

Arbeitsberichte des Instituts für Wirtschaftsinformatik

Herausgeber: Prof. Dr. J. Becker, Prof. Dr. H. L. Grob, Prof. Dr. S. Klein,
Prof. Dr. H. Kuchen, Prof. Dr. U. Müller-Funk, Prof. Dr. G. Vossen

Arbeitsbericht Nr. 74

**Die asymptotische Verteilung der Spannweite
bei Zufallsgrößen mit
paarweise identischer Korrelation**

Ingolf Terveer

ISSN 1438-3985

Institut für Wirtschaftsinformatik der Westfälischen Wilhelms-Universität Münster,
Leonardo-Campus 3, 48149 Münster
Februar 2002

Inhaltsverzeichnis

1	Einleitung	2
2	Modell und Grundannahmen	2
3	Methodische Grundlagen der asymptotischen Spannweitenverteilung	6
3.1	Die Verteilung der Spannweite im u.i.v-Fall	6
3.2	Einige Eigenschaften der Matrizen $M(a, b)$	7
3.3	Normalapproximation	13
3.4	Anderson´s Theorem	13
3.5	Ordnungseigenschaften der Kovarianzmatrix im Modell (G)	15
3.6	Abschätzung der Tails der Spannweitenverteilung im Modell (G)	16
3.7	Die approximative Verteilung der Spannweite im Modell (G)	18
4	Numerische Untersuchungen und Simulationen	19
4.1	Zur Schärfe der Anderson-Abschätzung	20
4.2	Multinomialexperimente	21
4.3	Urnenmodelle	25
4.4	Spannweitentest versus Chi-Quadrat-Test	28
5	Resümee	31
	Literaturverzeichnis	32
	Anhang: Fraktile zur asymptotischen Spannweitenverteilung	33

Zusammenfassung

Ein k -variates Einzelexperiment wird n -mal unabhängig wiederholt; die Ergebnisse werden summiert. Gesucht ist die Wahrscheinlichkeitsverteilung für die Spannweite der k Summen. Eine derartige Fragestellung ergibt sich etwa bei statistischen Tests zu Multinomial- oder Urnenexperimenten, wenn Alternativ-Verfahren zu den klassischen Chi-Quadrat-Tests gesucht werden. Die Untersuchung von Spannweitenverteilungen beschränkt sich allerdings in der Literatur vornehmlich auf den Fall, dass die der Spannweitenbildung zugrunde liegenden Zufallsvariablen stochastisch unabhängig und identisch verteilt sind. Im vorliegenden Arbeitsbericht werden einige Überlegungen für den Fall durchgeführt, dass eine einfache Abhängigkeitsstruktur innerhalb des Einzelexperimentes vorliegt. Es zeigt sich, dass asymptotisch bis aus einen Skalenfaktor die selbe Spannweitenverteilung wie im u.i.v-Fall vorliegt.

1 Einleitung

Die Spannweite von Daten ist als Instrument der statistischen Qualitätskontrolle verbreitet, was zum einen an ihrer leichten Berechenbarkeit liegt, zum anderen aber daran, dass sie in dem Standardansatz der Qualitätskontrolle – nämlich die Annahme normalverteilter Daten – der Stichprobenstandardabweichung hinsichtlich Effizienz nur wenig nachsteht. Für den Fall von u.i.v.-Modellen ist die Verteilung der Spannweite zudem aus derjenigen der Ordnungsstatistik leicht zu bestimmen. Im übrigen ist die Spannweite nur selten von Interesse, zumal sie eine extreme Reduktion der Daten bedeutet und anfällig gegen Ausreißer ist. Ein weiterer Grund wird die schwierige Berechnung von Spannweitenverteilungen sein, sobald die Unabhängigkeit oder die identische Verteilung der Daten nicht mehr gegeben ist.

In Modellen mit kategoriellen Merkmalen kann jedoch die Spannweite ein wichtiges Hilfsmittel zur Aufdeckung von Asymmetrien darstellen. Nimmt man beispielsweise an, dass alle Ausprägungen eines kategoriellen Merkmales gleich wahrscheinlich sind, so sollte bei unabhängiger Beobachtung dieses Merkmales jede Kategorie ungefähr gleich oft auftreten.¹ Eine besonders hohe Spannweite der Kategoriehäufigkeiten deutet also auf eine Verletzung der Gleichverteilungsannahme hin. Klassischerweise wird eine derartige Verletzung mittels Chi-Quadrat-Tests nachgeprüft. Diese verfügen aber nicht über eine zufriedenstellende Güte. Auf der Spannweite basierende Tests dienen hier als Alternative. Nachteilig ist bei diesen allerdings, dass die Nullverteilung, d.h. die Verteilung der Spannweite unter Annahme gleicher Kategorie-Wahrscheinlichkeiten nur auf dem Wege der – zumindest teilweisen – Enumeration der Elementarwahrscheinlichkeiten oder näherungsweise durch Simulation gewonnen werden können. Beide Wege sind extrem aufwendig, der erstere gar für realistische Größenordnungen nicht mehr durchführbar. Von Interesse sind daher geeignete Approximationen der Spannweitenverteilung für den Fall abhängiger Daten.

Im vorliegenden Arbeitsbericht wird auf Basis eines einfachen, aber generell gehaltenen Abhängigkeitsmodells, für das Multinomial- und Urnenmodelle mit gleichen Eintrittswahrscheinlichkeiten Instanzen sind, die Spannweitenverteilung approximiert.

2 Modell und Grundannahmen

Mit dem Zufallsvektor

$$X^{(m)} = \left(X_1^{(m)}, \dots, X_k^{(m)} \right) \in \mathbb{R}^k, \quad m = 1, \dots, n$$

¹Gleiches gilt in Urnenmodellen mit identischen Kugelhäufigkeiten

sei das Ergebnis des m -ten Einzelexperimentes bezeichnet. Wir nehmen im folgenden an, dass $X = (X_1, \dots, X_k), X^{(1)}, \dots, X^{(n)}$ dem Modell (G) genügen, d.h. dass folgende beiden Eigenschaften erfüllt sind:

- (G1) $X, X^{(1)}, \dots, X^{(n)}$ sind jeweils stochastisch unabhängig und besitzen dieselbe Verteilung.
- (G2) die ZV X_1, \dots, X_k besitzen dieselbe Varianz $\sigma^2 := \text{Var}(X_1)$ und paarweise stets dieselbe Korrelation $\rho := \text{Corr}(X_i, X_j) \in [-1; 1]$ für $i \neq j$.

Anmerkung 2.1 Wir wollen im folgenden den Extremfall $\rho = 1$ ausschließen, da in diesem Falle $X_i - X_j$ konstant und die weiter unten eingeführte Spannweitenverteilung ausgeartet wäre.

Da die Korrelationsmatrix von X positiv semidefinit ist, folgt mit Lemma 3.1e) (s.u.), dass $\rho \geq -\frac{1}{k-1}$.

Beispiele für Verteilungen mit der Eigenschaft (G2) sind etwa Multinomialverteilungen mit identischen Zellwahrscheinlichkeiten oder multivariate hypergeometrische Verteilungen, siehe Beispiele 2.2 und 2.3. In diesen Fällen wird die extremale Korrelation $\rho = -\frac{1}{k-1}$ realisiert.

Auch im Fall austauschbarer X_1, \dots, X_k ist die Eigenschaft (G2) erfüllt. ■

Im folgenden sei für $a, b \in \mathbb{R}$ und $k \in \mathbb{N}$

$$M(a, b) := \begin{bmatrix} a & b & \cdots & b & b \\ b & a & & b & b \\ \vdots & & \ddots & \vdots & \\ b & \cdots & b & a & b \\ b & b & \cdots & b & a \end{bmatrix} = (a - b)I_k + b\mathbf{1}_{k \times k}$$

Dabei bezeichnet I_k die $k \times k$ -Einheitsmatrix und $\mathbf{1}_{k \times k}$ die $k \times k$ -Matrix mit lauter Einträgen = 1. Unter den Annahmen (G1) und (G2) besitzt X die Kovarianzmatrix $\sigma(X) = \sigma^2 M(1 - \rho, \rho)$, d.h. die Korrelationsmatrix

$$M(1 - \rho, \rho) = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho & \rho \\ \rho & 1 & \rho & \cdots & \rho & \rho \\ \rho & \rho & 1 & \cdots & \rho & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho & \rho & \rho & \cdots & 1 & \rho \\ \rho & \rho & \rho & \cdots & \rho & 1 \end{bmatrix} = (1 - \rho)I_k + \rho\mathbf{1}_{k \times k}$$

Setzt man nun $S^{(n)} = (S_1^{(n)}, \dots, S_k^{(n)})$ als

$$S^{(n)} := \sum_{m=1}^n X^{(m)}$$

so ist die Spannweite $W^{(n)}$ erklärt als

$$W^{(n)} := \max_i(S_i^{(n)}) - \min_i(S_i^{(n)}) = \max_{i \neq j} |S_i^{(n)} - S_j^{(n)}|$$

Beispiel 2.2 : Ein Multinomialmodell wird durch die diskreten Wahrscheinlichkeiten

$$\mathbb{P}(X = x) = \frac{r!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

für $x = (x_1, \dots, x_k) \in \mathcal{M}_r^{k,l} := \{(x_1, \dots, x_k) \in \mathbb{N}_0^k : \sum_i x_i = r\}$ festgelegt. Dabei bezeichnen $p_1, \dots, p_k \geq 0$ die Wahrscheinlichkeiten bei einem einmaligen Experiment ein Ergebnis aus der Kategorie C_1, \dots, C_k zu erzielen. Es ist $p_1 + \dots + p_k = 1$. Das Einzelexperiment wird r -mal unabhängig wiederholt (Bezeichnung für die Verteilung: $\mathcal{M}(r, p_1, \dots, p_k)$ bzw. $\mathcal{M}(r, p)$ für $p = (p_1, \dots, p_k)$)

Hier lauten die Momente (siehe [2], S. 33f):

$$\mathbb{E}(X_i) = rp_i$$

$$\text{Var}(X_i) = rp_i(1 - p_i)$$

sowie für $i \neq j$

$$\sigma(X_i, X_j) = -rp_i p_j$$

$$\rho(X_i, X_j) = -\sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}}$$

Falls alle Kategorien gleich wahrscheinlich sind, d.h. $p_i = \frac{1}{k}$ so ergibt sich

$$\mathbb{E}(X_i) = \frac{r}{k}$$

$$\text{Var}(X_i) = \frac{r(k-1)}{k^2}$$

$$\sigma(X_i, X_j) = -\frac{r}{k^2}$$

$$\rho(X_i, X_j) = -\frac{1}{k-1}$$

und die Voraussetzungen (G1), (G2) sind erfüllt mit der Kovarianzmatrix

$$\sigma(X) = \frac{r(k-1)}{k^2} M\left(1, -\frac{1}{k-1}\right)$$

Beispiel 2.3 Im Urnenmodell werden bei einem Einzelexperiment aus einer Urne, die mit $l = l_1 + \dots + l_k$ Kugeln gefüllt ist, von denen jeweils l_i Kugeln die Nummer i tragen, $i = 1, \dots, k$, insgesamt r Kugeln ohne/mit Zurücklegen gezogen. Mit

$$X^{(m)} = \left(X_1^{(m)}, \dots, X_k^{(m)} \right)$$

sei das Ziehungsergebnis der m -ten Wiederholung bezeichnet. Dabei gilt $X^{(m)} \in \mathcal{M}_r^{k,l} = \{ (x_1, \dots, x_k) \in \mathbb{N}_0^k : \sum_i x_i = r \}$

Falls jede gezogene Kugel sofort in die Urne zurückgelegt wird, so besitzt $X^{(m)}$ eine Multinomialverteilung $\mathcal{M}(r, p_1, \dots, p_k)$, wobei $p_i = \frac{l_i}{l}$.

