

Neue Formalismen zum Schätzen von Messunsicherheiten – Ein Beitrag zum Verknüpfen und Fortpflanzen von Messfehlern

A new Approach for the Estimation of Measurement Uncertainties – A Contribution to the Combination and Propagation of Errors of Measurement

Michael Grabe, Universität Braunschweig

Manuskripteingang: 08. Juni 2001; zur Veröffentlichung angenommen: 30. Dezember 2001.

Auf Empfehlung der *International Organization for Standardization (ISO)* wurde mit dem *Guide to the Expression of Uncertainty in Measurement* ein Konzept zum Schätzen von Messunsicherheiten vorgelegt. Der *Guide* wurde seitens der nationalen metrologischen Staatsinstitute der Welt und zahlreicher anderer internationaler Organisationen als verbindlich erklärt.

In einem Übersichtsartikel *Gedanken zur Revision der Gauß'schen Fehlerrechnung* hat der Verfasser die Formalismen des *Guide* kritisch analysiert und ihnen ein alternatives Fehlermodell gegenübergestellt. Dessen wesentliche Kennzeichen sind nichterwartungstreue Schätzer – bedingt durch so genannte unbekannte systematische Messfehler – und der Rückgriff auf die Dichten von Fisher, Wishart und Hotelling. Die vorliegende Abhandlung detailliert die zum *Guide* alternativen Ideen zum Verknüpfen und Fortpflanzen von Messfehlern.

On behalf of the *International Organization for Standardization (ISO)*, the *Guide to the Expression of Uncertainty in Measurement* has been presented. The *Guide* has been declared mandatory on the part of the national metrology institutes of the world and numerous other international organizations.

In a review article „*On the Revision of the Gaussian Error Calculus*“, the author has critically analysed the *Guide's* formalisms. At the same time, he has presented an alternative error model, the essential features of which are biased estimators – caused by so-called unknown systematic errors – and the exploitation of the densities of Fisher, Wishart and Hotelling. Part 1 of the discourse at hand gives full particulars of the ideas to combine and propagate measurement errors in a manner alternative to the *Guide*.

Schlagwörter: Revision der Fehlerfortpflanzung, Messunsicherheiten, Varianz-Analyse, Key Comparisons, ISO-Guide

1 Einführung

Der Aufsatz „*Neue Formalismen zum Schätzen von Messunsicherheiten*“ konkretisiert die im einleitenden Übersichtsartikel *Gedanken zur Revision der Gauß'schen Fehlerrechnung* [1] angesprochenen und zum *ISO-Guide* alternativen Ideen. Aufgrund der Verbreitung und Stellung des

Guide innerhalb der Metrologie erübrigt es sich, auf seine Prinzipien einzugehen [2]. Im Gegensatz zum *Guide* stützen sich die an dieser Stelle zu diskutierenden Formalismen auf Annahmen, die (zum gegenwärtigen Zeitpunkt) weder in der Messtechnik noch in der Experimentalphysik üblich sind. So wird verlangt, dass jede der in die Auswerteprozeduren einfließenden Messgrößen über *einheitlich viele*

Wiederholungsmessungen verfüge. Des Weiteren haben so genannte unbekannte systematische Messfehler den Status *zeitkonstanter* Größen. Schließlich stützen sich Messunsicherheiten ausschließlich auf empirische, experimentell realisierbare Schätzer.

In Konsequenz der Forderung nach einheitlich vielen Wiederholungsmessungen für alle beteiligten Messgrößen sind *empirische* Varianz-Kovarianz-Matrizen definiert. Letztere eröffnen die Möglichkeit, auf die (noch immer) jenseits der Metrologie stehenden Verteilungsdichten von Fisher, Wishart und Hotelling zurückzugreifen – sofern unterstellt werden kann, dass die zufälligen Messfehler normalverteilt sind.

Das Zeitalter der Tischrechner und des Internet hat die technischen Voraussetzungen dafür geschaffen, dass, selbst wenn unterschiedliche Laboratorien an Messungen beteiligt sind, jedes der Labore dieselbe Zahl von Wiederholungsmessungen durchführt. In der Vergangenheit wären derartige Verabredungen schwierig, wenn nicht unmöglich gewesen. Wir gehen der Frage nach, welche metrologischen Konsequenzen sich mit der Forderung nach gleich vielen Wiederholungsmessungen verbinden.

Bereits K. F. Gauß hatte auf die Existenz unbekannter systematischer Messfehler hingewiesen [3]. Zeitbedingt sah er jedoch davon ab, sie innerhalb seiner Formalismen zum Tragen zu bringen – die Gauß'schen Prozeduren verarbeiten ausschließlich zufällige Messfehler. In unserer Zeit haben vor allem C. Eisenhart [4] und S. Wagner [5] den Einfluss unbekannter systematischer Fehler auf die Richtigkeit von Messresultaten untersucht.

Da unbekannte systematische Messfehler im Rahmen des hier interessierenden Fehlermodells die Rolle zeitkonstanter Größen spielen, verlieren alle Erwartungswerte, die Fehler dieses Typs einschließen, ihre Erwartungstreue. Dieser Umstand zieht beträchtliche Änderungen der Formalismen nach sich, u. a. den Zusammenbruch des Gauß-Markoff'schen Theorems und den der Varianz-Analyse.

Unsere Messunsicherheiten werden sich ausschließlich auf *empirische* Schätzer stützen. Nach Meinung des Verfassers liegt hier ein Problem, dem im Rahmen der Gauß'schen Formalismen zu wenig Aufmerksamkeit geschenkt worden ist.

Die zur Lösung des skizzierten Aufgabenkatalogs dienenden Prozeduren zur Analyse und Auswertung von Messdaten seien – zwecks Abgrenzung von gegenwärtig üblichen Vorgehensweisen – unter dem Arbeitstitel *Metrologische Statistik* [6] zusammengefasst.

Physikalisch bedingt, kann die Metrologie, *die Kunst des Messens*, nicht mit mathematischer Schärfe schließen. Dessen ungeachtet fällt der Metrologie, zwischen Theorie und Experiment stehend, die Aufgabe zu, die „wahren“ Werte der physikalischen Größen zu lokalisieren. – Die Terminologie unterstellt, die Natur an sich sei „wahr“, ohne dass dabei physikalische gegen philosophische Gesichtspunkte abzuwägen wären.

2 Stationäre Messprozesse

Zu messen sei die physikalische Größe x . Wir indizieren den wahren Wert¹ von x mit „0“. Hiernach ist x_0 der wahre Wert.

Eine ideale Messapparatur arbeitet statistisch gesehen stationär, d. h. sie zeigt während der Messzeit weder driftartige noch sprunghafte Änderungen ihrer messtechnischen Eigenschaften. Bezeichnen die

$$x_1, x_2, \dots, x_n \quad (2.1)$$

eine Folge von n Wiederholungsmessungen unter gleichbleibenden Messbedingungen, so dürfen wir davon ausgehen, dass die x_l , über ihrem Index l aufgetragen, statistisch regulär um eine gedanklich vorstellbare horizontale Gerade streuen, Bild 1. Ein wie auch immer kurvenförmiger, gebogener oder auch springender Verlauf jener gedachten Mittellinie würde signalisieren, dass *zeitabhängige* systematische Einflüsse die Messwerte veranlasst haben, zu driften oder zu springen.

