

Wahrscheinlichkeiten und Nachweisgrenzen

Ein kleiner Exkurs in die Zählstatistik (I)

In Urteilsbegründungen oder Zeitungsberichten lesen wir oft Formulierungen wie „... mit an Sicherheit grenzender Wahrscheinlichkeit ist ...“. Offensichtlich akzeptieren wir in diesem Zusammenhang quasi „sichere“ Aussagen. Im Strahlenschutz hingegen versehen wir Aussagen über Messwerte meistens nur mit einem Etikett „95 % Sicherheit“ oder, wenn es hoch kommt, mit „99 % Sicherheit“. Wir wissen, sobald wir Messwerte mit an Sicherheit grenzender Wahrscheinlichkeit angeben wollen, wächst die Unschärfe unseres Messwertes über alle Grenzen – wir können offenbar Messergebnisse nur vernünftig angeben, wenn wir „etwas“ Unsicherheit zulassen.

Dieser Umgang mit der Sicherheit bzw. Unsicherheit erzeugt oft ein gewisses Unbehagen. Auch in Vorlesungen an der Berufsakademie Karlsruhe zu diesem Thema spürt man dieses Unbehagen, z. B. in der Art der studentischen Fragen. Egal, ob mehr von einfacher oder auch von kryptischer Art, meistens waren die Fragen nicht so aus dem Stegreif zu beantworten und erforderten etwas Nachdenken. Ich möchte in diesem Beitrag exemplarisch auf einige dieser Fragen eingehen. Vielleicht können auch wir noch einige Einsichten gewinnen, auch wenn wir (oder gerade weil wir?) schon so lange im Strahlenschutz-Business tätig sind. Beginnen wir also „irgendwo ganz vorne“.

Aktivität und Wahrscheinlichkeit
Klar! Die Aktivität A einer Stoffmenge ist die Anzahl der Zerfälle pro Zeiteinheit, d. h. $A = -dN/dt$, N sei die Zahl der Atomkerne. Da die

Teilchenzahl abnimmt, garantiert das Minuszeichen, dass A eine positive Größe ist. Stellen wir uns jedoch jetzt einmal vor, dass wir nur ganz wenige Zerfälle beobachten, wie bei Low-Le-

vel-Messungen z. B. mit etwa nur einem Zerfall pro Stunde, oder dass alle Teilchen bis auf eines zerfallen seien. Gilt unsere Formel jetzt auch noch für alle Zeiten (Abb. 1)? Wie groß ist die Aktivität, solange das Teilchen noch nicht zerfallen ist ($t < t'$)? Welchen Wert nimmt denn die Aktivität im Augenblick des Zerfalls an?

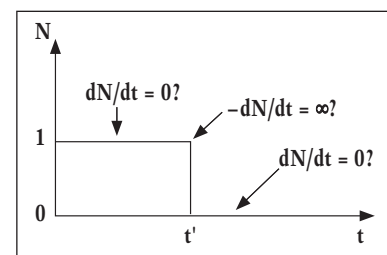


Abb. 1: Teilchenzahl als Funktion der Zeit für $N = 1$, Zerfallszeitpunkt t'

Der Praktiker sagt, dass man in solchen Fällen eben einen geeigneten Mittelwert bilden muss. Was heißt geeignet, wenn definitiv nur ein Teilchen vorhanden ist? Wir stellen entweder einen Zerfall fest oder auch nicht. Hilfe kommt hier von theoretischer Seite. Das statistische Verhalten eines Teilchens mit „to be or not to be“ wird exakt durch die sog. Binomialverteilung beschrieben (Gl. (1)): Wir haben eine Wahrscheinlichkeit p , dass das Teilchen in unserer Messzeit t zerfällt, bzw. die Wahrscheinlich-

keit $1 - p$, dass das Teilchen nicht zerfällt. Verallgemeinert auf N Teilchen erhalten wir für die Wahrscheinlichkeit $P(z)$, dass z von diesen N Teilchen zerfallen:

$$P(z) = P_{\text{Binom}}(z; N, p) = \binom{N}{z} \cdot p^z (1-p)^{N-z}$$

mit $z = 0, 1, 2, \dots, N$ (1)

Wir blättern im Statistik-Buch (z.B. L. Sachs, Angewandte Statistik) und erhalten für den Erwartungswert der Zerfälle z im Messzeitraum $\mu = E(z) = N \cdot p$ und für die Varianz oder das Quadrat der Standardabweichung $\sigma^2 = E((z - \mu)^2) = N \cdot p \cdot (1 - p)$. Diese Ergebnisse sind grundlegend. Alle folgenden Aussagen zur Zählstatistik werden auf Formel (1) aufbauen. Zur vollständigen Antwort jedoch auf die Frage nach der Aktivität eines Teilchens fehlt uns noch der genaue Wert der Wahrscheinlichkeit $p(t)$ für das Messintervall t .

