ISSN 1013-4506
FS-2010-153-T

Fachverband für Strahlenschutz e.V.

Mitgliedsgesellschaft der International Radiation Protection Association (IRPA) für die Bundesrepublik Deutschland und die Schweiz

Publikationsreihe FORTSCHRITTE IM STRAHLENSCHUTZ

Publication Series PROGRESS IN RADIATION PROTECTION

NATÜRLICHE UND KÜNSTLICHE RADIONUKLIDE IN UNSERER UMWELT

42. Jahrestagung des Fachverbandes für Strahlenschutz e. V.

26. bis 30. September 2010
Borkum

Bandherausgeber:
Alfred Neu
Anton Bayer
Thomas Steinkopff

NEUE ENTWICKLUNGEN IN DER METROLOGIE: BAYES ODER FREQUENTISTISCHE STATISTIK?

NEW DEVELOPMENTS IN METROLOGY: BAYESIAN OR FREQUENTISTIC STATISTICS?

R. Michel

Institut für Radioökologie und Strahlenschutz, Leibniz Universität Hannover, Deutschland

Zusammenfassung
Messunsicherheiten sind unverzichtbare Bestandteile der Qualitätskontrolle in allen Bereichen der Metrologie. Dies schließt auch Quantifizierung von Risiken in epidemiologischen Studien ein. In der Metrologie werden Messunsicherheiten auf der Grundlage des ISO Guide to the Expression of Uncertainty in Measurement (GUM) und eines kürzlich erschienenen Supplements zum GUM ermittelt. Da unter den Anwendern vielfach Verwirrung über die statistischen Grundlagen dieser Methodik herrscht, werden diese Grundlagen hier in einfacher Form vorgestellt. Insbesondere wird dargestellt, dass nur eine Bayes'sche Theorie der Messunsicherheiten alle Quellen der Unsicherheit berücksichtigen kann. Die frequentistische Statistik kann dies nicht. Ursache ist der Unterschiedlichkeit des Wahrscheinlichkeitsbegriffs in den beiden Schulen der Statistik.

Summary
Measurement uncertainties are indispensable elements of quality control in all fields of metrology. This includes also the quantification of risks in epidemiological studies. In metrology, measurement uncertainties are derived on the basis of the ISO Guide to the Expression of Uncertainty in Measurement (GUM) and its recently published Supplement 1. Frequently, there is confusion among the users about the statistical foundation of the GUM. Therefore, the basics are described here in a simple form. In particular, it is explained that only a Bayesian theory of measurement uncertainties is capable to handle all sources of uncertainty. The frequentistic statistics cannot achieve this goal. The reasons are the differing concepts of probability in the two schools of statistics.

Schlüsselwörter *Messunsicherheiten, Bayes Statistik, frequentistische Statistik, Wahrscheinlichkeit*
Keywords *measurement uncertainties, Bayesian statistics, frequentistic statistics, probability*

1. Einleitung

Während der nahezu drei Jahrzehnte der Arbeit des AK SIGMA haben sich im Messwesen dramatische Entwicklungen hinsichtlich der Behandlung von Messunsicherheiten ergeben. Nachdem Messunsicherheiten in Deutschland bereits durch DIN 1319 standardisiert worden waren, erschien im Jahr 1995 der ISO „Guide to the expression of uncertainty in measurement" (GUM) [1], der heute als international verbindliche Grundlage zur Bestimmung von Messunsicherheiten gilt. Die Methodik des GUM ist allgemein anerkannt und wurde vielfach adaptiert. Von den statistischen Grundlagen her war der GUM jedoch nicht frei von Widersprüchen, da frequentistische und Bayes-statistische Ansätze miteinander vermengt wurden. Erst im Jahr 2008 erschien eine Ergänzung zum GUM [2], in der eine Erweiterung der GUM-

Methodik auf nicht-linearisierbare Modelle erfolgte und gleichzeitig klargestellt wurde, dass die GUM-Methodik nur mittels Bayes-Statistik begründbar ist. Dies soll in dieser Arbeit, die lediglich eine Erläuterung darstellt zur Bayes Theorie der Messunsicherheiten [3,4], deutlich gemacht werden.