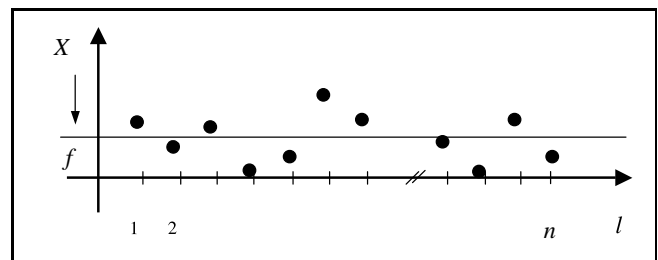


Bild 1: Wiederholungsmessungen unter gleichen Messbedingungen; f symbolisiert den unbekannt systematischen Messfehler.

Wir unterstellen, die den Messwerten überlagerten zufälligen Messfehler ε_l ($l = 1, 2, \dots, n$) seien normalverteilt und voneinander unabhängig. Des Weiteren bezeichne f den zeitkonstanten, nach Betrag und Vorzeichen unbekannt systematischen Messfehler, von dem wir annehmen, er liege in einem zu null symmetrischen Intervall der Form

$$-f_s \leq f \leq f_s; \quad f = const. \quad (2.2)$$

Sollte *definitiv* bekannt sein, dass das Fehlerintervall für f unsymmetrisch zu null liegt, so ließe sich dieses *Vorwissen* nutzen, um das zunächst unsymmetrische Intervall durch Subtraktion der halben Summe aus unterer und oberer Schranke zu symmetrisieren. – Derselbe Zahlenwert wäre dann auch von den Messdaten zu subtrahieren. In Bild 1 legt der unbekannte, zeitkonstante Wert $f = const.$ die Position der horizontalen Geraden fest, um die die Messwerte streuen.

Die *Fundamentalgleichung der Metrologie* [6] verknüpft die Größen x_l, x_0, ε_l und f ,

$$x_l = x_0 + \varepsilon_l + f \quad (l = 1, \dots, n). \quad (2.3)$$

¹ Eine Verwechslung von x_0 mit den Messdaten x_1, x_2, \dots, x_n dürfte insofern kaum möglich sein, als eine 0-te Messung nicht existiert.

Symbolisiere X eine Zufallsvariable und sei $p_X(x)$ die zugehörige Verteilungsdichte. Dann ist der Erwartungswert von X durch

$$E\{X\} = \int_{-\infty}^{\infty} x p_X(x) dx$$

definiert. Wir werden Erwartungswerte im Folgenden stets in der Kurzform $E\{\}$ notieren, d.h. die explizite Darstellung der Integrale unterdrücken. In der Klammer erscheint die jeweils in Frage stehende Zufallsvariable als Großbuchstabe. Nehmen wir jetzt an, die Variable X sei normalverteilt und ihre Realisierungen seien jene in (2.1) angegebenen x_l ($l = 1, \dots, n$), d.h.

$$X = \{x_1, x_2, \dots, x_n\}. \quad (2.4)$$

Dann liefert (2.3)

$$\mu_x = E\{X\} = \int_{-\infty}^{\infty} x p_X(x) dx = x_0 + f. \quad (2.5)$$

Offensichtlich spielt der systematische Fehler f die Rolle einer „Vorlast“ oder eines *Bias* bezüglich des wahren Wertes x_0 . Solange $f \neq 0$, ist $E\{X\}$ nicht erwartungstreu bezüglich x_0 . Der Parameter μ_x ist der theoretische, seitens der Verteilungsdichte definierte Mittelwert, der experimentell allerdings *nicht* realisierbar ist. Jedoch erfüllt er die wichtige Funktion, Bezugsgröße zu sein. Experimentell darstellbar hingegen ist der arithmetische Mittelwert

$$\bar{x} = \frac{1}{n} \sum_{l=1}^n x_l \quad (2.6)$$

als *empirischer Schätzer* des *theoretischen Mittels* μ_x . Einsetzen von (2.5) in (2.3) zeigt

$$x_l = \mu_x + \varepsilon_l. \quad (2.7)$$

D.h. die Folge der Wiederholungsmessungen streut nicht um den wahren Wert x_0 , sondern um den Parameter μ_x . Dabei kann μ_x sich nach (2.2) und (2.5) um bis zu $\pm f_s$ von x_0 unterscheiden. Diese Situation ist der Gauß'schen Fehlerrechnung durchaus fremd. Wie (2.5) und (2.7) beschreiben, separiert die Messapparatur, solange sie während der Messzeit stationär arbeitet, von sich aus zufällige und systematische Messfehler. Folgerichtig unterstreicht das arithmetische Mittel

$$\bar{x} = x_0 + \frac{1}{n} \sum_{l=1}^n \varepsilon_l + f, \quad (2.8)$$

dass noch so viele Wiederholungsmessungen nicht in der Lage sind, die Vorlast von \bar{x} bezüglich des wahren Wertes x_0 zu reduzieren.

Wir notieren (2.3) und (2.8) in Form der Identitäten,

$$x_l = x_0 + (x_l - \mu_x) + f, \quad (2.9)$$

und

$$\bar{x} = x_0 + (\bar{x} - \mu_x) + f, \quad (2.10)$$

die die Separierung zufälliger und systematischer Fehler in einer der Anwendung unmittelbar dienlichen Form zeigen. Überdies erspart uns (2.10) das wiederholte Hinschreiben der Summe der zufälligen Fehler. Beide Darstellungen zeigen gleichermaßen

$$E\{X\} = E\{\bar{X}\} = \mu_x = x_0 + f.$$

3 Verknüpfen von Messfehlern

Wie (2–10) unterstreicht, muss der Versuch, von \bar{x} ausgehend den wahren Wert x_0 zu lokalisieren, einer statistischen und einer nichtstatistischen Fehlerkomponente Rechnung tragen. Erstere könnte sich bereits in

$$E\left\{(\bar{X} - \mu_x)^2\right\} = \frac{\sigma_x^2}{n} \quad (3.1)$$

ausdrücken. Definitionsgemäß ist $\sigma_x^2 = E\{(X - \mu_x)^2\}$ die (experimentell allerdings unzugängliche) theoretische Varianz. Indessen können wir letztere durch eine empirische Größe, die empirische Varianz s_x^2 , approximieren,

$$\sigma_x^2 \approx s_x^2 = \frac{1}{n-1} \sum_{l=1}^n (x_l - \bar{x})^2. \quad (3.2)$$

Den Faktor n im Nenner von (3.1) finden wir in Gestalt von \sqrt{n} wieder, wenn wir mittels der Student'schen Dichte ein Vertrauensintervall für den Parameter μ_x festlegen [9]

$$\bar{X} - \frac{t_P(n-1)}{\sqrt{n}} S_x \leq \mu_x \leq \bar{X} + \frac{t_P(n-1)}{\sqrt{n}} S_x. \quad (3.3)$$

Wie bekannt, lokalisiert das Intervall (3.3) den Parameter μ_x mit der Wahrscheinlichkeit P ; $t_P(n-1)$ ist der auf eben dieses Vertrauensniveau Bezug nehmende Student-Faktor und S_x die s_x entsprechende Zufallsvariable.