$P(t)$ hängt im Prinzip nur von einer grundlegenden Aussage über unsere Atomkerne ab: Kerne zerfallen zwar, aber sie „altern“ nicht (Der Quantenmechanik sei Dank!). Was heißt das? Wenn wir z.B. ein ^{137}Cs -Atom, das gerade im Reaktor entstanden ist, zusammen betrachten mit einem ^{137}Cs -Atom, das schon vor 27 Jahren gebildet wurde, dann haben beide Atome die gleiche Wahrscheinlichkeit $p(t)$, in unserem Messzeitraum t zu zerfallen, unabhängig von ihrem Alter. Und falls sie in dieser Messzeit nicht zerfallen sind und wir eine erneute Messung starten, dann gilt für die neue

Messung dasselbe $p(t)$! Mit ein klein wenig Differenzial- und Wahrscheinlichkeitsrechnung können wir das Ergebnis präsentieren¹⁾:

$$p(t) = 1 - e^{-\lambda t} \text{ mit } \lambda = \ln 2/t_{1/2} \text{ und } t_{1/2} \text{ der Halbwertszeit des Nuklids} \quad (2)$$

Die Aktivität $A(t)$ ergibt sich als zeitliche Änderung des Erwartungswertes μ der Anzahl der Zerfälle:

$$A(t) = d\mu/dt = N \cdot dp/dt, \text{ d.h. } A(t) = N \cdot \lambda \cdot e^{-\lambda t} = (N \cdot e^{-\lambda t}) \cdot \lambda \approx N(t) \cdot \lambda \quad (3)$$

– unser bekanntes exponentielles Zerfallsgesetz (Abb. 2). Wir sehen aber auch, dass die tatsächliche Teilchenzahl $N(t)$ in $A(t)$ immer nur als Schätzwert für die Aktivität stehen kann. Bei großen Teilchenzahlen ist der Schätzwert natürlich hervorragend und wir merken nichts von den Schwankungen des Ergebnisses wie im Falle $N = 1$.

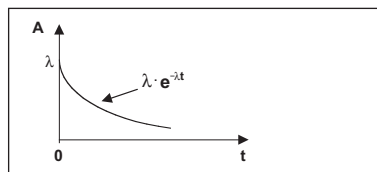


Abb. 2: Aktivitätsverlauf für ein Teilchen mit der Zerfallskonstanten λ und $N = 1$

Binomialverteilung

Beim Arbeiten beispielsweise mit der Norm zu Nachweis- und Erkennungsgrenzen (DIN 25482) trifft man oft auf den Begriff der Poisson-Prozesse und auf die näherungsweise Beschreibung dieser Prozesse durch Normal- oder Gauss-Verteilungen. Liegen Messwerte vor, so bleibt die konkrete Frage, muss ich (darf ich?) mit der Gauss- oder mit der Poisson-Verteilung oder vielleicht sogar mit einer anderen Verteilung arbeiten?

Wir haben gehört, dass wir mit der Binomialverteilung (1) immer die richtige Wahl treffen. Einige Beispiele für Ergebnisse der Binomialverteilung zeigt Abb. 3. Beispiele aus unserer täglichen Praxis im Strahlenschutz jedoch finden wir selten. Warum? – Ein Grund hierfür ist mit Sicherheit, dass die Binomialverteilung

im praktischen Umgang äußerst unhandlich ist.

Stellen Sie sich vor, Sie arbeiten mit 10^4 Bq ^{137}Cs , das sind ca. 10^{14} Atomkerne, und Sie müssten jetzt nach Formel (1) berechnen, wie wahrscheinlich z.B. 10^4 Zerfälle sind. Wetten, dass Sie mit einem Binomialkoeffizient 10^{14} über 10^4 jede Standardsoftware in die Knie zwingen werden! Zudem kennt man bei Strahlenschutzmessungen in den wenigsten Fällen die Teilchenzahl N , man wird sie umgekehrt erst über die Anzahl der beobachteten Zerfälle berechnen können. In (1) müssen wir aber mit diesen großen numerischen Werten für die Teilchen- und Zerfallszahlen arbeiten, da die Binomialverteilung als sog. diskrete Verteilung nur für ganzzahlige Zufallsvariable z erklärt ist und die praktische Verwendung von Zählraten $R = z/t$ nicht zulässt.

Die Binomialverteilung liefert trotzdem einige beachtenswerte Informationen. Aus (1) folgt z.B., dass immer $\sigma < \mu$! Wenn wir zudem die zeitliche Abhängigkeit von p beachten, erhalten wir sogar die Aussage, dass für Messzeiten $t \gg t_{1/2}$ die Standardabweichung σ gegen 0 geht. Was heißt das in der Praxis? Wir messen z.B. auf einem Luftfilter die kurzlebigen ^{222}Rn -Folgeprodukte ($t_{1/2} \sim 0,5$ h) und lassen das Messgerät über Nacht ($t \sim 16$ h) zählen. Angenommen, wir hätten ca. 10^4 Zerfälle, dann besagt $\sigma \rightarrow 0$, dass die Standardabweichung σ nicht 10^2 Zerfälle beträgt, sondern dass $\sigma = 0$ ist – die Zählstatistik liefert in diesem Falle keinen Beitrag zu evtl. Schwankungen des Messergebnisses!