2. Messung, Messwert, Messunsicherheit und Wahrscheinlichkeit

Am Anfang einer Messung steht die Festlegung der Messgröße (engl. measurand) X, die einen physikalischen (oder anderen) Effekt beschreibt. Ziel einer Messung ist es, den wahren Wert $\tilde{x}$ der Messgröße in einer aktuellen Situation zu bestimmen. „Bestimmen" ist hier allerdings das falsche Wort, da der wahre Wert einer Messgröße prinzipiell unbekannt ist und durch Messungen nur geschätzt werden kann [3]. Durch eine Messung erhält man einen Messwert x, der einen Schätzwert für den wahren Wert $\tilde{x}$ der Messgröße darstellt. Wir haben dazu der Größe X eine Zufallsvariable zugeordnet, deren Wert $\tilde{x}$ durch die Messung geschätzt wird. Die metrologische Aufgabe ist nun, diese Schätzung zu quantifizieren, indem die Frage gestellt wird: Gegeben der Messwert x, wie wahrscheinlich ist es, dass der wahre Wert $\tilde{x}$ ist? Mathematisch ist dies die Frage nach der bedingten Wahrscheinlichkeit $f(\tilde{x}|x)$. Die Frage ist beantwortet, wenn man die Wahrscheinlichkeitsdichte $f(\tilde{x}|x)$ (Abb. 1) angeben kann.

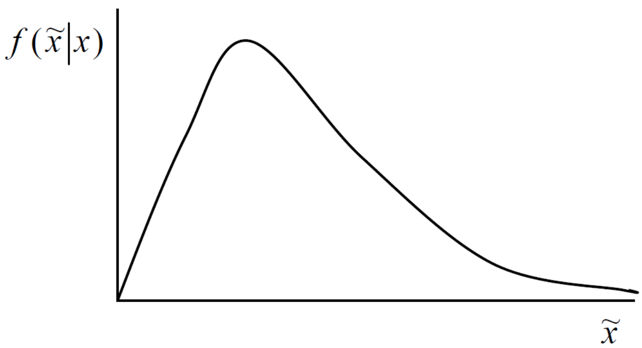

Abb. 1: Wahrscheinlichkeitsdichte (engl.: probability density function; abgekürzt PDF) $f(\tilde{x}|x)$ für die bedingte Wahrscheinlichkeit, dass der wahre Wert $\tilde{x}$ ist, gegeben den Messwert x. Die Dichtefunktion hängt von der jeweils verfügbaren Information ab

Wenn $f(\tilde{x}|x)$ bekannt ist, ergeben sich der beste Schätzwert $\hat{x}$ für den wahren Wert und die ihm zugeordnete Unsicherheit $u(\hat{x})$ als Erwartungswert bzw. aus der Varianz von $f(\tilde{x}|x)$ (Gl. 1 und 2).

$$\hat{x} = \mathrm{E}(f(\tilde{x}|x)) = \int_{-\infty}^{+\infty} \tilde{x} \cdot f(\tilde{x}|x)\, \mathrm{d}\tilde{x} \tag{1}$$

$$u^2(\hat{x}) = \mathrm{Var}(f(\tilde{x}|x)) = \int_{-\infty}^{+\infty} (\tilde{x} - \hat{x})^2 \cdot f(\tilde{x}|x)\, \mathrm{d}\tilde{x} \tag{2}$$

Nun kommen die Messunsicherheiten ins Spiel. Jede Messung ist mit Unsicherheiten behaftet. Die Messunsicherheit wird nach dem GUM [1] mit dem Symbol $u(x)$ bezeichnet. Sie ist jedoch keine Funktion von x, wie die Schreibweise leider suggeriert. Sie wird sprachlich als die dem Messwert zugeordnete (Standard-)Messunsicherheit bezeichnet.

Sie kann abgeschätzt werden durch mehrfach wiederholte Messungen, aus Erfahrungswerten über das Messverfahren oder durch Informationen aus anderen Quellen. Der GUM [1] unterscheidet zwei Arten der Ermittlung von Messunsicherheiten: Typ A und Typ B. Auch wenn häufig von Typ A und Typ B Messunsicherheiten gesprochen wird, sind diese als gleichwertig zu betrachten, da der GUM lediglich die Art ihrer Ermittlung unterscheidet. Messunsicherheiten enthalten im Allgemeinen viele Komponenten. Einige von diesen können über einen Typ A Ermittlung der Standardunsicherheit aus der statistischen Verteilung von wiederholten Messungen als Standardabweichungen erhalten werden. Die anderen Komponenten, die nach Typ B ermittelt werden, können auch durch Standardabweichungen charakterisiert und aus Wahrscheinlichkeitsdichten (engl. probability density functions, PDFs) auf der Grundlage der Erfahrung oder anderer Information erhalten werden. Messunsicherheiten enthalten auch Komponenten systematischer Effekte, die mit Korrektionen verbunden sind. Manchmal werden bekannte systematische Einflüsse nicht korrigiert, sondern als Komponenten der Messunsicherheit behandelt.