Das Vertrauensintervall (3.3) liefert uns eine Schätzung der möglichen Differenz zwischen \bar{X} und μ_x . Nach (2–5) fehlt allerdings noch eine Schätzung der Differenz zwischen μ_x und x_0 . Sind wir bereit, die Worst-Case-Schätzung $|\mu_x - x_0|_{\max} = f_s$ zu akzeptieren, so ist

$$u_{\bar{x}} = \frac{t_P(n-1)}{\sqrt{n}} s_x + f_s \quad (3.4)$$

die Messunsicherheit des Messresultates, das wir in der Form

$$\bar{x} \pm u_{\bar{x}} \quad (3.5)$$

notieren. Im Rahmen der *Metrologischen Statistik* ist der unbekannt systematische Fehler f keine Zufallsgröße, sondern eine nach Betrag und Vorzeichen unbekannt Konstante. Für diese Interpretation gibt es Gründe [1], die nahe legen, davon abzusehen, die Differenz zwischen μ_x und x_0 quadratisch zu schätzen.

4 Empirische Kovarianzen

Fehlerfortpflanzung ist in erster Linie eine Frage der Linearisierung der Verknüpfungsfunktion. Inwieweit das zulässig ist, hängt von der Funktion selbst und von den Unsicherheiten der zu verknüpfenden Messgrößen ab. Betrachten wir der Einfachheit halber zunächst eine Funktion $\phi(x, y)$ zweier Messgrößen x und y und entwickeln sie an der unbekanntem, wahren Stelle (x_0, y_0) bezüglich der gemessenen Datenpaare (x_l, y_l) , $l = 1, 2, \dots, n$,

$$\begin{aligned} \phi(x_l, y_l) &= \phi(x_0, y_0) + \frac{\partial \phi}{\partial x_0} (x_l - x_0) \\ &+ \frac{\partial \phi}{\partial y_0} (y_l - y_0) + \dots; \\ &\left(\frac{\partial \phi}{\partial x_0} \equiv \frac{\partial \phi}{\partial x} \Big|_{x=x_0, y=y_0}, \text{ etc.} \right) \end{aligned}$$

oder, mit (2–9),

$$\begin{aligned} \phi(x_l, y_l) &= \phi(x_0, y_0) + \frac{\partial \phi}{\partial x_0} (x_l - \mu_x) + \frac{\partial \phi}{\partial y_0} (y_l - \mu_y) \\ &+ \frac{\partial \phi}{\partial x_0} f_x + \frac{\partial \phi}{\partial y_0} f_y + \dots \end{aligned} \quad (4.1)$$

Der Erwartungswert

$$\mu_\phi = E\{\phi(X, Y)\} = \phi(x_0, y_0) + \frac{\partial \phi}{\partial x_0} f_x + \frac{\partial \phi}{\partial y_0} f_y \quad (4.2)$$

ist Träger des fortgepflanzten systematischen Fehlers

$$f_\phi = \frac{\partial \phi}{\partial x_0} f_x + \frac{\partial \phi}{\partial y_0} f_y. \quad (4.3)$$

Die Darstellung

$$\phi(x_l, y_l) = \mu_\phi + \frac{\partial \phi}{\partial x_0} (x_l - \mu_x) + \frac{\partial \phi}{\partial y_0} (y_l - \mu_y) + \dots \quad (4.4)$$

zeigt erneut, dass die Folge der $\phi(x_l, y_l)$ um den Erwartungswert μ_ϕ streut und nicht um den wahren Wert $\phi(x_0, y_0)$. Andererseits können wir $\phi(x, y)$ auch an der wahren Stelle (x_0, y_0) bezüglich der Mittel (\bar{x}, \bar{y}) entwickeln. Mit (2–10) finden wir

$$\begin{aligned} \phi(\bar{x}, \bar{y}) &= \phi(x_0, y_0) + \frac{\partial \phi}{\partial x_0} (\bar{x} - \mu_x) + \frac{\partial \phi}{\partial y_0} (\bar{y} - \mu_y) \\ &+ \frac{\partial \phi}{\partial x_0} f_x + \frac{\partial \phi}{\partial y_0} f_y + \dots \end{aligned} \quad (4.5)$$

Da die partiellen Ableitungen $\partial \phi / \partial x_0$ und $\partial \phi / \partial y_0$ nicht spezifizierbar sind, müssen wir sie durch Ableitungen $\partial \phi / \partial \bar{x}$ und $\partial \phi / \partial \bar{y}$ approximieren. Des Weiteren werden wir, wie angekündigt, die Reihenentwicklungen nach den linearen Gliedern abbrechen. Subtrahieren von (4.1) und (4.5) liefert dann

$$\phi(x_l, y_l) = \phi(\bar{x}, \bar{y}) + \frac{\partial \phi}{\partial \bar{x}} (x_l - \bar{x}) + \frac{\partial \phi}{\partial \bar{y}} (y_l - \bar{y}), \quad (4.6)$$

wenn wir noch, um die Notation so übersichtlich wie möglich zu halten, auf das an sich notwendige \approx -Zeichen verzichten.

An dieser Stelle greifen wir auf die Idee zurück, gleich viele Wiederholungsmessungen, sagen wir, *wohldefinierte Messbedingungen*, für die Messgrößen x_l und y_l zu verlangen. Dann nämlich gilt im Rahmen der Linearisierung

$$\bar{\phi} = \phi(\bar{x}, \bar{y}) = \frac{1}{n} \sum_{l=1}^n \phi(x_l, y_l). \quad (4.7)$$

Des Weiteren sind wir in der Lage, eine empirische Varianz

$$s_\phi^2 = \frac{1}{n-1} \sum_{l=1}^n [\phi(x_l, y_l) - \phi(\bar{x}, \bar{y})]^2 \quad (4.8)$$

zu definieren; sie zeigt in der Ausarbeitung

$$s_\phi^2 = \left(\frac{\partial \phi}{\partial \bar{x}} \right)^2 s_x^2 + 2 \left(\frac{\partial \phi}{\partial \bar{x}} \right) \left(\frac{\partial \phi}{\partial \bar{y}} \right) s_{xy} + \left(\frac{\partial \phi}{\partial \bar{y}} \right)^2 s_y^2. \quad (4.9)$$

Hierin bezeichnen

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} \sum_{l=1}^n (x_l - \bar{x})^2; & s_y^2 &= \frac{1}{n-1} \sum_{l=1}^n (y_l - \bar{y})^2; \\ s_{xy} &= \frac{1}{n-1} \sum_{l=1}^n (x_l - \bar{x})(y_l - \bar{y}) \end{aligned} \quad (4.10)$$

die empirischen Varianzen und die empirische Kovarianz. Von dem sich in (4.7) und (4.9) ausdrückenden statistischen Modell gleich vieler Wiederholungsmessungen für die beteiligten Messgrößen, hier x und y , macht die Standardliteratur nur dann Gebrauch, wenn die Messsituation von sich aus hierfür sorgt, wie bei der Konstruktion einer Ausgleichsgeraden oder eines Ausgleichskreises.