Poisson-Verteilung

Für ein sehr großes Spektrum von Messaufgaben stellt die Poisson-Verteilung (ebenfalls eine diskrete Verteilung) eine sehr gute Näherung der Binomialverteilung dar (die meisten Statistikbücher führen diese Näherung auch explizit vor):

¹⁾ Die Wahrscheinlichkeit, dass ein Kern im Zeitraum $t + \Delta t$ zerfällt, ist gleich der Wahrscheinlichkeit, dass er entweder im Intervall t zerfällt oder im Intervall t nicht zerfällt, aber dafür in Δt . In Formeln:
 $p(t + \Delta t) = p(t) + (1 - p(t)) \cdot p(\Delta t)$ oder $p(t + \Delta t) - p(t) = (1 - p(t)) \cdot p(\Delta t)$.
 Dividieren wir durch Δt , dann erhalten wir die Differenzialgleichung $dp/dt + dp(0)/dt \cdot p = dp(0)/dt$, mit der Lösung $p(t) = 1 - e^{-dp(0)/dt \cdot t}$. Für $p(t_{1/2}) = 0,5$ entpuppt sich $dp(0)/dt$ als unsere gewohnte Zerfallskonstante $\lambda = \ln 2/t_{1/2}$ mit der Halbwertszeit $t_{1/2}$.

$$P_{\text{Poisson}}(z; \mu) = \frac{e^{-\mu} \cdot \mu^z}{z!}$$

$$E(z) = \mu, \quad \sigma^2 = \mu, \quad z = 0, 1, 2, \dots \quad (4)$$

Heureka! Die Teilchenzahl N taucht nicht mehr explizit auf, nur noch der Parameter μ , den wir aber als Erwartungswert der Zerfälle aus unserem Experiment bestimmen können! Die Voraussetzungen für diese Näherung bestehen darin, dass die Messzeit $t \ll t_{1/2}$ sehr klein gegenüber der Halbwertszeit der zu messenden Radionuklide sein muss, d. h., dass in (1) der Faktor $(1 - p)$ gegen 1 und p gegen 0 strebt, gleichzeitig aber die Teilchenzahl N noch so groß ist, dass der Erwartungswert für die Anzahl der Zerfälle $\mu = N \cdot p$ einen endlichen Wert annimmt. Dies ist aber in den meisten Fällen gegeben²⁾. Wir entdecken in (4) ebenfalls die uns vertraute Relation $\sigma = \mu^{1/2}$. Ein Vergleich der Ergebnisse von Binomial- bzw. Poisson-Verteilung ist in Abb. 3b) aufgetragen.

Aber in Statistik-Büchern finden wir noch eine weitere für uns beruhigende Tatsache: Haben wir etwa bei der Messung eines Nuklids 1 eine Zählrate z_1 und einen Erwartungswert μ_1 , und entsprechend bei Nuklid 2 z_2 und μ_2 , dann ist auch die Summe $z = z_1 + z_2$ wieder nach Poisson verteilt mit dem Erwartungswert $\mu = \mu_1 + \mu_2$. Dies bedeutet, dass wir auch Nuklidgemische mit verschiedenen Halbwertszeiten in einer Gesamtmessung erfassen können – der Messwert gehorcht wieder einer Poisson-Verteilung! Diese Eigenschaft liegt für die Binomialverteilung i. A. nicht vor.

Gauss- oder Normalverteilung

Die Gauss-Verteilung ist omnipotent! Sie ergibt sich wegen des sog. Zentralen Grenzwertsatzes als kontinuierliche

²⁾ Obwohl man wegen $p \rightarrow 0$ in der Statistik von seltenen Ereignissen spricht, können wir bei großem N durchaus sehr große Zerfallszahlen vorliegen haben!

