Grundprinzipien der Bayes-Statistik

Karl-Rudolf Koch

Zusammenfassung:
In drei wesentlichen Punkten unterscheidet sich die Bayes-Statistik von der traditionellen Statistik. Zunächst beruht die Bayes-Statistik auf dem Bayes-Theorem, mit dessen Hilfe unbekannte Parameter zu schätzen, Konfidenzregionen für die Parameter anzugeben und Hypothesen für die Parameter zu prüfen sind. Ferner nimmt die Bayes-Statistik eine Erweiterung des Wahrscheinlichkeitsbegriffs vor, indem die Wahrscheinlichkeit von Aussagen definiert wird, wobei die Wahrscheinlichkeit ein Maß für die Plausibilität der Aussage gibt. Schließlich sind die unbekannten Parameter der Bayes-Statistik Zufallsvariable, was aber nicht bedeutet, daß sie keine Konstanten repräsentieren dürfen. Auf vielfältige Anwendungen der Bayes-Statistik bei geodätischen Problemstellungen wird hingewiesen.

1 Einführung

Im Mittelpunkt der Bayes-Statistik steht das Bayes-Theorem. Mit seiner Hilfe lassen sich unbekannte Parameter schätzen, Konfidenzregionen für die unbekannten Parameter festlegen und die Prüfung von Hypothesen für die Parameter ableiten. Dieser einfache und anschauliche Weg der Herleitung ist der traditionellen Statistik versperrt, da sie sich nicht auf das Bayes-Theorem gründet. Insofern besitzt die Bayes-Statistik einen wesentlichen Vorteil gegenüber der traditionellen Statistik.

Für die Bayes-Statistik wird eine Erweiterung des Begriffs der Wahrscheinlichkeit vorgenommen, indem die Wahrscheinlichkeit für Aussagen definiert wird. Dagegen beschränkt sich die traditionelle Statistik auf die Wahrscheinlichkeit von zufälligen Ereignissen. In der Bayes-Statistik wird also die Wahrscheinlichkeit nicht nur für zufällige Ereignisse, sondern ganz allgemein für Aussagen eingeführt. Die Wahrscheinlichkeit gibt die Plausibilität von Aussagen an, durch sie wird der Zustand des Wissens über eine Aussage ausgedrückt. Für die Wahrscheinlichkeit von Aussagen lassen sich durch logisches und konsistentes Schließen drei Gesetze ableiten, aus denen die gesamte Wahrscheinlichkeitstheorie entwickelt werden kann.

Im Gegensatz zur traditionellen Statistik, bei der die unbekannten Parameter feste Größen repräsentieren, sind die unbekannten Parameter in der Bayes-Statistik Zufallsvariable. Das bedeutet aber nicht, daß die unbekannten Parameter nicht Konstanten repräsentieren dürfen, wie beispielsweise die Koordinaten eines festen Punktes. Mit dem Bayes-Theorem erhalten die unbekannten Parameter Wahrscheinlichkeitsverteilungen, aus denen die Wahrscheinlichkeit, also die Plausibilität von Werten der Parameter folgt. Die Parameter selbst können daher feste Größen darstellen und brauchen nicht aus Zufallsexperimenten zu resultieren.

Im folgenden soll auf die Gesetze der Wahrscheinlichkeit und auf die Überlegungen zur Parameterschätzung, zur Festlegung von Konfidenzbereichen und zum Test von Hypothesen eingegangen werden. Außerdem werden Anwendungen genannt. Nur einige Ergebnisse können präsentiert werden, Details sind bei Koch (1990;1999) nachzulesen.