Im Realfall allerdings steht der Experimentator vor folgender Situation: Stehen ihm, sagen wir, $n_x = 10$ Wiederholungsmessungen für x und $n_y = 8$ Wiederholungsmessungen für y zur Verfügung, so wertet er seine Daten unter der Prämisse $n_x \neq n_y$ aus. Er tut das, weil er es als unvorteilhaft ansehen würde, nutzbare Information zu verschenken – wobei verschenken hieße, den Überschuss von zwei Messungen für x zu ignorieren, um in Konsequenz dieses Verzichts mit *gleich vielen* Wiederholungsmessungen für x und y arbeiten zu können.

Vordergründig gesehen, verarbeitet die Situation $n_x \neq n_y$ tatsächlich ein Mehr an Information. Allerdings dürfte es schwierig sein, eine Verteilungsdichte zu definieren, die unterschiedlich große Stichproben für die Zufallsvariablen X und Y zuließe. Insbesondere wären Datenpaare (x_l, y_l) , mittels derer empirische Kovarianzen gebildet werden könnten, nicht definiert. Dessen ungeachtet, selbst unter der Prämisse $n_x \neq n_y$, lassen sich nach den Überlegungen von Welch [8] zumindest approximativ Vertrauensbereiche konstruieren. Leider aber ist der Welch'sche Formalismus an die Unabhängigkeit der Messgrößen gebunden und darüber hinaus doch auch relativ unhandlich. So interessant dieses Konzept auch sein mag, seine Anwendungsmöglichkeiten dürften sich auf Sonderfälle beschränken. Ob es vor diesem Hintergrund Sinn hat, dem Welch'schen Konzept im Rahmen einer revidierten Fehlerrechnung Raum zu geben, entscheide die alternative Überlegung, der Rückgriff auf

das *Standardmodell normalverteilter Zufallsvariablen*. Im zweidimensionalen Fall ist das die Fisher'sche, im mehrdimensionalen Fall die Wishart'sche Dichte [9; 10]. Beide Dichten schließen die *empirischen* Varianz-Kovarianz-Matrizen der beteiligten Messgrößen ein und setzen damit *gleich viele* Wiederholungsmessungen voraus.

Diese Dichten werden uns die Möglichkeit geben, auch im Rahmen der Fehlerfortpflanzung mit Student'schen Vertrauensbereichen zu arbeiten. Dabei erweitert der *t*-Faktor den auf (normalverteilte) zufällige Messfehler zurückgehenden Anteil der Gesamtunsicherheit auf statistisch wohldefinierter Basis – je nach Anzahl vorliegender Wiederholungsmessungen und gewähltem Vertrauensniveau. Im Prinzip verläuft alles so wie im Falle einer einzigen Messgröße. Bemerkenswerterweise bleibt es dabei gleichgültig, ob die Messgrößen voneinander abhängen oder nicht. Das Standardmodell der Statistik für normalverteilte Zufallsvariablen leistet das eben. – Vor diesem Hintergrund erscheint die Welch'sche Approximation kaum noch attraktiv.

Die Fisher'sche Dichte beschreibt das statistische Verhalten der fünf empirischen Variablen

$$\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}.$$

Die Verteilungsdichte, deren explizite Form wir hier nicht hinschreiben wollen [9; 10], faktorisiert in der Form

$$p(\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}) = p_1(\bar{x}, \bar{y}) p_2(s_x^2, s_y^2, s_{xy}),$$

da \bar{x}, \bar{y} unabhängig von s_x^2, s_y^2, s_{xy} sind. Sind x und y selbst unabhängig voneinander, so faktorisiert p_1 nochmals, p_2 bemerkenswerterweise *nicht*. Diese letztere Eigenschaft deckt eine Abhängigkeit zwischen den Variablen s_x^2, s_y^2 und s_{xy} auf, d. h. p_2 verlangt von uns, die empirische Kovarianz s_{xy} mitzunehmen, und zwar auch dann, wenn die Messgrößen x und y unabhängig voneinander sind, der Erwartungswert der empirischen Kovarianz also verschwindet.

Was aber könnte s_{xy} angesichts der Unabhängigkeit der Zufallsvariablen X und Y bedeuten? Die Antwort lautet: Die empirische Kovarianz ist unverzichtbarer Bestandteil des Standardmodells – selbst wenn sie dort lediglich in Gestalt einer Rechengröße auftritt. Sie zu ignorieren hieße, das Standardmodell zu verletzen. Natürlich kann man das. Jedoch sollten wir die Konsequenzen hinterfragen. Betrachten wir hierzu ein aufschlussreiches Beispiel:

Der Experimentator habe zu entscheiden, ob die Differenz zweier unabhängiger arithmetischer Mittel \bar{x} und \bar{y} , die ein und derselben physikalischen Größe gelten, signifikant ist oder nicht. Uns interessiere dabei *allein* die den Test realisierende Zufallsvariable – nichts weiter. Wir nehmen an, die Zufallsvariablen X und Y genügen derselben Normalverteilung $N(\mu, \sigma)$. Die dem Autor bekannten Textbücher setzen *ungleich viele* Wiederholungsmessungen n_x und n_y

für die Messgrößen x und y voraus,

$$\bar{x} = \frac{1}{n_x} \sum_{l=1}^{n_x} x_l; \quad \bar{y} = \frac{1}{n_y} \sum_{l=1}^{n_y} y_l;$$

$$s_x^2 = \frac{1}{n_x - 1} \sum_{l=1}^{n_x} (x_l - \bar{x})^2; \quad s_y^2 = \frac{1}{n_y - 1} \sum_{l=1}^{n_y} (y_l - \bar{y})^2.$$

In Konsequenz der Voraussetzung ist die empirische Kovarianz nicht definierbar. Die Standardliteratur, z. B. [9], führt eine Zufallsvariable der Form

$$T^* = \frac{\bar{X} - \bar{Y}}{\sqrt{(n_x - 1) S_x^2 + (n_y - 1) S_y^2}} \sqrt{\frac{n_x n_y}{n_x + n_y} (n_x + n_y - 2)}$$

ein und unterstellt, T^* sei mit $(n_x + n_y - 2)$ Freiheitsgraden *t*-verteilt. Natürlich müsste diese Aussage auch für $n_x = n_y = n$ gelten, d. h. für

$$T^* = \frac{\bar{X} - \bar{Y}}{\sqrt{S_x^2 + S_y^2}} \sqrt{n}.$$

Bemerkenswerterweise führt unsere Prämisse $n_x = n_y = n$ zu einem anderen Resultat. Die Differenzen $Z_l = X_l - Y_l$ ($l = 1, \dots, n$) erzeugen nämlich

$$\bar{Z} = \frac{1}{n} \sum_{l=1}^n Z_l;$$

$$S_Z^2 = \frac{1}{n - 1} \sum_{l=1}^n (Z_l - \bar{Z})^2 = S_x^2 - 2S_{xy} + S_y^2,$$

woraus, *per definitionem*, die exakt *t*-verteilte Zufallsvariable

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_x^2 - 2S_{xy} + S_y^2}} \sqrt{n}$$

folgt. – *Wohldefinierte Messbedingungen* räumen dem Experimentator die Möglichkeit ein, mit an sich durchaus bekannten Tools der Statistik zu arbeiten, wobei T nicht schwieriger zu konstruieren ist als jenes, letztlich auf R.A. Fisher zurückgehende T^* .