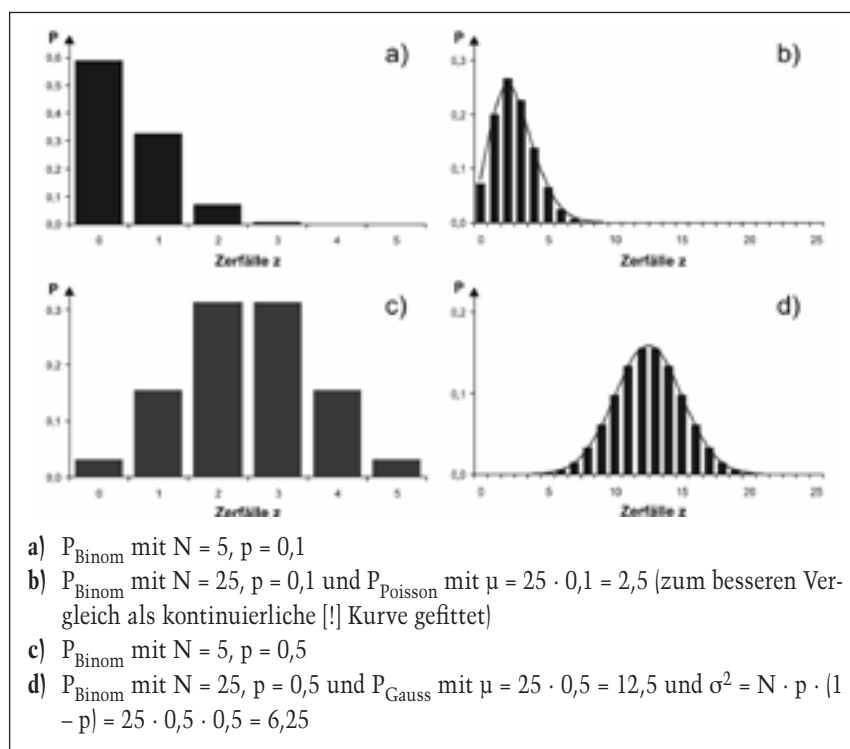


Abb. 3: Wahrscheinlichkeiten nach Binomial-, Poisson- und Gauss-Verteilung

liche Verteilung mit 2 Parametern im Grenzfall vieler Verteilungen. Im vorliegenden Fall tritt sie als Näherung der Poisson-Verteilung (4) wie auch der Binomialverteilung (1) für größere Werte von μ auf:

$$P_{\text{Gauss}}(z; \mu; \sigma) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

$$-\infty < z < +\infty \quad (5)$$

Die Parameter μ und σ sind allerdings nicht mehr frei wählbar, sondern hängen miteinander gemäß der Ausgangsverteilung zusammen, d. h. als Grenzfall der Poisson-Verteilung mit $\sigma^2 = \mu$ (wohl häufigster Fall im Strahlenschutz) und als Grenzfall der Binomialverteilung mit $\sigma^2 = \mu(1 - p)$! Natürlich sind beide Grenzfälle für $t \ll t_{1/2}$ identisch. Je nach Genauigkeitsansprüchen kann man Gauss-Verteilungen näherungsweise schon ab $\mu \geq 10$ für die Poisson- oder Binomial-Verteilung verwenden (siehe z. B. Abb. 3d).

Die Gauss-Verteilung bietet uns im

Vergleich zur Poisson- bzw. Binomialverteilung auch noch einen gewissen Komfort – sie ist z. B. symmetrisch um den Erwartungswert, eine Eigenschaft, die sich insbesondere für eine übersichtliche Darstellung und Angabe von Vertrauensintervallen und Quantilen als Vorteil erweist. Als kontinuierliche Verteilung schafft sie es auch, Zählraten zu „verarbeiten“. Man kann zeigen³⁾, dass Zählraten $R = z/t$ ebenfalls Gauss-verteilt sind, und zwar mit dem Erwartungswert $E(R) = \rho = \mu/t$ und der Varianz $E((R - \rho)^2) = \rho/t$.

Natürlich gilt auch für die Gauss-Verteilung, dass die Summe, hier sogar auch die Differenz, zweier Impulszahlen $z = z_1 \pm z_2$ ebenfalls wieder einer Gauss-Verteilung mit $\mu = \mu_1 \pm \mu_2$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2$ gehorchen. Dieser Komfort hat aber auch seine Tücken. Eine Gauss-Verteilung ist per se definiert für alle reellen Werte

³⁾ Die Transformationsregel von einer Verteilung $P(z)$ auf eine Verteilung $P(R)$ mit $R = z/t$ lautet $P'(R) = |dz/dR| \cdot P(z)$.

$-\infty < R < +\infty$, d.h. also auch für sinnlose negative Zählraten. Dieses Handicap spielt meistens keine Rolle, solange ρ einen größeren Wert annimmt, ist die Wahrscheinlichkeit für diese möglichen negativen Werte praktisch null. Falls wir aber Messwerte an der Nachweisgrenze bestimmen wollen, dann kommen wir nach Abzug unseres Nulleffektes ρ_0 mit der Netto-Zählrate ρ sehr nahe an $\rho = 0$, und

die Flanke der Gauss-Verteilung ragt in diesen „verbotenen“ negativen Bereich hinein. Um diesen Fehler wieder zu korrigieren, wird z.B. bei der Bestimmung von Nachweisgrenzen über den Bayes'schen Ansatz (Teil 10, DIN 25482) durch Einführung des sog. Modell-Priors ξ_0 dieser Bereich chirurgisch wieder von dem erlaubten positiven Wertebereich getrennt. Dieser Exkurs kann und will auch

nicht in irgendeiner Form vollständig sein. Einige Fragen zur Wahrscheinlichkeit, insbesondere bei der Anwendung von Nachweisgrenzen, blieben noch unerwähnt und werden „wahrscheinlich“ in einem später folgenden Exkurs (II) besprochen.