Falls aber die Kugeln nicht zurückgelegt werden, so besitzt $X^{(m)}$ eine multivariate hypergeometrische Verteilung, d.h. es gilt für alle $x = (x_1, \dots, x_k) \in \mathcal{M}_r^{k,l}$:

$$\mathbb{P}(X = x) = \frac{\binom{l_1}{x_1} \dots \binom{l_k}{x_k}}{\binom{l}{r}}$$

Hierzu gehören die Kenngrößen (siehe [2]. S. 173)

$$\begin{aligned} \mathbb{E}(X_i) &= \frac{r}{l} l_i \\ \text{Var}(X_i) &= \frac{r(l-r)}{l^2(l-1)} l_i(l-l_i) \end{aligned}$$

sowie für $i \neq j$

$$\begin{aligned} \sigma(X_i, X_j) &= -\frac{r(l-r)}{l^2(l-1)} l_i l_j \\ \rho(X_i, X_j) &= -\sqrt{\frac{l_i l_j}{(l-l_i)(l-l_j)}} \end{aligned}$$

Somit ist die Voraussetzung (G2) erfüllt, falls $l_1 = \dots = l_k = \frac{l}{k}$. In diesem Fall gilt ähnlich wie bei der Multinomialverteilung (s.o.)

$$\begin{aligned} \text{Var}(X_i) &= \frac{r(k-1)}{k^2} \frac{l-r}{l-1} \\ \rho(X_i, X_j) &= -\frac{1}{k-1} \end{aligned}$$

und X besitzt die Kovarianzmatrix

$$\sigma(X) = \frac{r(k-1)}{k^2} \frac{l-r}{l-1} M\left(1, -\frac{1}{k-1}\right)$$

Beispiel 2.4 (Abhängigkeit durch Kontrollvariate) Es seien U_1, \dots, U_k und V stochastisch unabhängige Zufallsvariablen, U_1, \dots, U_k besitzen dieselbe Verteilung mit Varianz $\sigma_1^2 = \text{Var}(U_1) > 0$, und es sei $\sigma_2^2 := \text{Var}(V) > 0$. Dann erfüllen die Zufallsvariablen

$X_i := U_i + V$ die Voraussetzung (G2) und können daher ebenfalls als Grundmodell iteriert werden. Es gilt nämlich

$$\text{Var}(X_i) = \sigma_1^2 + \sigma_2^2$$

sowie für $i \neq j$

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \text{Cov}(U_i, U_j) + \text{Cov}(U_i, V) + \text{Cov}(V, U_j) + \text{Cov}(V, V) \\ &= \sigma_2^2 \end{aligned}$$

d.h.

$$\text{Corr}(X_i, X_j) = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \rho > 0$$

3 Methodische Grundlagen der asymptotischen Spannweitenverteilung

In diesem Abschnitt leiten wir die asymptotische Verteilung der reskalierten Spannweite

$$\frac{W^{(n)}}{\sigma \sqrt{n(1-\rho)}}$$

im Modell (G) her. Dabei untersuchen wir zunächst die Möglichkeiten einer Abschätzung der Verteilung auf Basis eines klassischen Theorems von Anderson. Später wird sich dann zeigen, dass die obere Schranke nach Anderson bereits mit der asymptotischen Verteilung übereinstimmt. Die Fraktile dieser Verteilung können dann heran gezogen werden, um statistische Tests auf Basis der Spannweite zu bilden, deren Null-Hypothesen gerade Modelle vom Typ (G) beschreiben. Wir erhalten somit im Multinomialmodell den Spannweitentest auf identische Kategoriewahrscheinlichkeiten und in Urnenmodellen mit identischen Kugelhäufigkeiten den Spannweitentest auf Vorliegen eines Laplace-Experimentes. Die wesentlichen Hilfsmittel bei der Herleitung der asymptotischen Verteilung sind der zentrale Grenzwertsatz, die Transformationseigenschaften in Normalverteilungsmodellen sowie Quadratwurzelzerlegungen von Korrelationsmatrizen vom Typ $M(a, b)$, die wir ausnutzen werden, um eine geeignete lineare Transformation ausgehend von u.i.v-Modellen durchzuführen, welche die Spannweitenverteilung nur um einen Skalenfaktor modifiziert.

3.1 Die Verteilung der Spannweite im u.i.v-Fall

Seien Y_1, \dots, Y_k stochastisch unabhängige, identisch verteilte Zufallsvariablen mit Dichte f , Verteilungsfunktion F . Weiter bezeichnen $Y_{(1)}, \dots, Y_{(k)}$ die (nach Wert aufsteigend)

geordneten Zufallsvariablen. Ein klassisches Resultat über Ordnungsstatistiken besagt, dass die Verteilung der Spannweite

$$W := Y_{(k)} - Y_{(1)} = \max_{i \neq j} (Y^{(i)} - Y^{(j)})$$

die Dichte

$$g(w) = k(k-1) \int_{-\infty}^{\infty} f(x)f(x+w) (F(x+w) - F(x))^{k-2} dx$$

für $w > 0$ besitzt. Da die Spannweite ein lineares Funktional der Ordnungsstatistik ist, ergibt sich die genannte Dichte beispielsweise aus der Randverteilung $\mathcal{L}(Y_{(1)}, Y_{(k)})$ der Ordnungsstatistik $(Y_{(1)}, \dots, Y_{(k)})$, deren Dichte die Gestalt $f_{Y_{(1)}, Y_{(k)}}(y_1, y_k) = k(k-1)(F(y_k) - F(y_1))^{k-2} f(y_1)f(y_k)$ hat.² Durch Integration erhält man die Verteilungsfunktion der Spannweite für $w > 0$ als

$$G(w) = k \int_{-\infty}^{\infty} f(x) (F(x+w) - F(x))^{k-1} dx$$

Für den Spezialfall, dass die Y_i standardnormalverteilt sind, ist die Verteilungsfunktion G vielfach vertafelt worden, z.B. in [3] bis $n = 20$.

Falls die Zufallsvariablen Y_i nicht stochastisch unabhängig sind, sind derartige Ausdrücke für die Verteilungsfunktion der Spannweite nicht bekannt.

3.2 Einige Eigenschaften der Matrizen $M(a, b)$

Wir fassen nachstehend einige Eigenschaften der Menge $\mathcal{M} = \{M(a, b) : a, b \in \mathbb{R}\} \subset \mathbb{R}^{k \times k}$ zusammen: Hierzu sei $\Delta(x) := xI_k$ die Diagonalmatrix mit sämtlichen Einträgen identisch x auf der Hauptdiagonale, $x \in \mathbb{R}$.

Lemma 3.1 *Die Menge \mathcal{M} ist hinsichtlich Matrixaddition und Matrixmultiplikation eine \mathbb{R} -Unteralgebra von $\mathbb{R}_{sym}^{k \times k}$. Insbesondere haben die Matrizen in \mathcal{M} die folgende Eigenschaften (für $a, b, c, d, x \in \mathbb{R}$)*

a) $M(a, b)M(c, d) = M(ac + (k-1)bd, ad + bc + (k-2)bd)$

b) Für $M_1, M_2 \in \mathcal{M}$ gilt $M_1M_2 = M_2M_1$

c) $\det(M(a, b)) = (a + (k-1)b)(a-b)^{k-1}$

d) $M(a, b) - \Delta(x) = M(a-x, b)$

²Zur Verteilung der Ordnungsstatistik und zu deren Randverteilungen cf. [5], 562ff.

e) $M(a, b)$ hat die Eigenwerte $a - b$ (mit Vielfachheit $k - 1$) und $a + (k - 1)b$, ist also insbesondere positiv semidefinit für $a \geq b \geq -\frac{a}{k-1}$

f) Falls $a \neq b$ und $a \neq -(k - 1)b$, so ist $M(a, b)$ invertierbar mit $M(a, b)^{-1} = M(c, d)$, wobei

$$c = \frac{a + (k - 2)b}{(a - b)(a + (k - 1)b)}$$

$$d = \frac{a}{(b - a)(a + (k - 1)b)}$$

g) Falls $d \neq 0$ und $c \geq d \geq -\frac{c}{k-1}$, so hat $M(c, d)$ eine Quadratwurzel in \mathcal{M} ; genauer gilt für

$$b = \frac{\pm\sqrt{c-d} \pm \sqrt{c+(k-1)d}}{k}$$

$$a = \frac{d - (k-2)b^2}{2b}$$

die Gleichung $M(a, b)^2 = M(c, d)$. Insbesondere gilt dann $(a - b)^2 = c - d$ und $2ab + (k - 2)b^2 = d$

Beweis. a) und b) ergeben sich durch einfaches Nachrechnen; c) folgt, da $M(a, b)$ durch Anwendung elementarer Zeilen- und Spaltenumformungen auf die Gestalt

$$\begin{bmatrix} a + (k - 1)b & b & \cdots & b & b \\ a + (k - 1)b & a & & b & b \\ \vdots & & \ddots & \vdots & \\ a + (k - 1)b & \cdots & b & a & b \\ a + (k - 1)b & b & \cdots & b & a \end{bmatrix} \rightarrow \begin{bmatrix} a + (k - 1)b & b & \cdots & b & b \\ 0 & a - b & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & a - b & 0 \\ 0 & 0 & \cdots & 0 & a - b \end{bmatrix}$$

gebracht werden kann, ohne dass sich die Determinante ändert. d) ist sofort offensichtlich. Aus d) und c) ergeben sich sofort die Eigenwerte von $M(a, b)$, d.h. es folgt e). Zur Berechnung der inversen Matrix in f) kann man unmittelbar [5], Anmerkung 5.131a) (Umkehrformel) verwenden. Für einen direkten Beweis setzt man an

$$M(a, b)M(c, d) = M(1, 0)$$

$$\Leftrightarrow M(ac + (k - 1)bd, ad + bc + (k - 2)bd) = M(1, 0)$$

$$\Leftrightarrow ac + (k - 1)bd = 1, ad + bc + (k - 2)bd = 0$$

Subtrahiert man beim zuletzt erhaltenen Gleichungssystem die zweite von der ersten Gleichung, so erhält man die Gleichung

$$(a - b)(c - d) = 1 \Leftrightarrow c - d = \frac{1}{a - b} \Leftrightarrow c = d + \frac{1}{a - b}$$

Eingesetzt in die erste Gleichung erhält man

$$\begin{aligned} a \left(d + \frac{1}{a-b} \right) + (k-1)bd &= 1 \\ \Leftrightarrow (a + (k-1)b)d &= 1 + \frac{a}{b-a} = \frac{b}{b-a} \\ \Leftrightarrow d &= \frac{b}{(b-a)(a + (k-1)b)} \end{aligned}$$

Rücksubstitution liefert dann

$$\begin{aligned} c &= d + \frac{1}{a-b} \\ &= \frac{b}{(b-a)(a + (k-1)b)} + \frac{1}{a-b} \\ &= \frac{b - (a + (k-1)b)}{(b-a)(a + (k-1)b)} \\ &= \frac{a + (k-2)b}{(a-b)(a + (k-1)b)} \end{aligned}$$

Zur Berechnung der Quadratwurzel in g) setzt man an

$$\begin{aligned} M(a, b)^2 &= M(c, d) \\ \Leftrightarrow M(a^2 + (k-1)b^2, 2ab + (k-2)b^2) &= M(c, d) \\ \Leftrightarrow a^2 + (k-1)b^2 = c, 2ab + (k-2)b^2 &= d \\ \Leftrightarrow a^2 + (k-1)b^2 - 2ab - (k-2)b^2 = c - d, 2ab + (k-2)b^2 &= d \\ \Leftrightarrow (a-b)^2 = c - d, 2ab + (k-2)b^2 &= d \end{aligned}$$

Betrachtet man das zuletzt erhaltene Gleichungssystem und löst die zweite Gleichung nach a auf, so ergibt sich

$$a = \frac{d - (k-2)b^2}{2b}$$

und nach Einsetzen in die erste Gleichung

$$\begin{aligned} \left(\frac{d - (k-2)b^2}{2b} - b \right)^2 &= c - d \Leftrightarrow \left(\frac{d - kb^2}{2b} \right)^2 = c - d \\ \Leftrightarrow \frac{d - kb^2}{2b} &= \pm \sqrt{c - d} \\ \Leftrightarrow b^2 \pm 2b \frac{\sqrt{c-d}}{k} &= \frac{d}{k} \\ \Leftrightarrow \left(b \pm \frac{\sqrt{c-d}}{k} \right)^2 &= \frac{d}{k} + \frac{c-d}{k^2} = \frac{c + (k-1)d}{k^2} \end{aligned}$$

Hieraus ergibt sich die Lösung

$$b = \pm \frac{\sqrt{c-d}}{k} \pm \frac{\sqrt{c+(k-1)d}}{k} = \frac{\pm\sqrt{c-d} \pm \sqrt{c+(k-1)d}}{k}$$

■

Zur Bestimmung der Spannweitenverteilung im Modell (G) wollen wir die Quadratwurzelzerlegung der Matrix $M(1, \rho)$ genauer untersuchen. In diesem Fall sind wegen $1 \geq \rho \geq -\frac{1}{k-1}$ die Voraussetzungen von Lemma 3.1g) erfüllt und es ergeben sich die Quadratwurzeln $M(a(\rho), b(\rho))$, wobei gilt

$$b(\rho) = \frac{\pm\sqrt{1-\rho} \pm \sqrt{1+(k-1)\rho}}{k}$$

$$a(\rho) = \frac{\rho - (k-2)b(\rho)^2}{2b(\rho)}$$

Wir setzen

$$\delta = \begin{cases} 1 & \text{falls } b(\rho) = \pm \frac{\sqrt{1-\rho} + \sqrt{1+(k-1)\rho}}{k} \\ -1 & \text{sonst} \end{cases}$$

und erhalten in diesem Fall für die Differenz $a(\rho) - b(\rho)$ die Darstellung³

$$\begin{aligned} a(\rho) - b(\rho) &= \frac{\rho - (k-2)b(\rho)^2}{2b(\rho)} - b(\rho) \\ &= \frac{\rho - kb(\rho)^2}{2b(\rho)} \\ &= \frac{k\rho - \left(\pm\sqrt{1-\rho} \pm \sqrt{1+(k-1)\rho}\right)^2}{2kb(\rho)} \\ &= \frac{k\rho - (1-\rho) - (1+(k-1)\rho) - 2\delta\sqrt{(1-\rho)(1+(k-1)\rho)}}{2kb(\rho)} \\ &= \frac{2(\rho-1) - 2\delta\sqrt{(1-\rho)(1+(k-1)\rho)}}{2kb(\rho)} \\ &= \frac{-(1-\rho) - \delta\sqrt{(1-\rho)(1+(k-1)\rho)}}{\pm\sqrt{1-\rho} \pm \sqrt{1+(k-1)\rho}} \\ &= \frac{-\sqrt{1-\rho} - \delta\sqrt{1+(k-1)\rho}}{\pm\sqrt{1-\rho} \pm \sqrt{1+(k-1)\rho}} \sqrt{1-\rho} \end{aligned}$$

Damit ergibt sich nach Definition von δ für $a(\rho) - b(\rho)$

³Aus Lemma 3.1 g) folgt zwar sofort $(a(\rho) - b(\rho))^2 = 1 - \rho$ und somit $a(\rho) - b(\rho) = \pm\sqrt{1-\rho}$, wir benötigen aber später die positive Lösung $\sqrt{1-\rho}$, weshalb die nachstehenden Überlegungen nicht überflüssig (wenn auch trivial) sind.

