

Parametrische Verteilungsmodellierung von Stichproben diskontinuierlicher Schallmesswerte

Alois Heiß

Independent Researcher, Ehemals Bayerisches Umweltministerium,
85748 Garching, E-Mail: alois-heiss@t-online.de

Einleitung

Oft liegt die Verteilung von Schallmess- oder Kennwerten als Kollektiv von diskreten Einzelwerten vor. Für deren weitere Verarbeitung kann es zweckmäßig sein, diese Einzelmesswerte durch eine kontinuierliche Summenhäufigkeitsfunktion (SH) [1], Kap. I 2.3, zu interpolieren, etwa um den kontinuierlichen Verlauf der Verteilungsdichte und somit deren Struktur zu ermitteln, zwischenliegende Perzentilwerte zu bestimmen oder prognostisch die Summenhäufigkeit simultan überlagerter, in der Regel deutlich fluktuierender (Umgebungs-)Geräusche darzustellen. Diese Interpolation zwischen diskreten Einzelwerten kann zwar grundsätzlich auch sowohl mit den gängigen Standard-Interpolationsverfahren, etwa nach Newton oder Lagrange oder per Spline [2], Kap. 19.6, als auch mit den Standard-Verteilungstypen, wie etwa der Normal-, Weibull-, Exponential-Verteilung etc. [1][3] versucht werden. Der Nachteil dabei ist, dass damit ein reales, in der Regel ohnehin nur näherungsweise darstellbares Messergebnis, insbesondere in den Randbereichen der SH-Funktion nicht ausreichend realitätsgerecht dokumentiert und interpretiert werden kann. Hingegen hat das hier gezeigte Verfahren den Vorteil, dass die Interpolationskurve für die SH nicht nur die Messpunkte mitenthält, sondern insbesondere auch für den obersten Wertebereich selbstregulierend auf 1 normiert ist.

Methode

Die Summenhäufigkeitsfunktion ist bekanntlich so definiert, dass die Einzelwerte einer Verteilung in aufsteigender Reihenfolge sortiert werden und die Gesamtsumme des anteiligen Auftretens der Einzelwerte, beginnend mit dem niedrigsten Wert, auf 1 normiert wird [1].

Für das im Folgenden dargestellte Interpolationsmodell entscheidend sind sowohl die definitorische Eingrenzung auf den Bereich zwischen Null und 1 sowie der monotone, d. h. hier nichtnegative Anstieg jedes Summenhäufigkeitsverlaufs. Dazu bietet sich die bei Merkmalswert Null beginnende Exponentialfunktion, subtrahiert von 1 an, mit einer geeigneten Funktion des Merkmalswertes, z. B. Schalldruckpegel, ausschließlich im Exponenten.

Von einer Verteilung des Merkmalswertes der Art x , wie z. B. Online-Schallpegelmesswerte, instantane mittlere Schalldruckquadrate, Schallpegelkennwerte, wie etwa Tages-Beurteilungspegel etc. sei eine Stichprobe (Kollektiv) gegeben, bestehend aus n stochastisch voneinander unabhängigen Einzelwerten x_k , $k = 1 \dots n$, aus derselben Grundgesamtheit [1], Kap. V 1, etwa abgekürzt dargestellt durch den Vektor

$$\vec{x} := \{x_1, x_2, \dots, x_k, \dots, x_n\},$$

mit in aufsteigender Reihenfolge geordneten Komponenten. Daraus werde die Summenhäufigkeit $F_n(x_k)$

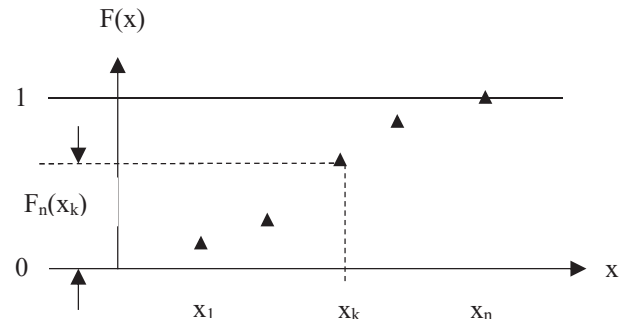


Abbildung 1: Primärdatenanzordnung und Bezeichnungen zur Bildung der Summenhäufigkeit aus diskontinuierlich verteilten Stichprobenelementen.

gebildet, wie in Abbildung 1 schematisch dargestellt ist. Der jeweilige Überschreitunganteil (Exzess) an der Stelle x ergibt sich damit definitionsgemäß zu

$$E_n(x_k) := 1 - F_n(x_k) \quad (1)$$

Nun wird, als entscheidender Schritt, eine *kontinuierlich* verlaufende Modellfunktion anhand des Ansatzes

$$E_n(x) := \exp \left\{ - \sum_{i=1}^n a_i \cdot x^i \right\} \quad (2)$$

gebildet. Dabei ist jeweils an den Stellen x_i der Überschreitunganteil $E_n(x_i)$ anhand der Messwerte x_k vorgegeben.