Die $\phi(x_l, y_l)$ mit $l = 1, 2, \dots, n$ in (4.6) sind voneinander unabhängig, wenn jedes der x_l von jedem anderen x_l unabhängig ist und gleiches für die y_l gilt. Sind die x_l, y_l darüber hinaus normalverteilt, so sind auch die $\phi(x_l, y_l)$ normalverteilt, und zwar gleichgültig, ob die Zufallsvariablen X und Y voneinander unabhängig sind oder nicht [11]. Dann aber ist

$$\phi(\bar{X}, \bar{Y}) - \frac{t_P(n-1)}{\sqrt{n}} S_\phi \leq \mu_\phi \leq \phi(\bar{X}, \bar{Y}) + \frac{t_P(n-1)}{\sqrt{n}} S_\phi \quad (4.11)$$

ein Vertrauensintervall für den experimentell unzugänglichen Erwartungswert

$$\begin{aligned} \mu_\phi &= E\{\phi(X, Y)\} = E\{\phi(\bar{X}, \bar{Y})\} \\ &= \phi(x_0, y_0) + \frac{\partial \phi}{\partial \bar{x}} f_x + \frac{\partial \phi}{\partial \bar{y}} f_y. \end{aligned} \quad (4.12)$$

Unter der Prämisse gleich vieler Wiederholungsmessungen für die Messgrößen x und y , d. h. unter Wahl des Modells der Fisher'schen Dichte, kann also auch im Falle zweier Variabler auf die Student'sche Technik zurückgegriffen werden, Vertrauensbereiche zu definieren. Dabei unterstreicht (4.9) die Rolle der empirischen Kovarianz. Führen wir noch den Spaltenvektor der partiellen Ableitungen

$$\mathbf{b} = \left(\frac{\partial \phi}{\partial \bar{x}} \quad \frac{\partial \phi}{\partial \bar{y}} \right)^T \quad (4.13)$$

und des Weiteren die Notation

$$\mathbf{s} = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{pmatrix}, \quad (4.14)$$

$$s_{xx} \equiv s_x^2; \quad s_{xy} = s_{yx}; \quad s_{yy} \equiv s_y^2$$

der empirischen Varianz-Kovarianz-Matrix der Momente s_{xx}, s_{xy}, s_{yy} ein, so geht (4.9) über in

$$s_\phi^2 = \mathbf{b}^T \mathbf{s} \mathbf{b}.$$

Wir sollten zwischen Kovarianzen unterscheiden, die durch die Wahl der Variablen bedingt und in dieser Eigenschaft eliminierbar sind, und zwischen Kovarianzen, die auf interne, seitens des Experimentators weder einsehbar noch beeinflussbare Abhängigkeiten innerhalb des physikalischen Geschehens zurückgehen. Für den ersteren Fall ein Beispiel:

Neben x fließe eine weitere Messgröße y in $\phi(x, y)$ ein, dabei gelte $y = \psi(x)$. Zur Illustration diene $y = ax$. Zu bilden sei s_ϕ^2 . Einsetzen von

$$y_l = a\bar{x} + a(x_l - \bar{x})$$

in

$$\phi(x_l, y_l) = \phi(\bar{x}, \bar{y}) + \frac{\partial \phi}{\partial \bar{x}} (x_l - \bar{x}) + \frac{\partial \phi}{\partial \bar{y}} (y_l - \bar{y})$$

liefert zunächst direkt

$$s_\phi^2 = \left(\frac{\partial \phi}{\partial \bar{x}} + a \frac{\partial \phi}{\partial \bar{y}} \right)^2 s_x^2.$$

Sollten wir andererseits

$$s_\phi^2 = \left(\frac{\partial \phi}{\partial \bar{x}} \right)^2 s_x^2 + 2 \frac{\partial \phi}{\partial \bar{x}} \frac{\partial \phi}{\partial \bar{y}} s_{xy} + \left(\frac{\partial \phi}{\partial \bar{y}} \right)^2 s_y^2$$

bilden wollen oder müssen, so führen uns $s_y^2 = a^2 s_x^2$ und $s_{xy} = a s_x^2$ ebenfalls zu

$$s_\phi^2 = \left(\frac{\partial \phi}{\partial \bar{x}} \right)^2 s_x^2 + 2a \frac{\partial \phi}{\partial \bar{x}} \frac{\partial \phi}{\partial \bar{y}} s_x^2 + a^2 \left(\frac{\partial \phi}{\partial \bar{y}} \right)^2 s_x^2$$

$$= \left(\frac{\partial \phi}{\partial \bar{x}} + a \frac{\partial \phi}{\partial \bar{y}} \right)^2 s_x^2$$

zurück.

5 Gesamtunsicherheiten

Wenn wir das Student'sche Vertrauensintervall linear mit der Worst-Case-Abschätzung

$$f_{s,\phi} = \left| \frac{\partial \phi}{\partial \bar{x}} \right| f_{s,x} + \left| \frac{\partial \phi}{\partial \bar{y}} \right| f_{s,y} \quad (5.1)$$

des systematischen Fehleranteils (4–3) verknüpfen, dürfen wir davon ausgehen, dass eine Gesamtunsicherheit der Form

$$u_{\bar{\phi}} = \frac{t_P (n-1)}{\sqrt{n}}$$

$$\times \sqrt{\left(\frac{\partial \phi}{\partial \bar{x}} \right)^2 s_x^2 + 2 \left(\frac{\partial \phi}{\partial \bar{x}} \right) \left(\frac{\partial \phi}{\partial \bar{y}} \right) s_{xy} + \left(\frac{\partial \phi}{\partial \bar{y}} \right)^2 s_y^2}$$

$$+ \left| \frac{\partial \phi}{\partial \bar{x}} \right| f_{s,x} + \left| \frac{\partial \phi}{\partial \bar{y}} \right| f_{s,y} \quad (5.2)$$

sicher ist – vielleicht sagen wir besser „quasi sicher“. D. h. wir unterstellen, dass das Intervall

$$\phi(\bar{x}, \bar{y}) - u_{\bar{\phi}} \leq \phi(x_0, y_0) \leq \phi(\bar{x}, \bar{y}) + u_{\bar{\phi}} \quad (5.3)$$

den wahren Wert $\phi(x_0, y_0)$ lokalisiert. Bevor wir diese Aussage diskutieren, einige Anmerkungen bezüglich der Approximationsfehler in den partiellen Ableitungen. Betrachten wir das Beispiel $\phi(x, y) = x/y^2$. Seien die relativen Unsicherheiten $u_{\bar{x}/\bar{x}}$ und $u_{\bar{y}/\bar{y}}$ der beiden Mittel \bar{x} und \bar{y} jeweils 10^{-6} . Dann finden wir

$$\frac{1}{\bar{y}^2} (1 - 2 * 10^{-6}) \leq \frac{\partial \phi}{\partial x_0} \leq \frac{1}{\bar{y}^2} (1 + 2 * 10^{-6})$$

und

$$-2 \frac{\bar{x}}{\bar{y}^3} (1 + 4 * 10^{-6}) \leq \frac{\partial \phi}{\partial y_0} \leq -2 \frac{\bar{x}}{\bar{y}^3} (1 - 4 * 10^{-6}).$$

Gegebenenfalls könnten die Schwankungen der partiellen Ableitungen also schon von Bedeutung sein. Wenigstens aber unterstreicht das Beispiel, dass Messunsicherheiten nicht zu niedrig gehalten werden sollten. Ein konkretes Beispiel für die Lokalisierungsfähigkeit von Messunsicherheiten findet sich in [7].

