TECHNISCHE UNIVERSITÄT DRESDEN
Fakultät Wirtschaftswissenschaften

Dresdner Beiträge zu
Quantitativen Verfahren

Nr. 64/16

Stetigkeit in der Statistik

von

Stefan Huschens

Herausgeber:
Die Professoren der
Fachgruppe Quantitative Verfahren

ISSN 0945-4802

Stetigkeit in der Statistik

Stefan Huschens*

Fassung vom 16. August 2016

Zusammenfassung

Es werden verschiedene Stetigkeitskonzepte, die in der statistischen Theorie und Methodik eine Rolle spielen, erläutert.
Schlüsselwörter: Stetige Verteilungsfunktion, stetige Zufallsvariable, stetige Wahrscheinlichkeitsverteilung, absolut stetig, stetiges Merkmal, λ-stetiges Maß, Stetigkeitskorrektur

1 Stetigkeitskonzepte

Bemerkung 1.1 (Stetiges Merkmal) In der deskriptiven Statistik wird ein Merkmal als stetig oder kontinuierlich bezeichnet, wenn die Menge aller Merkmalsausprägungen durch $\mathbb{R}$ oder durch ein Teilintervall von $\mathbb{R}$ gegeben ist.

Bemerkung 1.2 (Stetigkeit einer Verteilungsfunktion) Die Verteilungsfunktion

$$F_X : \mathbb{R} \to [0,1], \quad x \mapsto F_X(x) = P(X \leq x)$$

einer Zufallsvariablen X ist an jeder Stelle $x \in \mathbb{R}$ rechtsseitig stetig, aber nicht notwendig auch linksseitig stetig.

Bemerkung 1.3 (Stetigkeit einer Zufallsvariablen)

1. Eine Zufallsvariable X heißt **stetig im weiteren Sinn**, wenn die Verteilungsfunktion stetig ist. Eine Zufallsvariable X heißt **stetig im engeren Sinn**, wenn die Verteilungsfunktion an jeder Stelle durch ein Integral bezüglich einer Dichtefunktion f_X berechnet werden kann,
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t, \quad x \in \mathbb{R}.$$
2. Die Bezeichnungsweise „stetige Zufallsvariable“ in der statistischen Literatur ist schwankend.

*Bitte senden Sie Hinweise auf Fehler an `stefan.huschens@tu-dresden.de`. Unter `https://www.stefan-huschens.de/statistik/statistische-miszellen/` können Sie sich informieren, ob es eine aktuellere Fassung gibt, und diese gegebenenfalls herunterladen.

(a) Häufig werden Zufallsvariablen, für die eine Dichtefunktion existiert, die also stetige Zufallsvariablen im engeren Sinn sind, als stetige Zufallsvariablen bezeichnet.

(b) Manchmal wird der Begriff stetige Zufallsfallsvariable für eine Zufallsvariable mit stetiger Verteilungsfunktion verwendet, z. B. [2, S. 32].

Bemerkung 1.4 (Stetigkeit einer Wahrscheinlichkeitsverteilung) Die Stetigkeit einer Zufallsvariablen im weiteren und im engeren Sinn sind Eigenschaften der Wahrscheinlichkeitsverteilung der Zufallsvariablen und nicht Eigenschaften der Zufallsvariablen als Funktion $X : \Omega \to \mathbb{R}$. Daher vermeiden es einige Autoren völlig, von stetigen Zufallsvariablen zu sprechen, sondern sprechen nur von Zufallsvariablen mit stetiger oder absolut stetiger Wahrscheinlichkeitsverteilung. Die Wahrscheinlichkeitsverteilung einer Zufallsvariablen X ist das durch

$$P_X(B) \stackrel{\text{def}}{=} P(X \in B), \quad B \in \mathbb{B}$$

definierte Wahrscheinlichkeitsmaß P_X auf dem Messraum $(\mathbb{R}, \mathbb{B})$, wobei $\mathbb{B}$ die Menge der Borelmengen ist. Die Wahrscheinlichkeitsverteilung P_X heißt stetig, wenn $P_X(\{x\}) = 0$ für alle $x \in \mathbb{R}$ gilt, und heißt absolut stetig, wenn $P_X(B) = 0$ für alle $B \in \mathbb{B}$ mit $\lambda(B) = 0$ gilt, wobei λ das Lebesgue-Maß bezeichnet.