Anmerkung 3.2 Falls $1 \geq \rho \geq -\frac{1}{k-1}$, so gilt in der Quadratwurzelzerlegung $M(1, \rho) = M(a(\rho), b(\rho))^2$:

$$b(\rho) = \frac{\pm\sqrt{1-\rho} \pm \sqrt{1+(k-1)\rho}}{k}$$

und

$$a(\rho) - b(\rho) = \begin{cases} -\sqrt{1-\rho} \leq 0 & \text{falls } b(\rho) = \frac{\sqrt{1-\rho} + \sqrt{1+(k-1)\rho}}{k} \\ -\sqrt{1-\rho} \leq 0 & \text{falls } b(\rho) = \frac{\sqrt{1-\rho} - \sqrt{1+(k-1)\rho}}{k} \\ \sqrt{1-\rho} \geq 0 & \text{falls } b(\rho) = \frac{-\sqrt{1-\rho} + \sqrt{1+(k-1)\rho}}{k} \\ \sqrt{1-\rho} \geq 0 & \text{falls } b(\rho) = \frac{-\sqrt{1-\rho} - \sqrt{1+(k-1)\rho}}{k} \end{cases}$$

Im Urnenmodell kann die nachstehende Aussage als Hilfsmittel bei der Skalierung der Zufallsvektoren $X^{(m)}$ verwendet werden, wenn man die „Fairness“ der Urne mittels eines Chi-Quadrat-Tests überprüfen will. Wir verzichten auf die k -te Komponente des Zufallsvektors, da sich diese aus den übrigen ermitteln lässt und betrachten die reduzierte Kovarianzmatrix, die dann invertierbar ist. Es sei

$$\begin{aligned} \tilde{M}(k) &:= \frac{k}{k-1} I_{k-1} - \frac{1}{k-1} \mathbf{1}_{(k-1) \times (k-1)} \in \mathbb{R}^{(k-1) \times (k-1)} \\ &= \begin{bmatrix} 1 & -\frac{1}{(k-1)} & \cdots & -\frac{1}{(k-1)} \\ -\frac{1}{(k-1)} & 1 & \cdots & -\frac{1}{(k-1)} \\ \vdots & & \ddots & \vdots \\ -\frac{1}{(k-1)} & -\frac{1}{(k-1)} & & 1 \end{bmatrix} \in \mathbb{R}^{(k-1) \times (k-1)} \end{aligned}$$

Lemma 3.3 $\tilde{M}(k)$ ist positiv definit und ihre inverse Matrix ist

$$\tilde{M}(k)^{-1} = \frac{k-1}{k} (I_{k-1} + \mathbf{1}_{(k-1) \times (k-1)})$$

Diese hat die (positiv definite) Quadratwurzel

$$\tilde{M}(k)^{-\frac{1}{2}} = \sqrt{\frac{k-1}{k}} I_{k-1} + \left(\sqrt{\frac{1}{k-1}} - \sqrt{\frac{1}{k(k-1)}} \right) \mathbf{1}_{(k-1) \times (k-1)}$$

Dabei bezeichnet I_{k-1} die $(k-1) \times (k-1)$ -Einheitsmatrix und $\mathbf{1}_{(k-1) \times (k-1)}$ die $(k-1) \times (k-1)$ -Matrix mit lauter Einträgen = 1. Diese Matrix hat insbesondere die Eigenschaft

$$\tilde{M}(k)^{-\frac{1}{2}} \tilde{M}(k) \tilde{M}(k)^{-\frac{1}{2}} = I_{k-1}$$

Beweis. Dies ist eine unmittelbare Folgerung aus [5], Anmerkung 5.131a), (Umkehrformel). Zum direkten Beweis überlegt man sich, dass $\tilde{M}(k)^{-1} = aI_{k-1} + b\mathbf{1}_{(k-1)\times(k-1)}$ mit geeigneten Konstanten a, b sein könnte und bestimmt aus der Gleichung

$$(aI_{k-1} + b\mathbf{1}_{(k-1)\times(k-1)}) \left(\frac{k}{k-1}I_{k-1} - \frac{1}{k-1}\mathbf{1}_{(k-1)\times(k-1)} \right) = I_{k-1}$$

durch Koeffizientenvergleich die Konstanten

$$a = b = \frac{k-1}{k}$$

Dabei wurde ausgenutzt, dass gilt

$$\mathbf{1}_{(k-1)\times(k-1)}\mathbf{1}_{(k-1)\times(k-1)} = (k-1)\mathbf{1}_{(k-1)\times(k-1)}$$

Setzt man ebenso mit Konstanten c, d

$$(cI_{k-1} + d\mathbf{1}_{(k-1)\times(k-1)}) (cI_{k-1} + d\mathbf{1}_{(k-1)\times(k-1)}) = \frac{k-1}{k} (I_{k-1} + \mathbf{1}_{(k-1)\times(k-1)})$$

an, so ergibt der Koeffizientenvergleich

$$c = \sqrt{\frac{1}{k-1}}, d = \sqrt{\frac{1}{k-1}} - \sqrt{\frac{1}{k(k-1)}}$$

und damit die zweite gesuchte Aussage. Die dritte Aussage ergibt sich dann durch elementares Nachrechnen. ■

Konsequenz dieser Überlegung: setzt man

$$\tilde{X}^{(m)} = \left(\tilde{X}_1^{(m)}, \dots, \tilde{X}_{k-1}^{(m)} \right)$$

d.h. lässt man die Information weg, ob die Kugel k gezogen wurde, so besitzt der Zufallsvektor

$$\tilde{Z}^{(m)} = \left(\tilde{Z}_1^{(m)}, \dots, \tilde{Z}_{k-1}^{(m)} \right) := \sqrt{\frac{1}{\frac{r(k-r)}{k^2}}} \tilde{M}(k)^{-\frac{1}{2}} \left(\tilde{X}^{(m)} - \frac{r}{k}\mathbf{1}_{k-1} \right)$$

standardisierte und unkorrelierte Komponenten. Entsprechend kann man für die kumulierten Kugelhäufigkeiten $S^{(n)}$ verfahren und

$$\tilde{S}^{(n)} := \sum_{m=1}^n \tilde{X}^{(m)}$$

standardisieren.

3.3 Normalapproximation

Der zentrale Grenzwertsatz von Lindeberg-Levy besagt

$$\mathcal{L}\left(\frac{S^{(n)} - \mathbb{E}(S^{(n)})}{\sqrt{n}}\right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma(X))$$

Für hinreichend großes n lässt sich also die Verteilung von $S^{(n)}$ durch eine k -variate Normalverteilung, nämlich die $\mathcal{N}(n \frac{r}{k} \mathbf{1}_{k \times 1}, n\sigma(X))$ -Verteilung approximieren.

Entsprechend kann man im Urnenmodell (cf Beispiel 2.3) zeigen, dass

$$\mathcal{L}\left(\sqrt{\frac{1}{\frac{r(k-r)}{k^2}}} \tilde{M}(k)^{-\frac{1}{2}} \frac{\tilde{S}^{(n)} - \mathbb{E}(\tilde{S}^{(n)})}{\sqrt{n}}\right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, I_{k-1})$$

Die solchermaßen reskalierten Kugelhäufigkeiten der ersten $k - 1$ Kugeln verhalten sich also approximativ wie u.i.v.-standardnormalverteilte Zufallsvariablen. Im Urnenmodell kann man also nach dieser Transformation die Daten den klassischen Tests in Normalverteilungsmodellen unterziehen, wenn man etwa die „Fairness“ der Urne prüfen will.

3.4 Anderson’s Theorem

Für den Spezialfall normalverteilter Zufallsvariablen kann man ein Resultat von Anderson ausnutzen, um zumindest eine Abschätzung für die Verteilungsfunktion der Spannweite zu erhalten.

Theorem 3.4 *Es sei Y ein k -dimensionaler $\mathcal{N}(0, \Sigma)$ -verteilter Zufallsvektor und Σ_1, Σ_2 zwei positiv definite $k \times k$ -Matrizen. Weiter sei $A \subseteq \mathbb{R}^k$ eine ursprungssymmetrische und konvexe Menge (A heißt ursprungssymmetrisch, wenn für alle $x \in \mathbb{R}^k$ gilt: $x \in A \Leftrightarrow -x \in A$). Falls in der Loewner-Ordnung gilt $\Sigma_1 \preceq \Sigma_2$, so folgt*

$$\mathbb{P}_{\Sigma=\Sigma_1}(Y \in A) \geq \mathbb{P}_{\Sigma=\Sigma_2}(Y \in A)$$

Zum Beweis siehe etwa [4], Seite 56. Anderson’s Theorem kann als multivariate Verallgemeinerung des Grundsatzes angesehen werden, dass die „Tails“ von Verteilungen um so mehr Wahrscheinlichkeitsmasse erhalten, je größer die Varianz der Verteilung ist.

Anmerkung 3.5 *Anderson’s Theorem ist auch gültig, wenn die Matrizen Σ_1, Σ_2 nur positiv semidefinit sind und die konvexe Menge A einen topologischen Rand mit Lebesgue-Maß Null hat. In diesem Fall gilt die Ungleichung zunächst beim Übergang von Σ_1, Σ_2 zu $\Sigma_1 + \varepsilon I_k, \Sigma_2 + \varepsilon I_k$ mit $\varepsilon > 0$. Aufgrund von Verteilungskonvergenz folgt die Ungleichung dann auch mit $\varepsilon \rightarrow 0$. ■*

Setzt man nun für $t > 0$

$$A(t) := \{y = (y_1, \dots, y_k)^T \in \mathbb{R}^k : y_{(k)} - y_{(1)} < t\}$$

so hat $A(t)$ die in Anderson's Theorem verlangten Eigenschaften:

- $A(t)$ ist konvex. Hierzu reicht es zu zeigen, daß

$$h : \mathbb{R}^k \rightarrow \mathbb{R}, h(y) = y_{(k)} - y_{(1)}$$

eine konvexe Funktion ist. Denn dann ist

$$A(t) = \{y \in \mathbb{R}^k : h(y) < t\}$$

als Schnitt durch den Epigraph einer konvexen Funktion eine konvexe Menge. Für $x, y \in \mathbb{R}^k$ und $\alpha \in]0; 1[$ gilt aber:

$$(\alpha x + (1 - \alpha)y)_{(k)} \leq \alpha x_{(k)} + (1 - \alpha)y_{(k)}$$

$$(\alpha x + (1 - \alpha)y)_{(1)} \geq \alpha x_{(1)} + (1 - \alpha)y_{(1)}$$

und somit

$$\begin{aligned} h(\alpha x + (1 - \alpha)y) &= (\alpha x + (1 - \alpha)y)_{(k)} - (\alpha x + (1 - \alpha)y)_{(1)} \\ &\leq \alpha x_{(k)} + (1 - \alpha)y_{(k)} - (\alpha x_{(1)} + (1 - \alpha)y_{(1)}) \\ &= \alpha (x_{(k)} - x_{(1)}) + (1 - \alpha) (y_{(k)} - y_{(1)}) \\ &= \alpha h(x) + (1 - \alpha)h(y) \end{aligned}$$

- $A(t)$ ist nullpunktsymmetrisch, denn wenn $x \in A(t)$, so gilt

$$x_{(k)} - x_{(1)} < t$$

und wegen

$$(-x)_{(k)} = -x_{(1)} \quad \forall x$$

folgt

$$(-x)_{(k)} - (-x)_{(1)} = -x_{(1)} - (-x_{(k)}) = x_{(k)} - x_{(1)} < t$$

also auch $-x \in A(t)$

- Schließlich gilt auch $\lambda(\partial(A(t))) = 0$, denn $\partial(A(t)) = \{y \in \mathbb{R}^k : y_{(k)} - y_{(1)} = t\}$ ist eine 2-kodimensionale Teilmenge von \mathbb{R}^k , so dass man Anmerkung 3.5 ausnutzen kann.

