Anhang A Testverteilungen

In der Statistik treten Wahrscheinlichkeitsverteilungen in zweierlei Formen auf:

- Bei Zufallsexperimenten trifft man eine Annahme über die Verteilung der zugrundeliegenden Grundgesamtheiten (Modellbildung). Dabei können Verteilungen betrachtet werden wie die Normalverteilung, Binomialverteilung, Poissonverteilung, Gammaverteilung etc.

- Bei statistischen Tests benötigt man sogenannte *Prüfgrößen* oder *Teststatistiken*. Setzt man eine *normalverteilte Grundgesamtheit* voraus, dann besitzen solche Teststatistiken Verteilungen, die in den folgenden Abschnitten eingeführt werden.

A.1 Chi–Quadrat–Verteilung χ_n^2

Die erste Verteilung, die wir betrachten wollen, ist die *Chi–Quadrat–Verteilung*, welche eine spezielle Gammaverteilung, nämlich die $\gamma\left(\frac{n}{2}, \frac{1}{2}\right)$–Verteilung darstellt.

Definition A1

$$h_n(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)}\, x^{n/2-1}e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{1}$$

ist die Dichte einer *Chi–Quadrat–Verteilung mit n Freiheitsgraden* (χ_n^2–Verteilung).

Bemerkung

Für die wichtigsten Kenngrössen gilt

$$E(X) = n,\ Var(X) = 2n,\ \gamma_1(X) = \sqrt{\tfrac{8}{n}},\ \gamma_2(X) = \tfrac{12}{n}.$$

Satz A.1 Sind die Zufallsvariablen $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,1)$–verteilt, dann ist

$$Y = \sum_{i=1}^{n} X_i^2 \qquad \chi_n^2\text{–verteilt.} \tag{2}$$

Beweis (Übung).♠

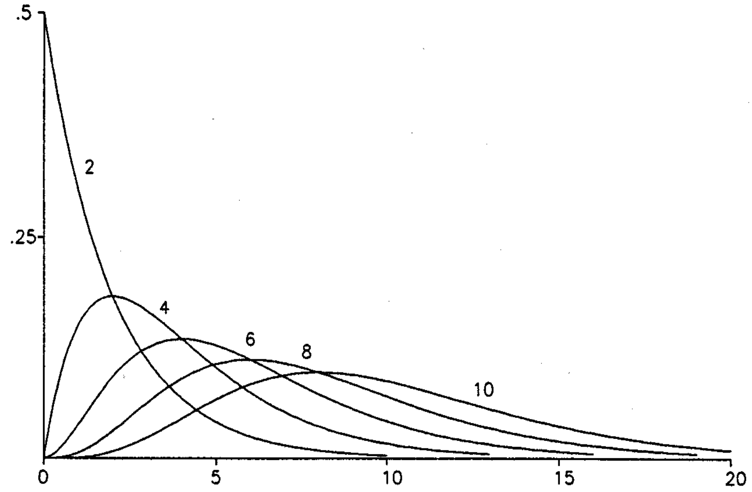

Abb. A1: Dichte der χ^2_n–Verteilung für $n = 2, 4, 6, 8$ und 10.

A.2 Student–t–Verteilung t_n

Die *Student–t–Verteilung* wurde von W.S. GOSSET (1908) gefunden, als er Angestellter der Brauerei Guinness war. Da sein Arbeitgeber seine statistischen Tätigkeiten geheimhalten wollte, um die Konkurrenz nicht aufzuscheuchen, wählte er für die Veröffentlichung das Pseudonym *'Student'*.

Satz A.2 Falls X und Y unabhängige Zufallsvariable mit $X \sim N(0,1)$ und $Y \sim \chi^2_n$, dann hat die Zufallsvariable

$$T = \frac{X}{\sqrt{Y/n}}$$

die Dichte

$$t_n(t) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + t^2/n\right)^{-(n+1)/2}, \quad -\infty < t < \infty . \tag{3}$$

Diese Verteilung heißt *Student–t–Verteilung mit* n *Freiheitsgraden* (t_n–Verteilung).

Hinweise zum Beweis. Für den Beweis dieses Satzes benötigt man den Transformationssatz für Dichten.