Bemerkung 1.5 (Stetigkeit eines Wahrscheinlichkeitsmaßes) Die sogenannte „Stetigkeit der Wahrscheinlichkeit“ als Maß betrifft eine Eigenschaft der Mengenfunktion $P : \mathbb{A} \to [0, 1]$, wobei $(\Omega, \mathbb{A}, P)$ ein Wahrscheinlichkeitsraum ist. Für P gelten die beiden folgenden Stetigkeitseigenschaften, z. B. [3, S. 30].

1. Stetigkeit von unten: Für eine monoton zunehmende Folge $(A_n)_{n\in\mathbb{N}}$ von Ereignissen $A_n \in \mathbb{A}$, d. h. $A_n \subseteq A_{n+1}$, und $A \stackrel{\text{def}}{=} \bigcup_{n\in\mathbb{N}} A_n$ gilt
$$\lim_{n\to\infty} P(A_n) = P(A).$$

2. Stetigkeit von oben: Für eine monoton abnehmende Folge $(A_n)_{n\in\mathbb{N}}$ von Ereignissen $A_n \in \mathbb{A}$, d. h. $A_{n+1} \subseteq A_n$, und $A \stackrel{\text{def}}{=} \bigcap_{n\in\mathbb{N}} A_n$ gilt
$$\lim_{n\to\infty} P(A_n) = P(A).$$

Bemerkung 1.6 (Stetige Abbildung und Konvergenz) Im Zusammenhang mit verschiedenen Konvergenzbegriffen für Zufallsvariablen und Zufallsvektoren spielt die Stetigkeit einer Funktion $g : \mathbb{R}^k \to \mathbb{R}^m$ eine Rolle, wenn aus der Konvergenz einer Folge $(X_n)_{n\in\mathbb{N}}$ gegen X auf die Konvergenz der Folge $(g(X_n))_{n\in\mathbb{N}}$ gegen $g(X)$ geschlossen werden soll.

Bemerkung 1.7 (Stetigkeitskorrektur) Die sogenannte Stetigkeitskorrektur kommt zum Einsatz, wenn die Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariable Z mit Werten in $\mathbb{Z}$ durch eine Wahrscheinlichkeitsverteilung mit stetiger Verteilungsfunktion F approximiert wird. Dabei wird z. B. die diskrete Wahrscheinlichkeit $P(Z = z)$ nicht auf den Punkt $\{z\}$ bezogen, sondern dem Intervall $[z - 1/2, z + 1/2]$ zugeordnet und dann die Wahrscheinlichkeit $P(Z = z)$ durch $F(z + 1/2) - F(z - 1/2)$ approximiert. Ein häufiger Anwendungsbereich ist die Approximation einer Wahrscheinlichkeit einer binomialverteilten Zufallsvariablen durch eine Wahrscheinlichkeit aus einer Normalverteilung.

2 Weiterführendes

Bemerkung 2.1 (Zur Stetigkeit eines Merkmals)

1. Ein stetiges Merkmal hat folgende Eigenschaften:

   (a) Die Anzahl der Merkmalsausprägungen ist überabzählbar.

   (b) Zu zwei Merkmalsausprägungen x, y mit $x < y$ existiert immer eine Merkmalsausprägung z mit $x < z < y$.

   (c) Zu zwei Merkmalsausprägungen x und y ist jede Zahl z mit $x < z < y$ eine Merkmalsausprägung.

   (d) Wenn $(z_n)_{n \in \mathbb{N}}$ eine Folge von Merkmalsausprägungen ist, die gegen $z \in \mathbb{R}$ konvergiert,
   $$\lim_{n \to \infty} z_n = z,$$
   und wenn $x < z < y$ für zwei Merkmalsausprägungen x und y gilt, dann ist auch z eine Merkmalsausprägung.