Kraut, Karlsruhe

E-Mail: kraut@ba-karlsruhe.de □

Wahrscheinlichkeiten und Nachweisgrenzen

Ein kleiner Exkurs in die Zählstatistik (II)

Nachdem im Exkurs (I) (Heft 4/2003, S. 79) die mathematischen Grundlagen von Wahrscheinlichkeitsangaben erläutert wurden, wollen wir uns in diesem Beitrag mit der Überprüfung statistischer Aussagen bei der Bestimmung von Erkennungs-, Nachweis- und Vertrauensgrenzen befassen (DIN 25482, Teil 1 bzw. 10 für Messungen mit Zeitvorwahl). Welche konkrete Bedeutung hat es z. B. für uns, wenn wir eine Nachweisgrenze mit Fehlern 1. und 2. Art von jeweils 5 % festlegen? Drückt die Wahrscheinlichkeitsaussage „5 %“ nur unsere persönliche Einschätzung über das Messverfahren aus oder müssen wir nicht auch eine operationale Umsetzung in der Praxis gewährleisten und gegebenenfalls auch eine Überprüfungsmöglichkeit angeben können? Dazu erfahren Sie mehr, wenn Sie sich dem – zugegebenermaßen nicht gerade leichten – Studium des nachfolgenden Textes unterziehen.

Wahrscheinlichkeit und Wahrheitsgehalt

Meteorologen untermauern Prognosen wie z. B. „Die Regenwahrscheinlichkeit für den nächsten Tag beträgt 95 %“ mit bisher aufgezeichneten Häufigkeiten: In 95 % ähnlicher Wetterlagen hatte es z. B. am nächsten Tag geregnet. Im Strahlenschutz hingegen führen wir üblicherweise (konstante Messbedingungen vorausgesetzt) nur *eine* (!) Nulleffektmessung durch und danach unsere Probenmessungen. Natürlich kennen wir das statistische Verhalten unseres Systems recht genau (Exkurs (I)) und müssen den Poisson-Prozess des Zerfalls nicht jedes Mal erneut verifizieren. Eine konkrete Prüfung allerdings des Wahrheitsgehaltes unserer Aussagen, wie z. B. Probeneffekt ja/nein, können wir aber nur an einem uns bekannten System (Standard) durchführen.

Dazu ein Experiment

Zu diesem Zweck stellen wir uns ein kleines Experiment vor: Die PTB besitzt einen „Nulleffekt-Standard“ (S0) mit einer exakten Impulsrate von $\rho_0 = 1$ IPS und einen „Netto-Probeneffekt-Standard“ (SN) von $\rho_n = 1$ IPS, bzw. in Kombination beider Standards einen „Brutto-Probeneffekt-Standard“ (SB) mit $\rho_b = 2$ IPS (wir interessieren uns nur für Impulszahlen und setzen den Zählwirkungsgrad gleich 1). Alles, was Rang und Namen in der Strahlenmesstechnik hat, nimmt an einem „Ringversuch“ bei der PTB teil.

Aufgaben

Es sind folgende Messaufgaben zu lösen:

Aufgabe 1

Bestimmung des Nulleffekts R_0 an S0 mit $t_b = 1.000$ s und die Festlegung von Erkennungs- und Nachweisgrenzen für eine Probenmess-

zeit $t_b = 100$ s und Fehlerparameter $\alpha = 5\%$, $\beta = 5\%$

Aufgabe 2

Kontrolle der Einhaltung des Fehlers 1. Art durch eine Messreihe an S0 – welcher Anteil der Messwerte wird entsprechend den Festlegungen in Aufgabe 1 fälschlicherweise als Probeneffekt klassifiziert?

Aufgabe 3

Probenmessung an SB mit $t_b = 100$ s und Bestimmung der Vertrauensgrenzen mit $\gamma/2 = 2,5\%$, Prüfung ob der wahre Wert ρ_n in diesen Grenzen liegt.

Die gewählten Messbedingungen könnten z. B. bei einer Filterauswertung vorliegen. Für unsere weiteren Überlegungen übernimmt die Rolle des Messgerätes ein Zufallsgenerator, der exakt Gauss-verteilte Messwerte für Null-Effekt und Brutto-Effekt liefert und damit die verschiedenen Messergebnisse der Teilnehmer unseres „Ringversuches“ simuliert. Abb. 1 zeigt beispielhaft Messergebnisse als 2D-Scatter-Plot für Null-Effekt und Brutto-Effekt.

Ergebnisse

Die PTB kennt die wahren Werte der Standards durch Messungen mit (beliebig) langen Messzeiten und erhält die exakten Bestimmungsgrenzen für eine Brutto-Messzeit $t_b = 100$ s aus den Formeln A(1), A(1a), A(2), A(2b) der Norm (Hinweis auf Formeln der DIN 25482 mit vorangestelltem A, siehe Kasten auf S. 55) im Grenzfall $t_0 \rightarrow \infty$.