2 Gesetze der Wahrscheinlichkeit

Die Bayes-Statistik arbeitet ausschließlich mit bedingten Wahrscheinlichkeiten, denn im allgemeinen hängt eine Aussage davon ab, ob eine weitere Aussage wahr ist. Man schreibt $A|B$, um auszudrücken, daß A wahr ist unter der Bedingung, daß B wahr ist. Die Wahrscheinlichkeit von $A|B$, auch bedingte Wahrscheinlichkeit genannt, wird mit

$$P(A|B) \tag{2.1}$$

bezeichnet. Sie gibt ein Maß für die Plausibilität der Aussage $A|B$ an. Bedingte Wahrscheinlichkeiten sind geeignet, empirisches Wissen auszudrücken, denn durch $P(A|B)$ wird die Wahrscheinlichkeit angegeben, mit der vorhandenes Wissen für weiteres Wissen relevant ist.
Wie von Cox (1946) und Jaynes (1995) gezeigt, lassen sich durch logisches und konsistentes Schließen das Produktgesetz der Wahrscheinlichkeit

$$P(AB|C) = P(A|C)P(B|AC) = P(B|C)P(A|BC) \tag{2.2}$$

mit

$$P(S|C) = 1 \tag{2.3}$$

und das Summengesetz

$$P(A|C) + P(\bar{A}|C) = 1 \tag{2.4}$$

ableiten. Hierin bedeuten A, B und C allgemeine Aussagen, S die sichere Aussage und $\bar{A}$ die Negation der Aussage A.
Aus diesen drei Gesetzen lassen sich alle Gesetze der Wahrscheinlichkeitstheorie ableiten, die für die Bayes-Statistik benötigt werden. Berücksichtigt man die Bedingung C im Produktgesetz (2.2) nicht und löst nach $P(A|B)$ auf, erhält man die Definition der bedingten Wahrscheinlichkeit der traditionellen Statistik. Sie wird dort durch die relative Häufigkeit erklärt, was im Gegensatz zu der Ableitung, die auf (2.2) führt, weniger einleuchtend ist.
Löst man das Produktgesetz (2.2) nach $P(A|BC)$ auf, erhält man das Bayes-Theorem

$$P(A|BC) = \frac{P(A|C)P(B|AC)}{P(B|C)} \,. \tag{2.5}$$

Bei den üblichen Anwendungen des Bayes-Theorems bedeutet A die Aussage über ein unbekanntes Phänomen, B die Aussage, die Information über das unbekannte Phänomen enthält, und C eine Aussage über zusätzliches Wissen. Man bezeichnet $P(A|C)$ als Priori-Wahrscheinlichkeit, $P(A|BC)$ als Posteriori-Wahrscheinlichkeit und $P(B|AC)$ als Likelihood. Die Priori-Wahrscheinlichkeit der Aussage über das unbekannte Phänomen wird also durch die Likelihood modifiziert, die Information über das Phänomen enthält, um die Posteriori-Wahrscheinlichkeit zu erhalten.
Im folgenden wird noch das verallgemeinerte Bayes-Theorem angegeben, das für Verteilungen gilt.

3 Verteilungen

Die Aussagen sollen sich im folgenden auf die numerischen Werte von Variablen in Form von reellen Zahlen beziehen. Diese Variablen werden in der traditionellen Statistik Zufallsvariablen genannt, da ihre Werte aus Zufallsexperimenten stammen. Diese Einschränkung besteht in der Bayes-Statistik nicht, die Aussagen können die Werte beliebiger Variablen beinhalten. Dennoch wird die Bezeichnung Zufallsvariable beibehalten.
Gegeben sei eine diskrete Zufallsvariable X mit den diskreten Werten $x_i \in \mathbb{R}$ für $i \in \{1, \ldots, m\}$. Die Wahrscheinlichkeit $P(X = x_i|C)$, daß X den Wert x_i unter der Bedingung der Aussage C annimmt, die zusätzliche Information enthält, wird mit

$$p(x_i|C) = P(X = x_i|C) \quad \text{für} \quad i \in \{1, \ldots, m\} \tag{3.1}$$

bezeichnet. Man nennt $p(x_i|C)$ die diskrete Wahrscheinlichkeitsdichte oder auch diskrete Verteilung für die diskrete Zufallsvariable X.
Die diskrete Dichte der n-dimensionalen Zufallsvariablen $X_1, \ldots, X_n$ ist (3.1) entsprechend gegeben durch

$$p(x_{1j_1}, \ldots, x_{nj_n}|C) = P(X_1 = x_{1j_1}, \ldots, X_n = x_{nj_n}|C) \\ \text{mit} \quad j_k \in \{1, \ldots, m_k\},\ k \in \{1, \ldots, n\}\ . \tag{3.2}$$