Transformationssatz für Dichten Sei $\mathbf{X}^T = (X_1, \ldots, X_p)$ ein p-dimensionaler Zufallsvektor mit der Dichte $f_{\mathbf{X}}(x_1, \ldots, x_p)$, $g : \mathbb{R}^p \longrightarrow \mathbb{R}^p$ eine bijektive Abbildung mit existierender Jacobi-Matrix

$$\mathbf{J} = \begin{pmatrix} \frac{\partial g_1^{-1}}{\partial y_1} & \cdots & \frac{\partial g_1^{-1}}{\partial y_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_p^{-1}}{\partial y_1} & \cdots & \frac{\partial g_p^{-1}}{\partial y_p} \end{pmatrix} .$$

Dann ist die Dichte des p-dimensionalen Zufallsvektors $\mathbf{Y}^T = (Y_1, \ldots, Y_p) = g(X_1, \ldots, X_p)$ gegeben durch

$$f_{\mathbf{Y}}(y_1, \ldots, y_p) = f_{\mathbf{X}}(g^{-1}(y_1, \ldots, y_p)) |\det(\mathbf{J})|$$

In unserem Fall ist $p = 2$ und $g(x_1, x_2) = g(x, y) = (g_1(x), g_2(y)) = \left(\frac{x}{\sqrt{y/n}}, y\right) = (t, y)$. Die gemeinsame Dichte von X und Y ist wegen der Unabhängigkeit von X und Y und den Voraussetzungen von Satz A.2: $f_{\mathbf{X}}(x, y) = f_X(x) f_Y(y) = \phi(x) h_n(y)$. Der Transformationssatz liefert die gemeinsame Dichte der Zufallsvariablen T und Y, woraus sich die Randdichte $f_T(t) = t_n(t)$ bestimmen lässt.
Detaillierter Beweis (Übung). ♠

Bemerkung

(i) Die Dichte der t_1–Verteilung (*CAUCHY–Verteilung*)

$$t_1(t) = \frac{1}{\pi(1 + t^2)}, \quad -\infty < t < \infty,$$

besitzt keinen Erwartungswert.

(ii) Für $n > k$ existieren alle Momente $E(T^j)$, $j \leq k$. Speziell gilt

$$E(T) = 0,\ n > 1,\ Var(T) = \tfrac{n}{n-2},\ n > 2,\ \gamma_2(T) = \tfrac{6}{n-4},\ n > 4. \tag{4}$$

(iii) Für große n nähert sich die t_n–Verteilung immer mehr der standardisierten Normalverteilung $N(0, 1)$, die wegen

$$\lim_{n \to \infty} (1 + t^2/n)^{-(n+1)/2} = e^{-t^2/2}$$

den Grenzfall $\phi = t_\infty$ darstellt.

(iv) Für kleine n weicht die t_n–Verteilung jedoch vor allem in den Tails noch sehr stark von der Normalverteilung ab (siehe Abbildung A.2 und Tabelle A.1).

Tab. A.2: Quantile $t_{n;\alpha}$ der t_n–Verteilung

α	$t_{5;\alpha}$	$t_{10;\alpha}$	$t_{30;\alpha}$	$t_{60;\alpha}$	$t_{100;\alpha}$	$t_{200;\alpha}$	$t_{\infty;\alpha} = z_\alpha$
0.950	2.015	1.812	1.697	1.671	1.660	1.653	1.645
0.975	2.571	2.228	2.042	2.000	1.984	1.972	1.960
0.990	3.365	2.764	2.457	2.390	2.364	2.345	2.326
0.995	4.032	3.169	2.750	2.660	2.626	2.601	2.576