(1) (Netto-)Erkennungsgrenze

$$R_n^* = 0,1645 \text{ IPS}$$

(2) (Netto-)Nachweisgrenze

$$\rho_n^* = 0,355 \text{ IPS}$$

Aufgabe 1

In Abb. 2 (a, b, c) bzw. Abb. 3 (a, b) tragen wir die von den Teilnehmern bestimmten Erkennungs- bzw. Nach-

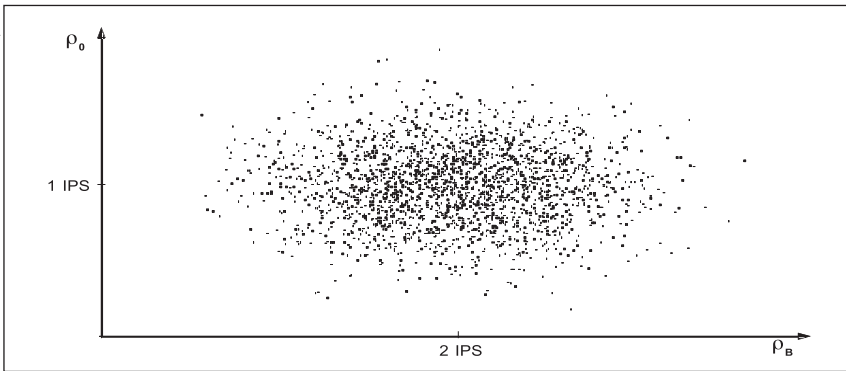


Abb. 1: 2D-Scatter-Plot; „Messwerte“ aus dem Zufallsgenerator für Null-Effekt und Brutto-Effekt

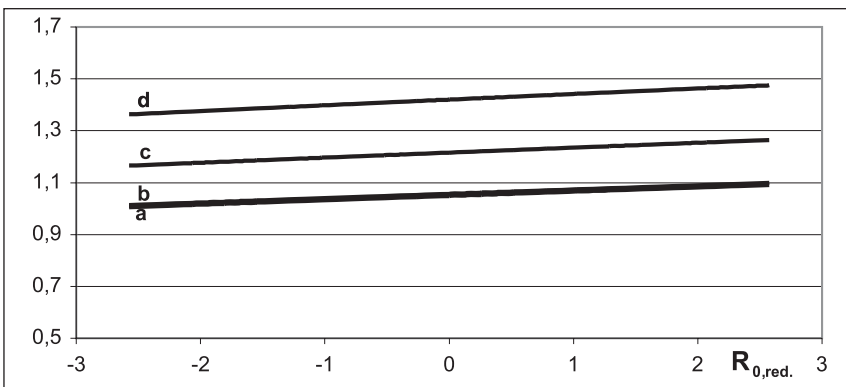


Abb. 2: Verhältnis Erkennungsgrenzen für $t_b = 100$ s, $t_0 = 1.000$ s, $\alpha = 5\%$ zum exakten Wert (1) als Funktion des Null-Effekts a) Formel A(1), b) Formel A(1a), c) „2-Sigma-Regel“, d) $(R^* - R_0) / 0,1645$ nach A)

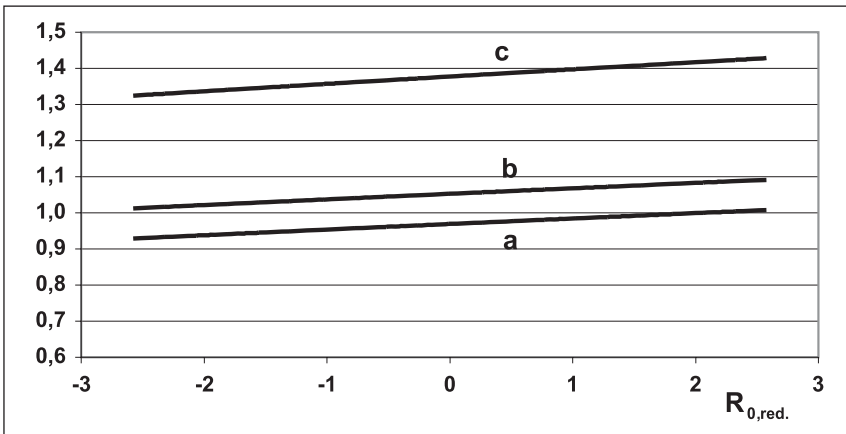


Abb. 3: Verhältnis Nachweisgrenzen für $t_b = 100$ s, $t_0 = 1.000$ s und $\alpha = \beta = 5\%$ zum exakten Wert (2) als Funktion des Null-Effekts a) Formel A(2a), b) Formel A(2), c) nach Formel (6) in C)

weisgrenzen als Verhältnis zum exakten PTB-Wert (1) bzw. (2) auf. Die Darstellung der Werte über der sog. reduzierten Variablen des Nulleffekts

$$(3) \quad R_{0,red.} = \frac{R_0 - \rho_0}{\sigma_0} = \frac{R_0 - \rho_0}{\sqrt{\rho_0 / t_0}}$$

hat den Vorteil, dass wir uns auf den

„wesentlichen“ Teil der R_0 -Werte konzentrieren können. Der gewählte Wertebereich $[-2,576, +2,576]$ umfasst z. B. 99% der Wahrscheinlichkeitsverteilung für R_0 . Die Erkennungsgrenzen nach Formel A(1) bzw. nach der einfacheren Formel A(1a) unterscheiden sich für unsere Parameter kaum

merkbar. Wir können von „konservativen“ Ergebnissen sprechen, beide Kurven liegen etwas über dem exakten Wert (1). Die dem Praktiker geläufige „2-Sigma-Regel“ wurde für „das Auge“ ebenfalls mit aufgetragen. Sie entspricht zwar nicht speziell einem Fehler 1. Art von 5%, wird aber sicher noch oft zu Rate gezogen.

