



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

Stochastik für Lehramt

Übungsblatt 8

Abgabe: Montag 21. Juni 2010, bis spätestens 10.00 Uhr; Briefkasten: Stochastik für Lehramt

Die Aufgaben können auch in Gruppen bearbeitet und abgegeben werden! Für jede Aufgabe gibt es 5 Punkte. Lösungen in R können in Form eines Skriptes (Textdatei) per email an die jeweiligen Übungsgruppenleiter geschickt werden!

Grundlegendes:

- Was ist ein erwartungstreuer Schätzer? Was ist ein asymptotisch erwartungstreuer Schätzer?
- Was ist $\int_{-\infty}^{\infty} f(x) \cdot x^2 dx$, wenn f die Dichte von X bezeichnet?
- Für welche Werte ϵ und $V[X]$ ist die Tschebychew-Ungleichung sinnvoll?

-
- Eine einzelne Vuvuzela erreicht in einem Meter Entfernung einen Schalldruckpegel von 120 dB. Eine Verdopplung des Abstands resultiert in einer Minderung von 6 dB. Zur Vereinfachung sei angenommen, dass Fußballspieler P. sich zu einem Zeitpunkt t genau n_t gleichzeitig lärmenden Vuvuzelas im Abstand von 128m gegenüber sehe. Der Schalldruckpegel beträgt zu einem Zeitpunkt t insgesamt $W_t = 10 \cdot \log_{10}(n_t) + L$, wobei L der Grundpegel einer Vuvuzela ist.
 - Spieler P. will seinem Teamkollegen S. zum Zeitpunkt t etwas zurufen. In welchem Bereich liegt zu 95% der Lärmpegel, den er überbieten muss, wenn $n_t \sim N(\mu = 1000, \sigma = 10)$?
 - Mit welcher Wahrscheinlichkeit liegt der Lärmpegel über 100 dB?
 - Die Anzahl n_t der gleichzeitig geblasenen Vuvuzelas kann nicht beobachtet werden. Stattdessen wird der Lärmpegel gemessen und ein Durchschnitt von 105 dB sowie eine Standardabweichung von 5dB notiert. Schätzen Sie die Wahrscheinlichkeit ab, dass der Durchschnitt von $m_1 = 100$ bzw. $m_2 = 1000$ Messzeitpunkten \bar{W} um mehr als 5 dB von seinem Erwartungswert abweicht...
 - ...mit der Ungleichung von Tschebychew.
 - ...mit dem zentralen Grenzwertsatz.Welche Abschätzung ist die bessere?
 - Laden Sie den Datensatz `Berkeley` in `Mondrian`. Der Datensatz enthält drei Variablen zu den Studienzulassungen an der Universität Berkeley. Die Variablen `Department` und `Admission` bezeichnen die Fakultät und ob die Person zum Studiengang zugelassen wurde. `Gender` bezeichnet das Geschlecht.
 - Erstellen Sie Barcharts und/oder Spineplots für die drei Variablen und untersuchen Sie deren paarweise Zusammenhänge. Wurden Frauen benachteiligt?
 - Selektieren Sie `Department` und `Gender` und erstellen Sie einen Mosaikplot. Wie ist der Plot aufgebaut? Beschreiben Sie!

(c) Welche Funktionen haben die Pfeiltasten?

3. (Fortsetzung von Aufgabe 2)

- (a) Was lässt sich im Mosaikplot aus Aufgabe 2c) durch CTRL+R erreichen?
Was durch CTRL+SHIFT+R ?
- (b) Legen Sie beide Variablen im Mosaikplot auf die x-Achse in der Reihenfolge *Department* > *Gender*.
- (c) Selektieren Sie die zugelassenen Studenten in einem Barchart und betrachten Sie den Mosaikplot. Was ist zu erkennen?
- (d) Ist das Ergebnis konsistent mit Ihren Schlüssen aus Aufgabe 2a) ? Wenn nicht, wie können Sie sich dies erklären?

4. X_1, \dots, X_n seien unabhängige, identisch verteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 . Der Parameter μ soll geschätzt werden. Hierzu stehen folgende Schätzfunktionen zur Verfügung:

i. $T_1 = \frac{1}{2}(X_1 + X_2)$

ii. $T_2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} X_i + \frac{1}{2}X_n$

iii. $T_3 = \frac{1}{n+1} \sum_{i=1}^n X_i$

iv. $T_4 = \frac{1}{n} \sum_{i=1}^n X_i$

- (a) Bestimmen Sie die Erwartungswerte der vier Schätzfunktionen und gegebenenfalls den Bias.
- (b) Bestimmen Sie die Varianzen der vier Schätzfunktionen.
- (c) Untersuchen Sie die Schätzfunktionen auf Konsistenz.



Abbildung 1: <http://richworks.in/wp-content/uploads/2010/04/243.jpg>

5. Untersuchen Sie die folgende Grafik über das Glück der Menschen:

- Beurteilen Sie die Grafik in Bezug auf Suggestivität, Eindeutigkeit und Ästhetik.
- Was wird dargestellt? Welche Variablen fließen in die Grafik ein? Ist die Darstellung gelungen?
- Was halten Sie von der Darstellung? Was würde ein fröhliches Smiley bedeuten, was ein trauriges?
- Was halten Sie allgemein von der Verwendung von Gesichtern, um Daten darzustellen (z.B. anhand der Größe der Augen, Ohren, usw.)?
- Welche Hauptaussagen würden Sie basierend auf dieser Grafik treffen und wie hätten Sie selbst die Daten dargestellt?

- R
- Was ist der Unterschied zwischen `replicate(100, rpois(10, lambda = 25))` und `replicate(10, rpois(100, lambda = 25))`?
 - Was ist der Unterschied zwischen `rep(rpois(1:3, lambda = 25), 10)` und `rpois(rep(1:3,10), lambda = 25)`?
 - Gilt `mean(median(x)) = median(mean(x))`?