kommentare bitte an <gs@statlab.uni-heidelberg.de>.

Hypothesen-Tests

Statistische Tests sind Entscheidungsverfahren. Einfachstes Beispiel sind Verfahren, um zu entscheiden, ob eine Stichprobe $X_1, \ldots, X_n$ dem entspricht, was von unabhängigen Beobachtungen aus einer hypothetischen Verteilung P_0 zu erwarten ist. Die Entscheidung ist: wir akzeptieren $X_1, \ldots, X_n$ als zu P_0 passend, oder wir verwerfen die Hypothese, daß $X_1, \ldots, X_n$ entsprechend der Verteilung P_0 generiert ist. Anstelle einfacher Hypothesen $P=P_0$ können auch zusammengesetzte Hypothesen $P \in \mathbf{P}_0$ mit einer Menge $\mathbf{P}_0$ von Verteilungen betrachtet werden, zu denen die Verteilung P der Stichprobe hypothetisch gehört. Wir konzentrieren uns zunächst auf einfache Hypothesen.

Als Anwender muß man sich mit der konkreten Einzelfallentscheidung befassen. Die Statistik hingegen befaßt sich mit Verfahren. Sie stellt Entscheidungsverfahren bereit und enwickelt Methoden, um diese Verfahren zu beurteilen. Dadurch ermöglicht sie es, unter verschiedenen Entscheidungsverfahren diejenigen zu wählen, von denen zu erwarten ist, daß sie auch in der konkreten Einzelfallentscheidung zu verläßlicheren Schlüssen führen.

Entscheidungsverfahren sollen in Situationen angewandt werden, in denen wir die tatsächlich der Stichprobe zugrunde liegende Verteilung nicht kennen. Die Qualität der Verfahren wird untersucht, indem wir bestimmte Verteilungen unterstellen und unter dieser Annahme das Verfahren studieren. Ein Minimalkriterium für Entscheidungsverfahren ist, daß zumindest mit hinreichend hoher Rate eine Stichprobe mit angenommener Verteilung $P=P_0$ nicht dazu führt, P_0 zu verwerfen. Entscheidungsverfahren haben oft Modifikationsmöglichkeiten, wie Wahl des Stichprobenumfangs, Wahl von Toleranzen etc. Diese Modifikationsmöglichkeiten können benutzt werden, um sicherzustellen, dass dieses Minimalkriterium erfüllt wird.

Tests werden hier als Entscheidungsproblem eingeführt.
Während für Anwender die kritische Interpretation der Testresultate im Vordergrund stehen wird, ist es Aufgabe des Mathematikers/Statistikers, die Tests als Methode zu bewerten. Eine globale Bewertung setzt sehr viel statistisches Wissen (etwa um Optimalitätsschranken und Verteilungsklassen) voraus und kann erst am Ende eines Kurses stehen. Eine Bewertung im Sinne eines Vergleichs von Verfahren ist jedoch bereits zu diesem frühen Stadium des Kurses möglich. Dadurch gewinnen

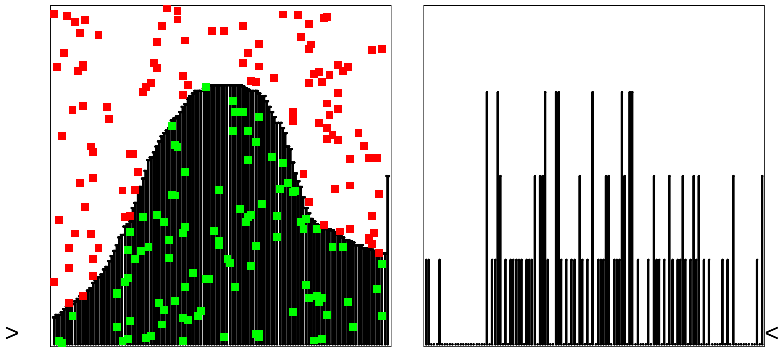
zunächst abstrakte Kriterien wie Niveau und Macht eine pragmatische Bedeutung.

Voraussetzung ist jedoch, dass mindestens zwei potentiell konkurrierende Verfahren hinreichend vertraut sind.

Welche Verfahren dies sind, ist für den Aufbau des Nachfolgenden nicht kritisch. Hier sind informelle Anpassungstests via empirischer Verteilung bzw. empirischer Verteilungsfuntkion als Beispiel gewählt.