Abb. 3 zeigt, dass die Teilnehmer, je nach gemessener Nullrate, „etwas bessere“ bzw. „etwas schlechtere“ Nachweisgrenzen ausweisen. Die Ergebnisse der einfacheren Formel A(2a) liegen größtenteils unter dem exakten Wert (2), Formel A(2) hingegen ergibt für unser Beispiel „konservative“ Nachweisgrenzen.

Aufgabe 2

Aufgrund der Aufgabenstellung erwarten wir, dass unsere Messungen an dem „Nulleffekt-Standard“ bei dem vorgegebenen Parameter $\alpha = 5\%$ auch 5% der Messwerte fälschlicherweise als Probeneffekt klassifizieren werden. Tatsächlich zeigt aber Abb. 4, dass schon für Nulleffektwerte geringfügig unterhalb $R_{0,red} = -0,258$ das 5%-Kriterium überschritten wird. Dies bedeutet für unseren „Ringversuch“, dass ca. $\delta = 40\%$ der Teilnehmer aufgrund ihrer Nulleffekt-Messreihe mehr als 5% falsch positive Aussagen über einen Probeneffekt machen!

Wie ist es möglich, dass wir einerseits nach Abb. 2 (a, b) eine konservative Netto-Erkennungsgrenze feststellen, andererseits aber nach Abb. 4 die Klassifikation der Werte versagt? Nun, wir entscheiden uns für „Probeneffekt vorhanden“, falls

$$(4) \quad R_b - R_0 \geq R_n^*$$

Nach Abb. 2 (a, b) ist die rechte Seite von (4) in ihrer Abhängigkeit von R_0 zwar konservativ, aber die Schwankungen der Messwerte R_0 betreffen auch direkt die linke Seite von (4), so dass $R_b - R_0$ gerade für kleinere Werte R_0 zu häufig die Bedingung (4) erfüllt. Unsere Formeln sind offenbar nicht in der Lage, die Varianz von R_0

vollständig im vorgegebenen Fehler 1. Art aufzufangen. Wir können uns aus diesem Dilemma nur befreien, wenn wir R_0 „noch länger“ messen, d.h. die Streubreite verringern oder die Erkennungsgrenze R_n^* „irgendwie“ größer machen.

Aufgabe 3

Wir sehen in Abb. 5, dass die Variation der möglichen Messwerte R_0 und R_b maßgeblich auch die Treffersicherheit unseres Vertrauensintervalls nach Formel A(3) beeinflusst. Der wahre Wert $\rho_b - \rho_0 = 1$ IPS liegt deutlich häufiger als jeweils $\gamma/2 = 2,5\%$ außerhalb der Grenzen. Bei unserem Versuch geben ca. 35% der Teilnehmer aufgrund ihrer Messung von R_b und R_0 eine zu kleine obere Vertrauensgrenze und ca. 60% eine zu große untere Grenze an. Wir können auch hier, ähnlich Aufgabe 2, die Varianz der Differenz $R_b - R_0$ nicht vollständig durch die vorgegebenen Vertrauensgrenzen auffangen.

Auswege?

Die beschriebenen Schwierigkeiten lassen sich durch folgende Überlegungen vermeiden:

A) Erkennungsgrenze R^* als obere Toleranzgrenze der Nulleffektverteilung:

In den Formeln nach der Norm werden die gemessenen Werte R_0 bzw. R_b als Schätzwerte der wahren Werte ρ_0 bzw. ρ_b fixiert. Alle Schlussfolgerungen aus diesen Formeln beruhen implizit auf der Annahme, dass die wahren Werte bekannt sind. Die tatsächliche Varianz der Verteilung von R_0 bzw. R_b wird dadurch aber, insbesondere bei kürzeren Messzeiten, nicht adäquat berücksichtigt.