In abgekürzter Schreibweise erhält man

$$p(x_1, \ldots, x_n|C) = P(X_1 = x_1, \ldots, X_n = x_n|C) \tag{3.3}$$

oder für den $n \times 1$ Zufallsvektor $\boldsymbol{x}$, dessen Werte ebenfalls mit $\boldsymbol{x}$ bezeichnet werden,

$$\boldsymbol{x} = |x_1, \ldots, x_n|'$$

die Dichte

$$p(\boldsymbol{x}|C)\ . \tag{3.4}$$

Auf eine entsprechende Form läßt sich auch die Dichte einer n-dimensionalen stetigen Zufallsvariablen $X_1, \ldots, X_n$ mit den Werten $x_1, \ldots, x_n \in \mathbb{R}$ in den Intervallen $-\infty < x_k < \infty$ mit $k \in \{1, \ldots, n\}$ bringen, so daß (3.4) die Dichte eines diskreten oder stetigen Zufallsvektors $\boldsymbol{x}$ bezeichnet.
Aus dem Produktgesetz (2.2) folgt die bedingte diskrete oder stetige Dichte für den diskreten oder stetigen Zufallsvektor $\boldsymbol{x}_1$ unter der Bedingung gegebener Werte für den diskreten oder stetigen Zufallsvektor $\boldsymbol{x}_2$ und der zusätzlichen Bedingung C mit

$$p(\boldsymbol{x}_1|\boldsymbol{x}_2, C) = \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2|C)}{p(\boldsymbol{x}_2|C)}\ . \tag{3.5}$$

Über die bedingte Dichte wird die bedingte Unabhängigkeit von Zufallsvariablen eingeführt. Sind $\boldsymbol{x}_i, \boldsymbol{x}_j$ und $\boldsymbol{x}_k$ diskrete oder stetige Zufallsvektoren, dann sind $\boldsymbol{x}_i$ und $\boldsymbol{x}_j$ genau dann voneinander unabhängig, falls gilt

$$p(\boldsymbol{x}_i|\boldsymbol{x}_j, \boldsymbol{x}_k, C) = p(\boldsymbol{x}_i|\boldsymbol{x}_k, C)\ . \tag{3.6}$$

Sind $\boldsymbol{x}$ und $\boldsymbol{y}$ diskrete oder stetige Zufallsvektoren, erhält man mit (3.5)

$$p(\boldsymbol{x}|\boldsymbol{y}, C) = \frac{p(\boldsymbol{x}, \boldsymbol{y}|C)}{p(\boldsymbol{y}|C)}\ , \tag{3.7}$$

worin der Vektor $\boldsymbol{y}$ gegebene Werte enthält, oder entsprechend

$$p(\boldsymbol{y}|\boldsymbol{x}, C) = \frac{p(\boldsymbol{x}, \boldsymbol{y}|C)}{p(\boldsymbol{x}|C)}\ . \tag{3.8}$$

Löst man (3.7) und (3.8) nach $p(\boldsymbol{x}, \boldsymbol{y}|C)$ auf und setzt die sich ergebenden Ausdrücke gleich, erhält man das verallgemeinerte Bayes-Theorem

$$p(\boldsymbol{x}|\boldsymbol{y}, C) = \frac{p(\boldsymbol{x}|C)p(\boldsymbol{y}|\boldsymbol{x}, C)}{p(\boldsymbol{y}|C)}\ . \tag{3.9}$$

Da der Vektor $\boldsymbol{y}$ feste Werte enthält, ist $p(\boldsymbol{y}|C)$ konstant. Das Bayes-Theorem wird daher häufig in der Form angewendet

$$p(\boldsymbol{x}|\boldsymbol{y}, C) \propto p(\boldsymbol{x}|C)p(\boldsymbol{y}|\boldsymbol{x}, C)\ , \tag{3.10}$$

in der $\propto$ das Proportionalitätszeichen bedeutet.