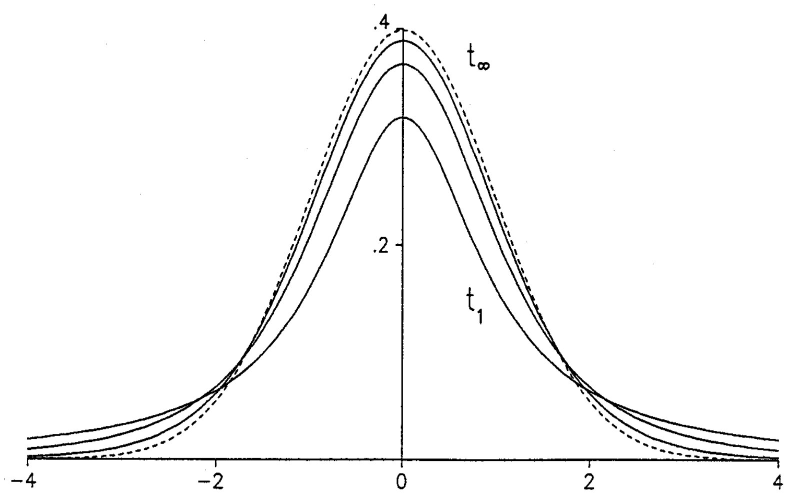

Abb. A.2: Dichte der t_n–Verteilung für $n = 1, 3, 10$, und ihr Grenzfall $\phi = t_\infty$ ($N(0,1)$–Verteilung).

A.3 Fishers F–Verteilung $F_{m,n}$

Satz A.3 X und Y seien unabhängige χ^2_m bzw. χ^2_n verteilte Zufallsvariable. Die Zufallsvariable

$$F = \frac{X/m}{Y/n} \tag{5}$$

hat die Dichte

$$f_{m,n}(t) = \begin{cases} \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right) \left(\frac{m}{n}t\right)^{m/2-1} \left(1+\frac{m}{n}t\right)^{-(m+n)/2} & t > 0 \\ 0 & t \le 0 \end{cases} \tag{6}$$

Diese Verteilung heißt F–*Verteilung* mit (m,n)–*Freiheitsgraden* ($F_{m,n}$–Verteilung).
Beweis: (Übung) ♠

Lemma A.1 Eigenschaften der $F_{m,n}$–Verteilung.
1. Für Erwartungswert und Varianz der Zufallsvariablen $F \sim F_{m,n}$ gilt:

$$E(F) = \frac{n}{n-2}\,,\ n > 2,\ \ Var(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}\,,\ n > 4. \tag{7}$$

2. Ist $F \sim F_{m,n}$ dann ist $1/F \sim F_{n,m}$.
3. Ist $T \sim t_n$, so gilt $T^2 \sim F_{1,n}$.

Bemerkung
Für das α–Quantil schreiben wir $F_{m,n;\alpha}$, d.h. $P_F(F \le F_{m,n;\alpha}) = \alpha$.
Wegen Lemma A.1(2.) hat man die Beziehung

$$P_F(F \le F_{m,n;\alpha}) = P_{1/F}(1/F \ge 1/F_{m,n;\alpha}) = P_{1/F}(1/F \ge F_{n,m;1-\alpha}) = \alpha.$$

Dies bedeutet, daß für die Quantile gilt:

$$F_{m,n;\alpha} = 1/F_{n,m;1-\alpha}\,. \tag{8}$$

Aus diesem Grund sind im **Anhang B** in der Tabelle B4 nur die Quantile $F_{m,n;\alpha}$ für $\alpha \geq 0.9$ angegeben.

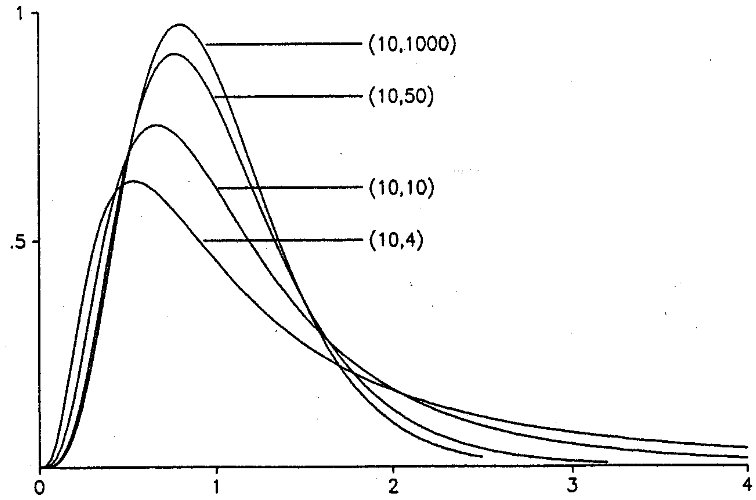

Abb. A.3: Dichte der $F_{m,n}$–Verteilung für $m = 10\,;\; n = 4, 10, 50, 1000$.