Nach dem VIM [5] wird die Messunsicherheit $u(x)$ – etwas kryptisch – als ein Parameter bezeichnet, der die mögliche Dispersion der Werte $\tilde{x}$ einer Messgröße beschreibt. Das bedeutet im Klartext: $u(x)$ beschreibt auf der Grundlage der verfügbaren Informationen über das Messverfahren und den Messwert x die Unsicherheit der Schätzung des wahren Wertes der Messgröße $\tilde{x}$. Die Grundfrage der Metrologie ist: Wie kommt man an die PDF $f(\tilde{x}|x)$? Die Antwort gibt die Statistik. Und da beginnt das Problem, das in der Metrologie einen Paradigmenwechsel verursacht hat.

3. Wahrscheinlichkeit und Statistik

Die Anwender des GUM leben noch in zwei verschiedenen Welten, der Welt der Bayes Statistik und der Welt der konventionellen oder frequentistischen Statistik. Die beiden Welten unterscheiden sich eklatant: **der Begriff Wahrscheinlichkeit hat unterschiedliche Bedeutungen in der Bayes Statistik und der frequentistischen Statistik**.

Die Verwendung und teilweise Vermischung der beiden Statistiken mit unterschiedlichen Wahrscheinlichkeitsbegriffen hat zu großer Verwirrung der Begriffe, der Grundlagen und der Sprache geführt. Obwohl viele (aber bei weitem nicht alle) Ergebnisse der beiden Statistiken praktisch gleich sind, dürfen sie nicht vermischt werden. Aus Gründen, die hier erläutert werden, sollte in der Metrologie nur noch die Bayes Statistik verwendet werden; siehe hierzu auch [6].

Die Sicht der konventionellen oder frequentistischen Statistik ist: Wahrscheinlichkeit ist der stochastische Grenzwert relativer Häufigkeiten in Wiederholungsexperimenten. Die Kritik an diesem frequentistischen Ansatz entzündet sich daran, dass einerseits nicht jede Häufigkeitsverteilung eine Wahrscheinlichkeit repräsentiert und es andererseits Wahrscheinlichkeiten gibt, die nicht durch Häufigkeitsverteilungen in wiederholten Experimenten beschrieben werden können.

Die Sicht der Bayes Statistik ist: Wahrscheinlichkeit ist ein Maß für das Vertrauen, das ein Individuum in eine unsichere Aussage hat. Die Bayes Statistik geht davon aus, dass die mathematische Wahrscheinlichkeitstheorie anwendbar ist zur Quantifizierung des Vertrauens, das der Einzelne in eine unsichere Aussage haben kann. Bayesianer sind der Ansicht, dass mathematische Methoden, u. a. das Bayes Theorem, die Produktregel und das Theorem der totalen Wahrscheinlichkeit, benutzt werden können, um früher ermittelte Wahrscheinlichkeitsaussagen im Lichte neuer, relevanter Information auf einen neuen Stand zu bringen.

Diese Schlussweise nennt man in der Statistik auch Bayes'sches Schließen (engl. Bayesian inference). Thomas Bayes (* 1702, † 1761), ein englischer Kleriker beschrieb in seinem 1763 posthum erschienenen Essay "Towards Solving a Problem in the Doctrine of Chances" [7] das, was wir heute eine Bayes Schätzung nennen: „*calculating the probability of the validity of a proposition on the basis of a prior estimate of its probability and new relevant evidence*".

Wahrscheinlichkeit wird dabei benutzt in dem Sinne, wie bei der Frage nach der Wahrscheinlichkeit, mit einem Würfel eine sechs zu werfen, bevor man gewürfelt hat. Dem Bayesianer fällt die Antwort „1/6" leicht, der Frequentist muss antworten: Ich weiß es nicht, ich habe noch nicht gewürfelt.

Für die Metrologie haben diese unterschiedlichen Wahrscheinlichkeitsbegriffe gravierende Konsequenzen. Typ B Unsicherheiten können nicht durch Wiederholungsexperimente beschrieben werden. Daher kann die frequentistische Statistik Typ B Unsicherheiten nicht berücksichtigen. Dies geht nur mittels Bayes Statistik. Ganz allgemein und abstrakt ausgedrückt kann die frequentistische Statistik nur Wahrscheinlichkeiten der Form $f(x|\tilde{x})$, nicht aber $f(\tilde{x}|x)$ bestimmen; die Bayes Statistik kann sowohl Aussagen über $f(\tilde{x}|x)$ als auch über $f(x|\tilde{x})$ machen.