Die Verteilungsdichte $f(x)$ ergibt sich aufgrund ihrer Definition

$$f(x) := dF(x)/dx \quad (:= -dE(x)/dx \text{ lt. Gl. (1)}) \quad (3)$$

in der Statistikliteratur [1], Kap. II 6, aus dem Modellansatz der Gl. (2) wie folgt:

$$f_n(x) := \left[\sum_{i=1}^n i \cdot a_i \cdot x^{i-1} \right] \cdot \exp \left\{ - \sum_{i=1}^n a_i \cdot x^i \right\} \quad (4)$$

Das bedeutet: Bei bekannten Koeffizienten a_i ist auch die Verteilungsdichte darstellbar.

Der Ansatz Gl. (2) ist so strukturiert, dass genauso viele unbekannte Koeffizienten a_i berechnet werden können, als Messwerte x_k bei den jeweils zugeordneten Überschreitungswerten $E_n(x_k)$ vorhanden sind.

Die Gl. (2) sollte nun zur Vereinfachung der Kalkulation linearisiert werden. Dazu bietet sich an, die Gleichung zu logarithmieren, wodurch sie transformiert wird zu

$$-\ln(E_n(x)) := b(x) := \sum_{i=1}^n a_i \cdot x^i \quad (5)$$

Bis hierher wurde eine kontinuierliche Variable x zugrunde gelegt. Nun bedarf es jedoch der Koppelung mit der Stichprobe, dem Messwertekollektiv. Dazu werden in Gl. (5) die ermittelten n Messwerte eingesetzt. Mit der Abkürzung

$$b(x_k) := b_k \quad (6)$$

erhält man dadurch das lineare Gleichungssystem

$$\sum_{i=1}^n a_i \cdot (x_k)^i := \sum_{i=1}^n a_i \cdot x_k^i := b_k. \quad (7)$$

Die Matrix der Gl. (7) besteht also aus den Elementen

$$M_{k,i} = (x_k)^i \quad (8)$$

mit k als Zeilen- und Messwert-Nummer in aufsteigender Reihenfolge sowie mit i als Spalten-Zähler und Term-Nummer bzw. Potenz in dem Polynom im Exponenten des Modellansatzes Gl. (2).

In der üblichen abgekürzten Schreibweise lautet Gl. (7)

$$M \cdot \vec{a} = \vec{b} \quad (9)$$

wobei \vec{a} und \vec{b} als Vektoren zu verstehen sind. Die Matrixelemente Gl. (8) sind bekannt und der Vektor \vec{b} aus den nach Gl. (5) den Messwerten zuzuordnenden Überschreitungsanteilen ebenfalls. Das Gleichungssystem (7) bzw. (9) beinhaltet auch, dass die Modellfunktion alle Messpunkte mitenthält. Damit unterscheidet sich diese von einer Modellierung anhand des LMS-Verfahrens [2], Kap. 19.6.2. Ferner haben das Newtonsche Interpolationsverfahren und das nach Lagrange den Nachteil, dass – zumindest am oberen Verteilungsrand – keine Normierung auf 1 garantiert ist.

Nach der Lösung des linearen Gleichungssystems (9) werden zum Abschluss des Modellierungsprozesses die Koeffizienten in den mit kontinuierlich verlaufenden Merkmalsvariablen formulierten Polynom-Ansatz nach Gl. (2), eingesetzt und durch eine dafür geeignete Rechnersoftware verarbeitet. Damit können der Überschreitungsanteil bzw. die Summenhäufigkeit und auch die Verteilungsdichtefunktion als kontinuierlich verlaufende Funktionen bestimmt und dargestellt werden.