Der Zufallsvektor $\boldsymbol{x}$ enthalte unbekannte Parameter. Die Werte, die $\boldsymbol{x}$ annehmen kann, werden, wie bereits erwähnt, ebenfalls mit $\boldsymbol{x}$ bezeichnet. Die Menge der Werte $\boldsymbol{x}$ bezeichnet man als Parameterraum $\mathcal{X}$, also $\boldsymbol{x} \in \mathcal{X}$. Der Zufallsvektor $\boldsymbol{y}$ repräsentiere Daten. Die Dichte $p(\boldsymbol{x}|C)$ enthält Information über die Parameter $\boldsymbol{x}$, bevor die Daten $\boldsymbol{y}$ erhoben wurden, also Vorinformation. Man nennt daher $p(\boldsymbol{x}|C)$ die Priori-Dichte. Mit Berücksichtigung der Beobachtungen $\boldsymbol{y}$ folgt die Dichte $p(\boldsymbol{x}|\boldsymbol{y}, C)$, die als Posteriori-Dichte für die Parameter $\boldsymbol{x}$ bezeichnet wird. Da die Daten $\boldsymbol{y}$ vorliegen, wird $p(\boldsymbol{y}|\boldsymbol{x}, C)$ nicht als Funktion der Daten $\boldsymbol{y}$, sondern als Funktion der Parameter $\boldsymbol{x}$ interpretiert. Die Dichte $p(\boldsymbol{y}|\boldsymbol{x}, C)$ wird daher als Likelihoodfunktion bezeichnet. Die Daten modifizieren also durch die Likelihoodfunktion die Priori-Dichte und führen auf die Posteriori-Dichte für die Parameter.
Im Bayes-Theorem wird der Vektor $\boldsymbol{x}$ der unbekannten Parameter als Zufallsvektor definiert, dem eine Priori- und Posteriori-Dichte zugeordnet wird. Das bedeutet jedoch nicht, daß der Parametervektor $\boldsymbol{x}$ keine Konstanten repräsentieren dürfe wie beispielsweise die Koordinaten eines festen Punktes. Durch die Priori- und die Posteriori-Dichte wird die Wahrscheinlichkeit bestimmt, daß die Werte der Parameter in gewissen Bereichen liegen. Die Wahrscheinlichkeit gibt die Plausibilität dieser Aussagen an, also die Plausibilität von Werten der Parameter. Die Parameter können daher konstante Größen repräsentieren, sie brauchen nicht als Ergebnisse von Zufallsexperimenten interpretiert zu werden.
Die Kenntnis der Posteriori-Dichte $p(\boldsymbol{x}|\boldsymbol{y}, C)$ genügt, um die unbekannten Parameter zu schätzen, um Hypothesen für die unbekannten Parameter zu testen und um Bereiche anzugeben, in denen die Parameter mit vorgegebener Wahrscheinlichkeit liegen. Hierauf wird im folgenden Kapitel eingegangen.

4 Parameterschätzung, Konfidenzregionen und Hypothesenprüfung

Um Parameter zu schätzen oder Hypothesen zu testen, sind verschiedene Wege möglich, und für einen muß man sich entscheiden. Die Entscheidung ist zu beurteilen, um zu wissen, ob eine gute Entscheidung getroffen wurde. Dies hängt von dem wahren Zustand des Systems ab, in dem die Entscheidung zu treffen ist. Das System werde durch den Zufallsvektor $\boldsymbol{x}$ der unbekannten Parameter repräsentiert. Daten mit Information über das System existieren, sie seien in dem Zufallsvektor $\boldsymbol{y}$ zusammengefaßt. Um eine Entscheidung zu fällen, wird die Entscheidungsregel $\delta(\boldsymbol{y})$ aufgestellt, die bestimmt, welche Aktion in Abhängigkeit von den Daten $\boldsymbol{y}$ gestartet wird. Mit den Kosten der durch $\delta(\boldsymbol{y})$ ausgelösten Aktion soll die Entscheidung beurteilt werden. In Abhängigkeit von $\boldsymbol{x}$ und $\delta(\boldsymbol{y})$ wird die Kostenfunktion $L(\boldsymbol{x}, \delta(\boldsymbol{y}))$ eingeführt. Betrachtet werden die a posteriori zu erwartenden Kosten, die mit der Posteriori-Dichte $p(\boldsymbol{x}|\boldsymbol{y}, C)$ berechnet werden,