Korollar 3.6 *Es sei Y ein k -dimensionaler $\mathcal{N}(0, \Sigma)$ -verteilter Zufallsvektor und Σ_1, Σ_2 zwei positiv semidefinite $k \times k$ -Matrizen. Falls in der Loewner-Ordnung gilt $\Sigma_1 \preceq \Sigma_2$, so folgt für alle $t > 0$*

$$\mathbb{P}_{\Sigma=\Sigma_1}(Y_{(k)} - Y_{(1)} \geq t) \leq \mathbb{P}_{\Sigma=\Sigma_2}(Y_{(k)} - Y_{(1)} \geq t)$$

3.5 Ordnungseigenschaften der Kovarianzmatrix im Modell (G)

Im folgenden sei stets

$$M(\rho) := M(1, \rho)$$

Um die Spannweitenverteilung im unabhängigen Fall ausnutzen zu können, ist es erforderlich, die Matrix $M(\rho)$ im Sinne der Loewner-Ordnung durch Diagonalmatrizen zu beschränken. Idealerweise sollte dies in beide Richtungen erfolgen. Daher stellt sich die Frage, ob bzw. für welche $\rho \in [-\frac{1}{k-1}, 1]$ es $x, y \in \mathbb{R}$ gibt derart, dass

$$\Delta(x) \preceq M(\rho) \preceq \Delta(y)$$

Aus Lemma 3.1 folgt, dass $M(\rho) - \Delta(x)$ die Eigenwerte $1 - \rho - x$ und $1 + (k - 1)\rho - x$ hat. Daher gilt für $x, y \in \mathbb{R}$

$$\Delta(x) \preceq M(\rho) \preceq \Delta(y)$$

genau dann, wenn

$$\begin{aligned} x &\leq \min(1 - \rho, 1 + (k - 1)\rho) \\ y &\geq \max(1 - \rho, 1 + (k - 1)\rho) \end{aligned}$$

Aus Sicht der Interpretation von x, y als Varianzen muss ferner $x \geq 0$ gelten. Zusammengefasst ergeben sich je nach Art der Korrelation (positiv, negativ) zwei Fälle:

Korollar 3.7 (a) *Falls $\rho > 0$, so gilt $\Delta(x) \preceq M(\rho) \preceq \Delta(y)$ genau dann, wenn $x \leq 1 - \rho$, $y \geq 1 + (k - 1)\rho$.*

(b) *Falls $\rho < 0$, so gilt $\Delta(x) \preceq M(\rho) \preceq \Delta(y)$ genau dann, wenn $0 \leq x \leq 1 + (k - 1)\rho$ und $y \geq 1 - \rho$*

Die Ungleichungen sollte man selbstverständlich so weit wie möglich schärfen unter Verwendung extremerer Werte für x und y , d.h. im Fall $\rho > 0$ sollte man $x = 1 - \rho$ und $y = 1 + (k - 1)\rho$ bzw. im Fall $\rho < 0$ für $x = 1 + (k - 1)\rho$ und $y = 1 - \rho$ setzen. In

den beiden später behandelten Beispielen gilt allerdings jeweils $\rho = -\frac{1}{k-1}$, weshalb sich eine untere Loewner-Schranke auf diesem Wege leider nicht bestimmen lässt. Für positiv korrelierte Daten steht jedoch mit Korollar 3.7 ein Werkzeug zur beidseitigen Abschätzung zur Verfügung.

Anmerkung 3.8 *Wegen*

$$M(\rho) - \rho \mathbf{1}_{k \times k} = \Delta(1 - \rho)$$

gilt stets die Abschätzung

$$\rho \mathbf{1}_{k \times k} \preceq M(\rho)$$

Diese kann jedoch, wie weiter unten erläutert wird, nicht zufriedenstellend genutzt werden.

3.6 Abschätzung der Tails der Spannweitenverteilung im Modell (G)

Mit den oben eingeführten Notationen kann man nun folgendermaßen approximativ die Wahrscheinlichkeit für die Tails der Spannweitenverteilung abschätzen: Für $t > 0$ gilt nämlich

$$\mathbb{P}\left(\frac{W^{(n)}}{\sigma\sqrt{n}} \geq t\right) = \mathbb{P}\left(\frac{S^{(n)} - \mathbb{E}(S^{(n)})}{\sigma\sqrt{n}} \in A(t)^c\right)$$

Letztere Wahrscheinlichkeit konvergiert aber nach dem zentralen Grenzwertsatz für $n \rightarrow \infty$ gegen

$$\mathbb{P}_{\Sigma=\sigma(X)}\left(\frac{Y}{\sigma} \in A(t)^c\right) = \mathbb{P}_{\Sigma=\sigma(X)}(Y_{(k)} - Y_{(1)} \geq \sigma t)$$

Falls nun $\rho \leq 0$, so folgt aufgrund der Anordnungseigenschaft $\sigma(X) \preceq \sigma^2 \Delta(1 - \rho)$ nach der Folgerung 3.6 aus Anderson's Theorem

$$\mathbb{P}_{\Sigma=\sigma(X)}(Y_{(k)} - Y_{(1)} \geq \sigma t) \leq \mathbb{P}_{\Sigma=\sigma^2 \Delta(\rho)}(Y_{(k)} - Y_{(1)} \geq \sigma t)$$

d.h.

$$\mathbb{P}_{\Sigma=\sigma(X)}(Y_{(k)} - Y_{(1)} \geq \sigma t) \leq \mathbb{P}_{\Sigma=I_k}\left(Y_{(k)} - Y_{(1)} \geq \frac{t}{\sqrt{1-\rho}}\right)$$

Da nunmehr die Y_i unter der Voraussetzung $\Sigma = I_k$ stochastisch unabhängig und standardnormalverteilt sind, folgt

$$\mathbb{P}_{\Sigma=I_k}\left(Y_{(k)} - Y_{(1)} \geq \frac{t}{\sqrt{1-\rho}}\right) = 1 - k \int_{-\infty}^{\infty} \varphi(x) \left(\Phi\left(x + \frac{t}{\sqrt{1-\rho}}\right) - \Phi(x)\right)^{k-1} dx$$

Zusammengefasst ergibt sich, indem man t durch $t\sqrt{1-\rho}$ ersetzt, die folgende Aussage:

Korollar 3.9 *Im Modell (G) gilt unter der Voraussetzung $\rho \leq 0$ die approximative Abschätzung*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{W^{(n)}}{\sigma \sqrt{n(1-\rho)}} \geq t\right) \leq 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx$$

für alle $t > 0$, wobei φ die Dichte und Φ die Verteilungsfunktion der Standardnormalverteilung ist.

Diese Abschätzung ist auf den ersten Blick unabhängig von σ und ρ . Man erkennt allerdings sofort, dass die Abschätzung um so besser ist, je kleiner σ ist und je näher ρ bei 0 liegt, d.h. mit zunehmender Korrelation wird die Ungleichung unschärfer. Weiter ist festzuhalten, dass die Ungleichung nur asymptotisch korrekt ist, und über die Abweichungen aufgrund der Approximation zunächst keine Erkenntnisse vorliegen. Wir werden daher einige numerische Untersuchungen anschließen.

Anmerkung 3.10 *Völlig entsprechend kann man unter Verwendung der Ungleichungen aus Korollar 3.7 zeigen, dass für alle $t > 0$ weitere Abschätzungen gelten.*

(a) Falls $\rho > 0$, so ist wegen $\Delta(1-\rho) \preceq M(\rho) \preceq \Delta(1+(k-1)\rho)$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{W^{(n)}}{\sigma \sqrt{n(1-\rho)}} \geq t\right) &\geq 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx \\ \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{W^{(n)}}{\sigma \sqrt{n(1+(k-1)\rho)}} \geq t\right) &\leq 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx \end{aligned}$$

Dies kann etwa zur beidseitigen Abschätzung im Beispiel 2.4 herangezogen werden.

(b) Falls $-\frac{1}{k-1} < \rho < 0$, so ist wegen $\Delta(1+(k-1)\rho) \preceq M(\rho)$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{W^{(n)}}{\sigma \sqrt{n(1+(k-1)\rho)}} \geq t\right) \geq 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx$$

Die Abschätzung in (b) nach unten ist in den vorliegenden Beispielen unbrauchbar, da dort jeweils $\rho = -\frac{1}{k-1}$ gilt. Der Versuch, die Anordnungseigenschaft $\rho \mathbf{1}_{k \times k} \preceq M(\rho)$ auszunutzen, schlägt allerdings ebenso fehl, da sich dann für $\mathbb{P}\left(\frac{W^{(n)}}{\sigma \sqrt{n}} \geq t\right)$ lediglich die triviale untere Schranke

$$\begin{aligned} \mathbb{P}_{\Sigma = \sigma^2 \rho \mathbf{1}_{k \times k}}(Y_{(k)} - Y_{(1)} \geq \sigma t) &= \mathbb{P}_{\Sigma = \mathbf{1}_{k \times k}}\left(Y_{(k)} - Y_{(1)} \geq \frac{t}{\sqrt{1-\rho}}\right) \\ &= \mathbb{P}_{\Sigma = I_k}\left(\frac{1}{\sqrt{k}} \mathbf{1}_{k \times k} \cdot Y \in A\left(\frac{t}{\sqrt{1-\rho}}\right)^c\right) = 0 \end{aligned}$$

ergibt, denn $\mathbf{1}_{k \times k} \cdot Y$ besitzt identische Komponenten $Y_1 + \dots + Y_n$, so dass $\max(\mathbf{1}_{k \times k} \cdot Y) - \min(\mathbf{1}_{k \times k} \cdot Y) = 0$. Problematisch ist hier, dass die untere Loewner-Schranke $\sigma^2 \rho \mathbf{1}_{k \times k}$ eine Matrix mit dem Rang 1 ist, während $M(\rho)$ höchstens einen Rangdefekt der Größe 1 hat. Vor allem im Fall $\rho = -\frac{1}{k-1}$ muss eine im Sinne der Loewner-Ordnung dazwischen liegende Matrix gefunden werden, die aber wegen des Rangdefektes von $M(-\frac{1}{k-1})$ schon keine Diagonalmatrix mehr sein kann.

3.7 Die approximative Verteilung der Spannweite im Modell (G)

Im Modell (G) kann man zeigen, dass die Abschätzung nach Anderson scharf ist und dass die Spannweite nach geeigneter Reskalierung approximativ wieder die selbe Verteilung besitzt wie im u.i.v.-Fall mit Normalverteilungsannahme.

Dazu wählen wir nach Anmerkung 3.2 eine Quadratwurzelzerlegung

$$M(1, \rho) = M(a(\rho), b(\rho))^2$$

derart, dass $a(\rho) - b(\rho) = \sqrt{1 - \rho}$. Es sei $Z = (Z_1, \dots, Z_k)^T$ ein k -dimensionaler Zufallsvektor mit unabhängig und identisch $\mathcal{N}(0, 1)$ -verteilten Komponenten Z_i . Für den Zufallsvektor $V = (V_1, \dots, V_k)^T$,

$$V := \sigma M(a(\rho), b(\rho))Z$$

gilt daher: V ist zentriert und normalverteilt mit Kovarianzmatrix

$$\begin{aligned} \sigma(V) &= \sigma^2 M(a(\rho), b(\rho)) M(a(\rho), b(\rho))^T \\ &= \sigma^2 M(a(\rho), b(\rho))^2 \\ &= \sigma^2 M(1, \rho) \\ &= \sigma(X) \end{aligned}$$

Für $t > 0$ sei

$$A(t) := \{y = (y_1, \dots, y_k)^T \in \mathbb{R}^k : y_{(k)} - y_{(1)} < t\}$$

Dann gilt

$$\mathbb{P}\left(\frac{W^{(n)}}{\sigma\sqrt{n}} \geq t\right) = \mathbb{P}\left(\frac{S^{(n)} - \mathbb{E}(S^{(n)})}{\sigma\sqrt{n}} \in A(t)^c\right)$$

Letztere Wahrscheinlichkeit konvergiert aber nach dem zentralen Grenzwertsatz für $n \rightarrow \infty$ gegen

$$\mathbb{P}\left(\frac{V}{\sigma} \in A(t)^c\right) = \mathbb{P}(V_{(k)} - V_{(1)} \geq \sigma t)$$

Weiter ist

$$\begin{aligned} V &= \sigma ((a(\rho) - b(\rho))I_k + b(\rho)\mathbf{1}_{k \times k}) Z \\ &= \sigma \sqrt{1 - \rho} Z + kb(\rho)\bar{Z}_k \mathbf{1}_k \end{aligned}$$

wobei $\bar{Z}_k = \frac{1}{k} \sum_{i=1}^k Z_i$ und $\mathbf{1}_k = (1, \dots, 1)^T \in \mathbb{R}^k$. Also unterscheiden sich V_i und $\sigma \sqrt{1 - \rho} Z_i$ nur um den zwar stochastischen, aber von $i \in \{1, \dots, k\}$ unabhängigen Wert $kb(\rho)\bar{Z}_k$, sie besitzen also die selbe Spannweite. Darüber hinaus lässt sich der gemeinsame positive Skalenfaktor $\sigma \sqrt{1 - \rho}$ aus der Spannweite faktorisieren. Damit folgt:

$$\begin{aligned} \mathbb{P}(V_{(k)} - V_{(1)} \geq \sigma t) &= \mathbb{P}\left(\sigma \sqrt{1 - \rho} (Z_{(k)} - Z_{(1)}) \geq \sigma t\right) \\ &= \mathbb{P}\left(Z_{(k)} - Z_{(1)} \geq \frac{t}{\sqrt{1 - \rho}}\right) \end{aligned}$$

Da nunmehr die Z_i stochastisch unabhängig und standardnormalverteilt sind, folgt

$$\mathbb{P}(V_{(k)} - V_{(1)} \geq \sigma t) = 1 - k \int_{-\infty}^{\infty} \varphi(x) \left(\Phi\left(x + \frac{t}{\sqrt{1 - \rho}}\right) - \Phi(x) \right)^{k-1} dx$$

wobei φ die Dichte und Φ die Verteilungsfunktion der Standardnormalverteilung ist. Zusammengefasst ergibt sich, indem man t durch $t\sqrt{1 - \rho}$ ersetzt, die folgende Aussage:

Theorem 3.11 *Im Modell (G) gilt*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{W^{(n)}}{\sigma \sqrt{n(1 - \rho)}} \geq t\right) = 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x + t) - \Phi(x))^{k-1} dx$$

für alle $t > 0$, wobei φ die Dichte und Φ die Verteilungsfunktion der Standardnormalverteilung ist.