Insbesondere für kleine Stichprobenumfänge lässt sich der untere Rand der Modellverteilung – in Extrapolation vom kleinsten Messwert x_1 nach unten – durch den Wert, hier bezeichnet mit x_{\min} schätzen, der ausgehend von Gl. (2) – und Weglassen des Index n durch

$$E(x_{\min}) = 1 \quad (10)$$

bzw.

$$b(x_{\min}) = 0 \quad (11)$$

bestimmt ist. Zur Ermittlung des Quantils/Perzentils x_p zu einem vorgegebenem Überschreitungsanteil p % ist, ausgehend von

$$E_n(x_p) = \frac{p}{100} \quad (12)$$

die Gleichung

$$b(x_p) = -\ln\left(\frac{p}{100}\right) \quad (13)$$

zu lösen.

Mittelwert und Varianz: Anhand der Definitionen für den Mittelwert \bar{x} und die Varianz σ^2 [1] Kap. 8.2, nämlich

$$\bar{x} = \int x \cdot f(x) dx \quad (14)$$

und

$$\sigma^2 = \overline{x^2} - \bar{x}^2 \quad (15)$$

wobei

$$\overline{x^2} = \int x^2 \cdot f(x) dx, \quad (16)$$

ergibt sich mit Verwendung der Gln. (3) und (2)

$$\bar{x} = \int E(x) dx \quad (17)$$

und

$$\overline{x^2} = \int x \cdot E(x) dx. \quad (18)$$

Unsicherheit: Bei endlichem Stichprobenumfang besteht eine naturgemäße Unsicherheit im Verlauf der Summenhäufigkeitsfunktion. Dieser Zusammenhang ist anhand des Vertrauensbereichs auf der Basis des Binomialsatzes darstellbar [1], soll aber hier im Rahmen der Modellierung nicht mit behandelt werden.

Anwendungsbeispiele zu Standardverteilungen

Exponentialverteilung: Dieser Verteilungstyp [1], Kap. II.6), beinhaltet einen Spezialfall der Gl. (2), nämlich indem dort mit Ausnahme von a_1 alle sonstigen Koeffizienten Null sind. Eine Modellierung dieser Verteilung nach dem hier beschriebenen Verfahren erübrigt sich also.

Standard-Normalverteilung: Die Modellierung dieser Verteilung nach dem hier beschriebenen Verfahren soll – im Sinne einer Validierung, wie auch im folgenden Beispiel der Gleichverteilung – zunächst aufzeigen, inwieweit dieses Vorgehen mit einer Genauigkeit durchführbar ist, die für praktische Anwendungen ausreicht.

Die Summenhäufigkeit $\Phi(x)$ der Standard-Normalverteilung ist nicht explizit, sondern nur durch das Gaußsche Integral

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du \quad (19)$$

bestimmt. Die Funktion der der Modellierung zugrundezulegenden „Messwerte“ haben hier die Standard-Tabellenwerte dieses Verteilungstyps. Sie sind praktisch in jedem Statistikhandbuch, beispielsweise in [1] und [3] zu finden.

Zur hier durchzuführenden beispielhaften Modellierung der mit Gl. (19) gegebenen SH-Funktion wird diese zweckmäßigerweise um 3 Einheiten in positive Richtung verschoben, und u durch $u - \mu$, mit $\mu = 3$ ersetzt. Als „Messwerte“-Vektor wird

$$\vec{x} = \{1; 2; 3; 4; 5\},$$

gewählt, mit den in der gleichen Reihenfolge zugeordneten Tabellenwerten des Gaußschen Integrals als Summenhäufigkeit

$$\vec{F} = \{0.0228; 0.1587; 0.5; 0.8413; 0.9772\},$$

bzw. mit den zu 1 komplementären Überschreitungsanteilen E. Mit Anwendung der Gl. (5) ergibt sich der Vektor \vec{b} zu

$$\vec{b} = \{0.0231; 0.1728; 0.6931; 1.8407; 3.781\},$$

Unter Beachtung von Gl. (8) ergibt die numerische Lösung des linearen Gleichungssystems (9) den Koeffizientenvektor

$$\vec{a} = \begin{Bmatrix} 0.017 \\ -9.456 \cdot 10^{-3} \\ 7.1408 \cdot 10^{-3} \\ 9.1958 \cdot 10^{-3} \\ -8.6651 \cdot 10^{-4} \end{Bmatrix}.$$

Damit lässt sich die kontinuierlich verlaufende Überschreitungsfunktion laut Gl. (2) und infolge der Gl. (1) die kontinuierlich verlaufende SH-Funktion $F(x)$ sowie nach Gl. (4) die Dichtefunktion $f(x)$ erzeugen. Die Modellfunktion ist in der folgenden Abbildung 2 dargestellt.