Legen wir dem auf zufällige Fehler zurückgehenden Unsicherheitsanteil eine Wahrscheinlichkeit von $P = 95\%$ zugrunde, so dürfen wir davon ausgehen, dass der Erwartungswert μ_ϕ durch (4–11) lokalisiert wird. Die verbleibenden 5% Ungewissheit werden in aller Regel seitens der empirischen Beobachtung aufgefangen, dass experimentell erzeugte zufällige Fehler quasi-normalverteilt sind, d. h. dass die extremen Ausläufer der Dichte praktisch nicht realisiert werden. Die Worst-Case-Abschätzung der unbekannt systematischen Fehler birgt zwar die Gefahr einer Überbewertung des Einflusses systematischer Fehler, andererseits aber wissen wir nichts Definitives über die tatsächlichen Werte der Fehlergrößen f_x und f_y , und jeder Versuch, (5.2) zu reduzieren, würde dazu führen, dass der wahre Wert nicht mehr lokalisiert werden würde, wenn f_x und f_y ihre Fehlerintervalle ausschöpften. Wollen wir den wahren Wert mit Sicherheit lokalisieren, sollten wir die

Möglichkeit des ungünstigsten Falles zulassen. Nicht zuletzt haben wir in (5.2) die Schwankungen der partiellen Ableitungen der Verknüpfungsfunktion $\phi(x, y)$ in den Intervallen $\bar{x} \pm u_{\bar{x}}$ und $\bar{y} \pm u_{\bar{y}}$ vernachlässigt und überdies die Reihenentwicklungen nach den linearen Gliedern abgebrochen.

6 Die Toolbox der „Metrologischen Statistik“

Wir zeigen, dass die Verknüpfungsfunktionen bausteinähnlich auf beliebig viele Variable ausdehnbar sind. In Frage stehe die Unsicherheit u_ϕ einer Größe $\phi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m)$. Jedes der m Mittel \bar{x}_i umfasse genau n Wiederholungsmessungen,

$$\bar{x}_i = \frac{1}{n} \sum_{l=1}^n x_{il} = x_{0,i} + (\bar{x}_i - \mu_i) + f_i \quad (i = 1, \dots, m). \tag{6.1}$$

Sicherheitshalber stellen wir die Forderung nach gleich vielen Wiederholungsmessungen noch einmal anhand einer Graphik dar.

$$\begin{array}{ccccccc} x_{11} & x_{12} & \dots & x_{1n} & \rightarrow & \bar{x}_1 & \\ x_{21} & x_{22} & \dots & x_{2n} & \rightarrow & \bar{x}_2 & \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ x_{m1} & x_{m2} & \dots & x_{mn} & \rightarrow & \bar{x}_m & \end{array} \Rightarrow \phi(x_1, x_2, \dots, x_m)$$

Im Falle von mehr als zwei Variablen tritt die Wishart'sche Dichte an die Stelle der Fisher'schen Dichte. Natürlich legt uns auch diese Dichte nahe, die empirischen Kovarianzen s_{ij} zu berücksichtigen [9]; d.h. das Standardmodell normalverteilter Messwerte verlangt von uns, mit einer empirischen Varianz-Kovarianz Matrix der Form

$$s = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1m} \\ s_{21} & s_{22} & \dots & s_{2m} \\ \dots & \dots & \dots & \dots \\ s_{m1} & s_{m2} & \dots & s_{mm} \end{pmatrix}, \tag{6.2}$$

$$s_{ij} = \frac{1}{n-1} \sum_{l=1}^n (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j)$$

zu arbeiten. In (6.2) gelte wieder $s_{ii} \equiv s_i^2$; des Weiteren natürlich $s_{ij} = s_{ji}$. Solange die Linearisierungsfehler der Verknüpfungsfunktion $\phi(x_1, x_2, \dots, x_m)$ hinsichtlich der Unsicherheiten

$$u_{\bar{x}_i} = \frac{t_P(n-1)}{\sqrt{n}} s_i + f_{s,i} \quad (i = 1, \dots, m) \tag{6.3}$$

der Mittel \bar{x}_i hinreichend klein bleiben, haben wir im Wesentlichen die gleiche Situation wie im Falle zweier Variabler: Die linearisierte Taylor'sche Entwicklung generiert die empirische Varianz

$$s_\phi^2 = \sum_{i,j} \frac{\partial \phi}{\partial \bar{x}_i} \frac{\partial \phi}{\partial \bar{x}_j} s_{ij} = \mathbf{b}^T \mathbf{s} \mathbf{b} \tag{6.4}$$

der zufälligen Fehler, worin

$$\mathbf{b} = \left(\frac{\partial \phi}{\partial \bar{x}_1} \quad \frac{\partial \phi}{\partial \bar{x}_2} \quad \dots \quad \frac{\partial \phi}{\partial \bar{x}_m} \right)^T \tag{6.5}$$

den Spaltenvektor der partiellen Ableitungen bezeichnet und des Weiteren den fortgepflanzten systematischen Fehler oder Bias

$$f_\phi = \sum_{i=1}^m \frac{\partial \phi}{\partial \bar{x}_i} f_i \tag{6.6}$$

der m systematischen Fehler

$$-f_{s,i} \leq f_i \leq f_{s,i} \quad (i = 1, \dots, m). \tag{6.7}$$

Nach allem lassen sich Gesamtunsicherheit schematisch bilden. Im Sinne der Argumentation (5–2) liefern (6.4) und (6.6)

$$u_\phi = \frac{t_P(n-1)}{\sqrt{n}} \sqrt{\sum_{i,j} \frac{\partial \phi}{\partial \bar{x}_i} \frac{\partial \phi}{\partial \bar{x}_j} s_{ij}} + \sum_{i=1}^m \left| \frac{\partial \phi}{\partial \bar{x}_i} \right| f_{s,i}, \tag{6.8}$$

wobei wir annehmen, das Intervall

$$\phi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m) \pm u_\phi \tag{6.9}$$

lokalisieren den wahren Wert $\phi(x_{0,1}, x_{0,2}, \dots, x_{0,m})$.

Die so definierte Unsicherheit (6.8) ließe sich nur durch Hinzunehmen weiterer Annahmen verringern. Entschlüsse man sich dazu, verlöre die so definierte Unsicherheit ihre wesentliche Eigenschaft, auch dann noch sicher zu sein, wenn alle systematischen Fehler f_i ihre Intervalle (6.7) ausschöpfen. Wie wahrscheinlich das ist, wissen wir nicht – aber auch über (6.8) verkürzende Möglichkeiten wissen wir nichts.