   Die erste Eigenschaft unterscheidet zwar ein stetiges Merkmal von einem diskreten Merkmal, dessen Menge von Merkmalsausprägungen abzählbar ist, ist aber zur Definition eines stetigen Merkmals ungeeignet. So ist z. B. die Cantor-Menge[1] (auch Cantor'sches Diskontinuum oder Cantorstaub genannt) eine überabzählbare Menge, die kein noch so kleines, nicht zu einem Punkt degeneriertes, Intervall als Teilmenge enthält. Auch die Menge $\mathbb{R} \setminus \mathbb{Q}$ der irrationalen Zahlen ist überabzählbar, aber als Menge der Merkmalsausprägungen eines stetigen Merkmals ungeeignet.

   Die zweite Eigenschaft wird auch von der abzählbaren Menge der rationalen Zahlen geteilt und ist daher kein Unterscheidungskriterium zwischen stetigen und diskreten Merkmalen. Wenn x und y rationale Zahlen mit $x < y$ sind, dann ist auch $z = (x + y)/2$ eine rationale Zahl und es gilt $x < z < y$.

   Die dritte Eigenschaft charakterisiert Teilintervalle der reellen Zahlen und ist als definierende Eigenschaft eines stetigen Merkmals geeignet, wenn vorausgesetzt wird, dass das Merkmal numerisch ist, also die Menge der Merkmalsausprägungen eine Teilmenge von $\mathbb{R}$ ist.

   Die vierte Eigenschaft folgt weder aus der ersten noch aus der zweiten Eigenschaft, aber aus der dritten Eigenschaft.

2. Bei Anwendungen erhält in der Regel eine statistische Variable, die die Merkmalsausprägungen eines Merkmals an verschiedenen Einheiten misst, denselben Namen wie das Merkmal. Wenn M die Menge der möglichen Merkmalsausprägungen eines bestimmten Merkmals ist und I die Menge der statistischen Einheiten, dann ist $X : I \to M$ eine statistische Variable, die das Auftreten der Merkmalsausprägungen an den statistischen Einheiten charakterisiert. Dabei ist $X(i) \in M$ für $i \in I$ die Merkmalsausprägung der i-ten Untersuchungseinheit. Es gibt dann in einer Untersuchung beispielsweise die Variablen ‚Geschlecht' und ‚Körpergröße', die zu den entsprechenden Merkmalen gehören. Dabei wird teilweise die Begrifflichkeit von den

[1] `https://de.wikipedia.org/wiki/Cantor-Menge`

Merkmalen auf die Variablen übertragen und in diesem speziellen Sinn wird auch von einer **stetigen Variablen** gesprochen, wenn das gemessene Merkmal stetig ist. Dies kann missverständlich sein, da es dann nicht um eine Stetigkeitseigenschaft der Funktion $X : I \to M$, sondern um eine Eigenschaft der Menge M, des Wertevorrats der Funktion X, geht. Der Bildbereich der Funktion X, d. h. die Menge $X(I) = \{X(i) \in M \mid i \in I\} \subseteq M$, ist für ein stetiges Merkmal immer abzählbar, wenn I eine abzählbare Menge ist, und kann auch dann eine abzählbare Menge sein, wenn I und M überabzählbare Mengen sind.

Bemerkung 2.2 (Zur Stetigkeit der Verteilungsfunktion)

1. Die rechtsseitige Stetigkeit einer Verteilungsfunktion ergibt sich aus der Stetigkeit der Wahrscheinlichkeit, siehe Bemerkung 1.5.

2. Wenn die Verteilungsfunktion F_X einer Zufallsvariablen X an der Stelle $x_0 \in \mathbb{R}$ nicht linksseitig stetig ist, dann liegt eine Sprungstelle vor und es gilt $P(X = x_0) > 0$.

3. Die Zufallsvariablen mit stetiger Verteilungsfunktion bilden für viele Aussagen der statistischen Theorie eine wichtige Teilmenge aller Zufallsvariablen.

4. Für eine Zufallsvariable X mit stetiger Verteilungsfunktion $F_X : \mathbb{R} \to [0, 1]$ gilt $F_X(X) \sim \text{uni}(0, 1)$.

Bemerkung 2.3 (Zur Stetigkeit der Wahrscheinlichkeit)

1. Die beiden Stetigkeitseigenschaften 1. und 2. aus Bemerkung 1.5 hängen wesentlich von der Eigenschaft der σ-Additivität ab. Für einen Wahrscheinlichkeitsinhalt, d. h. eine nichtnegative, normierte und endlich-additive Mengenfunktion auf einem Messraum, ist die zusätzliche Eigenschaft der σ-Additivität äquivalent zu jeder der beiden Eigenschaften 1. und 2. aus Bemerkung 1.5, [1, S. 16, 21].