Zur Beurteilung der Güte dieses Modellansatzes mit (nur) 5 Stützpunkten ist in Abb. 3 die Differenzfunktion zwischen der Modell-SH und dem „exakten“ Verlauf dargestellt. Die absolute Abweichung beträgt etwa 0,003, die bei $x = 0,5$ auftretende maximale relative Abweichung liegt bei ca. 0,04, d. h. bei 4 %. Für die Modellierung von Stichproben im

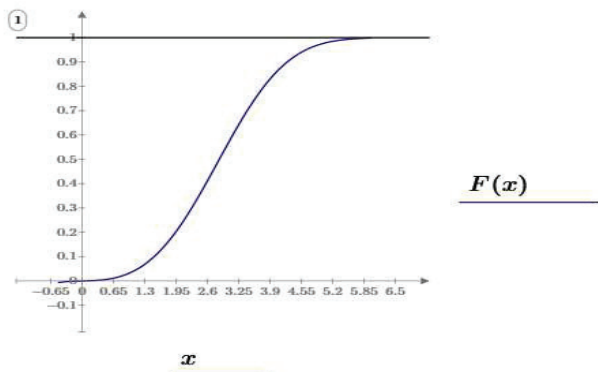


Abbildung 2: Modellfunktion für die SH der Standard-Normalverteilung gemäß den Gln. (1) und (2).

akustischen Messwesen ist diese Unsicherheit völlig irrelevant. Sie käme erst zum Tragen bei mehreren hundert statistisch *unabhängigen* Stichprobenelementen. Die aus Abb. 3 ersichtlichen Schwingungen sind als ein durch den Polynomansatz bedingter Artefakt zu deuten.

Somit ist die Modellierung der Normalverteilung anhand des Koeffizientenvektors \vec{a} mit hoher Präzision möglich. Diese Art Vektor kann also ganz allgemein als ein „Fingerabdruck“ einer zugehörigen Verteilung betrachtet werden.

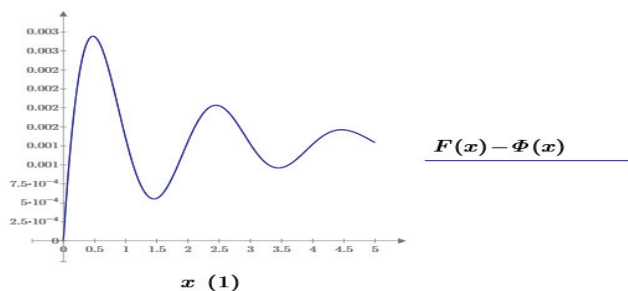


Abbildung 3: Differenzfunktion zwischen Modell und „exaktem“ Verlauf der SH der Standard-Normalverteilung.

Gleichverteilung: Die SH-Funktion der Gleichverteilung steigt definitionsgemäß linear von einem unteren Anfangspunkt bis zu einem vorgegebenen Endpunkt an, wo sie den Wert 1 erreicht. Siehe z. B. [1], IV 2. Die Gleichverteilung beginnt beispielsweise beim Wert $x = 0$ und hat die Breite 5. Die 5 „Messwerte“ seien

$$0,05; 1,25; 2,5; 3,75 \text{ und } 4,95,$$

und die entsprechenden, auf 1 normierten SH-Werte

$$0,01; 0,25; 0,5; 0,75 \text{ und } 0,99.$$

Der Koeffizientenvektor dafür lautet

$$\vec{a} = \begin{Bmatrix} 0.0225 \\ -0.104 \\ 0.148 \\ -0.059 \\ 0.00791 \end{Bmatrix}.$$

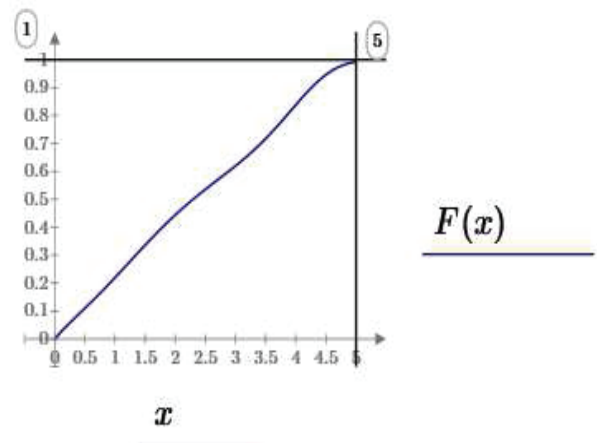


Abbildung 4: Modellfunktion für die SH der Gleichverteilung, hier mit der Breite 5, gestützt auf 5 „Messpunkte“.

