Siegmund Brandt

Datenanalyse

Mit statistischen Methoden und Computerprogrammen

4. Auflage

mit CD-ROM

Spektrum Akademischer Verlag Heidelberg · Berlin

Autor:
Prof. Dr. Siegmund Brandt
Fachbereich Physik
Universität Siegen
57968 Siegen
e-mail: brandt@physik.uni-siegen.de

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Brandt, Siegmund:
Datenanalyse [Medienkombination] ; mit statistischen Methoden und Computerprogrammen / Siegmund Brandt. – Heidelberg ; Berlin : Spektrum, Akad. Verl.
1. Aufl. im Verl, Bibliographisches Inst., Mannheim, Wien, Zürich.
– 2. und 3. Aufl. im BI-Wiss.-Verl., Mannheim, Leipzig, Wien, Zürich
ISBN 3-8274-0158-5

Buch. - 4. Aufl. -1999
CD-ROM [zur 4. Aufl.]. - 1999

© 1999 Spektrum Akademischer Verlag GmbH Heidelberg · Berlin

Alle Rechte, insbesondere die der Übersetzung in fremde Sprachen, sind vorbehalten. Kein Teil des Buches darf ohne schriftliche Genehmigung des Verlages photokopiert oder in irgendeiner anderen Form reproduziert oder in eine von Maschinen verwendbare Sprache übertragen oder übersetzt werden.

Der Verlag und der Autor haben alle Sorgfalt walten lassen, um vollständige und möglichst fehlerfreie Informationen in diesem Buch und auf der beigefügten CD-ROM zu publizieren. Verlag und Autor übernehmen jedoch weder Garantie noch die juristische Verantwortung oder irgendeine Haftung für die Nutzung dieser Informationen, für deren Wirtschaftlichkeit oder fehlerfreie Funktion für einen bestimmten Zweck. Ferner können Verlag und Autor für Schäden, die auf eine Fehlfunktion von Programmen oder ähnliches zurückzuführen sind, nicht haftbar gemacht werden. Eine telefonische oder schriftliche Beratung durch Verlag oder Autor über den Einsatz der Programme ist nicht möglich. Verlag und Autor übernehmen keine Gewähr dafür, daß die beschriebenen Verfahren, Programme usw. frei von Schutzrechten Dritter sind. Mit dem Erwerb des Buches ist das Recht verbunden, den Inhalt der CD-ROM auf die Festplatte eines Rechners herunterzuladen. Die Herstellung von Kopien der CD-ROM ist untersagt. Das Recht zur Nutzung der Programme ist auf persönliche Zwecke und Zwecke von Forschung und Lehre beschränkt. Nutzung für kommerzielle Zwecke bedarf der schriftlichen Zustimmung des Verlages.

Lektorat: Dr. Georg W. Botz, Jutta Liebau (Ass.)
Umschlaggestaltung: Kurt Bitsch, Birkenau
Druck und Verarbeitung: Franz Spiegel Buch GmbH, Ulm

im Vergleich zu

$$E = \frac{\pi R^2}{2} \cdot \frac{1}{2R^2} = \frac{\pi}{4}$$

in Beispiel 4.4. ■

4.10 Erzeugung normalverteilter Zufallszahlen

Die mit Abstand wichtigste Verteilung für die Datenanalyse ist die Normalverteilung, die wir im Abschnitt 5.7 diskutieren werden. Schon hier wollen wir ein Verfahren und ein Programm angeben, mit dem Zufallszahlen x_i erzeugt werden können, die der standardisierten Normalverteilung mit der Wahrscheinlichkeitsdichte

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \tag{4.10.1}$$

folgen. Die zugehörige Verteilungsfunktion $F(x)$ läßt sich nur numerisch berechnen und umkehren (Anhang C). Damit scheidet die einfache Transformationsmethode des Abschnitts 4.9.1 aus. Die im folgenden beschriebene *Polar-Methode* von BOX und MULLER [5] kombiniert auf eindrucksvolle Weise Rückweisung mit Transformation. Der Algorithmus besteht aus den folgenden Teilschritten.