2. Wegen des Bezugs zur σ-Additivität heißt die Stetigkeit von oben bzw. unten auch σ-Stetigkeit von oben bzw. unten.

Bemerkung 2.4 (Zur Stetigkeit von Zufallsvariablen)

1. Jede stetige Zufallsvariable im engeren Sinn ist auch eine stetige Zufallsvariable im weiteren Sinn.

2. Eine Zufallsvariable ist genau dann stetig im engeren Sinn, wenn die Verteilungsfunktion eine absolut stetige[2] Verteilungsfunktion hat.

3. Ein Beispiel einer Zufallsvariablen mit stetiger Verteilungsfunktion, die nicht absolut stetig ist, ist eine Zufallsvariable mit der sogenannten Cantor-Verteilung[3].

Bemerkung 2.5 (Zur Stetigkeit einer Wahrscheinlichkeitsverteilung) Die Stetigkeit einer Wahrscheinlichkeitsverteilung im Sinn von Bemerkung 1.4 ist ein Spezialfall eines allgemeineren Stetigkeitsbegriffs aus der Maßtheorie:

[2] https://de.wikipedia.org/wiki/Absolut_stetige_Funktion

[3] https://de.wikipedia.org/wiki/Cantor-Verteilung

1. Unter der Stetigkeit eines Maßes ν bezogen auf eine anderes Maß μ, wobei beide Maße auf demselben Messraum $(\Omega, \mathbb{A})$ definiert sind, versteht man folgende Eigenschaft:
$$\mu(A) = 0 \implies \nu(A) = 0, \quad \text{für alle } A \in \mathbb{A}.$$
Man sagt, das Maß ν ist *absolut stetig bezüglich* μ, ist μ*-stetig*, oder ist *dominiert von* μ. Man schreibt dafür auch $\nu \ll \mu$.

2. Diese Stetigkeit eines Maßes spielt eine wesentliche Rolle beim Satz von Radon-Nikodým[4] über die Existenz einer Dichtefunktion.

3. Ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathbb{B})$ ist genau dann absolut stetig bezüglich des Lebesgue-Maßes λ oder λ-stetig, wenn die Verteilungsfunktion $F(t) = P((-\infty, t])$, $t \in \mathbb{R}$ eine absolut stetige Funktion ist.

Bemerkung 2.6 (Zur Konvergenzerhaltung bei stetigen Abbildungen) Das Theorem über stetige Abbildungen (*continuous mapping theorem*)[5], das auch Mann-Wald-Theorem genannt wird, hat folgende Aussage: Wenn die Funktion $g : \mathbb{R}^k \to \mathbb{R}^m$ stetig ist, dann gilt
$$X_n \to X \implies g(X_n) \to g(X),$$
wobei $\to$ für einen der folgenden Konvergenzbegriffe für Zufallsvariablen stehen kann: Konvergenz in Wahrscheinlichkeit, Konvergenz in Verteilung und fast sichere Konvergenz. Diese Aussage bleibt erhalten, wenn allgemeiner die Funktion g Unstetigkeitsstellen hat und für die Menge D_g der Unstetigkeitsstellen $P(X \in D_g) = 0$ gilt.

Literatur

[1] Borovkov, A. A.: Probability Theory. Springer: London 2013.

[2] Casella, G., Berger, R. L.: Statistical Inference. 2. Aufl. Duxbury Press: Belmont 2002.

[3] Proschan, M. A., Shaw, P. A.: Essentials of Probability Theory for Statisticians. CRC Press: Boca Raton 2016.

[4] https://de.wikipedia.org/wiki/Satz_von_Radon-Nikod%C3%BDm
[5] https://en.wikipedia.org/wiki/Continuous_mapping_theorem

Dresdner Beiträge zu Quantitativen Verfahren (ISSN 0945-4802)

Ältere Ausgaben (1/94 – 43/04): http://wwqvs.file3.wcms.tu-dresden.de/f-db.htm

44/05 S. Höse, K. Vogl: Modeling and Estimating the Credit Cycle by a Probit-AR(1)-Process.
Erschienen in: *From Data and Information Analysis to Knowledge Engineering,* Hrsg: M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nürnberger, W. Gaul, Springer, Berlin, 2006, S. 534-541.