4 Numerische Untersuchungen und Simulationen

Die Güte der Approximation der Spannweitenverteilung hängt lediglich von der Güte der Approximation im zentralen Grenzwertsatz ab, welche durch geeignete Korrekturterme ggf. noch verbessert werden kann. Problematisch ist hierbei, wie solche Korrekturterme für Ereignisse vom Typ $A(t)$ aussehen könnten. Dies soll an dieser Stelle nicht untersucht werden. Statt dessen wird zunächst überprüft, wie gut die beidseitigen Abschätzungen nach Anderson im Falle normalverteilter Daten sind – da dann der Approximationsfehler entfällt. Wir werden danach die Güte der Approximation in Multinomialmodellen durch Vergleich mit der exakten Verteilung der Spannweite und in Urnenmodellen durch Vergleich mit Simulationsresultaten untersuchen. Schließlich vergleichen wir im Fall von Multinomialmodellen den Spannweitentest mit dem Chi-Quadrat-Test, indem wir die Gütefunktionen beider Tests mittels Simulation näherungsweise bestimmen.

4.1 Zur Schärfe der Anderson-Abschätzung

Wir nehmen im folgenden an, dass die $X, X^{(1)}, \dots, X^{(n)}$ stochastisch unabhängig und zentriert k -variat normalverteilt sind mit Korrelationsmatrix $M(\rho)$, wobei $\rho > 0$, und – der Einfachheit halber – der Varianz $\sigma = 1$. Zufallsvariablen mit einem derartigen Modell lassen sich sehr einfach unter Verwendung von Beispiel 2.4 konstruieren. Der Vorteil bei der Beschränkung auf normalverteilte Modelle besteht nun darin, dass die Ungleichung nach Anderson nunmehr nicht mehr approximativ, sondern exakt ist, da bereits finit die Normalverteilungseigenschaft gegeben ist. Außerdem besitzt $S^{(n)} := \frac{X^{(1)} + \dots + X^{(n)}}{\sqrt{n}}$ dieselbe Verteilung wie X , so dass wir gleich die Spannweite $W := X_{(k)} - X_{(1)}$ betrachten können. Schließlich erhalten wir durch die Wahl von $\rho > 0$ die Möglichkeit, die Wahrscheinlichkeitsverteilung der Spannweite nach beiden Seiten hin abzuschätzen. So kann man wie in Anmerkung 3.10, Abschnitt 3.6, folgern, dass für alle $t > 0$ gilt

$$\begin{aligned} \mathbb{P}\left(\frac{W}{\sqrt{(1-\rho)}} \geq t\right) &\geq 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx \\ \mathbb{P}\left(\frac{W}{\sqrt{(1+(k-1)\rho)}} \geq t\right) &\leq 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx \end{aligned}$$

Es ergibt sich also die folgende Ungleichung für die Spannweitenverteilung:

$$\begin{aligned} &1 - k \int_{-\infty}^{\infty} \varphi(x) \left(\Phi\left(x + \frac{t}{\sqrt{1-\rho}}\right) - \Phi(x) \right)^{k-1} dx \\ &\leq \mathbb{P}(W \geq t) \\ &\leq 1 - k \int_{-\infty}^{\infty} \varphi(x) \left(\Phi\left(x + \frac{t}{\sqrt{1+(k-1)\rho}}\right) - \Phi(x) \right)^{k-1} dx \end{aligned}$$

Es ist klar, dass diese Ungleichung nur für kleine Werte von ρ und k ausreichend scharf sein kann. Dass hierbei ρ deutlich kleiner als $\frac{1}{k-1}$ sein sollte, um noch einigermaßen zufriedenstellende Ungleichungen zu erhalten, belegen auch die nachfolgenden numerischen Beispiele.

Beispiel 4.1 Für $k = 5$ und $\rho = \frac{1}{10} \frac{1}{k-1}$ wurden 10000 Realisierungen eines $\mathcal{N}(\bar{0}, M(\rho))$ -verteilten Zufallsvektors erzeugt⁴ und die Spannweite berechnet. In Abbildung 1 finden sich die Anderson-Abschätzungen zusammen mit der empirischen Spannweitenverteilung aus der Simulation. Man erkennt eine verhältnismäßig gute Übereinstimmung der Simulation mit der Unterschranke, während die Oberschranke vor allem im oberen Bereich der Verteilung eher mäßige Approximationsgüte besitzt.

⁴Zur Simulation wurde das Programm Mathematica 3.0 auf einem Windows-PC eingesetzt.

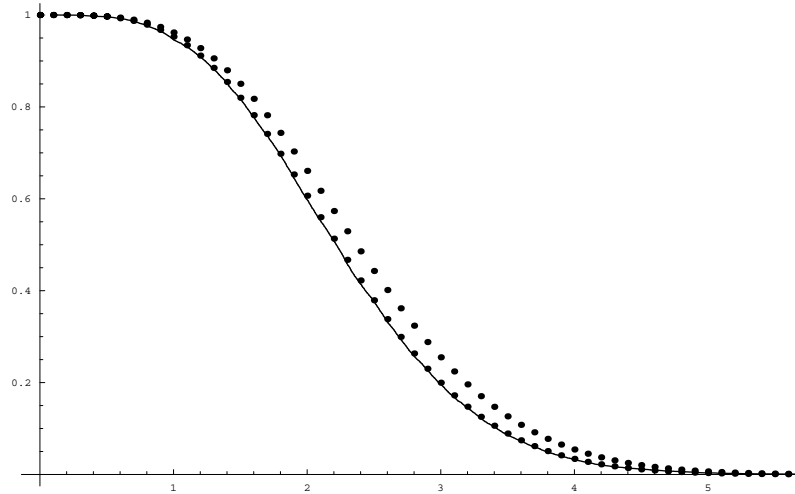


Abbildung 1: Vergleich der Anderson-Schranken mit den simulierten Spannweiten im Fall multivariat normalverteilter Zufallsvektoren mit $k = 5$ und $\rho = \frac{1}{40}$

Beispiel 4.2 Für $k = 10$ und $\rho = \frac{1}{k-1}$ wurden wiederum 10000 Realisierungen eines $\mathcal{N}(\bar{0}, M(\rho))$ -verteilten Zufallsvektors erzeugt und die Spannweite berechnet. In Abbildung 2 finden sich die Anderson-Abschätzungen zusammen mit der empirischen Spannweitenverteilung aus der Simulation. Wieder ist die Übereinstimmung der empirischen Verteilung mit der Unterschranke nach Anderson zufriedenstellend. Die Oberschranke nach Anderson liefert aber keine ausreichende Approximation.

4.2 Multinomialexperimente

Wir betrachten ein Multinomialmodell $\mathcal{M}(1, p_1, \dots, p_k)$ mit $p_1 = \dots = p_k = \frac{1}{k}$. Hier gilt:

$$\sigma^2 = \frac{k-1}{k^2}$$

$$1 - \rho = \frac{k}{k-1}$$

Dann erhält die Approximation nach Theorem 3.11 die Form

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{W^{(n)}}{\sqrt{\frac{n}{k}}} \geq t\right) = 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx$$

Im folgenden soll diese Approximation für moderate Werte von n und k verglichen werden mit den exakten Wahrscheinlichkeiten der Spannweitenverteilung. Bekanntlich folgt aus den Faltungseigenschaften der Multinomialverteilung, dass $S^{(n)}$ eine $\mathcal{M}(n, p_1, \dots, p_k)$ -Verteilung besitzt mit den Punktwahrscheinlichkeiten

$$p_{x_1, \dots, x_k} = \frac{n!}{x_1! \dots x_k!} \frac{1}{k^n}$$

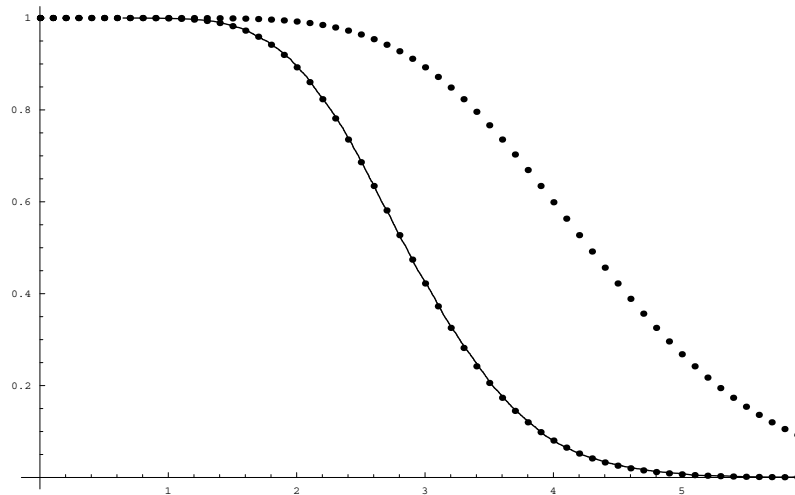


Abbildung 2: Vergleich der Anderson-Schranken mit den simulierten Spannweiten im Fall multivariat normalverteilter Zufallsvektoren mit $k = 10$ und $\rho = \frac{1}{9}$

Zur Steigerung der Effizienz bei der Berechnung der Spannweitenverteilung ist festzustellen, dass

$$p_{x_1, \dots, x_k} = p_{x_{(1)}, \dots, x_{(k)}}$$

wobei $x_{(1)}, \dots, x_{(k)}$ das gemäß der Ordnungsstatistik aus x_1, \dots, x_k gebildete Tupel bezeichnet. Eine solches angeordnetes k -Tupel wird als Partition bezeichnet, wobei die 0-Einträge noch fortgelassen werden können. Sämtliche, wesentlich verschiedene, d.h. verschiedene Elementarwahrscheinlichkeiten erzeugende Partitionen von n in k -Tupel können mittels eines von Freeman ([1]) stammenden Algorithmus erzeugt werden. Dabei wird eine Partition in absteigender Anordnung geschrieben. Wir haben den Algorithmus von Freeman nur geringfügig modifiziert, um die Verteilung der Spannweite berechnen zu können. In dieser modifizierten Form lautet er:

1. Initialisierung:

$K \geq 2$: Anzahl der Kästen, $N \geq 1$: Anzahl der Kugeln

$N_{\min} \in \{0, \dots, N\}$, $N_{\text{up}} := K^N$

$S := \frac{N + K - 1}{K}$,

$P(i) := 0$ für $i = 1, \dots, N - N_{\min} + 1$

2. Start-Partition: (Länge 1): $m := 1$, $a_1 := N$

3. Eine neue antitone Partition ist in $a = (a_1, \dots, a_m)$ abgelegt, wobei

$$a_1 = \dots = a_{j_1} > a_{j_1+1} = \dots a_{j_2} > a_{j_2+1} = \dots = a_{j_{r-1}} > a_{j_{r-1}+1} = \dots = a_{j_r} (= a_m)$$

Die Anzahl der Permutationen dieser Partition wird berechnet und anschließend mit der Anzahl von Aufteilungen dieser Permutationen auf die Kästen multipliziert:

$$n(a) := \frac{N!}{a_1! \dots a_m!} \frac{K!}{j_1!(j_2 - j_1)! \dots (j_r - j_{r-1})!(K - m)!}$$

4. Die Häufigkeit $n(a)$, mit der die gegebene Partition auftritt, wird gemäß dem Wert $a_1 = \max(a) - \min(a)$ auf die zugehörige Wahrscheinlichkeit aufaddiert:

$$P(a_1 - N_{\min} + 1) := P(a_1 - N_{\min} + 1) + n(a)$$

5. Eine neue Partition $b = (b_1, \dots, b_\ell)$ wird aus der gegebenen $a = (a_1, \dots, a_m)$ berechnet.

(a) Gehe über zu $(a_1, \dots, a_m - 1)$

(b) Falls $a_1 < N_{\min}$, so kann keine weitere relevante Partition gefunden werden. Der Algorithmus fährt in Schritt 6 fort.

(c) Falls $m = k$, so kann das Tupel nicht verlängert werden, der Algorithmus fährt in Schritt (a) fort mit dem Tupel $a = (a_1, \dots, a_{m-1})$

(d) Man gehe über zum Tupel $b' = (a_1, \dots, a_m - 1)$

(e) Man gehe über zum Tupel $b = (b_1, \dots, b_\ell) = (b', n - \sum b'_i)$

(f) Falls $b_\ell \leq b_{\ell-1}$, so ist die neue Partition gefunden. Der Algorithmus fährt fort in Schritt 3 mit der Partition $a := b$.

Anderenfalls ist die neue Partition nicht isoton

(g) Ist $\ell = k$, so wird die neue Partition verkleinert auf $a = (b_1, \dots, b_{\ell-2})$. Der Algorithmus fährt in Schritt (a) fort.