(i) Wahl zweier unabhängiger Zufallszahlen $\mathsf{u}_1, \mathsf{u}_2$ aus einer Gleichverteilung zwischen 0 und 1. Umformungen $\mathsf{v}_1 = 2\mathsf{u}_1 - 1$, $\mathsf{v}_2 = 2\mathsf{u}_2 - 1$.

(ii) $\mathsf{s} = \mathsf{v}_1^2 + \mathsf{v}_2^2$.

(iii) Falls $\mathsf{s} \geq 1$, zurück zu Schritt (i).

(iv) $\mathsf{x}_1 = \mathsf{v}_1\sqrt{-(2/\mathsf{s})\ln \mathsf{s}}$ und $\mathsf{x}_2 = \mathsf{v}_2\sqrt{-(2/\mathsf{s})\ln \mathsf{s}}$ sind zwei unabhängige Zufallszahlen, die der standardisierten Normalverteilung folgen.

Die Zahlenpaare $(\mathsf{v}_1, \mathsf{v}_2)$, die in Schritt (i) vorliegen, sind die Koordinaten eine Punktmenge, die gleichförmig über das Innere des Einheitskreises verteilt ist. Der Punkt $(\mathsf{v}_1, \mathsf{v}_2)$ hat die Polarkoordinaten $\mathsf{v}_1 = \mathsf{r}\cos\theta$, $\mathsf{v}_2 = \mathsf{r}\sin\theta$ mit $\mathsf{r} = \sqrt{\mathsf{s}}$, $\theta = \arctan(\mathsf{v}_2/\mathsf{v}_1)$. Der Punkt $(\mathsf{x}_1, \mathsf{x}_2)$ hat die Polarkoordinaten

$$\mathsf{x}_1 = \cos\theta\sqrt{-2\ln \mathsf{s}}\,, \quad \mathsf{x}_2 = \sin\theta\sqrt{-2\ln \mathsf{s}}\,.$$

Wir fragen jetzt nach der Wahrscheinlichkeit

$$F(r) = P(\sqrt{-2\ln \mathsf{s}} \leq r) = P(-2\ln \mathsf{s} \leq r^2) = P(\mathsf{s} > \mathrm{e}^{-r^2/2})\,.$$

Da $\mathsf{s} = \mathsf{r}^2$ nach Konstruktion gleichverteilt zwischen 0 und 1 ist, gilt

$$F(r) = P(\mathsf{s} > \mathrm{e}^{-r^2/2}) = 1 - \mathrm{e}^{-r^2/2}\,.$$

Die Wahrscheinlichkeitsdichte von r ist

$$f(r) = \frac{\mathrm{d}F(r)}{\mathrm{d}r} = r\mathrm{e}^{-r^2/2}\,.$$

Die gemeinsame Verteilungsfunktion von x_1 und x_2,

$$\begin{aligned}
F(x_1, x_2) &= P(\mathsf{x}_1 \le x_1,\ \mathsf{x}_2 \le x_2) = P(\mathsf{r}\cos\theta \le x_1,\ \mathsf{r}\sin\theta \le x_2) \\
&= \frac{1}{2\pi}\iint_{(\mathsf{x}_1<x_1,\mathsf{x}_2<x_2)} r\mathrm{e}^{-r^2/2}\,\mathrm{d}r\,\mathrm{d}\varphi \\
&= \frac{1}{2\pi}\iint_{(\mathsf{x}_1<x_1,\mathsf{x}_2<x_2)} \mathrm{e}^{-(x_1^2+x_2^2)/2}\,\mathrm{d}x\,\mathrm{d}y \\
&= \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_1} \mathrm{e}^{-x_1^2/2}\,\mathrm{d}x_1\right)\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_2} \mathrm{e}^{-x_2^2/2}\,\mathrm{d}x_2\right),
\end{aligned}$$

ist das Produkt aus zwei Verteilungsfunktionen der standardisierten Normalverteilung. Das Verfahren ist in Bild 4.6 graphisch veranschaulicht.