4. Der Ansatz des GUM: ein minimalistischer Spezialfall

In der Bayes'schen Theorie der Messunsicherheiten [3, 4], die der GUM-Methodik zugrunde liegt, wird die gesuchte Funktion $f(\tilde{x}|x)$ mit Hilfe des Prinzips der maximalen Informationsentropie (PME) [8] bestimmt. Das PME sagt, dass auf der Grundlage der verfügbaren Information diejenige Wahrscheinlichkeitsdichte $f(\tilde{x}|x)$ die beste ist, die die Informationsentropie *S* nach Gl. 3 maximiert:

$$S = -\int f(\tilde{x}|x) \cdot \ln\big(f(\tilde{x}|x)\big)\,\mathrm{d}\tilde{x} = \max \tag{3}$$

Die so bestimmte Wahrscheinlichkeitsdichte ist diejenige, die nur die verfügbare Information nutzt und nichts „hinzufügt". Notwendiger Weise hängt das Ergebnis vom jeweiligen Informationsstand ab. Im Supplement 1 zum GUM [1] wird dies ausführlich dargestellt. Hier soll nur der Ansatz des GUM [1] exemplarisch vorgestellt werden. Der GUM geht davon aus, dass lediglich der Messwert *x* und die ihm zugeordnete Messunsicherheit *u*(*x*) bekannt sind. Diese Information reicht jedoch noch nicht aus, um die Maximierungsaufgabe in Gl. 3 zu lösen. Man braucht Nebenbedingungen. Der GUM nimmt als Nebenbedingungen $\mathrm{E}(f(\tilde{x}|x)) = x$ und $\mathrm{Var}(f(\tilde{x}|x)) = u^2(x)$. Mit diesen Nebenbedingungen kann Gl. 3 mit der Methode der Lagrange Multiplikatoren gelöst werden und man erhält das Ergebnis in Gl. 4.

$$f(\tilde{x}|x) = C \cdot \exp\big(-(\tilde{x}-x)^2/(2\cdot u^2(x))\big) \tag{4}$$

Was bedeuten die Nebenbedingungen und was das Ergebnis? $\mathrm{E}(f(\tilde{x}|x)) = x$ bedeutet schlicht, dass man, wenn nur ein Messwert x vorliegt, dieser notwendiger Weise der beste Schätzwert $\hat{x}$ ist. $\mathrm{Var}(f(\tilde{x}|x)) = u^2(x)$ bedeutet, dass die gesamte Unsicherheit über den wahren Wert $\tilde{x}$ auf der Messunsicherheit beruht und damit $u^2(x) = u^2(\hat{x})$ gilt. $f(\tilde{x}|x)$ ist eine Normalverteilung (Gl. 4), in der x fest und $\tilde{x}$ die unabhängige Variable ist.

Mit dem Bayes Theorem (Gl. 5)

$$f(x|\tilde{x}) \cdot f(\tilde{x}) = f(\tilde{x}|x) \cdot f(x) \tag{5}$$

ergibt sich auch $f(x|\tilde{x})$ als Normalverteilung (Gl. 6), in der x die unabhängige Variable ist.

$$f(x|\tilde{x}) = C \cdot \exp\left(-(\tilde{x} - x)^2 / (2 \cdot u^2(x))\right) \tag{6}$$

$f(x)$, die Wahrscheinlichkeit überhaupt einen Messwert x zu erhalten, kann unter unveränderten Experimentbedingungen als konstant angenommen werden. C ist eine Normierungskonstante.

Sind andere Informationen verfügbar als die hier und im GUM minimalistisch, aber allgemein angesetzte Information x und $u(x)$ ergeben sich andere Wahrscheinlichkeitsdichtefunktionen, wie sie im Supplement 1 zum GUM [2] ausführlich dargestellt sind.