(h) Ist $\ell < k$, so geht man über zum Tupel $b' = (b_1, \dots, b_{\ell-1}, b_{\ell-1})$ (noch keine Partition) und fährt fort mit Schritt (e)

6. Die gesuchten Wahrscheinlichkeiten ergeben sich zu $\frac{1}{KN}P(1), \dots, \frac{1}{KN}P(N - N_{\min} + 1)$

Auch wenn der Freeman-Algorithmus etliche Möglichkeiten zum Effizienzgewinn ausschöpft, muss man doch immer noch einen exponentiellen Rechenaufwand zur Bestimmung der verschiedenen Partitionen und damit der Spannweitenverteilung veranschlagen.

t	$\mathbb{P}(W^{(n)} \geq t)$	Approx.	t	$\mathbb{P}(W^{(n)} \geq t)$	Approx.
1	1	1	15	0.017876	0.0127661
2	1	0.999998	16	0.00748446	0.00505247
3	0.999979	0.999878	17	0.00295962	0.00185895
4	0.999247	0.998115	18	0.00110949	0.000637956
5	0.994246	0.987118	19	0.000395215	0.000204759
6	0.97098	0.948803	20	0.000133978	0.0000615965
7	0.908649	0.861947	21	0.0000432728	0.0000173966
8	0.790439	0.721015	22	0.0000133286	4.61883 10^{-6}
9	0.624535	0.546235	23	3.91832 10^{-6}	1.15391 10^{-6}
10	0.44274	0.372445	24	1.10026 10^{-6}	2.71447 10^{-7}
11	0.281113	0.228765	25	2.95308 10^{-7}	6.01557 10^{-8}
12	0.160778	0.127177	26	7.58114 10^{-8}	1.25627 10^{-8}
13	0.0836696	0.0643837	27	1.8627 10^{-8}	2.47275 10^{-9}
14	0.0400859	0.0298663	28	4.3828 10^{-9}	4.58794 10^{-10}

Tabelle 1: Vergleich der Spannweitenverteilung mit der Normal-Approximation aus Theorem 3.11 bei der $\mathcal{M}(100, \frac{1}{12}, \dots, \frac{1}{12})$ -Verteilung

Wir haben deshalb den Algorithmus nur für moderate Werte von k und n ausgeführt und die Wahrscheinlichkeiten

$$\mathbb{P}(W^{(n)} \geq t)$$

bestimmt. Bei der Berechnung wurde das Programm Mathematica 3.0 verwendet. Nachfolgend einige Tabellen, in denen die mit dem Algorithmus von Freeman bestimmten exakten Wahrscheinlichkeiten mit den mittels Theorem 3.11 gewonnenen Approximationen verglichen werden. Betrachtet man etwa ein Multinomialexperiment mit 12 verschiedenen Ausgängen und 100maliger Wiederholung, so ergeben sich bei Annahme einer Gleichverteilung die in Tabelle 1 angegebenen Wahrscheinlichkeiten für das Auftreten einer Spannweite größer oder gleich t , $t = 1, \dots, 28$. Die Approximationen nach Theorem 3.11 sind daneben angegeben. Man erkennt zwar, dass die relativen Abweichungen teilweise erheblich sind. Jedoch unterscheiden sich die Wahrscheinlichkeiten in keinem Fall um mehr als den Faktor 10. Eine Verwendung der Approximation zur Erstellung von statistischen Tests vor allem im Bereich der üblicherweise verwendeten Fraktile scheint daher nicht abwegig, zumal der Aufwand zur Berechnung der exakten Spannweitenverteilung in n und k exponentiell steigt⁵. Um den Einfluss des Stichprobenumfanges n auf die Güte

⁵Es sei erwähnt, dass der Mathematica-Kernel zur Berechnung von $\mathcal{L}(W^{(n)})$ in der Tabelle auf einem Pentium-II-400 MHz-Rechner mit 128MB RAM insgesamt 165919 Sekunden (über 46 Stunden) CPU-Zeit verbrauchte. Wenngleich Mathematica für diese Art von Berechnungen sicher nicht die effizienteste Programmierumgebung darstellt, dürften für realistischere Größenordnungen selbst Hochgeschwindigkeits-Implementierungen rasch ausgebremst werden. Genau hier bietet sich die Approximation als effiziente Alternative an.

der Approximation zu untersuchen, betrachten wir ein einfaches Multinomialexperiment mit $k = 3$ sowie den Werten $n = 50, 100, 500$ und 1000 . Die Graphen der Funktionen

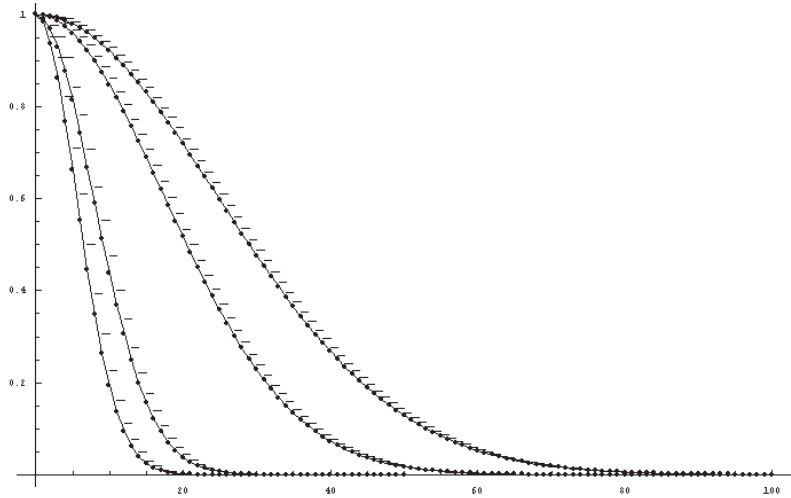


Abbildung 3: Werte von $\mathbb{P}(W^{(n)} \geq t)$ für die $\mathcal{M}(n, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ -Verteilung (Treppenfunktionen) und die Normal-Approximation aus Theorem 3.11 hierzu (Kurven mit berechneten Punkten) Dargestellt sind die Graphen für $n = 50, 100, 500$ und $n = 1000$ (von links nach rechts)

$\mathbb{P}(W^{(n)} \geq t)$ mit den Normal-Approximationen gemäß Theorem 3.11 finden sich in Abbildung 3. Man erkennt wie schon in Tabelle 1 eine vor allen in den Randbereichen der Verteilung gute Übereinstimmung der exakten Verteilung mit der Normal-Approximation. Mit wachsendem n ist die Approximation auch im mittleren Bereich der Verteilung recht zufriedenstellend⁶.

4.3 Urnenmodelle

Betrachten wir nun das Urnenmodell „Ziehen ohne Zurücklegen“, wobei $l_1 = \dots = l_k = 1$, d.h. es ergibt sich in jedem Einzelexperiment die Gleichverteilung über den $\binom{k}{r}$ möglichen Ziehungen. Es folgt

$$\sigma^2 = \frac{r(k-r)}{k^2}, \quad 1 - \rho = \frac{k}{k-1}$$

so dass die Approximation gemäß Theorem 3.11 die Form

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{W^{(n)}}{\sqrt{n \frac{r(k-r)}{k(k-1)}}} \geq t\right) = 1 - k \int_{-\infty}^{\infty} \varphi(x) (\Phi(x+t) - \Phi(x))^{k-1} dx$$

⁶Für $n = 1000$, $k = 3$ betrug der Rechenaufwand ca 20 Minuten CPU-Zeit, das zehnfache der benötigten CPU-Zeit für $n = 500$ und das etwa 200-fache der benötigten CPU-Zeit für $n = 100$.

t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$	
	simuliert	approximiert		simuliert	approximiert		simuliert	approximiert
4	1,0	0,99999	10	0,71286	0,633369	16	0,0184	0,0142992
5	0,99994	0,999693	11	0,52202	0,442806	17	0,0064	0,00547034
6	0,99868	0,996436	12	0,33904	0,274804	18	0,00254	0,00193912
7	0,98972	0,97867	13	0,19216	0,152274	19	0,00066	0,000639659
8	0,95486	0,922197	14	0,09992	0,0759871	20	0,00022	0,000197034
9	0,86608	0,805172	15	0,04606	0,0344484	21	0,00006	0,0000568281

Tabelle 2: Vergleich der simulierten Spannweitenverteilung mit der Normal-Approximation beim Urnenmodell 4 aus 20 mit 50 Ausspielungen

erhält. Zur Beurteilung der Güte der Approximation wurden die nach Theorem 3.11 berechneten Werte verglichen mit den Resultaten aus einer Simulation.⁷ Dabei wurden jeweils 50000 Simulationsläufe von 50 bzw. 100 bzw. 200 bzw. 1000 Wiederholungen der Ziehung „4 aus 20“ erzeugt und hieraus dann jeweils die 50000 Spannweiten berechnet. Durch die verschiedenen Wiederholungszahlen sollte geprüft werden, wie zuverlässig die Normal-Approximation bei kleinen, mittleren und großen Wiederholungszahlen ist, d.h. ab wann die Stabilisierung nach dem zentralen Grenzwertsatz ausreicht. Hierzu wurde das Programm Mathematica 3.0 und die hierin implementierte Zufallsfunktion zur Erzeugung von gleichverteilten Zufallszahlen auf $\{1, \dots, k\}$ unter Imitation des Ziehungsverfahrens aus einer Urne – ohne Zurücklegen – verwendet. Die Ergebnisse für durch die Simulation geschätzten Wahrscheinlichkeiten $\mathbb{P}(W^{(n)} \geq t)$ sowie die durch Theorem 3.11 gegebenen Approximationen dieser Wahrscheinlichkeiten sind in Abbildung 4 (Seite 29) und in den Tabellen 2 bis 5 wiedergegeben.

Deutlich erkennbar ist die Affinität der simulierten Spannweitenverteilung mit der durch Normal-Approximation. Diese scheint uns zur Verwendung in statistischen Tests auch schon bei kleinen Wiederholungszahlen (d.h. ab 100 Wiederholungen) bedingt geeignet. Bei $n = 1000$ ist die Übereinstimmung der Simulationsergebnisse mit den Werten, die die Normal-Approximation angibt, sogar im mittleren Bereich der Verteilung ziemlich überzeugend.

⁷Der Rückgriff auf Simulationsergebnisse erfolgte unter dem Aspekt, dass die Berechnung der tatsächlichen Spannweitenverteilung nach einem dem Freeman-Algorithmus verwandten Verfahren hier indiskutabel ist. Zwar macht weiterhin die Aufteilung in Partitionen Sinn, aber die Faltungseigenschaft der Multinomialverteilung steht nicht zur Verfügung; deshalb ist eine einfache Berechnung der Elementarwahrscheinlichkeiten wie in Schritt 3 des Freeman-Algorithmus nicht möglich.

t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$	
	simuliert	approximiert		simuliert	approximiert		simuliert	approximiert
5	1,0	0,999999	15	0,57892	0,517133	25	0,00328	0,00273074
6	0,99998	0,999974	16	0,44044	0,386107	26	0,0016	0,0012761
7	0,9999	0,999735	17	0,31756	0,271754	27	0,0007	0,000575679
8	0,99938	0,998327	18	0,21456	0,18074	28	0,00036	0,000250997
9	0,99654	0,992695	19	0,13838	0,113935	29	0,00016	0,000105869
10	0,98662	0,976269	20	0,084	0,0682909	30	0,00008	0,000043235
11	0,96026	0,939465	21	0,05004	0,0390385	32	0,00004	6,56155 10^{-6}
12	0,90814	0,873467	22	0,02678	0,0213428	35	0,00002	3,08573 10^{-7}
13	0,8255	0,775583	23	0,01374	0,0111868			
14	0,71328	0,652228	24	0,00708	0,00563355			

Tabelle 3: Simulierte Spannweitenverteilung im Urnenmodell 4 aus 20, 100 Ziehungen

t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$	
	simuliert	approximiert		simuliert	approximiert		simuliert	approximiert
8	1,0	0,99999	20	0,67732	0,633369	32	0,01754	0,0142992
9	0,99998	0,999935	21	0,58338	0,537613	33	0,0112	0,00893173
10	0,99986	0,999693	22	0,48906	0,442806	34	0,00702	0,00547034
11	0,99948	0,998849	23	0,39732	0,353976	35	0,00418	0,00328717
12	0,99786	0,996436	24	0,31078	0,274804	36	0,0022	0,00193912
13	0,99356	0,990641	25	0,23782	0,207374	37	0,0012	0,00112352
14	0,9855	0,97867	26	0,17688	0,152274	38	0,00066	0,000639659
15	0,96832	0,957019	27	0,12714	0,108924	39	0,00032	0,000358001
16	0,93956	0,922197	28	0,08952	0,0759871	40	0,00014	0,000197034
17	0,89654	0,871748	29	0,0609	0,0517535	41	0,00006	0,000106674
18	0,83644	0,805172	30	0,04228	0,0344484			
19	0,76244	0,724361	31	0,02744	0,0224303			

