

## Mathematische und rechentechnische Aufbereitung der Betaverteilung 1. Art für technologische Untersuchungen

### 1. Die Bedeutung der Betaverteilung 1. Art für technologische Untersuchungen

Die Familie der Betaverteilung nimmt einen wichtigen Platz in der Klasse aller stetigen Verteilungsfunktionen ein. Auf die genannte Verteilung lassen sich die t-Verteilung, die F-Verteilung, die negative Binomialverteilung, die Verteilung des Korrelationskoeffizienten, die U-förmige, die Dreieck- und die Gleichverteilung zurückführen [6, S. 333], [10, S. 457], [11, S. 22, 439].

Die Anwendung der Betaverteilung 1. Art innerhalb der mathematischen Statistik ist außerordentlich vielseitig. Sie kommt u. a. in der Schätz- und Testtheorie, in der Bedienungstheorie und in ökonomischen, technischen und technologischen Modellen zum Einsatz. Bei eisenbahntechnologischen Untersuchungen kommen oftmals stetige, unimodale und asymmetrische empirische Häufigkeitslinien vor, die mit dem positiven Teil der Abszissenachse zwei Schnittpunkte besitzen. Mit der Betaverteilung 1. Art können für diese Fälle Modelle aufgestellt werden, die es ermöglichen, sowohl positiv schiefe als auch negativ schiefe beobachtete Verteilungen zu beschreiben. In dem Maße, wie größere technologische Einheiten analysiert werden und wie die zu bearbeitende Datenmenge wächst, erscheint es sinnvoll, rechentechnische Hilfsmittel bereitzustellen und, damit verbunden, theoretische Untersuchungen zum Problem der Betaverteilung 1. Art anzustellen. Im Anschluß an die Arbeiten von Potthoff [8], [9] soll in diesem Artikel eine mathematisch-algorithmische und rechentechnische Aufbereitung der Betaverteilung 1. Art zur Aufstellung eines Rechenprogramms vorgenommen werden.

### 2. Definition der Betaverteilung 1. Art

Eine stetige Zufallsgröße  $Y$  besitzt Betaverteilung 1. Art (*Pearson*-Typ I-Verteilung) über dem offenen Intervall  $(a, b)$  mit den Parametern  $(p, q)$ , wobei  $p > 0$ ,  $q > 0$ , wenn ihre Dichtefunktion

$$f(Y; p, q) = \frac{(b - a)^{1-p-q}}{B(p, q)} (Y - a)^{p-1} (b - Y)^{q-1}, \quad a < Y < b \quad (1)$$

lautet [2, S. 186], [3, S. 48], [5, S. 2], [11, S. 21, 439].

Die Betaverteilung 1. Art über  $(a, b)$  kann durch Transformation

$$X = \frac{Y - a}{b - a}$$

auf die spezielle Betaverteilung 1. Art über  $(0, 1)$  zurückgeführt werden, so daß in den meisten Fällen die Betrachtung dieses Spezialfalls mit der Dichtefunktion

$$f(X; p, q) = \frac{X^{p-1} (1 - X)^{q-1}}{B(p, q)}, \quad 0 < X < 1 \quad (2)$$

(Bild 1) genügt.

Die zugehörige Verteilungsfunktion (Bild 2) hat die folgende Form:

$$F(X; p, q) = \begin{cases} 0 & x \leq 0 \\ \int_0^x f(X; p, q) dX & 0 < x < 1 \\ 1 & 1 \leq x \end{cases} \quad (3)$$

und stellt den Zusammenhang mit der unvollständigen Betafunktion ( $\sim$  B-Funktion), einem Sonderfall der *Gaußschen* Funktion, her:

$$B_x(p, q) = F(x; p, q) \cdot B(p, q). \quad (4)$$

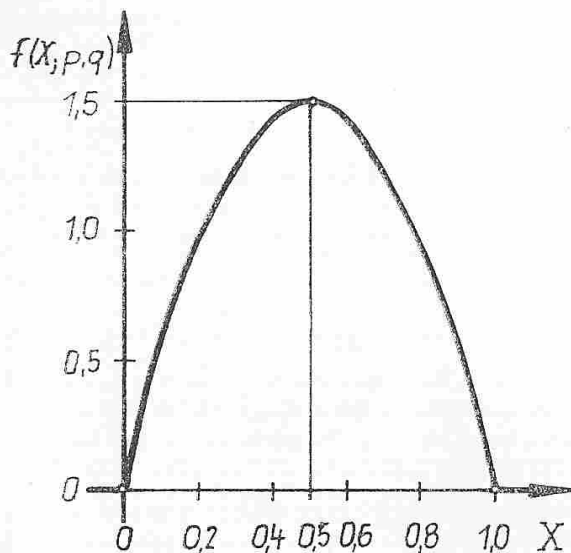


Bild 1  
Wahrscheinlichkeitsdichte bzw. Dichte bzw. Dichtefunktion der Beta-Verteilung  
1. Art für  $a = 0, b = 1, p = q = 2$ :  
 $f(X; p, q) = 6X(1 - X), 0 < X < 1$