$$E[L(\boldsymbol{x}, \delta(\boldsymbol{y}))] = \int_{\mathcal{X}} L(\boldsymbol{x}, \delta(\boldsymbol{y})) p(\boldsymbol{x}|\boldsymbol{y}, C) d\boldsymbol{x} \; . \tag{4.1}$$

Die Entscheidungsregel $\delta(\boldsymbol{y})$ wird nun derart festgelegt, daß die a posteriori zu erwartenden Kosten (4.1) minimal werden. Dies bezeichnet man als Bayes-Strategie.
Es sei $\hat{\boldsymbol{x}}$ die Schätzung des Vektors $\boldsymbol{x}$ der unbekannten Parameter, so daß $\delta(\boldsymbol{y}) = \hat{\boldsymbol{x}}$ gilt. Eine einfache Kostenfunktion ergibt sich mit der Quadratsumme $(\boldsymbol{x} - \hat{\boldsymbol{x}})'(\boldsymbol{x} - \hat{\boldsymbol{x}})$ der Fehler $\boldsymbol{x} - \hat{\boldsymbol{x}}$ der Schätzung, die noch durch die Inverse $\boldsymbol{\Sigma}^{-1}$ der positiv definiten Kovarianzmatrix $D(\boldsymbol{x}) = \boldsymbol{\Sigma}$ der Parameter $\boldsymbol{x}$ gewichtet wird, so daß die quadratische Kostenfunktion

$$L(\boldsymbol{x}, \hat{\boldsymbol{x}}) = (\boldsymbol{x} - \hat{\boldsymbol{x}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \hat{\boldsymbol{x}}) \tag{4.2}$$

erhalten wird. Die Bayes-Strategie führt dann auf die Bayes-Schätzung $\hat{\boldsymbol{x}}_B$ mit

$$\hat{\boldsymbol{x}}_B = E(\boldsymbol{x}|\boldsymbol{y}) \tag{4.3}$$

oder mit der Definition des Erwartungswertes auf

$$\hat{\boldsymbol{x}}_B = \int_{\mathcal{X}} \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, C) d\boldsymbol{x} \ . \tag{4.4}$$

Mit der Kostenfunktion der absoluten Fehler erhält man die Median-Schätzung und mit Null-Eins-Kosten die MAP-Schätzung $\hat{\boldsymbol{x}}_M$, die Maximum-A-Posteriori-Schätzung

$$\hat{\boldsymbol{x}}_M = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}, C) \ . \tag{4.5}$$

Wegen ihrer einfachen Berechnung wird sie häufig angewendet. Sie entspricht der Maximum-Likelihood-Schätzung der traditionellen Statistik.
Mit der Posteriori-Dichte $p(\boldsymbol{x}|\boldsymbol{y}, C)$ für den Vektor $\boldsymbol{x}$ der unbekannten Parameter aus dem Bayes-Theorem läßt sich mit

$$P(\boldsymbol{x} \in \mathcal{X}_u|\boldsymbol{y}, C) = \int_{\mathcal{X}_u} p(\boldsymbol{x}|\boldsymbol{y}, C) d\boldsymbol{x} \tag{4.6}$$