45/05 S. Höse, K. Vogl: Predicting the Credit Cycle with an Autoregressive Model.

46/06 S. Huschens, A. Karmann, D. Maltritz, K. Vogl: Country Default Probabilities: Assessing and Backtesting.
Erschienen in: *The Journal of Risk Model Validation*, Vol.1, Heft 2, 2007, S. 3-26.

47/08 S. Höse, S. Huschens, R. Wania: Rating Migrations.
Erschienen in: *Applied Quantitative Finance*, Hrsg.: W. K. Härdle, N. Hautsch, L. Overbeck, Springer, Berlin, 2009, S. 105-123.

48/08 S. Höse, S. Huschens: Ausfallrisiko.
Erschienen in: *Praxishandbuch Risikomanagement: Konzepte - Methoden - Umsetzung*, Hrsg: W. Gleißner, F. Romeike, Erich Schmidt Verlag, Berlin 2015, S. 305-324.

49/09 E. Lovász, B. Schipp: The Impact of HIV/AIDS on Economic Growth in Sub-Saharan Africa.
Erschienen in: *South African Journal of Economics*, Vol. 77, Nr. 2, 2009, S. 245-256.

50/09 S. Höse, S. Huschens: Confidence Intervals for Correlations in the Asymptotic Single Risk Factor Model.

51/10 S. Höse, S. Huschens: Confidence Intervals for Quantiles of a Vasicek-distributed Credit Portfolio Loss.

52/10 D. Tillich: Risikomaßzahlen für Kreditportfoliotranchen.
Erschienen in: *AStA Wirtschafts- und Sozialstatistisches Archiv,* Vol. 5, Nr. 1, 2011, S. 59-76.

53/10 S. Huschens: Kann es Rückzahlungswahrscheinlichkeiten von 100% geben?
Erschienen in: *bank und markt – Zeitschrift für Retailbanking*, 39. Jg., Heft 3/2010, S. 11.

54/11 S. Höse, S. Huschens: Confidence Intervals for Asset Correlations in the Asymptotic Single Risk Factor Model.
Erschienen in: *Operations Research Proceedings 2010,* Hrsg: B. Hu, K. Morasch, S. Pickl, M. Siegle, Springer, Berlin, 2011, S. 111-116.

55/11 S. Höse, S. Huschens: Stochastic Orders and Non-Gaussian Risk Factor Models.
Erschienen in: *Review of Managerial Science*. Vol. 7, Nr. 2, 2013, S. 99-140.

56/11 D. Tillich: Bounds for the Expectation of Bounded Random Variables.

57/12 S. Höse, S. Huschens: Credit Portfolio Correlations and Uncertainty.
Erschienen in: *Credit Securitisations and Derivatives – Challenges for the Global Markets,* Hrsg.: D. Rösch, H. Scheule, Wiley: Chichester, 2013, S. 53-70

58/12 S. Fischer: Ratio calculandi periculi - Ein analytischer Ansatz zur Bestimmung der Verlustverteilung eines Kreditportfolios

59/13 D. Tillich, D. Ferger: Estimation of Rating Classes and Default Probabilities in Credit Risk Models with Dependencies.
Erschienen in: *Applied Stochastic Models in Business and Industry.* DOI: 10.1002/asmb.2089

60/14 C. Lehmann, D. Tillich: Consensus Information and Consensus Rating – A Note on Methodological Problems of Rating Aggregation.
Erscheint in: *Operations Research Proceedings 2014,* Springer.

61/15 C. Lehmann, D. Tillich: Applied Consensus Information and Consensus Rating – A Simulation Study on Rating Aggregation.

62/16 C. Lehmann: Modellierung der Abhängigkeitsstruktur von Ausfallkörben – Eine Betrachtung für den Spezialfall des Duo-Baskets.

63/16 S. Huschens: Chance (*odd*) versus Wahrscheinlichkeit (*probability*).