Zum Abschluss soll noch der allgemeinere Fall einer Messgröße Y behandelt werden, die über ein Modell der Auswertung G mit m Eingangsgrößen X_i verknüpft ist (Gl. 7).

$$Y = G(X_1, X_2, \ldots, X_m) \tag{7}$$

Durch Messungen erhält man Schätzwerte x_i der wahren Werte der Eingangsgrößen X_i. Diese Schätzwerte x_i haben die zugeordneten Messunsicherheiten $u(x_i)$. Durch die Auswertung erhält man mit Gleichung 8 einen Schätzwert y des wahren Wertes $\tilde{y}$ der Messgröße Y und mit dem GUM [1] bzw. dem Supplement 1 zum GUM [2] die zugeordnete Messunsicherheit aus den $u(x_i)$ die $u(y)$.

$$y = G(x_1, \ldots, x_m)\,. \tag{8}$$

Außerdem sei Vorinformation $\Im$ verfügbar, z.B. dass die Messgröße Y nicht negativ sei. Man beachte, dass bei einer nicht negativen Messgröße die Messwerte y negativ sein können.

Gesucht wird die Wahrscheinlichkeitsdichte $f(\tilde{y}|y,\Im)$, die bedingte Wahrscheinlichkeit dafür, dass der Wert $\tilde{y}$ der Messgröße Y ist, gegeben der Messwert y und die Information $\Im$. Mit der Bayes'schen Theorie der Messunsicherheit nach Weise und Wöger [3, 4] wird das Problem behandelt mit dem Ansatz in Gleichung 9.

$$f(\tilde{y}|y,\Im) = f_0(\tilde{y}|y) \cdot f(\tilde{y}|\Im) \tag{9}$$

Dabei sind $f_0(\tilde{y}|y)$ der sog. Datenprior, der alle Information aus der aktuellen Messung enthält, und $f(\tilde{y}|\Im)$ der sog. Modellprior, der die vorher existierende Information beschreibt und damit bekannt ist. Die gesuchte Wahrscheinlichkeitsdichte $f(\tilde{y}|y,\Im)$ wird auch die posterior PDF genannt. Sind lediglich y und $u(y)$ bekannt, so erhält man mit dem PME analog zu Gl. 3 und 4, dass der Datenprior $f_0(\tilde{y}|y)$ eine Normalverteilung ist. Damit erhält man die Lösung:

$$f(\tilde{y}|y) = C \cdot f(\tilde{y}|\Im) \cdot \exp\left(-(\tilde{y}-y)^2/(2 \cdot u^2(y))\right) \tag{10}$$

Gleichungen 5 und 6 gelten analog. Für weiterführende Lektüre sei auf [9] verwiesen.

5. Schlussfolgerung

Mit der Bayes'schen Theorie der Messunsicherheiten liegt eine Methodik vor, die eine konsistente Behandlung von Messunsicherheiten erlaubt und Wahrscheinlichkeitsaussagen über den wahren Wert der Messgröße auf der Grundlage aller bekannten Quellen der Unsicherheit ermöglicht. Dies ist mit den Methoden der frequentistischen Statistik nicht möglich.

6. Literaturverzeichnis[1]

[1] ISO: Guide to the Expression of Uncertainty in Measurement. Geneva: ISO, 1995

[2] ISO: Guide to the expression of uncertainty in measurement (GUM) - Supplement 1: Propagation of distributions using a Monte Carlo method, JCGM 101, 2008

[3] Weise, K.; Wöger, W.A., Bayesian theory of measurement uncertainty. Meas. Sci. Technol. 4, 1 – 11, 1993

[4] Weise, K.; Wöger, W.: Messunsicherheit und Messdatenauswertung. Berlin: Wiley-VCH, 1999

[5] ISO: International Vocabulary of Basic and General Terms in Metrology (VIM) — Third Edition, Geneva: ISO/IEC, 2007

[6] G. D'Agostini, Bayesian Reasoning in Data Analysis. New Jersey, London, Singapore, Hong Kong: World Scientific, 2005

[7] Th. Bayes, Essay Towards Solving a Problem in the Doctrine of Chances, The Philosophical Transactions 53 (1763), 370-418; siehe hierzu auch: Biometrika 45 (3/4), 293 – 315, 1958

[8] E.T. Jaynes, Papers on probability, statistics and statistical physics. Hrsg. R.D. Rosenkrantz, Boston, New York, Dordrecht: Kluwe Publ., 1989

[9] Weise, K. G. Kanisch, R. Michel, M. Schläger, D. Schrammel, M. Täschner, Monte Carlo determination of the characteristic limits in measurement of ionising radiation – Fundamental and numerics, Radiation Protection Dosimetry 135 (3) (2009) 169 – 196; doi:10.1093/rpd/ncp105

[1] Die Referenzen [1, 2, 5] stehen unter http://www.bipm.org/en/publications/guides/ kostenlos zur Verfügung