A.4 Quadratische Formen und der Satz von Cochran

Bemerkung

Eine n-dimensionale quadratische Form (QF) ist eine Abbildung $x \to q(x)$ von $\mathbb{R}^n \to \mathbb{R}$, die sich folgendermassen darstellen lässt:

$$q(x) = \mathbf{x}^T\mathbf{B}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}x_ix_j$$

mit symmetrischer $(n \times n)$–Matrix $\mathbf{B}$ und $(n \times 1)$–Vektor $\mathbf{x} = (x_1, \ldots, x_n)^T$.
Zu jeder quadratischen Form gibt es eine orthogonale Matrix $\mathbf{T}$

$$(\mathbf{T}^T\mathbf{T} = \mathbf{E} = \mathbf{T}\mathbf{T}^T\,,\; \mathbf{T}^{-1} = \mathbf{T}^T\,,\; \det(\mathbf{T}) = \det(\mathbf{T}^T) = \det(\mathbf{T}^{-1}) = 1)\,,$$

so dass $\mathbf{T}^T\mathbf{B}\mathbf{T}$ die Diagonalform

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

besitzt. $\lambda_1, \ldots, \lambda_n$ sind die stets reellen Eigenwerte der Matrix $\mathbf{B}$. Offensichtlich gilt

$$q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i \, x_i^2 \,.$$

Die Anzahl der Eigenwerte $\lambda_i \neq 0$ heißt Rang der quadratischen Form $q(\cdot)$ ($Rang(q)$). Es gilt also per definitionem

$$Rang(q) = Rang(B)\,.$$

Wir brauchen folgende Eigenschaft der Normalverteilung.

Lemma A.2 Der Zufallsvektor $\mathbf{X}$ besteht aus den Komponenten $X_1, \ldots, X_n$, die unabhängig $N(0,1)$–verteilt sind. $\mathbf{T}$ sei eine orthogonale $(n \times n)$–Matrix. Dann sind die Komponenten $Y_1, \ldots, Y_n$, des Zufallsvektors

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^T = \mathbf{TX}$$

ebenfalls unabhängig $N(0,1)$–verteilt.

Beweis

Wegen $\mathbf{T}$ orthogonal folgt $\mathbf{T}^T = \mathbf{T}^{-1}$. Die Abbildung $\mathbf{T}^{-1}\mathbf{y}$ ist überall stetig partiell differenzierbar und es gilt wegen der Orthogonalität von $\mathbf{T}$

$$\left| \det \frac{d(\mathbf{T}^{-1}\mathbf{y})}{d\mathbf{y}} \right| = \left| \det \mathbf{T}^{-1} \right| = 1\,.$$

Die Dichte $f_{\mathbf{X}}(x_1, \ldots, x_n)$ des Zufallsvektors $\mathbf{X}$ ist nach Voraussetzung

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{1}{2\pi^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} x_i^2 \right) = \frac{1}{2\pi^{n/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{E} \mathbf{x} \right)\,.$$

Für die Dichte $f_{\mathbf{Y}}(y_1, \ldots, y_n)$ des Zufallsvektors $\mathbf{Y}$ gilt nach dem Transformationssatz für Dichten

$$\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{T}^{-1}\mathbf{y}) \left| \det \frac{d(\mathbf{T}^{-1}\mathbf{y})}{d\mathbf{y}} \right| &= \frac{1}{2\pi^{n/2}} \exp\left(-\frac{1}{2} \mathbf{T}^{-1}\mathbf{y}^T \mathbf{E} \mathbf{T}^{-1} \mathbf{y} \right) \\
&= \frac{1}{2\pi^{n/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \underbrace{\mathbf{T}\mathbf{E}\mathbf{T}^{-1}}_{E} \mathbf{y} \right) \\
&= \frac{1}{2\pi^{n/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \mathbf{E} \mathbf{y} \right) \\
&= \prod_{i=1}^{n} \frac{1}{2\pi^{1/2}} \exp\left(-\frac{1}{2} y_i^2 \right) = \prod_{i=1}^{n} f_{Y_i}(y_i)\,,
\end{aligned}$$