Statistische Fragestellungen:

❗"StatgdpTest0Dens.InstallTest0Dens"
❗"StatgdpInstall.Form('StatgdpTest0Dens.Test0Dens');StdCmds.SetMaskMode"



Sample **Ok** **No**

NrOk **0**

NrSamples **0**

Percent **Reset**

PossibleValues **128**

SampleSize **100** **Set**

Abb 3.1 Informeller Anpassungstest als Entscheidungsproblem

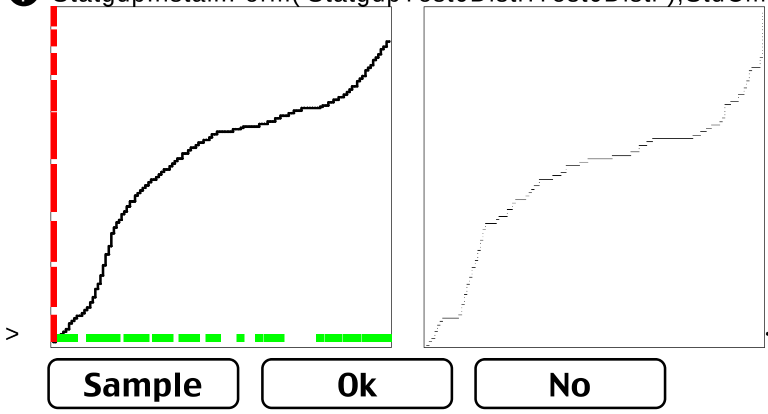
In der Simulation werden Stichproben aus der gegebenen Verteilung gezogen. Wie oft kommt es ihrem Eindruck nach vor, dass die empirische Verteilung nicht zur gegebenen Verteilung passt? Wie oft passt - ihrem Eindruck nach - die empirische Verteilungsfunktion nicht zur gegebenen Verteilung?

Verändern Sie den Stichprobenumfang so, daß in 75% aller Fälle ihrem Eindruck nach die empirischen Resultate zur gegebenen Verteilung passen. Welcher Stichprobenumfang ist im Vergleich dazu nötig, in 90% aller Fälle empirischen Resultate zu bekommen, die zur gegebenen Verteilung passen?

Statistische Fragestellungen:

❶"StatgdpTest0Distr.InstallTest0Distr"

❶"StatgdpInstall.Form('StatgdpTest0Distr.Test0Distr');StdCmds.SetMaskMode"



NrOk **0**

NrSamples **0**

Percent **Reset**

PossibleValues **128**

SampleSize **100** **Set**

Abb 3.2 Informeller Anpassungstest - hier mit Verteilungsfunktion

Das Minimalkriterium, eine zutreffende Hypothese nicht vorschnell zu verwerfen, können wir formal fassen. Wir benutzen wieder eine statistische Modellierung. Die Wahrscheinlichkeit, bei Vorliegen einer Stichprobe $X_1, \ldots, X_n$ die Hypothese zu verwerfen sei $\Phi(X_1, \ldots, X_n)$. Wir treffen diese Entscheidung nur aufgrund der Stichprobe $X_1, \ldots, X_n$, ohne die wahre, der Stichprobe zu–grunde liegende Verteilung zu kennen. Φ kann also als Funktion der Stichprobe modelliert werden, ohne Bezug auf die "wahre" Verteilung. Die "wahre" Verteilung P bestimmt das Mass, nachdem die Stichproben verteilt sind. Unter diesem Mass erwarten wir, daß im Mittel mit dem Anteil

$$E_P\, \Phi(X_1, \ldots, X_n)$$

gegen die Hypothese entschieden wird.

Für die mathematische Analyse sind die internen Details des Entscheidungsverfahrens zunächst unwichtig, und Φ wird mit dem entsprechenden Test identifiziert. Ein Test Φ hält *das Niveau* α für eine Hypothese $\mathbf{P}_0$ ein, wenn

$$E_P\, \Phi(X_1, \ldots, X_n) \leq \alpha \qquad \text{für alle Verteilungen } P \in \mathbf{P}_0$$

Diese Formalisierung kann nicht nur benutzt werden, um das Niveau (als Minimalkriterium) zu definieren, sondern auch um darüber hinaus die Güte von Tests zu vergleichen: wenn eine Ver–teilung P aus der Hypothese vorliegt, sollte die erwartete Verwerfungsrate $E_P\, \Phi(X_1, \ldots, X_n)$ möglichst niedrig sein. Wenn allerdings eine Verteilung P vorliegt, die nicht zur Hypothese gehört, sollte die Hypothese möglichst verworfen werden, also $E_P \Phi(X_1, ..., X_n)$ möglichst gross sein.

Die Funktion

$$P \to E_P \Phi(X_1, ..., X_n)$$

heißt Operationscharakteristik des Tests Φ. Die Einschränkung

$$P \to E_P \Phi(X_1, ..., X_n) \qquad \text{für Verteilungen } P \notin \mathbf{P}_0$$

heißt Gütefunktion des Tests Φ.