die Wahrscheinlichkeit berechnen, daß der Vektor $\boldsymbol{x}$ im Unterraum $\mathcal{X}_u$ des Parameterraums $\mathcal{X}$ mit $\mathcal{X}_u \subset \mathcal{X}$ liegt. Um als Konfidenzregion zu dienen, ist der Unterraum derart festzulegen, daß er für maximale Dichten einen großen Teil der Wahrscheinlichkeit enthält, zum Beispiel 95%. Als Konfidenzregion $\mathcal{X}_B$ mit $\mathcal{X}_B \subset \mathcal{X}$ wird daher eine Region höchster Posteriori-Dichte, auch H.P.D.-Region genannt, definiert durch

$$P(\boldsymbol{x} \in \mathcal{X}_B|\boldsymbol{y}, C) = \int_{\mathcal{X}_B} p(\boldsymbol{x}|\boldsymbol{y}, C) d\boldsymbol{x} = 1 - \alpha$$

und

$$p(\boldsymbol{x}_1|\boldsymbol{y}, C) \geq p(\boldsymbol{x}_2|\boldsymbol{y}, C) \quad \text{für} \quad \boldsymbol{x}_1 \in \mathcal{X}_B \ , \ \boldsymbol{x}_2 \notin \mathcal{X}_B \ . \tag{4.7}$$

Den Wert für $1 - \alpha$ bezeichnet man als Konfidenzniveau und wählt in der Regel $\alpha = 0,05$. Es läßt sich zeigen, daß das Hypervolumen der Konfidenzregion (4.7) minimal im Vergleich zu den Hypervolumen beliebiger Konfidenzregionen mit dem Konfidenzniveau $1 - \alpha$ ist.
Es seien $\mathcal{X}_0 \subset \mathcal{X}$ und $\mathcal{X}_1 \subset \mathcal{X}$ Unterräume des Parameterraums $\mathcal{X}$, und $\mathcal{X}_0$ und $\mathcal{X}_1$ seien disjunkt, also $\mathcal{X}_0 \cap \mathcal{X}_1 = \emptyset$. Die Annahme $\boldsymbol{x} \in \mathcal{X}_0$ bezeichnet man als Nullhypothese und $\boldsymbol{x} \in \mathcal{X}_1$ als Alternativhypothese. Die Nullhypothese H_0 ist gegen die Alternativhypothese H_1 zu testen, folglich

$$H_0 : \boldsymbol{x} \in \mathcal{X}_0 \quad \text{gegen} \quad H_1 : \boldsymbol{x} \in \mathcal{X}_1 \ . \tag{4.8}$$

Null-Eins-Kosten werden eingeführt, indem der richtigen Entscheidung für eine korrekte Nullhypothese H_0 oder eine korrekte Alternativhypothese H_1 keine Kosten aufgebürdet werden. Die Bayes-Strategie führt dann auf die Entscheidungsregel, falls

$$\frac{\int_{\mathcal{X}_0} p(\boldsymbol{x}|\boldsymbol{y}, C) d\boldsymbol{x}}{\int_{\mathcal{X}_1} p(\boldsymbol{x}|\boldsymbol{y}, C) d\boldsymbol{x}} > 1 \ , \ \text{akzeptiere} \quad H_0 \ . \tag{4.9}$$

Andernfalls ist H_1 anzunehmen. Ähnliche Entscheidungsregeln erhält man, falls spezielle Priori-Dichten für die Hypothesen eingeführt werden.