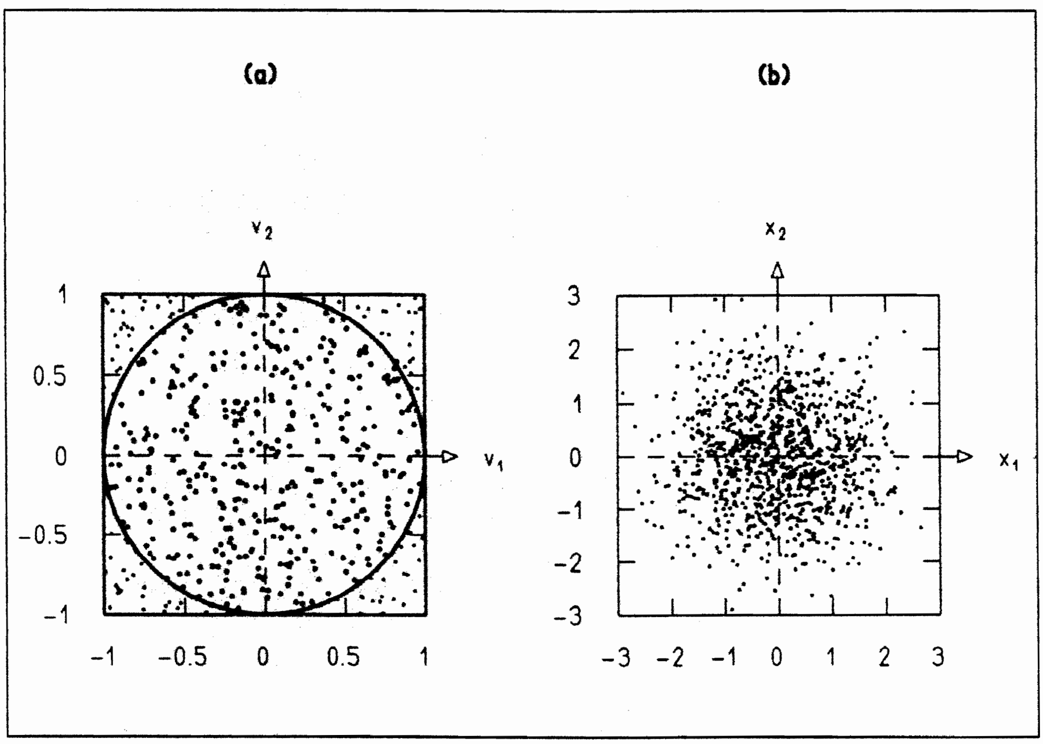

Bild 4.6: Illustration des Box-Muller-Verfahrens. (a) Es werden Zahlenpaare $(\mathsf{v}_1, \mathsf{v}_2)$ erzeugt, die das Quadrat gleichmäßig bevölkern. Anschließend werden diejenigen verworfen, die nicht innerhalb des Einheitskreises liegen (markiert durch kleine Punkte). (b) Anschließend erfolgt die Transformation $(\mathsf{v}_1, \mathsf{v}_2) \rightarrow (\mathsf{x}_1, \mathsf{x}_2)$.

Literatur

Im Text zitierte Literaturhinweise

[1] A. Kolmogorov, *Ergebn. Math.* **2** (1933) 3

[2] D.E. Knuth, *The Art of Computer Programming*, vol. 2, Addison-Wesley, Reading MA 1981

[3] P. L'Ecuyer, *Comm. ACM* **31** (1988) 742

[4] B.A. Wichmann and I.D. Hill, *Appl. Stat.* **31** (1982) 188

[5] G.E.P. Box and M.E. Muller, *Ann. Math. Stat.* **29** (1958) 611

[6] L. von Bortkiewicz, *Das Gesetz der kleinen Zahlen*, Teubner, Leipzig 1898

[7] D.J. De Solla Price, *Little Science, Big Science*, Columbia University Press, New York 1965

[8] M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 2, Charles Griffin, London 1968

[9] S.S. Wilks, *Ann. Math. Stat.*, **9** (1938) 60

[10] A.H. Rosenfeld, H.A. Barbero-Galtieri, W.J. Podolski, L.R. Price, P. Söding, Ch.G. Wohl, M. Roos and W.J. Willis, *Rev. Mod. Phys.* **33** (1967) 1

[11] Particle Data Group, *Physics Letters* **B204** (1988) 1

[12] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge 1986

[13] R.P. Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs NJ 1973

[14] J.A. Nelder and R. Mead, *Computer Journal* **7** (1965) 308

[15] M.J.D. Powell, *Computer Journal* **7** (1965) 155