Im allereinfachsten Fall besteht sowohl unser hypothetisches Modell $\mathbf{P}_0$ nur aus einer Verteilung, $\mathbf{P}_0 = \{P_0\}$, als auch die zur Diskussion stehende Alternative, d.h. $P \in \{P_0, P_1\}$ für ein Alternativmodell P_1. Das Niveau ist dann $E_{P_0} \Phi(X_1, ..., X_n)$ und die Gütefunktion ist die Güte an der einpunktigen Stelle $E_{P_1} \Phi(X_1, ..., X_n)$. Wenn wir zwei Tests vergleichen wollen, müssen wir sowohl Güte als auch Niveau miteinander vergleichen.

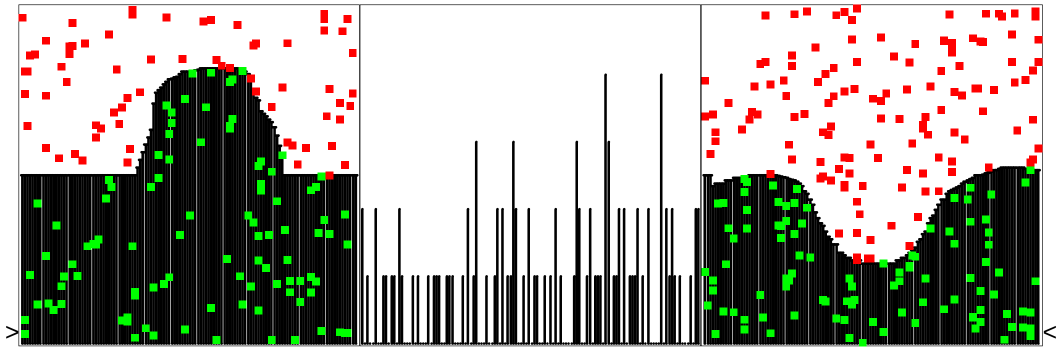
Diese Situation ist ein klassisches Dilemma. Ein konservatives, der Hypothese P_0 verhaftetes Verfahren wird die Hypothese P_0 nicht vorschnell verwerfen, aber es läuft Gefahr, die Hypothese P_0 auch dann zu akzeptieren, wenn die Stichprobe in Wirklichkeit aus P_1 generiert ist. Ein skeptisches Verfahren akzeptiert nur Stichproben, die deutlich zu P_0 passen. Es läuft dann aber Gefahr, P_0 voreilig zu verwerfen.

Es gibt keinen Ausweg aus diesem Dilemma. Selbst um ein einfaches Entscheidungsverfahren zu beurteilen, brauchen wir zwei Informationen: Wie ist die Qualität des Verfahrens, wenn die Stichprobe tatsächlich aus der hypothetischen Verteilung P_0 stammt? Wie ist die Qualität des Verfahrens, wenn die Stichprobe tatsächlich aus der Alternativ-Verteilung P_1 stammt?
In dieser Simulation können Sie zwei Verteilungen spezifizieren. Jede der Zufallsstichproben entstammt einer der beiden Verteilungen und Sie müssen sich für eine der beiden Verteilungen entscheiden.

Statische Fragestellung:

"StatgdpTest1Dens.InstallTest1Dens"

"StatgdpInstall.Form('StatgdpTest1Dens.Test1Dens');StdCmds.SetMaskMode"



H0 **Sample** **H1**

FalseReject	**0**	**GoodReject**	**0**
NrH0	**0**	**NrH1**	**0**
Alpha		**Power**	

Reset

Abb 3.3: Anpassungstest: einfache Hypothese gegen einfache Alternative, via empirischer Verteilung

Statistische Fragestellungen:

"StatgdpTest1Distr.InstallTest1Distr"
"StatgdpInstall.Form('StatgdpTest1Distr.Test1Distr');StdCmds.SetMaskMode"