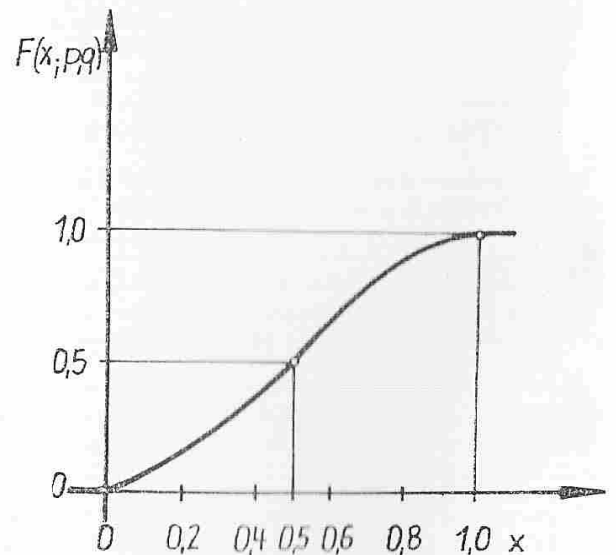


Bild 2  
Verteilungsfunktion der Beta-Verteilung  
1. Art für  $a = 0, b = 1, p = q = 2$ :  
 $F(x; p, q) = \begin{cases} 0 & \text{für } x \leq 0 \text{ u. } x \geq 1 \\ 6 \left( \frac{x^2}{2} - \frac{x^3}{3} \right) & \text{für } 0 < x < 1 \end{cases}$

In den Gleichungen (1), (2) und (4) tritt implizit  $B(p, q)$  auf.  $B(p, q)$  ist die Betafunktion (B-Funktion, *Eulersches* Integral erster Art), die wie folgt definiert ist:

$$B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)} \quad (5)$$

[2, S. 186], [3, S. 48]. In dieser Gleichung wird die B-Funktion mit der  $\Gamma$ -Funktion (dem *Eulerschen* Integral zweiter Art)

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt \quad (6)$$

verknüpft.

Um Schätzungen der Parameter (Schätzwerte, Punktschätzungen für die Parameter) der Beta-Verteilung 1. Art zu konstruieren, bedient man sich vor allem der Momentenmethode.

Für eine betaverteilte Zufallsgröße  $X$  1. Art existieren die  $k$ -ten Momente ( $k$  natürliche Zahl). Entsprechend [2, S. 186], [3, S. 48], [11, S. 21] gilt für die gewöhnlichen Anfangsmomente  $k$ -ter Ordnung

$$m_k = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^1 x^{p+k-1}(1-x)^{q-1} dx = \frac{B(k+p, q)}{B(p, q)} = \frac{p(p+1)\cdots(p+k-1)}{(p+q)(p+q+1)\cdots(p+q+k-1)} \quad (7)$$

Für die  $k$ -ten gewöhnlichen zentralen Momente bezüglich des Erwartungswertes  $EX = m_1$  erhält man

$$\mu_k = E(X - EX)^k. \quad (8)$$

Mit Hilfe der gewöhnlichen Momente können Parameter der Beta-Verteilung 1. Art bestimmt werden. Insbesondere findet man:

– für den Erwartungswert

$$\mu = EX = m_1 = \frac{p}{p+q}, \quad (9)$$

– für die Varianz

$$\sigma^2 = D^2 X = \mu_2 = \frac{pq}{(p+q)^2(p+q+1)}, \quad (10)$$

– für den Variationskoeffizienten

$$\mu = \frac{\sigma}{m_1} \sqrt{\frac{q}{p(p+q+1)}}, \quad (11)$$

– für die *Charliersche* Schiefe

$$\gamma_3 = S = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{2(q-p)}{p+q+2} \sqrt{\frac{p+q+1}{pq}}, \quad (12)$$

– für den Exzeß

$$\gamma_4 = E = \frac{\mu_4}{\mu_2^2} - 3 = \frac{6(p(p+1)(p-2q) + q(q+1)(q-2p))}{pq(p+q+2)(p+q+3)}, \quad (13)$$

– für den Modalwert

$$M = \frac{p-1}{p+q-2}, \quad p \geq 1, \quad q \geq 1, \quad p+q > 2 \quad (14)$$

– für  $p$  und  $q$

$$p = \frac{m_1^2 - m_1 m_2}{\mu_2}, \quad (15)$$

$$q = \frac{\mu_2}{m_1^3} p(p + m_1). \quad (16)$$

Im Sinne der Schätztheorie sind die Stichprobenmomente Punktschätzungen für die entsprechenden (theoretischen) Momente. Ersetzt man demzufolge in den Gleichungen (9) bis (13) und in den Gleichungen (15) und (16) die (theoretischen) Momente durch die empirischen Momente, so erhält man konkrete Punktschätzungen.