Die Zufallsvariablen Y_i sind also unabhängig $N(0,1)$–verteilt. ♠

Satz A.4 Satz von Cochran

Seien $X_i \overset{iid}{\sim} N(0,1)\,, i=1,\ldots,n;\; q_1,\ldots,q_k\,,\; k\geq 1\,;\; n$–dimensionale quadratische Formen mit $Rang(q_j)=n_j\,,\; n_j\geq 1\,,$ und den Eigenschaften

$$\sum_{j=1}^{k} Rang(q_j) := \sum_{j=1}^{k} n_j = n \tag{9}$$

$$q(x_1,\ldots,x_n) := \sum_{j=1}^{k} q_j(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i^2\,, \tag{10}$$

Dann gelten für die Zufallsvariablen $Q_j = q_j(X_1,\ldots,X_n)$ folgende Aussagen:

(a) Die Q_j sind voneinander unabhängig,

(b) Die Q_j sind $\chi^2_{n_j}$–verteilt (Chi-Quadrat-verteilt mit n_j Freiheitsgraden).

Beweis (vollständige Induktion nach k)

$k=1$: $Q_1 = \sum_{i=1}^{n} X_i^2$ ist χ^2_n-verteilt nach Definition.

$k-1 \to k$: Die Behauptung sei für $k-1$ bewiesen. Betrachtung von k quadratischen Formen q_j. Da q_1 nach Voraussetzung den Rang n_1 hat, gibt es eine orthogonale Transformation $\mathbf{x}=\mathbf{T}\mathbf{y}$, so dass gilt

$$q_1(\mathbf{x}) = q_1(\mathbf{T}\mathbf{y}) = \mathbf{T}\mathbf{y}^T\mathbf{B}\mathbf{T}\mathbf{y} = \mathbf{y}^T\mathbf{T}^T\mathbf{B}\mathbf{T}\mathbf{y} = \sum_{j=1}^{n_1} \lambda_j y_j^2 = q_1(\mathbf{y})\,.$$

Wegen $\mathbf{x}=\mathbf{T}\mathbf{y}$ gilt weiters

$$q(\mathbf{x}) = \mathbf{x}^T\mathbf{E}\mathbf{x} = \mathbf{y}^T\mathbf{T}^T\mathbf{E}\mathbf{T}\mathbf{y} = \mathbf{y}^T\mathbf{E}\mathbf{y} = q(\mathbf{y})\,,$$

also

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2\,.$$

Damit erhält man

$$\sum_{j=2}^{k} q_j(\mathbf{y}) \overset{(10)}{=} q(\mathbf{y}) - q_1(\mathbf{y}) = \sum_{i=1}^{n} y_i^2 - (\lambda_1 y_1^2 + \cdots + \lambda_{n_1} y_{n_1})^2 \tag{11}$$

$$= (1-\lambda_1)\, y_1^2 + \cdots + (1-\lambda_{n_1})\, y_{n_1}^2 + \sum_{n_1+1}^{n} y_i^2\,. \tag{12}$$

Aus $Rang(q_j)=n_j$ und (9) folgt

$$Rang\left(\sum_{j=2}^{k} q_j(\mathbf{y})\right) \leq n - n_1\,.$$

Damit dies kein Widerspruch zu (11) ist, muss gelten $\lambda_j = 1$ für $j = 1, \ldots, n_1$, und

$$Rang\left(\sum_{j=2}^{k} q_j(\mathbf{y})\right) = n - n_1 \,. \tag{13}$$

Man hat also

$$q_1(\mathbf{y}) = \sum_{i=1}^{n_1} y_i^2 \,, \tag{14}$$

$$\sum_{j=2}^{k} q_j(\mathbf{y}) = \sum_{i=n_1+1}^{n} y_i^2 \,. \tag{15}$$

Wegen Lemma A.2 sind die Komponenten Y_i des Zufallsvektors $\mathbf{T}^{-1}\mathbf{X}$ unabhängig $N(0,1)$-verteilt. Nach Induktionsvoraussetzung sind die Zufallsvariablen Q_j für $2 \leq j \leq k$ wegen (11) und (15) voneinander unabhängig und $\chi^2_{n_j}$-verteilt.