5 Anwendungen

Wendet man die Bayes-Schätzung (4.3) im linearen Modell an, ergeben sich Ergebnisse, die mit denen der traditionellen Statistik übereinstimmen, so daß die Bayes-Statistik die Resultate der traditionellen Statistik enthält. Die Ergebnisse sind daher allgemeiner, denn beispielsweise bei der Ableitung robuster Schätzverfahren mit Hilfe der Bayes-Statistik gewinnt man den Vorteil, daß Konfidenzbereiche angegeben oder Hypothesentests vorgenommen werden können (Koch und Yang 1998a), was mit der traditionellen Statistik nicht möglich ist. Auch das robuste Kalman-Filter ist mit der Bayes-Statistik leicht angebbar (Koch und Yang 1998b). Statistische Inferenz von Varianzkomponenten, deren Schätzung von Grafarend (1978) und Grafarend und D'Hone (1978) erstmalig für geodätische Problemstellungen umfassend untersucht wurden, läßt sich ebenfalls mit der Bayes-Statistik lösen (Koch 1988; Ou und Koch 1994). Das Modell der Prädiktion und Filterung macht erst dann wirklich Sinn, wenn es im Sinne der Bayes-Statistik interpretiert wird (Koch 1994). Die automatische Interpretation digitaler Bilder mit Hilfe von Markoff-Zufallsfeldern benötigt die Bayes-Statistik (Klonowski 1998; Köster 1995; Koch 1995b). Schließlich beruhen die Bayes-Netze, die Entscheidungen in Systemen mit Unsicherheit ermöglichen, auf der Bayes-Statistik (Koch 1999). Bayes-Netze eignen sich ebenfalls zur automatischen Interpretation digitaler Bilder (Koch 1995a; Kulschewski 1999) oder für Entscheidungen im Zusammenhang mit Informationssystemen (Stassopoulou et al. 1998). Bei der Analyse geodätisch relevanter Daten ist also die Bayes-Statistik nicht mehr fortzudenken.

Literatur

Cox, R.T. (1946) Probability, frequency and reasonable expectation. *American Journal of Physics*, 14:1–13.

Grafarend, E. (1978) Schätzung von Varianz und Kovarianz der Beobachtungen in geodätischen Ausgleichungsmodellen. *Allgemeine Vermessungs-Nachrichten*, 85:41–49.

Grafarend, E. und A. d'Hone (1978) *Gewichtsschätzung in geodätischen Netzen.* Reihe A, 88. Deutsche Geodätische Kommission, München.

Jaynes, E.T. (1995) Probability theory: The logic of science. *http://bayes.wustl.edu/pub/Jaynes/book.probability.theory.*

Klonowski, J. (1998) *Segmentierung und Interpretation digitaler Bilder mit Markoff-Zufallsfeldern.* Reihe C. Deutsche Geodätische Kommission, München (im Druck).

Koch, K.R. (1988) *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer, Berlin.

Koch, K.R. (1990) *Bayesian Inference with Geodetic Applications.* Springer, Berlin.

Koch, K.R. (1994) Bayessche Inferenz für die Prädiktion und Filterung. *Z Vermessungswesen*, 119:464–470.

Koch, K.R. (1995a) Bildinterpretation mit Hilfe eines Bayes-Netzes. *Z Vermessungswesen*, 120:277–285.

Koch, K.R. (1995b) Markov random fields for image interpretation. *Z Photogrammetrie und Fernerkundung*, 63:84–90, 147.

Koch, K.R. (1999) *Einführung in die Bayes-Statistik.* Manuskript. Institut für Theoretische Geodäsie der Universität, Bonn.

Koch, K.R. und Y. Yang (1998a) Konfidenzbereiche und Hypothesentests für robuste Parameterschätzungen. *Z Vermessungswesen*, 123:20–26.

Koch, K.R. und Y. Yang (1998b) Robust Kalman filter for rank deficient observation models. *J Geodesy*, 72:436–441.

KÖSTER, M. (1995) *Kontextsensitive Bildinterpretation mit Markoff-Zufallsfeldern.* Reihe C, 444. Deutsche Geodätische Kommission, München.

KULSCHEWSKI, K. (1999) *Modellierung von Unsicherheiten mit Bayes-Netzen zur qualitativen Gebäudeerkennung und -rekonstruktion.* Institut für Theoretische Geodäsie der Universität Bonn, Bonn (in Vorbereitung).

OU, Z. und K.R. KOCH (1994) Analytical expressions for Bayes estimates of variance components. *Manuscripta geodaetica*, 19:284–293.

STASSOPOULOU, A., M. PETROU und J. KITTLER (1998) Application of a Bayesian network in a GIS based decision making system. *Int J Geographical Information Science*, 12:23–45.