Die empirischen Momente  $k$ -ter Ordnung ergeben sich zu:

$$\begin{array}{ll} \text{(für gruppierte Daten)} & \text{(für ungruppierte Daten)} \\ \hat{m}_k = \frac{1}{n} \sum_{i=1}^m x_i^k n_i & = \frac{1}{n} \sum_{i=1}^n x_i^k, \end{array} \quad (17)$$

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^m (x_i - \bar{x})^k n_i = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, \quad (18)$$

wobei gilt

$x_i$  — Stichprobenwert

$n_i$  — Häufigkeit von  $x_i$  ( $i = 1(1) m$ )

$m$  — Anzahl der voneinander verschiedenen  $x_i$ -Werte

$\bar{x}$  — arithmetischer Mittelwert

$n$  — Stichprobenumfang.

Die Forderung nach hoher Effizienz ist der Grund dafür, weshalb *Koch* [6] Maximum-Likelihood-Schätzungen für Parameter der Betaverteilung anwendet. Bei diesem Schätzverfahren treten jedoch Schwierigkeiten bei der Berechnung von Schätzwerten auf, die ebenda mit Hilfe von elektronischen Rechenanlagen gelöst werden.

Zur Berechnung von  $\hat{a}$  und  $\hat{b}$  ist folgendes zu bemerken:  $\hat{p}$  und  $\hat{q}$  lassen sich nach der Transformation von  $Y$  in die auf  $(0, 1)$  verteilte Zufallsgröße  $X$  berechnen, nachdem auch  $\hat{a}$  und  $\hat{b}$  bestimmt worden sind. Für die pessimistische Schätzung [3, S. 47] von  $Y$ , die untere Grenze  $\hat{a}$ , kann der Stichprobenminimalwert bzw. die Reduktionslage (bei der Einteilung in Klassen) und für die optimistische Schätzung [3, S. 47], die obere Grenze  $\hat{b}$ , kann der Stichprobenmaximalwert bzw. der theoretische Maximalwert (der sich nach Einteilung in Klassen ergibt) angenommen werden.

Die Parameter  $a$  und  $b$  bestimmen die Lage der Funktion (1), während  $p$  und  $q$  Formparameter der Funktion (1) bzw. (2) sind. Die Dichtefunktionskurve der Betaverteilung 1. Art ist unimodal für  $p, q > 1$ , U-förmig für  $p, q < 1$  und J-förmig für  $p < 1, q \geq 1$  oder umgekehrt [5, S. 2]. Für  $1 < p < 2$  (bzw.  $1 < q < 2$ ) besitzt die Funktionskurve der Dichte im linken (bzw. im rechten) Endpunkt des Definitionsintervalls eine senkrechte Tangente, für  $p > 2$  (bzw.  $q > 2$ ) wird die Verteilungsfunktionskurve im linken (bzw. im rechten) Randpunkt des Definitionsbereiches zusammen mit ihrer Ableitung null [3, S. 48, 49], und für  $p = 1, q = 1$  ergibt sich eine stetige gleichmäßige Verteilung über  $(a, b)$  bzw.  $(0, 1)$  [11, S. 22].

### 3. Algorithmische Aufbereitung der Betaverteilung 1. Art zur rechnergestützten Anwendung

Soll mit empirischem Datenmaterial ein Anpassungstest an die Betaverteilung 1. Art vorgenommen werden, dann ist die Berechnung der Verteilungsfunktion erforderlich. Die Anwendung von Tabellen dieser Verteilungsfunktion wird dadurch erschwert, daß man erstens nicht ohne umfangreiche Interpolationen auskommt und zweitens bei der Berechnung auf einer elektronischen Rechenanlage die Speicherung großer Tabellenwerke meistens nicht realisieren kann. Um diese Schwierigkeit zu umgehen, wurde in [7, S. 97 bis 131] ein Algorithmus und, darauf aufbauend, ein Programm entwickelt, das — auch für nicht ganzzahlige Parameterwerte — mit verhältnismäßig geringem Rechenaufwand und ausreichender Genauigkeit die Verteilungsfunktion der Betaverteilung 1. Art berechnet. Dieser Algorithmus stützt sich auf die in [5] und in [12] entwickelte grundlegende Idee der Variablensubstitution und der Reihenentwicklung, die es eingehend zu behandeln und auszubauen galt. Nach der Variablensubstitution  $X = Z/(1 + Z)$  kann man die Verteilungsfunktion (3) in der Form

$$F(x; p, q) = \int_0^{\frac{x}{1-x}} \frac{Z^{p-1}(1+Z)^{-(p+q)}}{B(p, q)} dZ \quad \text{für } 0 < x < 1 \quad (19)$$

darstellen. Die Durchführung der partiellen Integration

$$\int u \cdot v' dx = uv - \int v u' dx$$

liefert

$$u \triangleq (1+Z)^{-(p+q)}/B(p, q)$$

$$v' \triangleq Z^{p-1}$$

$$uv \triangleq \frac{(1+Z)^{-p+q} Z^p}{pB(p, q)} \Big|_0^x = \frac{(1-x)^q x^p}{pB(p, q)}$$

$$-\int v u' dx \triangleq + \int_0^x \frac{(p+q) Z^p (1+Z)^{-(p+q+1)}}{pB(p, q)} dZ$$

$$= + \int_0^x \frac{Z^p (1+Z)^{-(p+q+1)}}{B(p+1, q)} dZ$$

$$F(x; p, q) = \frac{x^p (1-x)^q}{pB(p, q)} + \int_0^x \frac{Z^p (1+Z)^{-(p+q+1)}}{B(p+1, q)} dZ$$

$$= \frac{x(1-x)}{p} f(x; p, q) + F(x, p+1, q)$$

$$F(x; p, q) = \sum_{i=0}^{\infty} \frac{x(1-x)}{p+i} f(x; p+i, q). \quad (20)$$

Diese Funktionenreihe (20) ist zunächst hinsichtlich ihres Konvergenzverhaltens zu untersuchen.