Wegen (14) ist Q_1 aber $\chi^2_{n_1}$-verteilt. Auf Grund von (14) und (15) ist Q_1 auch von allen Q_j, $2 \leq j \leq k$, unabhängig. ♠

Korollar zu Satz A.4 Die Verteilung von $\overline{Z}$ und S_z^2.
Es seien $Z_1, ..., Z_n$ unabhängig $N(0,1)$–verteilt. Dann gilt:

(a) Die Zufallsvariable $Q_1 = (n-1)S_z^2$ ist χ^2_{n-1}-verteilt.

(b) Das empirische Mittel $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ und die empirische Varianz $S_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z})^2$ sind unabhängige Zufallsvariable.

Beweis

Es ist zu zeigen, dass die Bedingungen (9) und (10) von Satz A.4 erfüllt sind. Wir haben die zwei quadratischen Formen $q_1 = (n-1)s_z^2$ und $q_2 = n\overline{z}^2$.

Bedingung (10): $q_1 + q_2 = (n-1)s_z^2 + n\overline{z}^2 = \sum_{i=1}^{n}(z_i - \overline{z})^2 + n\overline{z}^2 = \sum_{i=1}^{n} z_i^2 = q$ (Anwendung des Verschiebungssatzes)

Bedingung (9): Es ist zu zeigen, dass $Rang(q_1) = n-1$, $Rang(q_2) = 1$.

Wegen

$$q_2(z_1, \ldots, z_n) = n\overline{z}^2 = \mathbf{z}^T\mathbf{B}\mathbf{z}$$

folgt

$$\mathbf{B}\mathbf{z} = \begin{pmatrix} \frac{1}{n}(z_1 + \cdots + z_n) \\ \vdots \\ \frac{1}{n}(z_1 + \cdots + z_n) \end{pmatrix} \Longrightarrow \mathbf{B} = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} .$$

Die Matrix $\mathbf{B}$ hat also $Rang(\mathbf{B}) = Rang(q_2) = 1$.

Wegen $q = q_1 + q_2$ und $Rang(q) = n$, $Rang(q_2) = 1$ gilt

$$n = Rang(q) = Rang(q_1 + q_2) \le Rang(q_1) + Rang(q_2) \Longrightarrow Rang(q_1) \ge n - 1\,.$$

Es bleibt noch nachzuweisen, dass $Rang(q_1) \le n - 1$. Aus

$$\begin{aligned} q_1(z_1, \ldots, z_n) &= \sum_{i=1}^{n} (z_i - \overline{z})^2 \quad \text{folgt} \\ \mathbf{Bz} &= \begin{pmatrix} z_1 - \overline{z} \\ \vdots \\ z_n - \overline{z} \end{pmatrix} = \begin{pmatrix} z_1(1 - \frac{1}{n}) - \frac{1}{n} z_2 - \cdots - \frac{1}{n} z_n \\ \vdots \\ -\frac{1}{n} z_1 - \frac{1}{n} z_2 - \cdots - z_n(1 - \frac{1}{n}) \end{pmatrix} \end{aligned}$$

und daraus

$$\mathbf{B} = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{pmatrix}.$$

Die Summation aller Zeilen ergibt eine Nullzeile. Damit ist $Rang(\mathbf{B}) = Rang(q_1) \le n-1$, andererseits ist $Rang(q_1) \ge n - 1$, woraus $Rang(q_1) = n - 1$ folgt. ♠

Bemerkung

Korollar A.4 lässt sich unmittelbar auf $N(\mu, \sigma)$-verteilte Zufallsvariable X_i übertragen. Es seien $X_i \overset{iid}{\sim} N(\mu, \sigma)$. Dann ist $Z_i = \frac{X_i - \mu}{\sigma} \overset{iid}{\sim} N(0, 1)$ und $\overline{Z} = \frac{\overline{X} - \mu}{\sigma}$, $S_z^2 = S_x^2/\sigma^2$ erfüllen Korollar A.4. Daraus folgen die Aussagen

- $Y = \frac{(n-1)}{\sigma^2} S_x^2$ ist χ^2_{n-1}–verteilt,
- Aus der Unabhängigkeit von $\overline{Z}$ und S_z^2 folgt die Unabhängigkeit von $\overline{X}$ und S_x^2.