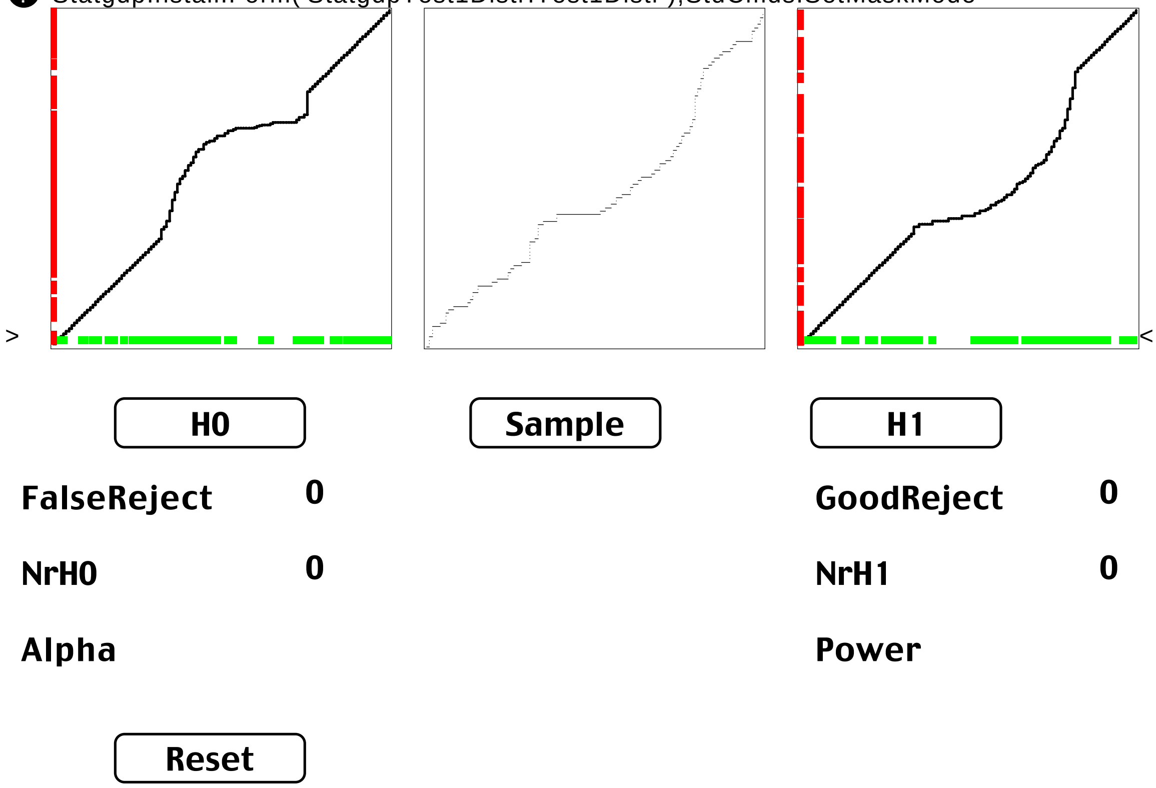


Abb 3.4 Anpassungstest: einfache Hypothese gegen einfache Alternative, via empirischer Verteilungsfunktion

Fehler erster und zweiter Art; unzulässige Verfahren

In der vereinfachten Situation, in der Hypothese und Alternative jeweils nur aus einer Verteilung bestehen, können Testverfahren einfach miteinander verglichen werden. In der Anwendung jedoch bestehen Hypothese und Alternative jedoch oft aus einer Familie von Verteilungen, und ein typischer Fall ist, daß zur Alternative alles gehört, was nicht zur Hypothese passt: $\mathbf{P}_1=\mathbf{P}_0^C$. Das bringt jedoch Schwierigkeiten: Im einfachsten Fall etwa ist $\mathbf{P}_0$ eine einfache Hypothese, besteht also aus nur einer Verteilung. Zur Alternative $\mathbf{P}_0^C$ gehören alle anderen Verteilungen, und diese können beliebig "nahe" an $\mathbf{P}_0$ liegen. Der Kern dieser Schwierigkeit ist, daß $\mathbf{P}_0$ dann eine abgeschlossene Menge ist, $\mathbf{P}_1=\mathbf{P}_0^C$ aber eine offene, die im Rand die Hypothese $\mathbf{P}_0$ trifft. Hypothese und Alternative sind nicht trennbar.

Diese Schwierigkeit ist keine Ausnahme, sondern ein typischer Fall. Sie führt dazu, dass der Fehler erster Art,d.h. $\mathbf{P}_0$ zu verwerfen, obwohl eine Verteilung aus $\mathbf{P}_0$ vorliegt, anders behandelt wird als der Fehler zweiter Art, d.h. $\mathbf{P}_0$ nicht zu verwerfen, obwohl eine Verteilung aus $\mathbf{P}_1$ vorliegt.

Für den Fehler erster Art wird in der Regel eine Schranke vorgegeben, d.h. ein Niveau, daß einzuhalten ist. Zur Konkurrenz stehen dann alle Verfahren, die dieses Niveau einhalten. Diese Verfahren werden dann nach ihrer Güte beurteilt, d.h. nach der Wahrscheinlichkeit, mit der sie zur Verwerfung der Hypothese führen, wenn eine Verteilung aus der Alternative vorliegt.

Ein Verfahren ist bei vorgegebenem Niveau *unzulässig*, wenn es ein konkurrierendes Verfahren gibt, dass das Niveau einhält, bei allen Verteilungen aus der Alternative mindestens gleich gut ist, und zumindest bei einer Verteilung aus der Alternative besser ist.

Im Falle einfacher Alternative reduziert sich dieses Kriterium. Ein Test ist unzulässig zu gegebenem Niveau, wenn es ein konkurriendes Verfahren gibt, das das Niveau einhält, aber auf der Alternative mit größerer Wahrscheinlichkeit verwirft.