$a_i$  sei das  $(i+1)$ -te Glied der Reihe.

$$a_i(x; p+i, q) = \frac{x(1-x)}{p+i} \cdot \frac{x^{p+i-1} (1-x)^{q-1}}{B(p+i, q)}$$

$$\begin{aligned} a_{i+1}(x; p+i+1, q) &= \frac{x(1-x)}{p+i+1} \cdot \frac{x^{p+i+1-1} (1-x)^{q-1}}{B(p+i+1, q)} \\ &= a_i(x; p+i, q) \cdot \frac{x(p+i+q)}{p+i+1} \end{aligned}$$

In anderer Schreibweise ergibt sich die Rekursionsformel

$$\frac{x(1-x)}{p+i+1} f(x; p+i+1, q) = \frac{x(1-x)}{p+i} f(x; p+i, q) \cdot \frac{x(p+i+q)}{p+i+1}. \quad (21)$$

Der Konvergenzradius  $\varrho (|x| < \varrho)$  der Reihe (20) läßt sich laut [1, S. 257] nach der Formel  $\frac{1}{\varrho} = \lim_{i \rightarrow \infty} \frac{|a_{i+1}|}{|a_i|}$  ermitteln und berechnet sich für die genannte Potenzreihe aus  $\frac{1}{\varrho} = \lim_{i \rightarrow \infty} \frac{p+i+q}{p+i+1} = 1$  zu  $\varrho = 1$ .

$\varrho = 1$  bedeutet, daß die Reihe absolut für alle  $x$ -Werte, die absolut genommen kleiner als 1 sind, konvergiert. Da nur mit der über  $(0, 1)$  verteilten Zufallsgröße und an dieser Stelle mit der Dichte- bzw. Verteilungsfunktion im Bereich  $0 < x < 1$  ge-

arbeitet wird, kann die Schlußfolgerung gezogen werden, daß die Reihe (20) stets konvergiert.

Die gemeinsame Grundlage für die Berechnung von Dichte- und Verteilungsfunktion der Beta-Verteilung 1. Art bildet das erste Glied der Reihenentwicklung (20)

$$\begin{aligned} a_0 &= \frac{x(1-x)}{p} f(x; p, q) = \frac{x^p(1-x)^q}{p B(p, q)} = \frac{x^p(1-x)^q \Gamma(p+q)}{p \Gamma(p) \Gamma(q)} \\ &= \frac{x^p(1-x)^q \Gamma(p+q)}{\Gamma(p+1) \Gamma(q)}, \end{aligned} \quad (22)$$

aus welchem mit Hilfe der Rekursionsformel (21) die nachfolgenden Glieder für  $i = 1, 2, \dots$  leicht berechnet werden können. Das letzte Glied der Reihenentwicklung sei

$$a_k = \frac{x(1-x)}{p+k} f(x; p+k, q). \quad (23)$$

Da bei der numerischen Rechnung die Reihenentwicklung zwangsläufig einmal abgebrochen werden muß, gilt es, ein Abbruchkriterium zu bestimmen. Der Fehler beim Weglassen der restlichen Glieder beträgt

$$\varepsilon = \sum_{i=k+1}^{\infty} a_i.$$

Wegen (20):

$$\begin{aligned} \sum_{i=k+1}^{\infty} a_i &= \frac{x(1-x)}{p+k+1} f(x; p+k+1, q) + \frac{x(1-x)}{p+k+2} f(x; p+k+2, q) + \dots \\ &\quad + \frac{x(1-x)}{p+k+j} f(x; p+k+j, q) + \dots \end{aligned}$$

und wegen (21):

$$\begin{aligned} \sum_{i=k+1}^{\infty} a_i &= \frac{x(1-x)}{p+k} f(x; p+k, q) \cdot \frac{x(p+k+q)}{p+k+1} \\ &\quad + \frac{x(1-x)}{p+k} f(x; p+k, q) \cdot \frac{x(p+k+q)}{p+k+1} \cdot \frac{x(p+k+1+q)}{p+k+2} \\ &\quad + \dots + \frac{x(1-x)}{p+k} f(x; p+k, q) \cdot \frac{x(p+k+q)}{p+k+1} \cdot \dots \cdot \frac{x(p+k+j-1+q)}{p+k+j} \end{aligned}$$

gilt

$$\begin{aligned} \varepsilon = \sum_{i=k+1}^{\infty} a_i &= a_k \left( \frac{x(p+k+q)}{p+k+1} + \frac{x(p+k+q)}{p+k+1} \cdot \frac{x(p+k+1+q)}{p+k+2} + \dots \right. \\ &\quad \left. + \frac{x(p+k+q)}{p+k+1} \cdot \dots \cdot \frac{x(p+k+j-1+q)}{p+k+j} + \dots \right) \end{aligned}$$

und damit

$$\varepsilon = \sum_{i=k+1}^{\infty} a_i = a_k \sum_{i=1}^{\infty} \prod_{j=1}^i x \frac{p+k-1+j+q}{p+k+j}. \quad (24)$$