Die mittlere Abweichung der Modellfunktion von der exakten Geraden beträgt hier ca. 0,03, d. h. 3%.

Anwendungsbeispiel zur Schallmessgrößenverteilung

Schallemission eines Pumpenaggregates in einem Chemiewerk: Während des stationären Betriebs wurden über eine Dauer von 1 ½ Minuten von dem verwendeten Messsystem die nachstehend aufgelisteten Schalldruckpegel in dB(A), einschließlich deren Vertrauensbereiche gemessen. Vertrauensbereiche unter ± 0.05 dB(A) sind digitalisierungsbedingt auf 0 abgerundet. Für dieses Anwendungsbeispiel stehen nicht, wie vorangehend, nur 5 diskontinuierliche Messwerte zur Interpolation der SH-Funktion zur Verfügung, sondern

- $L_{eq} = 75.7 \pm 0.0$
- $L_1 = 76.4 \pm 0.4$
- $L_5 = 76.2 \pm 0.1$
- $L_{10} = 76.1 \pm 0.0$
- $L_{30} = 75.8 \pm 0.1$
- $L_{40} = 75.7 \pm 0.1$
- $L_{50} = 75.6 \pm 0.1$
- $L_{70} = 75.4 \pm 0.1$
- $L_{90} = 75.2 \pm 0.0$
- $L_{95} = 75.1 \pm 0.0$
- $L_{99} = 74.9 \pm 0.1$

neben dem L_{eq} , 10 Stützstellen. Diese werden stets von dem hier verwendeten Messsystem ausgegeben. Für eine vereinfachte numerische Auswertung per Rechnersoftware werden hier die jeweils um 73.8 dB(A) verringerten Perzentilpegelwerte der voranstehenden Liste verwendet. Das führt auf den Koeffizientenvektor

$$\vec{a} = \begin{Bmatrix} 0.0183 \\ -0.118 \\ 0.333 \\ -0.541 \\ 0.555 \\ -0.355 \\ 0.166 \\ -0.0466 \\ 7.5224 \cdot 10^{-3} \\ -5.324 \cdot 10^{-4} \end{Bmatrix}$$

für das Exponentialpolynom. Mit Anwendung der Gln. (2) und (1) ergibt sich die folgende kontinuierlich verlaufende Summenhäufigkeitsfunktion $F(x)$ (blau) mit x als Pegelgröße:

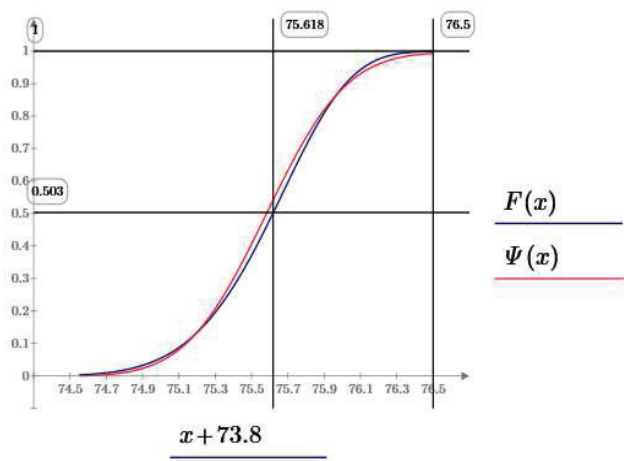


Abbildung 5: Modellfunktion $F(x)$ (blau) für die SH der Schallemissionspegel in 1,6 m Abstand zu einem stationär laufenden Pumpenaggregat mit 20 kW Leistung.

