

Prof. Dr. Antony Unwin, Dr. Ali Ünlü
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>

Stochastik II — Statistik I (Statistik und Datenanalyse, SS 2009)

Übungsblatt 3

Abgabe: Dienstag 12. Mai 2009, bis spätestens 12.00 Uhr; Briefkasten: Stochastik II

1. Seien X_1, X_2, \dots, X_n unabhängige, identisch verteilte Zufallsvariablen, deren Verteilung den Erwartungswert μ und die Varianz σ^2 besitzt.

Bearbeiten Sie die nachfolgenden Teilaufgaben, indem Sie hierbei jeden Schritt angeben und begründen.

- (a) Zeigen Sie, dass das arithmetische Mittel $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ ein erwartungstreuer Schätzer für μ ist. Berechnen Sie den mittleren quadratischen Fehler dieses Schätzers.
 - (b) Zeigen Sie, dass die Stichprobenvarianz $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ein erwartungstreuer Schätzer für σ^2 ist. Berechnen Sie den Bias des Schätzers $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ für σ^2 .
 - (c) Ist μ bekannt, so ist $\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ ein erwartungstreuer Schätzer für σ^2 .
2. Betrachtet wird der Datensatz `michelson` im Paket `MASS` in R. Der Datensatz beschreibt Messungen, die in dem 1879 Experiment von Michelson zur Messung der Lichtgeschwindigkeit erhoben wurden. Im Datensatz gibt es fünf Experimente mit jeweils 20 Durchführungen, mit gemessenen Lichtgeschwindigkeiten, in km/sec, minus 299,000. Der aktuell akzeptierte Wert auf dieser Skala beträgt 734.5.

Michelson's measurements of the velocity of light in air. In 1879, A. A. Michelson made 100 determinations of the velocity of light in air using a modification of a method proposed by the French physicist Foucault. The data are given here as reported by Stigler. The measurements are derived from sets of often widely disparate numbers of observations. The numbers are in km/sec, and have had 299,000 subtracted from them. The currently accepted 'true' velocity of light in vacuum is 299,792.5 km/sec. Stigler has applied the corrections used by Michelson and reports that the 'true' value appropriate for comparison to these measurements is 734.5.

Source: Stigler, S. M. (1977). Do robust estimators work with real data?. *The Annals of Statistics* 5:4, 1075.

- (a) Vergleichen Sie die fünf Experimente mittels Boxplots.
- (b) Extrahieren Sie aus dem Datensatz die Lichtgeschwindigkeitsmessungen für das erste Experiment und berechnen Sie den Mittelwert, den Median, das obere und untere Quartil, den Interquartilsabstand, die Spannweite und die empirische Streuung. Welche einzelne Kenngröße würden Sie zur Beschreibung dieses Teildatensatzes empfehlen? Warum? Wie würde Ihre Wahl ausfallen, wenn Sie zwei Kenngrößen benutzen dürfen?
- (c) Untersuchen Sie die Messungen aus dem ersten Experiment graphisch auf Ausreißer. Färben Sie gegebenenfalls solche ein.

3. Eine Online-Umfrage zum Thema Abtreibungspille 'Mifegyne' ergab, dass von 4,993 Befragten 1,214 die Einführung dieser Pille gut fanden, wohingegen 3,779 dies ablehnten (Quelle: <http://www.pro-leben.de/abtr/umfragen.php>).

(a) Diskutieren Sie, ob obiger Sachverhalt wie folgt formal gefasst werden darf: Eine Zufallsstichprobe x_1, x_2, \dots, x_n —unabhängig, identisch verteilt—stammt aus einer Bernoulli-Verteilung ($x \in \{0, 1\}$):

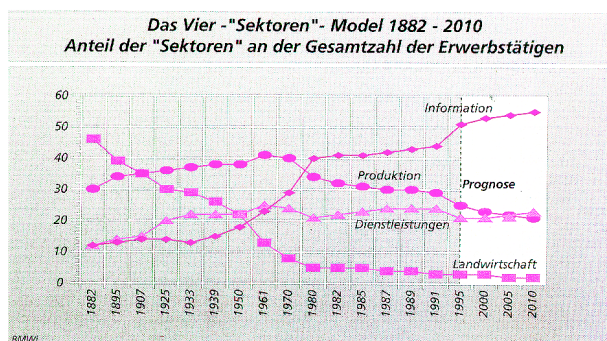
$$f(x; p) = p^x q^{(1-x)} \quad p + q = 1, p, q > 0.$$

Spezifizieren Sie hierbei explizit die Interpretationen aller mathematischen Größen des obigen Modells im Anwendungskontext der Abtreibungsumfrage.

- (b) Welchen Schätzer \hat{p} würden Sie intuitiv für p vorschlagen?
- (c) Leiten Sie den Maximum-Likelihood-Schätzer für p im offenen Intervall $(0, 1)$ ab. Für welche Umfragesituationen würde dieser Schätzer im Intervall $(0, 1)$ nicht existieren?
- (d) Plotten Sie in \mathbb{R} die entsprechende Likelihood-Funktion und machen Sie sich ein Bild über den Maximum-Likelihood-Schätzer.

4. Kommentieren Sie die nebenstehende Graphik.

- Was würden Sie für die Kernaussage der Graphik halten?
- Was genau wird dargestellt?
- Beurteilen Sie die Graphik in Bezug auf Übersichtlichkeit, Ästhetik, Suggestivität.
- Welchen Stellenwert nimmt die Prognose im Vergleich zu den Vergangenheitsdaten in der Graphik ein?
- Wie beurteilen Sie die 'Güte' der Prognose?
- Welche (statistischen) Eigenschaften sollte eine gute Prognose haben?



Entnommen aus Info 2000, Bericht der Bundesregierung, herausgegeben vom Bundesministerium für Wirtschaft, 1996.

5. Simulieren Sie in \mathbb{R} jeweils 100, 1,000 und 10,000 Realisationen einer Poisson-Verteilung mit dem Parameter $\lambda = 2$. Bestimmen Sie in jedem dieser drei Fälle den Maximum-Likelihood-Schätzer für den bekannten Populationsparameter λ und überprüfen Sie das asymptotische Verhalten dieses Schätzers bezüglich der Eigenschaften Bias, Varianz, mittlerer quadratischer Fehler und Konsistenz. Wiederholen Sie die obigen drei Simulationsfälle 100 mal, um in geeigneter Weise die Eigenschaft der Konsistenz für diesen Schätzer zu überprüfen.

6. Einzeiler-Aufgaben.

- (a) Sei $X \sim B(10, 0.25)$. Berechnen Sie die Wahrscheinlichkeiten $P(X \geq 8)$ und $P(X = 5)$.
- (b) Wie kann man Teilaufgabe 6 (b) auf Übungsblatt 2 einfacher in \mathbb{R} ausführen? Welche zentrale Eigenschaft der Sprache \mathbb{R} kommt hierbei zum Tragen?