Es kann eine Abgrenzung des Fehlers  $\varepsilon$  nach oben hin erfolgen. Wegen

$$\frac{p+k+q}{p+k} = 1 + \frac{q}{p+k} > 1 + \frac{q-1}{p+k+j} = \frac{p+k-1+j+q}{p+k+j}$$

ist eine Begrenzung durch

$$\varepsilon < \varepsilon_1 = a_k \sum_{i=1}^{\infty} \left( x \frac{p+k+q}{p+k} \right)^i$$

möglich. Die Reihe  $\varepsilon_1$  ist eine geometrische Reihe mit  $q_1 = x \frac{p+k+q}{p+k}$ , wobei  $q_1$  der Quotient zweier aufeinanderfolgender Glieder mit dem ersten Glied als Divisor und dem folgenden Glied als Dividend ist. Die Reihensumme der unendlichen geometrischen Reihe  $\varepsilon_1$  beträgt

$$\varepsilon_1 = \frac{a_{k+1}}{1 - q_1} = \frac{a_k}{\frac{p+k}{xp + xk + xq} - 1} \quad (25)$$

Der Konvergenzradius dieser Reihe errechnet sich aus

$$\frac{1}{\varrho} = \lim_{i \rightarrow \infty} \frac{|a_{i+1}|}{|a_i|} = \frac{p+k+q}{p+k}$$

zu

$$\varrho = \frac{p+k}{p+k+q}$$

Die Reihe  $\varepsilon_1$  konvergiert absolut für alle  $x$ -Werte, die absolut genommen kleiner als der Konvergenzradius  $\varrho$  sind:  $|x| < \varrho$ .

Es gilt

$$0 < x < 1,$$

woraus das vereinfachte Konvergenzkriterium

$$x < \varrho$$

und damit

$$x < \frac{p+k}{p+k+q}$$

$$x(p+k+q) < p+k$$

folgt.

Wegen

$p, q > 0$  und  $k =$  positiv ganzzahlig

ergibt sich weiter

$$x(p+q) < p$$

$$x < \frac{p}{p+q} \quad (26)$$

Es kann also festgestellt werden, daß der Fehler  $\varepsilon$  nach oben durch  $\varepsilon_1$  abgegrenzt werden kann, wobei  $x < \frac{p+k}{p+k+q}$  vorausgesetzt ist. Das ist mit Sicherheit

dann der Fall, wenn  $x < \frac{p}{p+q}$  gilt.

Für die Fehlerbestimmung im Falle  $x \geq \frac{p}{p+q}$  muß ein anderer Weg der Reihenentwicklung angewandt werden, um eine entsprechende Abbruchbedingung finden zu können. Aus diesem Grunde sollen die Symmetrieeigenschaften der Dichtefunktion (2) und der Verteilungsfunktion (3) untersucht werden.

Infolge Symmetrie gilt

$$f(X; p, q) = f(1 - X; q, p)$$

$$F(x; p, q) = 1 - F(1 - x; q, p)$$

$$\begin{aligned} 1 - F(x; p, q) &= \sum_{i=0}^{\infty} \frac{(1-x)x}{q+i} f(1-x; q+i, p) \\ &= \sum_{i=0}^{\infty} \frac{x(1-x)}{q+i} f(x; p, q+i). \end{aligned} \quad (27)$$

Analog der Reihe (20) liefert für die Reihe (27) die Untersuchung des Konvergenzverhaltens folgendes Ergebnis:

$$\begin{aligned} a_i(x; p, q+i) &= \frac{x(1-x)}{q+i} \cdot \frac{x^{p-1}(1-x)^{q+i-1}}{B(p, q+i)} \\ a_{i+1}(x; p, q+i+1) &= \frac{x(1-x)}{q+i+1} \cdot \frac{x^{p-1}(1-x)^{q+i}}{B(p, q+i+1)} \\ &= a_i(x; p, q+i) \cdot \frac{(1-x)(p+i+q)}{q+i+1}. \end{aligned}$$

Bei Umformung ergibt sich die Rekursionsformel

$$\frac{x(1-x)}{q+i+1} f(x; p, q+i+1) = \frac{x(1-x)}{q+i} f(x; p, q+i) \cdot \frac{(1-x)(p+i+q)}{q+i+1}. \quad (28)$$