Die rote Kurve $\Psi(x)$ in Abb. 5 gibt die SH der bestangepassten Normalverteilung (Gaußsches Integral) wieder, mit $\sigma = 0,35$ dB(A), in Übereinstimmung mit dem anhand der Gln. (14) bis (16) ermittelten Wert für die aus den Messwerten modellierte SH-Kurve $F(x)$. Dieses Ergebnis bestätigt die Größenordnung der in der Normung in der Regel für Schallemissionsdaten von stationär betriebenen Maschinen angenommenen Pegel-Streubreiten zwischen ca. 0,3 und 0,5 dB [4]. Die durchschnittliche Absolutwerte-Differenz zwischen den beiden Kurven beträgt 0,003. Die aus $F(x)$ in Abb. 5 anhand der Gl. (3) folgende Pegelverteilungsdichte ist um etwa 0,18 dB(A) leicht rechtslastig. Geht man hingegen für die SH $F(x)$ in Abb. 5 zur Antilog-Skala über und wendet darauf Gl. (3) an, so wird die in Abb. 6 dargestellte Verteilungsdichtefunktion um ihr Maximum herum symmetrisch. Sie stimmt dann mit der Normalverteilung innerhalb einer Unsicherheitsmarge von höchstens 0,1 dB, d. h. praktisch vollständig, überein!

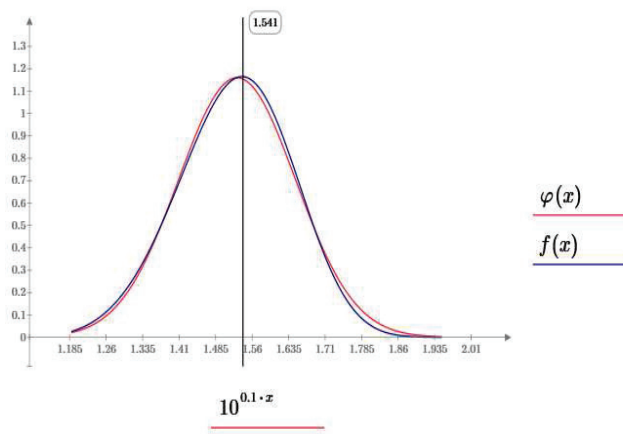


Abbildung 6: Verteilungsdichte $f(x)$ (blau) der Antilog-Werte zur SH $F(x)$ und dazu die bestangepasste Normalverteilung $\varphi(x)$ (rot).

Akustisch und im Hinblick auf den zentralen Grenzwertsatz der Statistik, [1], Kap. II 9 und [3], Kap. B 4.5, bedeutet das eine Reihe unabhängiger, jeweils für sich aktiver simultan wirkender Emissionsbereiche mit spezifischen Abstrahl-Verteilungen der mittleren Schalldruckquadrate („Antilog-Werte“), die sich zur gemessenen Emission aufaddieren.

Zusammenfassung

Liegen Schallmess- oder Kennwerte als Kollektiv von *diskreten* Einzelwerten vor, so lassen sich diese durch die hier aufgezeigte Vorgehensweise rechnergestützt anhand einer kontinuierlich verlaufenden Summenhäufigkeitsfunktion miteinander verbinden. Dabei berücksichtigt die Modellierung von selbst die Normierung der Summenhäufigkeitsfunktion auf ihren Endwert 1 für den oberen Variablenbereich. Dies lässt sich durch ein Exponential-Polynom erreichen. Dessen Koeffizienten können durch die in der Regel rechnergestützte Lösung eines linearen Gleichungssystems von der Ordnung des Stichprobenumfangs ermittelt werden. Mit der so numerisch bestimmten Summenhäufigkeitsfunktion können dann zweckdienliche statistische Operationen durchgeführt werden, wie z. B. die Bestimmung von Verteilungsparametern oder Anpassungstests zur Prüfung auf Vorliegen eines der Standard-Verteilungstypen aus der statistischen Literatur.

Literatur

- [1] Hartung, J. Statistik. Oldenbourg Verlag, München Wien, 13. Auflage 2002 (oder neuere Auflage)
- [2] Bronstein I. N., Semendjajew K. A., Musiol G. und Mühlig H. Taschenbuch der Mathematik. Verlag Harri Deutsch, Frankfurt/M, 2005 (oder neuere Auflage)
- [3] Rinne, H. Taschenbuch der Statistik. Verlag Harri Deutsch, Frankfurt/M, 1997 (oder neuere Auflage)
- [4] DIN SPEC 45660-1: 2014-05: Leitfaden zum Umgang mit der Unsicherheit in der Akustik und Schwingungstechnik — Teil 1: Unsicherheit akustischer Kenngrößen.