Tabelle 4: Simulierte Spannweitenverteilung im Urnenmodell 4 aus 20, 200 Ziehungen

t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$		t	$P(W^{(n)} \geq t)$	
	simuliert	approximiert		simuliert	approximiert		simuliert	approximiert
20	1	0,999941	47	0,55712	0,535781	74	0,00972	0,00853731
21	0,99998	0,999877	48	0,51498	0,492959	75	0,0075	0,00686727
22	0,99984	0,999757	49	0,47204	0,450838	76	0,00596	0,00550261
23	0,99972	0,999543	50	0,43034	0,409852	77	0,00472	0,00439237
24	0,99954	0,999175	51	0,39148	0,370383	78	0,00382	0,00349299
25	0,99912	0,99857	52	0,354	0,33275	79	0,00302	0,00276748
26	0,99858	0,997612	53	0,317	0,297208	80	0,00222	0,00218466
27	0,99746	0,996149	54	0,28254	0,263944	81	0,00174	0,00171835
28	0,99578	0,993988	55	0,25044	0,233084	82	0,0013	0,00134675
29	0,99318	0,990895	56	0,22052	0,204693	83	0,00104	0,00105179
30	0,98904	0,986598	57	0,19338	0,178783	84	0,00084	0,000818569
31	0,9841	0,980791	58	0,16926	0,155318	85	0,00072	0,000634863
32	0,97676	0,97315	59	0,14686	0,134227	86	0,00058	0,000490704
33	0,96766	0,963342	60	0,12682	0,115404	87	0,0005	0,000377996
34	0,9572	0,951048	61	0,10864	0,0987209	88	0,00038	0,000290199
35	0,94362	0,935978	62	0,09164	0,0840325	89	0,00034	0,000222055
36	0,9272	0,917895	63	0,07812	0,0711832	90	0,0003	0,000169352
37	0,90784	0,896628	64	0,06556	0,0600124	91	0,00024	0,000128736
38	0,88474	0,872093	65	0,05464	0,0503591	92	0,00018	0,0000975442
39	0,86008	0,844298	66	0,04592	0,0420656	93	0,00016	0,0000736718
40	0,8311	0,813353	67	0,038	0,0349805	94	0,00014	0,0000554641
41	0,79752	0,779464	68	0,03148	0,0289607	95	0,00012	0,0000416238
42	0,76424	0,742932	69	0,02576	0,0238732	96	0,00006	0,0000311386
43	0,72528	0,704136	70	0,0212	0,019596	98	0,00004	0,0000172635
44	0,68418	0,663521	71	0,01736	0,0160179	100	0,00002	$9,45286 \cdot 10^{-6}$
45	0,64192	0,621578	72	0,01462	0,0130394			
46	0,5993	0,578824	73	0,01156	0,0105719			

Tabelle 5: Simulierte Spannweitenverteilung im Urnenmodell 4 aus 20, 1000 Ziehungen

4.4 Spannweitentest versus Chi-Quadrat-Test

In Multinomialmodellen steht der Spannweitentest in unmittelbarer Konkurrenz zum klassischen Chi-Quadrat-Test. Beide liefern als asymptotische Tests nur näherungsweise das vorgegebene Signifikanzniveau, von daher sind beide Tests aufgrund der asymptotischen Sichtweise als einander gleichwertig anzusehen. Interessant ist daher die Frage, wie es beide Tests mit dem Fehler 2. Art halten. Bekanntlich ist die Güte des Chi-Quadrat-Tests bei speziellen Alternativen gering, insbesondere, wenn in nur einer Kategoriewahrscheinlichkeit von der Gleichverteilungsannahme abgewichen wird. Wir wollen daher untersuchen, wie sich Spannweiten- und Chi-Quadrat-Test verhalten, wenn von der Hypothese der Gleichverteilung abgewichen wird. Wir betrachten das Multinomialmodell $\mathcal{M}(1, p)$ aus Beispiel 2.2 und die Gleichverteilungsannahme

$$p = p_{Lapl} = (p_1, \dots, p_k) = \left(\frac{1}{k}, \dots, \frac{1}{k}\right)$$

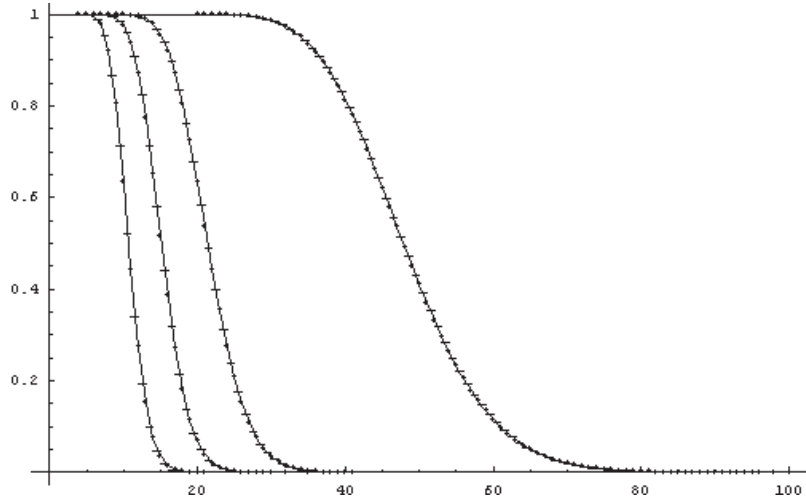


Abbildung 4: Simulationen von $\mathbb{P}(W^{(n)} \geq t)$ bei n Ziehungen im Urnenmodell 4 aus 20 (Treppenfunktionen) und Normal-Approximation hierzu (Kurven mit berechneten Punkten). Graphen für $n = 50, 100, 500$ und $n = 1000$ von links nach rechts.

als Hypothese sowie die speziellen Alternativen mit Kategorie-Verteilung

$$p = p_\varepsilon = \left(\frac{1}{k} + \varepsilon, \frac{1}{k} - \frac{\varepsilon}{k-1}, \dots, \frac{1}{k} - \frac{\varepsilon}{k-1} \right)$$

wobei $\varepsilon \in [-\frac{1}{k}; \frac{k-1}{k}]$. Zu der reskalierten Spannweite

$$\frac{W^{(n)}}{\sqrt{\frac{n}{k}}}$$

für die Summe $S^{(n)}$ der Zufallsvektoren $X^{(1)}, \dots, X^{(n)}$ gesellt sich die Chi-Quadrat-Statistik,

$$\hat{\chi}^2 = \sum_{i=1}^k \frac{(S_i^{(n)} - \frac{n}{k})^2}{\frac{n}{k}}$$

die für $n \rightarrow \infty$ unter der Hypothese p_{Lapl} asymptotisch $\chi^2(k-1)$ -verteilt ist. Um die Güte von Spannweiten-Test und Chi-Quadrat-Test zu vergleichen, wurden 10000 $\mathcal{M}(n, p_\varepsilon)$ -verteilte Zufallsvektoren simuliert und aus diesen jeweils die reskalierte Spannweite und die Chi-Quadrat-Statistik berechnet. Diese Statistiken wurden dann jeweils dahingehend geprüft, ob sie bei Annahme eines Signifikanzniveaus von 1% zur Ablehnung der Hypothese führen⁸. Schließlich wurde für den Spannweitenfall und den Chi-Quadrat-Fall die Anzahl der Ablehnungen der Hypothese in diesen 10000 Simulationsläufen bestimmt.

Es folgen die Simulationsergebnisse für den Fall $n = 100, k = 50$. Wir simulierten die Ablehnungszahlen für $\varepsilon \in \{-\frac{1}{50}, \dots, \frac{6}{50}\}$ mit der Schrittweite $\frac{1}{50}$. Die Werte sind in Tabelle 6 wiedergegeben und in Abbildung 5 dargestellt.

ε	Spw.	χ^2	ε	Spw.	χ^2	ε	Spw.	χ^2	ε	Spw.	χ^2
-0,02	142	194	0,062	4421	3087	0,144	9915	9659	0,226	9999	9999
-0,018	134	186	0,064	4695	3382	0,146	9919	9715	0,228	10000	9999
-0,016	131	168	0,066	4833	3488	0,148	9934	9780	0,23	10000	10000
-0,014	158	152	0,068	5190	3801	0,15	9940	9765	0,232	9999	9999
-0,012	131	115	0,07	5535	4030	0,152	9951	9799	0,234	10000	10000
-0,01	133	152	0,072	5797	4306	0,154	9958	9832	0,236	10000	10000
-0,008	130	151	0,074	6087	4572	0,156	9958	9830	0,238	10000	10000
-0,006	117	143	0,076	6314	4727	0,158	9964	9859	0,24	10000	10000
-0,004	117	151	0,078	6571	5072	0,16	9976	9864	0,242	10000	10000
-0,002	129	118	0,08	6772	5275	0,162	9972	9869	0,244	10000	10000
0	130	139	0,082	7050	5567	0,164	9978	9883	0,246	10000	10000
0,002	118	124	0,084	7238	5764	0,166	9987	9904	0,248	10000	10000
0,004	121	124	0,086	7577	6107	0,168	9984	9925	0,25	10000	10000
0,006	124	140	0,088	7648	6146	0,17	9990	9919	0,252	10000	10000
0,008	134	156	0,09	7864	6514	0,172	9982	9943	0,254	10000	10000
0,01	152	164	0,092	8080	6677	0,174	9989	9956	0,256	10000	10000
0,012	181	181	0,094	8224	6828	0,176	9991	9960	0,258	10000	10000
0,014	173	173	0,096	8334	7095	0,178	9993	9965	0,26	10000	10000
0,016	198	204	0,098	8505	7343	0,18	9993	9964	0,262	10000	10000
0,018	255	248	0,1	8664	7413	0,182	9998	9977	0,264	10000	10000
0,02	290	278	0,102	8765	7698	0,184	9998	9978	0,266	10000	10000
0,022	370	318	0,104	8889	7806	0,186	9997	9981	0,268	10000	10000
0,024	438	379	0,106	9041	7972	0,188	9996	9984	0,27	10000	10000
0,026	552	412	0,108	9130	8156	0,19	9999	9985	0,272	10000	10000
0,028	589	475	0,11	9181	8266	0,192	9999	9985	0,274	10000	10000
0,03	751	561	0,112	9272	8391	0,194	9998	9987	0,276	10000	10000
0,032	834	613	0,114	9342	8512	0,196	9999	9987	0,278	10000	10000
0,034	959	729	0,116	9425	8689	0,198	9999	9987	0,28	10000	10000
0,036	1131	826	0,118	9497	8817	0,2	9999	9990	0,282	10000	10000
0,038	1337	969	0,12	9530	8835	0,202	9999	9991	0,284	10000	10000
0,04	1523	1049	0,122	9616	8964	0,204	9999	9990	0,286	10000	10000
0,042	1774	1236	0,124	9645	9048	0,206	10000	9997	0,288	10000	10000
0,044	1974	1343	0,126	9709	9160	0,208	9998	9998	0,29	10000	10000
0,046	2213	1559	0,128	9738	9198	0,21	10000	9999	0,292	10000	10000
0,048	2338	1593	0,13	9753	9302	0,212	10000	9995	0,294	10000	10000
0,05	2669	1832	0,132	9794	9421	0,214	10000	9997	0,296	10000	10000
0,052	2957	1987	0,134	9806	9422	0,216	10000	10000	0,298	10000	10000
0,054	3266	2271	0,136	9843	9494	0,218	9999	9999	0,3	10000	10000
0,056	3500	2426	0,138	9855	9565	0,22	10000	10000			
0,058	3808	2655	0,14	9866	9582	0,222	10000	9998			
0,06	4055	2836	0,142	9880	9619	0,224	10000	9999			

Tabelle 6: Vergleich der Güte des Spannweitentests mit der des Chi-Quadrat-Tests im $\mathcal{M}(100, p)$ -Fall mit 50 Kategorien.

Auffällig ist zunächst, dass für $\varepsilon < 0$ beide Tests die Ablehnung der Hypothese nur sporadisch vornehmen. Selbst der Fall $\varepsilon = -\frac{1}{50}$, d.h. das Nichtauftreten von Kategorie 1, wird von beiden Tests nicht erkannt. Das vorgegebene Signifikanzniveau von 1% – für den Fall $\varepsilon = 0$ – halten beide Tests in den Simulationen nicht ein, wobei die Unterschiede zwischen den beiden Tests sicher simulationsbedingt zu bewerten sind. Wird die Kategorie 1 hingegen bevorzugt ausgewählt – d.h. im Fall $\varepsilon > 0$ – zeigen beide Tests ansteigende Güte, wobei der Spannweitentest einen deutlich steileren Anstieg der Gütefunktion aufweist, was in einer teilweise um 40% höheren Güte als beim Chi-Quadrat-Test resultiert. In diesem Bereich ist der Spannweitentest dem Chi-Quadrat-Test also deutlich überlegen. Natürlich ließe sich einwenden, dass der Chi-Quadrat-Test hier nicht praktikabel ist⁹, weshalb der

⁸durch Vergleich mit dem 1%-Fraktile der standardisierten Spannweitenverteilung im einen Fall bzw. dem 1%-Fraktile der $\chi^2(k-1)$ -Verteilung im anderen Fall.

⁹denn nach der Faustregel „ $\frac{n}{k} > 5$ “ sollte er hier nicht angewendet werden.

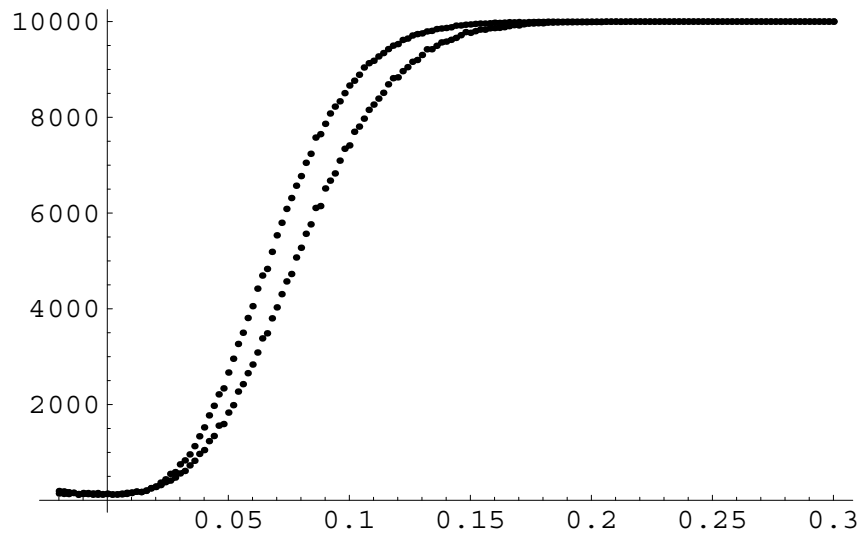


Abbildung 5: Graphische Darstellung der Werte aus Tabelle 6

Vergleich der beiden Tests nicht angebracht ist. Andererseits handelt es sich bei den gewählten Parametern $n = 100$ und $k = 50$ um durchaus relevante Größenordnungen¹⁰, was die Frage nach adäquaten Tests aufwirft. Der Spannweitentest profiliert sich m.E. hier deutlich vor dem Chi-Quadrat-Test.

