UNIVERSITÄT LEIPZIG

Eine Einführung in R: Dichten und Verteilungsfunktionen

Bernd Klaus, Verena Zuber
Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE),
Universität Leipzig

http://www.uni-leipzig.de/ zuber/teaching/ws11/r-kurs/

10. November 2011

Einschub: Zufallsvariablen

Eine Variable oder Merkmal X, dessen Werte die Ergebnisse eines Zufallsvorganges sind, heißt Zufallsvariable.

Notation:

- X: Die Zufallsvariable
- x: Eine Realisierung oder Beobachtung der Zufallsvariable

Induktive (Schließende) Statistik:

Mittels einer Stichprobe wird versucht Aussagen bezüglich einer Grundgesamtheit zu treffen.

- Grundgesamtheit: Menge aller für die Fragestellung relevanten Objekte
- Stichprobe: Tatsächlich untersuchte Teilmenge der Grundgesamtheit

Die Aussagen beziehen sich auf Merkmale der Grundgesamtheit.

- Merkmal: Die interessierende Größe oder Variable
- Merkmalsausprägung: Der konkret gemessene Wert an einem Objekt der Stichprobe

Das Model: Theoretische Ebene

- Statistische Analysen beruhen auf Modellannahmen.
- Ziel: Formalisierung eines reellen Sachverhaltes
    - Stetige Variablen mit Erwartungswert und Varianz
    - Diskrete Variablen mit Gruppenzugehörigkeiten
- Parametrischer Ansatz: Verteilungsannahmen, wie eine Zufallsvariable X ist normalverteilt mit Erwartungswert μ und Varianz σ^2
- Non-Parametrischer Ansatz: Ohne Verteilungsannahmen

Die beobachteten Daten: Die empirische Ebene

- Erwartungswert und Varianz einer Grundgesamtheit können nicht in der Realität beobachtet werden, sondern müssen aus der Stichprobe geschätzt werden.
- Beobachtet werden n Realisierungen $x_1, ..., x_n$ einer Zufallsstichprobe X.
- Notation:
  - Erwartungswert μ
  - Schätzer für den Erwartungswert $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$
- Gesetz der großen Zahlen: "Je mehr Realisierungen einer Zufallszahl beobachtet werden, desto besser approximiert der Mittelwert den Erwartungswert"
- Realisierungen einer Zufallsvariable folgen nicht exakt einer bestimmten Verteilung. Nur bei großer Stichprobenzahl nährt sich die empirische Dichte der theoretischen an.

Normalverteilung $N(\mu, \sigma)$

Die Normal- oder Gauß-Verteilung ist formalisiert durch Erwartungswert μ und Varianz σ^2:

$$f(x|\mu,\sigma) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

- Diese Funktion ist in R implementiert:
  `dnorm(x, mean=0, sd=1)`
  (Vorsicht: mean steht hier für den Erwartungswert)
- Erzeugen von n Realisierungen $x_1, \ldots, x_n$:
  `rnorm(n, mean=0, sd=1)`

Beispiel: Normalverteilung

- Darstellung: Gesetz der großen Zahlen

```
x10<-matrix(rnorm(100),nrow=10,ncol=10)
x1000<-matrix(rnorm(10000),nrow=10,ncol=1000)
apply(x10,MARGIN=1, mean)
```

-0.392 -0.309 0.195 -0.727 -0.150 0.327 0.142 0.020 0.069 0.594

```
apply(x1000,MARGIN=1, mean)
```

-0.018 -0.011 0.007 -0.011 -0.021 -0.013 0.036 0.026 0.074 0.010

Beispiel: Normalverteilung

- Anpassung der empirischen an die theoretische Verteilung: