

# Statistische Methoden der Datenanalyse

Vorlesung im Sommersemester 2008

H. Kolanoski  
*Humboldt-Universität zu Berlin*



# Inhaltsverzeichnis

<b>Literaturverzeichnis</b>	<b>iii</b>
<b>1 Grundlagen der Statistik</b>	<b>3</b>
1.1 Wahrscheinlichkeit . . . . .	3
1.1.1 Definition über die Häufigkeit . . . . .	3
1.1.2 Kombinatorische Definition . . . . .	3
1.1.3 Axiomatische Definition der Wahrscheinlichkeit . . . . .	4
1.2 Verteilungen von Zufallsvariablen . . . . .	7
1.2.1 Eigenschaften von Verteilungen . . . . .	8
1.2.2 Erwartungswerte . . . . .	11
1.2.3 Wahrscheinlichster Wert und Median . . . . .	12
1.2.4 Stichproben und Schätzwerte . . . . .	13
1.3 Simulation von Verteilungen . . . . .	14
1.3.1 Umkehrung der Verteilungsfunktion . . . . .	14
1.3.2 ‘Hit and Miss’ Methode . . . . .	16
<b>2 Spezielle Verteilungen einer Variablen</b>	<b>19</b>
2.1 Binomial-Verteilung . . . . .	19
2.2 Multinomial-Verteilung . . . . .	24
2.3 Poisson-Verteilung . . . . .	25
2.4 Gleichverteilung . . . . .	27
2.5 Normalverteilung . . . . .	30
2.5.1 Vertrauensintervalle: . . . . .	32
2.6 Zentraler Grenzwertsatz . . . . .	34
<b>3 Verteilungen mehrerer Variablen</b>	<b>37</b>
3.1 Eigenschaften von Verteilungen mehrerer Variablen . . . . .	37
3.1.1 Wahrscheinlichkeitsdichte, Verteilungsfunktion, Randverteilung	37
3.1.2 Bedingte Wahrscheinlichkeitsdichten, Selektionsschnitte . . . . .	38
3.2 Erwartungswerte . . . . .	39
3.3 Kovarianzmatrix . . . . .	40
3.3.1 Definition und Eigenschaften der Kovarianzmatrix . . . . .	40
3.3.2 Beispiel: Multi-dimensionale Gaussverteilung . . . . .	40
3.3.3 Kovarianzen von Stichproben . . . . .	41
3.3.4 Kovarianzmatrix von unabhängigen Variablen . . . . .	41
3.3.5 Korrelationen . . . . .	42
3.4 Lineare Funktionen von mehreren Zufallsvariablen . . . . .	44

3.5	Nicht-lineare Funktionen von Zufallsvariablen . . . . .	46
3.5.1	Eine Funktion von einem Satz von Zufallsvariablen . . . . .	46
3.5.2	Mehrere Funktionen von einem Satz von Zufallszahlen . . . . .	47
3.6	Transformationen von Zufallsvariablen . . . . .	50
<b>4</b>	<b>Stichproben und Schätzungen</b>	<b>53</b>
4.1	Stichproben, Verteilungen und Schätzwerte . . . . .	53
4.2	Eigenschaften von Schätzwerten . . . . .	54
4.3	Stichproben aus Normalverteilungen; $\chi^2$ -Verteilung . . . . .	56
<b>5</b>	<b>Monte-Carlo-Methoden</b>	<b>63</b>
5.1	Zufallszahlengeneratoren . . . . .	63
5.2	Monte-Carlo-Integration . . . . .	63
5.3	Monte-Carlo-Erzeugung von Ereignissen nach Zufallsverteilungen . . . . .	63
<b>6</b>	<b>Die Maximum-Likelihood-Methode</b>	<b>65</b>
6.1	Das Maximum-Likelihood-Prinzip . . . . .	65
6.2	Fehlerbestimmung für ML-Schätzungen . . . . .	68
6.2.1	Allgemeine Methoden der Varianzabschätzung . . . . .	68
6.2.2	Varianzabschätzung durch Entwicklung um das Maximum . . . . .	68
6.2.3	Vertrauensintervalle und Likelihood-Kontouren . . . . .	69
6.3	Eigenschaften von ML-Schätzungen . . . . .	71
<b>7</b>	<b>Methode der kleinsten Quadrate</b>	<b>75</b>
7.1	Prinzip der Methode der kleinsten Quadrate . . . . .	75
7.2	Lineare Anpassung . . . . .	76
7.2.1	Anpassung der Messwerte an eine Gerade . . . . .	76
7.2.2	Anpassung einer allgemeinen linearen Funktion der Parameter . . . . .	79
7.3	Anpassung nicht-linearer Funktionen der Parameter . . . . .	84
<b>8</b>	<b>Entfaltung</b>	<b>87</b>
8.1	Einführung . . . . .	87
8.2	Entfaltung von Histogrammen . . . . .	87
<b>9</b>	<b>Neuronale Netze zur Datenklassifikation</b>	<b>89</b>
9.1	Einleitung . . . . .	89
9.1.1	Attraktivität neuronaler Modelle . . . . .	89
9.2	Natürliche und künstliche neuronale Netze . . . . .	91
9.2.1	Natürliche neuronale Netze . . . . .	91
9.2.2	Künstliche neuronale Netze (KNN) . . . . .	94
9.3	Feed-Forward-Netzwerke . . . . .	96
9.3.1	Das einfache Perzeptron . . . . .	97
9.3.2	Das Mehrlagen-Perzeptron . . . . .	100
9.3.3	Lernen . . . . .	103
9.3.4	Typische Anwendungen für Feed-Forward-Netze . . . . .	109
9.3.5	BP-Lernen und der Bayes-Diskriminator . . . . .	112

# Literaturverzeichnis

- [1] S. Brandt: ‘Datenanalyse’, 4. Auflage, 1999, Spektrum Akademischer Verlag.
- [2] R.J. Barlow, ‘Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences’, Wiley, 1989.
- [3] V. Blobel und E. Lohrmann, ‘Statistische und numerische Methoden der Datenanalyse’, Teubner Studienbücher, 1998.
- [4] G. Bohm und G. Zech, ‘Einführung in Statistik und Messwertanalyse für Physiker’, Hamburg, DESY 2005; e-book:<<http://www-library.desy.de/preparch/books/vstatmp.pdf>>

## Zu “Neuronale Netze”:

- [5] D.E.Rumelhart and J.L.McClelland: ‘*Parallel Distributed Processing*’, MIT Press 1984 (9.Aufl. 1989).
- [6] J.Hertz, A.Krogh and R.G.Palmer: ‘*Introduction to the Theory of Neural Computation*’, Addison-Wesley Publishing Company, 1991.
- [7] R.Brause: ‘*Neuronale Netze*’, Teubner Verlag 1991.
- [8] R.Hecht-Nielsen: ‘*Neurocomputing*’, Addison-Wesley Publishing Company, 1987.
- [9] H.Ritter, T.Martinetz und K.Schulten: ‘*Neuronale Netze. Eine Einführung in die Neuroinformatik selbstorganisierender Netze*’, Addison-Wesley Publishing Company, 1991.
- [10] G.E.Hinton: ‘*Wie Neuronale Netze aus Erfahrung lernen*’, Spektrum der Wissenschaft, Nov. 1992.
- [11] T.Kohonen: ‘*Self-Organization and Associative Memory*’, Springer Verlag, 3.Auflage 1989.
- [12] A.Zell: ‘*Simulation Neuronaler Netze*’, Addison-Wesley, 1.Auflage 1994.
- [13] Scientific American: ‘*The Brain*’, Vol. 241, Sept. 1979.
- [14] Spektrum der Wissenschaft, Nov. 1992.

## “PDG”:

- [15] W.-M. Yao et al. (Particle Data Group), ‘Review of Particle Physics’, J. Phys. G33, 1 (2006); <http://pdg.lbl.gov> (Kapitel 31-32, reviews).