5 Resümee

Zusammenfassend kann festgestellt werden, dass die auf der Basis der Normalapproximation gewonnene asymptotische Verteilung der Spannweite in vielen Anwendungen eine gute Näherung der tatsächlichen Spannweitenverteilung im Modell (G) bedeutet. Diese Beobachtung wurde für Urnenmodelle und Multinomialmodelle gemacht, wobei bei kleinen Wiederholungszahlen und moderaten Kategorie-Anzahlen eine Berechnung der exakten Verteilung angeraten ist. Im Vergleich zum klassischen Chi-Quadrat-Test (der oftmals nicht ohne weiteres anwendbar ist) weist der asymptotische Spannweitentest bei Multinomialmodellen mit einer größeren Anzahl von Kategorien eine deutlich höhere Güte auf. Es scheint daher legitim, Tests auf Basis der Spannweite mit Hilfe von Schwellenwerten zu konstruieren, die mittels Normalapproximation gewonnen wurden. Zumindest aber sollte der Spannweitentest in solchen Fällen noch zusätzlich angewandt werden, in denen die Aussagekraft des Chi-Quadrat-Testes fragwürdig ist.

¹⁰etwa wenn statistische Verfahren zur Qualitätssicherung einer regelmäßig durchgeführten Lotterie eingesetzt werden sollen. Wenn die Verfahren als Indikatoren für einen notwendigen Tausch von Komponenten des Ziehungsmechanismus verwendet werden sollten, müssen auch kleinere Stichproben bei einer größeren Anzahl von Kategorien akzeptabel sein, will man nicht Kategorien zusammenfassen

Literatur

- [1] Freeman, P.R.: Exact distribution of the largest multinomial frequency. *Applied Statistics*, 28(1979), 333-336.
- [2] Johnson/Kotz/Balakrishnan: *Discrete multivariate distributions*. Wiley. New York 1997
- [3] Pearson/Hartley: *Biometrika tables for statisticians, Vol. 1*. Cambridge University Press. London 1966.
- [4] Y.L. Tong: *Probability inequalities in multivariate distributions*. Academic Press. London 1980.
- [5] H. Witting/U. Müller-Funk: *Mathematische Statistik II*. Teubner. Stuttgart 1995.

Anhang: Fraktile zur asymptotischen Spannweitenverteilung

Angegeben sind die Werte w_α mit der Eigenschaft $\mathbb{P}(W^{(k)} \geq w_\alpha) = \alpha$, wobei die Zufallsvariablen als standardnormalverteilt angenommen wurden. Die Berechnung erfolgte mit Hilfe der Funktion FindRoot des Programmpaketes Mathematica 3.0.

α k	0.1	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001
2	2.32617	2.77181	3.16982	3.64277	3.96975	4.65351	4.92253	5.49906
3	2.90238	3.31449	3.68227	4.1203	4.42424	5.06345	5.3164	5.86099
4	3.24045	3.63316	3.98401	4.4028	4.69409	5.3088	5.55286	6.08259
5	3.47828	3.85766	4.19703	4.60282	4.88559	5.48375	5.72177	6.23979
6	3.66072	4.03009	4.36091	4.75705	5.03348	5.61933	5.85285	6.36392
7	3.8081	4.16956	4.49362	4.88217	5.15362	5.72976	5.9597	6.46176
8	3.93135	4.28631	4.60486	4.98718	5.25455	5.82273	6.04976	6.54547
9	4.03703	4.38651	4.70041	5.07752	5.34144	5.90291	6.12748	6.6184
10	4.12935	4.47412	4.78403	5.15664	5.41762	5.97331	6.19574	6.68215
11	4.2112	4.55186	4.85829	5.22696	5.48537	6.036	6.25657	6.73873
12	4.28463	4.62165	4.92499	5.2902	5.54631	6.09247	6.31138	6.79101
13	4.35116	4.68492	4.9855	5.34761	5.60166	6.14384	6.36127	6.83771
14	4.41191	4.74273	5.04082	5.40011	5.65238	6.19084	6.40693	6.8807
15	4.46778	4.79593	5.09174	5.44848	5.69903	6.23422	6.44907	6.92031
16	4.51946	4.84516	5.1389	5.49329	5.74229	6.27445	6.48823	6.95965
17	4.56752	4.89095	5.18278	5.53502	5.7826	6.31196	6.52465	6.9928
18	4.6124	4.93375	5.22381	5.57405	5.82031	6.34711	6.55879	7.02399
19	4.65449	4.97389	5.26231	5.6107	5.85572	6.38009	6.5909	7.05371
20	4.6941	5.01169	5.29857	5.64521	5.8891	6.4112	6.62116	7.08235
21	4.7315	5.04738	5.33282	5.67785	5.92065	6.44062	6.6498	7.11029
22	4.76691	5.08119	5.36528	5.70877	5.95057	6.46854	6.67698	7.13486
23	4.80052	5.1133	5.39611	5.73815	5.979	6.4951	6.70282	7.15921
24	4.83249	5.14385	5.42545	5.76615	6.00609	6.52038	6.72746	7.18238
25	4.86299	5.173	5.45345	5.79285	6.03194	6.54454	6.75103	7.20445
26	4.89212	5.20085	5.48022	5.81839	6.05667	6.56766	6.77349	7.22549
27	4.92001	5.22753	5.50585	5.84285	6.08036	6.58983	6.79508	7.24558
28	4.94675	5.25309	5.53045	5.86632	6.10311	6.61108	6.8158	7.26478
29	4.97242	5.27766	5.55407	5.88888	6.12497	6.63153	6.83573	7.28316
30	4.99712	5.30129	5.5768	5.91059	6.146	6.65122	6.85492	7.30349
31	5.02089	5.32405	5.5987	5.93151	6.16628	6.67021	6.87342	7.32108
32	5.04382	5.346	5.61982	5.95169	6.18584	6.68854	6.89128	7.33807
33	5.06595	5.36719	5.64022	5.97119	6.20474	6.70625	6.90855	7.35448
34	5.08733	5.38769	5.65993	5.99004	6.22301	6.72338	6.92525	7.37034
35	5.10802	5.4075	5.67902	6.00829	6.2407	6.74	6.94142	7.38569
36	5.12806	5.4267	5.6975	6.02597	6.25785	6.75604	6.95719	7.40081
37	5.14747	5.44531	5.71543	6.04311	6.27447	6.77164	6.97237	7.41479
38	5.16631	5.46337	5.73282	6.05976	6.29061	6.78679	6.98712	7.42853
39	5.1846	5.4809	5.74971	6.07592	6.30628	6.8015	7.00146	7.44199
40	5.20237	5.49794	5.76613	6.09164	6.32152	6.81581	7.01541	7.45515
41	5.21965	5.51451	5.78209	6.10692	6.33635	6.82974	7.02899	7.46986
42	5.23646	5.53063	5.79763	6.12181	6.35079	6.8433	7.04222	7.48177
43	5.25283	5.54634	5.81278	6.1363	6.36485	6.85652	7.05542	7.49357
44	5.26878	5.56164	5.82753	6.15043	6.37856	6.8694	7.06769	7.5052
45	5.28433	5.57656	5.84192	6.16421	6.39194	6.88197	7.07996	7.51614
46	5.2995	5.59112	5.85595	6.17766	6.40499	6.89424	7.09193	7.52756
47	5.3143	5.60533	5.86966	6.19079	6.41774	6.90623	7.10362	7.5387

Fortsetzung der Tabelle								
α k	0.1	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001
48	5.32876	5.61921	5.88305	6.20361	6.43025	6.91794	7.11505	7.54957
49	5.34288	5.63277	5.89613	6.21615	6.44239	6.92948	7.12624	7.5602
50	5.35669	5.64602	5.90892	6.22841	6.45428	6.94066	7.13717	7.57259
51	5.37019	5.65899	5.92143	6.2404	6.46592	6.95161	7.14787	7.58223
52	5.3834	5.67168	5.93367	6.25214	6.47732	6.96232	7.15834	7.59178
53	5.39633	5.6841	5.94566	6.26363	6.48848	6.97281	7.1686	7.60122
54	5.40899	5.69627	5.9574	6.27489	6.49942	6.9831	7.17865	7.61055
55	5.42139	5.70819	5.96891	6.28592	6.51013	6.99318	7.18851	7.61974
56	5.43355	5.71987	5.98018	6.29673	6.52064	7.00306	7.19817	7.6288
57	5.44546	5.73132	5.99124	6.30733	6.53095	7.01276	7.20765	7.63772
58	5.45715	5.74255	6.00208	6.31774	6.54106	7.02228	7.21696	7.6465
59	5.46862	5.75357	6.01273	6.32794	6.55099	7.03163	7.22609	7.65725
60	5.47987	5.76439	6.02317	6.33797	6.56076	7.04081	7.23506	7.66502
61	5.49091	5.77501	6.03343	6.34781	6.57029	7.04982	7.24387	7.67285
62	5.50176	5.78544	6.0435	6.35747	6.57968	7.05867	7.25253	7.6807
63	5.51241	5.79568	6.0534	6.36697	6.58892	7.06737	7.26104	7.68855
64	5.52288	5.80575	6.06312	6.37631	6.59799	7.07593	7.2694	7.69395
65	5.53317	5.81564	6.07268	6.38548	6.60691	7.08434	7.27763	7.70275
66	5.54328	5.82537	6.08208	6.39454	6.61568	7.09261	7.28565	7.71108
67	5.55323	5.83494	6.09133	6.4034	6.62432	7.10074	7.2936	7.71902
68	5.56301	5.84435	6.10042	6.41213	6.63281	7.10875	7.30143	7.72665
69	5.57264	5.85362	6.10937	6.42072	6.64117	7.11664	7.30913	7.73658
70	5.58211	5.86273	6.11818	6.42918	6.64942	7.1244	7.31671	7.74183
71	5.59144	5.8717	6.12685	6.4375	6.65883	7.13204	7.32418	7.7475
72	5.60062	5.88054	6.13539	6.44571	6.66548	7.13957	7.33154	7.75631
73	5.60966	5.88924	6.1438	6.45379	6.67333	7.14698	7.33879	7.76237
74	5.61856	5.89781	6.15209	6.46175	6.68108	7.15429	7.34593	7.76936
75	5.62734	5.90625	6.16025	6.46958	6.68871	7.16149	7.35297	7.77642
76	5.63598	5.91458	6.1683	6.47732	6.69623	7.16859	7.3599	7.7836
77	5.6445	5.92278	6.17623	6.48494	6.70365	7.17559	7.36674	7.79095
78	5.6529	5.93087	6.18405	6.49246	6.71097	7.18249	7.37348	7.7953
79	5.66118	5.93884	6.19176	6.49987	6.71818	7.1893	7.38014	7.8016
80	5.66935	5.9467	6.19936	6.50718	6.7253	7.19602	7.3867	7.8078
81	5.6774	5.95446	6.20686	6.5144	6.73233	7.20264	7.39318	7.81389
82	5.68535	5.96211	6.21426	6.52153	6.73921	7.20918	7.39957	7.81984
83	5.69318	5.96966	6.22157	6.52852	6.74604	7.21564	7.40588	7.82565
84	5.70092	5.97711	6.22878	6.53545	6.75279	7.22201	7.41211	7.83129
85	5.70856	5.98447	6.23589	6.54229	6.75944	7.2283	7.41826	7.83674
86	5.71609	5.99173	6.24291	6.54904	6.76602	7.23452	7.42433	7.84196
87	5.72353	5.9989	6.24985	6.55571	6.77251	7.24065	7.43033	7.84962
88	5.73088	6.00597	6.2567	6.5623	6.77896	7.24672	7.43625	7.85536
89	5.73814	6.01297	6.26346	6.5688	6.78529	7.2527	7.4421	7.86107
90	5.7453	6.01987	6.27015	6.57523	6.79155	7.25862	7.44795	7.86678
91	5.75238	6.02669	6.27676	6.58158	6.79773	7.26447	7.45365	7.87252
92	5.75938	6.03343	6.28328	6.58785	6.80384	7.27025	7.45928	7.8783
93	5.76629	6.0401	6.28973	6.59406	6.80988	7.27597	7.46485	7.88416
94	5.77312	6.04668	6.29609	6.60019	6.81585	7.28154	7.47037	7.88769
95	5.77987	6.05319	6.30239	6.60625	6.82175	7.28712	7.47581	7.89295
96	5.78655	6.05962	6.30862	6.61224	6.82758	7.29264	7.4812	7.89818
97	5.79314	6.06598	6.31477	6.61816	6.83335	7.29809	7.48654	7.9034
98	5.79967	6.07227	6.32086	6.62402	6.83905	7.30348	7.49181	7.90865
99	5.80612	6.07849	6.32688	6.62981	6.84469	7.30882	7.49702	7.91395
100	5.8125	6.08464	6.33283	6.63554	6.85027	7.31409	7.50219	7.91935