Halten wir also fest: Unter realistischen Gesichtspunkten spezifiziert die Messunsicherheit (6.8) das kleinstmögliche Intervall (6.9), das den wahren Wert $\phi(x_{0,1}, x_{0,2}, \dots, x_{0,m})$ der Messgröße lokalisiert – wenn nicht mit mathematischer, so doch mit einer an den Erfordernissen der Praxis orientierten Sicherheit.

Der Einwand, (6.8) verlange die Einzelmessungen aller m Eingangsgrößen \bar{x}_i , trifft zu. Denn nur wenn sämtliche Wiederholungsmessungen $x_{i,l} (i = 1, \dots, m; l = 1, \dots, n)$ vorliegen, lässt sich die Varianz-Kovarianz-Matrix (6.2) darstellen. Indessen, im Zeitalter der Tischrechner und des Internet bereiten weder das Archivieren von Messdaten noch ihr wechselseitiger Austausch (über beliebige) Distanzen Probleme.

Grenzbetrachtung:

Verzichtet man auf kleinstmögliche Messunsicherheiten, so lassen (4–9) und (6–4) interessante Vereinfachungen zu. Wie bekannt, genügt die empirische Kovarianz s_{xy} der Einschränkung

$$-s_x s_y \leq s_{xy} \leq s_x s_y. \tag{6.10}$$

Hiermit können wir, falls wir das wollen, den Radikanden in (4–9) vergrößern,

$$\begin{aligned} & \left(\frac{\partial\phi}{\partial\bar{x}}\right)^2 s_x^2 + 2\left(\frac{\partial\phi}{\partial\bar{x}}\right)\left(\frac{\partial\phi}{\partial\bar{y}}\right) s_{xy} + \left(\frac{\partial\phi}{\partial\bar{y}}\right)^2 s_y^2 \\ & \leq \left(\frac{\partial\phi}{\partial\bar{x}}\right)^2 s_x^2 + 2\left|\frac{\partial\phi}{\partial\bar{x}}\right|\left|\frac{\partial\phi}{\partial\bar{y}}\right| s_x s_y + \left(\frac{\partial\phi}{\partial\bar{y}}\right)^2 s_y^2 \\ & = \left(\left|\frac{\partial\phi}{\partial\bar{x}}\right| s_x + \left|\frac{\partial\phi}{\partial\bar{y}}\right| s_y\right)^2, \end{aligned}$$

sodass

$$u_\phi = \frac{t_P(n-1)}{\sqrt{n}} \left(\left|\frac{\partial\phi}{\partial\bar{x}}\right| s_x + \left|\frac{\partial\phi}{\partial\bar{y}}\right| s_y \right) + \left|\frac{\partial\phi}{\partial\bar{x}}\right| f_x + \left|\frac{\partial\phi}{\partial\bar{y}}\right| f_y,$$

d. h.

$$u_\phi = \left|\frac{\partial\phi}{\partial\bar{x}}\right| u_{\bar{x}} + \left|\frac{\partial\phi}{\partial\bar{y}}\right| u_{\bar{y}}, \tag{6.11}$$

worin $u_{\bar{x}} = \frac{t_P(n-1)}{\sqrt{n}} s_x + f_{s,x}$ und $u_{\bar{y}} = \frac{t_P(n-1)}{\sqrt{n}} s_y + f_{s,y}$ die Unsicherheiten der Mittel \bar{x} und \bar{y} sind. Eine ähnliche Vereinfachung lässt (6–8) zu, wenn wir jeweils zwei Variable

$$(x_1, x_2), (x_1, x_3), \dots, (x_{m-1}, x_m)$$

separat betrachten,

$$u_\phi = \sum_{i=1}^m \left|\frac{\partial\phi}{\partial\bar{x}_i}\right| u_{\bar{x}_i}; \quad u_{\bar{x}_i} = \frac{t_P(n-1)}{\sqrt{n}} s_i + f_{s,i}. \tag{6.12}$$

7 Elementare Beispiele

Für die beiden Messreihen

$$\{x_1, x_2, \dots, x_n\} \text{ und } \{y_1, y_2, \dots, y_n\}$$

seien die Messunsicherheiten

$$\bar{x} \pm \left(\frac{t_P(n-1)}{\sqrt{n}} s_x + f_{s,x}\right) \text{ und } \bar{y} \pm \left(\frac{t_P(n-1)}{\sqrt{n}} s_y + f_{s,y}\right)$$

geschätzt worden. Wir formulieren $\phi(\bar{x}, \bar{y}) \pm u_\phi$, wenn $\phi(x, y)$ die Summe, die Differenz, das Produkt und den Quotienten der Messgrößen x und y bezeichnet. Zum Schluss analysieren wir das Schicksal der Varianz-Analyse, wenn der Experimentator a priori davon auszugehen hat, seinen Messdaten seien unbekannte systematische Messfehler überlagert.

Summe $\phi(x, y) = x + y$: Die Entwicklungen

$$\begin{aligned} \phi(x_l, y_l) &= x_0 + y_0 + (x_l - \mu_x) + (y_l - \mu_y) + f_x + f_y \\ \phi(\bar{x}, \bar{y}) &= x_0 + y_0 + (\bar{x} - \mu_x) + (\bar{y} - \mu_y) + f_x + f_y \end{aligned}$$

definieren die Differenz

$$\phi(x_l, y_l) - \phi(\bar{x}, \bar{y}) = (x_l - \bar{x}) + (y_l - \bar{y}),$$

sodass

$$s_\phi^2 = \frac{1}{n-1} \sum_{l=1}^n [(x_l - \bar{x}) + (y_l - \bar{y})]^2 = s_x^2 + 2s_{xy} + s_y^2.$$

Der fortgepflanzte systematische Messfehler ist

$$f_\phi = f_x + f_y,$$

d. h.

$$(\bar{x} + \bar{y}) \pm \frac{t_P(n-1)}{\sqrt{n}} \sqrt{s_x^2 + 2s_{xy} + s_y^2} + f_{s,x} + f_{s,y}.$$

Differenz $\phi(x, y) = x - y$: Zu entscheiden sei, ob die Mittel \bar{x} und \bar{y} „miteinander verträglich sind“, unter der Prämisse, dass beide Mittel derselben physikalischen Größe gelten.

Wegen

$$(\bar{x} - \bar{y}) \pm \frac{t_P(n-1)}{\sqrt{n}} \sqrt{s_x^2 - 2s_{xy} + s_y^2} + f_{s,x} + f_{s,y}$$

sind die beiden Mittel als widerspruchsfrei oder miteinander verträglich anzusehen, solange

$$|\bar{x} - \bar{y}| \leq \frac{t_P(n-1)}{\sqrt{n}} \sqrt{s_x^2 - 2s_{xy} + s_y^2} + f_{s,x} + f_{s,y}.$$

Ließen wir (6–11) zu, hätten wir stattdessen

$$|\bar{x} - \bar{y}| \leq u_{\bar{x}} + u_{\bar{y}}.$$

Im Rahmen so genannter *Key Comparisons* ist die Verträglichkeit der Messresultate der Teilnehmer gesichert, wenn deren Unsicherheiten eine sinnvolle „Schnittmenge“ definieren, d. h. wenn sich die Unsicherheitsangaben aller Teilnehmer wechselseitig überlappen.