Der Konvergenzradius der Potenzreihe (28) läßt sich aus  $\frac{1}{\varrho} = \lim_{i \rightarrow \infty} \frac{|-p-i-q|}{q+i+1}$  berechnen und beträgt  $\varrho = 1$ . Daraus folgt, daß die Reihe absolut für alle  $x$ -Werte, die der Ungleichung  $|x-1| < 1$  genügen, konvergiert. Wegen  $0 < x < 1$  ist festzustellen, daß die Reihe stets konvergiert. Somit kann die Reihenentwicklung (27) zur Berechnung der Verteilungsfunktion der Beta-Verteilung 1. Art benutzt werden. Das erste Glied dieser Reihe bildet

$$a_0 = \frac{x(1-x)}{q} f(x; p, q) = \frac{x^p(1-x)^q \Gamma(p+q)}{\Gamma(q+1) \Gamma(p)}, \quad (29)$$

und das letzte Glied sei

$$a_k = \frac{x(1-x)}{q+k} f(x; p, q+k). \quad (30)$$

Der Fehler beim Weglassen der restlichen Glieder beträgt dann

$$\varepsilon = \sum_{i=k+1}^{\infty} a_i = a_k \sum_{i=1}^{\infty} \prod_{j=1}^i (1-x) \frac{p+q+k+j-1}{q+k+j}. \quad (31)$$

Analog der Fehlerberechnung und Fehlerbegrenzung bei der Reihenentwicklung (20) kann auch hier eine Begrenzung des Fehlers  $\varepsilon$  nach oben hin erfolgen:

$$\begin{aligned} \varepsilon < \varepsilon_1 &= a_k \sum_{i=1}^{\infty} \left( (1-x) \frac{p+q+k}{q+k} \right)^i \\ &= \frac{a_k}{\frac{q+k}{p+q+k-xp-xq-xk} - 1}. \end{aligned} \quad (32)$$

Die Berechnung des Konvergenzradius der Reihe  $\varepsilon_1$  ergibt

$$\varrho = \frac{-q-k}{p+q+k}.$$



Falls  $|x - 1| < \rho$  bzw. wegen  $0 < x < 1 \rightarrow x - 1 > \rho$  und damit

$$x - 1 > \frac{-q - k}{p + q + k}$$

$$x > \frac{p}{p + q}$$

gilt, konvergiert diese Reihe  $\varepsilon_1$ .

Mit (20) und (26) bzw. mit (27) und (33) liegen somit zwei Möglichkeiten der Reihenentwicklung und zwei Möglichkeiten der Fehlerabschätzung vor, die jeweils in Abhängigkeit von den konkreten  $x$ -Werten anzuwenden sind. Noch offen ist der Fall,

wenn  $x = \frac{p}{p + q}$  ist. Da die Reihe in den Endpunkten des Konvergenzintervalls entweder konvergent oder divergent sein kann [1, S. 257], soll untersucht werden, ob die Reihe  $\varepsilon_1$ , die für  $x > \frac{p}{p + q}$  konvergiert, in ihrem Endpunkt  $x = \frac{p}{p + q}$  Konvergenz oder Divergenz aufweist.

Aus

$$x = \frac{p}{p + q}$$

folgt der Konvergenzradius

$$\begin{aligned} \rho &= (1 - x) \frac{p + q + k}{q + k} = \frac{p + q + k}{p + k} - \frac{p}{p + q} \cdot \frac{p + q + k}{q + k} \\ &= \frac{qp + qq + qk}{qp + qq + qk + kp} < 1. \end{aligned}$$

Durch Einsetzen von  $\frac{p}{p + q}$  in die Funktionenreihe erhält man eine numerische Reihe, die konvergiert, da der Konvergenzradius  $\rho = \lim_{i \rightarrow \infty} \frac{a_{i+1}}{a_i}$  kleiner als eins ist [1, S. 251]. Mithin ist bewiesen, daß die Reihenentwicklung (20) für den Fall  $x < \frac{p}{p + q}$  und die Reihenentwicklung (27) für die Fälle  $x \geq \frac{p}{p + q}$  angewendet werden kann.

#### 4. Die Handhabung der Gammafunktion für die Berechnungen der Beta-Verteilung I. Art

Die Berechnung der Gammafunktion kann nach vier Möglichkeiten erfolgen:

- Tabellenablesung,
- Anwendung der Rekursionsformen  $\Gamma(x + 1) = x\Gamma(x)$ ,
- Berechnung nach der *Stirlingschen* Formel

$$\Gamma(x + 1) = x! \approx \left(\frac{x}{e}\right)^x \sqrt{2\pi x} \left(1 + \frac{1}{12x} + \frac{1}{288x^2} + \dots\right),$$

- Zugrundelegung der Approximationsformel nach *Hastings* [4, S. 155].

Während die 1. und die 2. Möglichkeit umfangreiche Einlesearbeit am Automaten erfordert und diese Varianten somit nur bei kleinen  $x$ -Werten anzuwenden sind, ist die 3. Berechnungsmethode nur für größere Werte geeignet. Die Approximation nach *Hastings* ist dagegen sowohl bei kleinen als auch bei größeren Werten anwendbar. Die beiden ersten Verfahren liefern Fehler durch Interpolation. Bei der 3. und 4. Art

der Berechnung treten Näherungsfehler auf, wobei die Fehlerkurve der Approximationskurve nach *Hastings*, die in [4, S. 155] ausführlich dargestellt ist, maximale Fehler von  $\pm 0,00005$  aufweist. Die Anwendbarkeit der 4. Methode für jede Art von  $x$ -Werten, die niedrigen Fehlerwerte und die gute Programmierbarkeit sind Gründe dafür, weshalb diese Variante der Berechnung der Gammafunktion zugrunde gelegt wird.

Nach [1, S. 59] gilt

$$\Gamma(x) = \frac{\Gamma(x+1)}{x} \text{ und } \Gamma(x) = (x-1) \Gamma(x-1).$$

[4, S. 155] liefert für den Bereich  $0 \leq x \leq 1$  für  $\Gamma(1+x)$  folgende Approximationsformel:

$$\Gamma^*(x+1) = 1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5, \quad (34)$$

mit

$$a_1 = -0,5748646,$$

$$a_2 = 0,9512363,$$

$$a_3 = -0,6998588,$$

$$a_4 = 0,4245549,$$

$$a_5 = -0,1010678.$$

Unabhängig von den spezifischen Bedingungen der Beta-Verteilung 1. Art ( $0 < x < 1$ ) bildet folgender Algorithmus den Ausgangspunkt eines FUNCTION-Unterprogramms GAMMA, das für die Berechnungen der Beta-Verteilung 1. Art angewendet wird.

$G$  sei ein beliebig großer positiver  $x$ -Wert, den es in den Bereich  $0 \leq x \leq 1$  zu transformieren gilt,  $G_i$  ( $i = 1, 2, 3, \dots$ ) sei der um 1 verringerte  $G$ -Wert und  $G_n$  der im Bereich  $0 \leq x \leq 1$  liegende Wert. Somit wird

$$\Gamma(G) = \Gamma(G_1 + 1) = G_1 \Gamma(G_1)$$

$$G_1 \Gamma(G_1) = G_1 \Gamma(G_2 + 1) = G_1 G_2 \Gamma(G_2)$$

$\vdots$

$$G_1 \cdot \dots \cdot G_{n-1} \Gamma(G_{n-1}) = G_1 \cdot \dots \cdot G_{n-1} \Gamma(G_n + 1) =$$

$$= G_1 \cdot \dots \cdot G_n \Gamma(G_n)$$

$\rightarrow$

$$\Gamma(G) = G_1 \cdot \dots \cdot G_{n-1} \Gamma(G_n + 1) = \frac{G_1 \cdot \dots \cdot G_n}{G_n} \Gamma(G_n + 1)$$

$$= \frac{G_1 \cdot \dots \cdot G_n}{G_n} (1 + a_1 G_n + a_2 G_n^2 + a_3 G_n^3 + a_4 G_n^4 + a_5 G_n^5).$$

## 5. Bemerkungen zu den aufgestellten FORTRAN IV-Programmen

In [7, S. 101 bis 104, 117 bis 122, 124 bis 131] werden die Hauptprogramme B 1, B 2, B 3 und zwei Unterprogramme zur Berechnung der Beta-Verteilung 1. Art aufgestellt und beschrieben.

Das Rechenprogramm B 1 dient der Transformation der Zufallsgröße  $Y$  in  $X$ , der Momentenberechnung und schließlich der Parameterbestimmung. Im FORTRAN-Programm B 2, das als Hauptprogramm zur Durchführung des *Kolmogorov*-Tests dient, ist als Unterprogramm die Beta-Verteilung 1. Art eingefügt. Das Unterprogramm SUBROUTINE BETAV 1 (Tafel 1) ruft das FUNCTION-Unterprogramm GAMMA auf, das zur Berechnung der Gammafunktion, auf der Grundlage der Approximationsformel nach *Hastings*, dient. Das FORTRAN-Hauptprogramm B 3 (Tafel 2) stellt das eigentliche Rechenprogramm der Beta-Verteilung 1. Art dar. Während B 2 in jedem Fall nach der Rechnung B 1 abgearbeitet wird, sorgt ein Anfangstest in B 3

dafür, daß dieses Programm nur dann zur Abarbeitung kommt, wenn in B 2 der Ausdruck „MODELLANNAHME“ erfolgte. B 3, das auch wieder die SUBROUTINE BETAV 1 aufruft, dient dem Ausdruck der theoretischen Häufigkeiten für die in Klassen eingeteilten Werte.

Sämtliche Programme wurden am Minsk 22 mittels spezieller Steuerbefehle getestet.

Tafel 1 Maßgebende Schritte des Unterprogramms SUBROUTINE BETAV 1

```

:
C  BETAVERTeilUNG, BERECHNUNG DER DICHTe (DF) — UND
C  DER VERTEILUNGSFUNKTION (VF)
:
C  DF UND VF FUEr X.LE.0 UND FUEr X.GE.1
:
C  AUSWAHL DER GUENSTIGSTEN REIHENENTWICKLUNG
:
C  REIHENENTWICKLUNG NACH P
:
C  REIHENENTWICKLUNG NACH Q
:
C  BERECHNUNG DER DF UND DES ERSTEN REIHENGLIEDES
:
C  BERECHNUNG DER VERTEILUNGSFUNKTION

```

Tafel 2 Ausdruck des Hauptprogramms B3  
E K XGR SGR FK DF VF N×VF HK

Legende:

E	△	vorgegebener Wert, der von $\varepsilon_1$ zu erreichen bzw. zu unterschreiten ist
K	△	i, Zählgröße
XGR	△	Klassengrenze der Zufallsgröße Y
SGR	△	transformierte Klassengrenze der Zufallsgröße X
FK	△	Wert der kumulativen Häufigkeitsverteilung, empirische Häufigkeit
DF	△	Dichtefunktionswert
VF	△	Verteilungsfunktionswert
N	△	Stichprobenumfang
HK	△	theoretische Häufigkeit

## 6. Ausblick

An verbessernden rechentechnisch-organisatorischen Maßnahmen, die das Zusammenspiel der drei Hauptprogramme berühren und die sich u. a. mit einer Verringerung der Einlesearbeit, dem Ausdrucken weiterer Werte und einer eleganteren Programmzusammenarbeit der drei Hauptprogramme und der beiden Unterprogramme beschäftigen, wird zur Zeit vom Verfasser gearbeitet.

Bei der Errechnung des ersten Reihengliedes der Verteilungsfunktionsreihe sei darauf hingewiesen, daß die Funktion  $I(n)$  mit steigendem n sehr schnell anwächst. Es ist hierbei zu beachten, daß in Rechenanlagen nicht beliebig große Zahlen dargestellt werden können und daß somit eventuell eine Faktorenzerlegung vorgenommen werden muß.

## 7. Zusammenfassung

Da mancher Praktiker noch nicht genügend Gelegenheit hatte, sich mit der Betaverteilung 1. Art vertraut zu machen, soll mit diesem Beitrag ein Überblick hierüber sowie über bestimmte theoretische Probleme dieser Verteilung und deren Lösung gegeben werden. Gleichzeitig will der Autor mit diesem Artikel durch die Nutzung der elek-

tronischen Datenverarbeitung die Tätigkeit im Bereich der Forschung und Entwicklung intensivieren.

Nach der Definition der Beta-Verteilung 1. Art wird auf die algorithmische Aufbereitung der genannten Verteilung zur rechnergestützten Anwendung eingegangen. Die mit der Problematik in Verbindung stehende Gammafunktion wird dabei ebenfalls untersucht, und schließlich wird auf erarbeitete FORTRAN-IV-Programme (am Minsk 22 getestet) eingegangen.

#### Literatur

- [1] *Bronstein, I. N./Semendjajew, K. A.*: Taschenbuch der Mathematik für Ingenieure und Studenten der Technischen Hochschulen. Leipzig, BSB B. G. Teubner Verlagsgesellschaft, 1969, 10. Auflage.
- [2] *Fisz, M.*: Wahrscheinlichkeitsrechnung und mathematische Statistik. Berlin, VEB Deutscher Verlag der Wissenschaften, 1970, 5., erweiterte Auflage.
- [3] *Golenko, D. I.*: Statistische Methoden der Netzplantechnik. Leipzig, BSB B. G. Teubner Verlagsgesellschaft, 1972.
- [4] *Hastings, C.*: Approximation for digital computers (engl.). Princeton, New Jersey, Princeton University Press, 1955.
- [5] *Jöhnk, M.-D./Naeve, P./Nunner, O.*: Drei Arbeiten zur Beta- und Gamma-Verteilung, logarithmischen Normalverteilung und Verteilung des Spearman'schen Rangkorrelationskoeffizienten. Würzburg, Physica-Verlag, 1968.
- [6] *Koch, R.*: Maximum-Likelihood-Schätzungen von Parametern der Gamma- und Beta-Verteilung. In: Mathematische Operationsforschung und Statistik 6 (1974) 4/5, S. 333—355.
- [7] *Markstein, S.*: Betrachtungen zu Problemen der Güterwagenumlaufverkürzung, der Zugbildungsoptimierung und der Zweigruppenzugbildung. Diplomarb. Hochschule für Verkehrswesen „Friedrich List“ Dresden, 1976.
- [8] *Potthoff, G.*: Beta-verteilte Betriebsabläufe. In: DET — Die Eisenbahntechnik 21 (1973) 7, S. 317—319.
- [9] *Potthoff, G.*: Betriebstechnische Anwendung der Rayleigh-Verteilung. In: DET — Die Eisenbahntechnik 21 (1973) 10, S. 468—470.
- [10] *Sachs, L.*: Angewandte Statistik. Berlin, Heidelberg, New York, Springer Verlag, 1974, 4., neu bearbeitete und erweiterte Auflage der „Statistischen Auswertungsmethoden“ mit neuer Bibliographie.
- [11] Wahrscheinlichkeitsrechnung u. Mathematische Statistik. Lexikon der Stochastik. Berlin, Akademie-Verlag, 1975, 2., bearbeitete und wesentlich erweiterte Auflage.
- [12] Verfahrenorientiertes Programmpaket (VOPP) — STATISTIK für Robotron 21 (Algorithmen und Programme).

Eingegangen am 31. Januar 1977

Verfasser: Dipl.-Ing. *Sigrüd Markstein*, Sektion Technische Verkehrskybernetik, Wissenschaftsbereich Technologie des Eisenbahntransports, Hochschule für Verkehrswesen „Friedrich List“ Dresden.