Produkt $\phi(x, y) = xy$: Aus

$$\begin{aligned} \phi(x_l, y_l) &= x_0 y_0 + \bar{y}(x_l - \mu_x) + \bar{x}(y_l - \mu_y) + \bar{y} f_x + \bar{x} f_y \\ \phi(\bar{x}, \bar{y}) &= x_0 y_0 + \bar{y}(\bar{x} - \mu_x) + \bar{x}(\bar{y} - \mu_y) + \bar{y} f_x + \bar{x} f_y \end{aligned}$$

folgen

$$\begin{aligned} s_\phi^2 &= \frac{1}{n-1} \sum_{l=1}^n [\bar{y}(x_l - \bar{x}) + \bar{x}(y_l - \bar{y})]^2 \\ &= \bar{y}^2 s_x^2 + 2\bar{x}\bar{y} s_{xy} + \bar{x}^2 s_y^2 \end{aligned}$$

und

$$f_\phi = \bar{y} f_x + \bar{x} f_y,$$

sodass

$$\begin{aligned} \bar{x}\bar{y} \pm \frac{t_P(n-1)}{\sqrt{n}} \sqrt{\bar{y}^2 s_x^2 + 2\bar{x}\bar{y} s_{xy} + \bar{x}^2 s_y^2} \\ + |\bar{y}| f_{s,x} + |\bar{x}| f_{s,y}. \end{aligned}$$

Quotient $\phi(x, y) = x/y$:

$$\begin{aligned} \frac{\bar{x}}{\bar{y}} \pm \frac{t_P(n-1)}{\sqrt{n}} \sqrt{\frac{1}{\bar{y}^2} s_x^2 - 2\frac{\bar{x}}{\bar{y}^3} s_{xy} + \frac{\bar{x}^2}{\bar{y}^4} s_y^2} \\ + \frac{1}{|\bar{y}|} f_{s,x} + \frac{|\bar{x}|}{\bar{y}^2} f_{s,y}. \end{aligned}$$

Varianz-Analyse: Im Rahmen der *Key Comparisons* messen mehrere Laboratorien an ein und demselben Objekt ein und dieselbe physikalische Größe. Um entscheiden zu können, ob die arithmetischen Mittelwerte der Teilnehmer

„miteinander verträglich“ sind, werden die Messdaten einer Varianz-Analyse unterworfen. Wir gehen von m beteiligten Laboratorien aus, die jeweils n Wiederholungsmessungen durchgeführt haben,

$$\begin{aligned} \text{Labor 1} & x_{11}, x_{12}, \dots, x_{1n} \\ \text{Labor 2} & x_{21}, x_{22}, \dots, x_{2n} \\ & \dots\dots\dots \\ \text{Labor } m & x_{m1}, x_{m2}, \dots, x_{mn}. \end{aligned}$$

Wir unterstellen, die x_{il} seien normalverteilt, voneinander unabhängig und ließen sich überdies ein und derselben Grundgesamtheit zuordnen. Aus den $N = m \cdot n$ Messdaten bilden wir dann zunächst das „große Mittel“

$$\bar{x} = \frac{1}{N} \sum_{i=1}^m \sum_{l=1}^n x_{il}$$

und die zu analysierende „große Varianz“

$$s^2 = \frac{1}{N-1} \sum_{i=1}^m \sum_{l=1}^n (x_{il} - \bar{x})^2.$$

Das Aufbrechen letzterer mittels der Identität

$$x_{il} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{il} - \bar{x}_i)$$

liefert

$$(N-1) s^2 = n \sum_{i=1}^m (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^m \sum_{l=1}^n (x_{il} - \bar{x}_i)^2,$$

da der gemischte Term verschwindet. Zunächst folgen wir der traditionellen Notation und stellen die beiden Summanden der rechten Seite als „empirische Varianzen“ dar,

$$(N-1) s^2 = (m-1) s_1^2 + (N-m) s_2^2,$$

mit

$$s_1^2 = \frac{n}{m-1} \sum_{i=1}^m (\bar{x}_i - \bar{x})^2$$

und

$$s_2^2 = \frac{1}{N-m} \sum_{i=1}^m \sum_{l=1}^n (x_{il} - \bar{x}_i)^2.$$

In s_2^2 heben sich die systematischen Fehler heraus, also haben wir $E \{S_2^2\} = \sigma^2$. Für s_1^2 indessen gilt

$$\begin{aligned} E \left\{ \frac{n}{m-1} \sum_{i=1}^m (\bar{X}_i - \bar{X})^2 \right\} \\ = \sigma^2 + \frac{n}{(m-1)} \sum_{i=1}^m \left(f_i - \frac{1}{m} \sum_{j=1}^m f_j \right)^2, \end{aligned}$$

d. h. s_1^2 ist keine empirische Varianz. Nach allem erübrigt es sich, aus s_1^2 und s_2^2 die Testgröße $F = S_1^2/S_2^2$ der Varianz-Analyse zu bilden, wenn a priori davon auszugehen ist, dass den Messdaten unbekannte systematische Messfehler überlagert sind.

Würde man dennoch versuchen, unbekannte systematische Fehler mittels einer Varianz-Analyse aufzuspüren, würde man sich auf eine undurchschaubare Mischung der Einflüsse statistischer Fehler und unbekannter systematischer Fehler einlassen.

Literatur

- [1] Grabe, M., *Gedanken zur Revision der Gauß'schen Fehlerrechnung*, tm-Technisches Messen **67** (2000) 283–288.
- [2] Kessel, W., *The ISO/BIPM-Guide for the Evaluation of Measurement Uncertainty*, tm-Technisches Messen **68** (2001) 5–13.
- [3] Gauß, K.F., *Werke*, Vol. 4, Dietrich, Göttingen 1880.
- [4] Eisenhart, C., *The Reliability of Measured Values – Part I, Fundamental Concepts*, Photogrammetric Engineering **18** (1952) 543–561.
- [5] Wagner, S., *Zur Behandlung systematischer Messfehler bei der Angabe von Messunsicherheiten*, PTB-Mitt. **79** (1969) 343–347.
- [6] Grabe, M., *Principles of „Metrological Statistics“*, metrologia **23** (1986/87) 213–219.
- [7] Grabe, M., *Estimation of measurement uncertainties, an alternative to the ISO-Guide*, metrologia **38** (2001) 97–106.
- [8] Welch, B. L., *The Generalization of Student's Problem when Several Different Population Variances are Involved*, Biometrika **37** (1950) 149–157.
- [9] Fisz, M., *Probability theory and mathematical statistics*, John Wiley&Sons, New York 1967.
- [10] Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton 1966.
- [11] Graybill, F.A., *An Introduction to Linear Statistical Models*, McGraw-Hill, New York 1961.

Dr. Michael Grabe, Am Hasselteich 5, D-38104 Braunschweig, E-Mail: michael.grabe@ptb.de, www.uncertainty.de