Zur Abhilfe machen wir eine „Anleihe“ bei der Statistik: Mit einem Stichprobenwert x einer Verteilung $f(x)$ können wir nicht nur einzelne Parameter (z.B. Erwartungswert) schätzen, sondern man kann auch ganze Bereiche der Verteilung näher bestimmen. Wollen wir z.B. mit Wahrscheinlichkeit δ' einen Bereich der Verteilung $f(x)$ festlegen, dessen Fläche unter $f(x)$

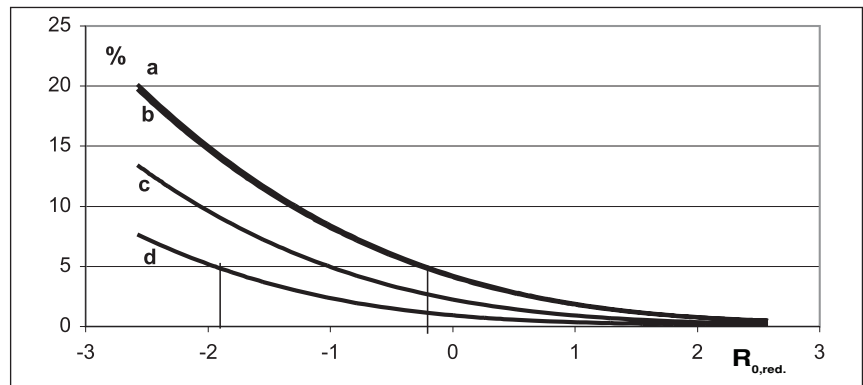


Abb. 4: Häufigkeit der Klassifikation von Nulleffektmesswerten größer Erkennungsgrenze als Probeneffekt, als Funktion des zu Beginn gemessenen Nulleffekts a) nach Formel A(1), b) Formel A(1a), c) „2-Sigma-Regel“, d) mit R^* nach A)

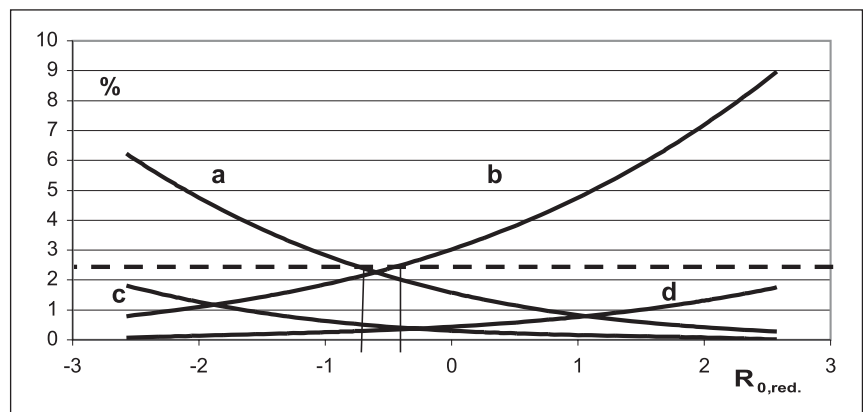


Abb. 5: Häufigkeit der Lage des Wahren Wertes $\rho_N = 1$ IPS außerhalb der vorgegebenen einseitigen Vertrauensintervalle mit $\gamma/2 = 2,5\%$, a) außerhalb oberer Vertrauensgrenze nach A(3), b) außerhalb unterer Vertrauensgrenze nach A(3), c) außerhalb unterer Vertrauensgrenze nach B), d) außerhalb oberer Vertrauensgrenze nach B)

mindestens α' entspricht, dann nennt man die Grenzen $L^-(x)$ bzw. $L^+(x)$ dieses Bereichs untere bzw. obere Toleranzgrenze und das Intervall $[L^-(x), L^+(x)]$ ein sog. (δ', α') -Toleranzintervall¹⁾.

In Abb. 2 sahen wir z.B., dass nur ca. $\delta' = (1-\delta) = 60\%$ der Teilnehmer aufgrund ihrer Erkennungsgrenze die Nulleffektwerte korrekt mit Wahrscheinlichkeit $\alpha' = (1-\alpha) = 95\%$ zugeordnet. Wollen wir erreichen, dass z.B. mindestens $\delta' = 97,5\%$ der Teilnehmer die Klassifikation korrekt mit $\alpha' = 95\%$ durchführen, dann müssen wir nur eine Erkennungsgrenze R^* als obere (97,5%, 95%)-Toleranzgrenze für $f(R_0)$ festlegen. Die Festlegung erfolgt in zwei Schritten. Das Konstruktionsprinzip in Abb. 6 und Abb. 7 veranschaulicht schneller den Sachver-

halt als die der Darstellung zugrunde liegende Theorie [Zacks]²⁾.

Schritt 1, Abb. 6: Mithilfe der gemessenen Nullrate R_0 werden die Vertrauensgrenzen ρ_0^- und ρ_0^+ für den wahren Wert ρ_0 gebildet (einseitiges Vertrauensniveau δ'). Für eine Gauss-Verteilung lassen sich die Ergebnisse analytisch angeben als Lösung der quadratischen Gleichung $\rho_0^\pm = R_0 \pm k_{1-\delta} \sqrt{\rho_0^\pm / t_0}$. Vertrauensgrenzen für eine Poisson-Verteilung bestimmt man am einfachsten numerisch (bei einem Bayes-statistischen Ansatz für die Verteilung

1) Konfidenz- oder Vertrauensintervalle überdecken mit vorgegebener Wahrscheinlichkeit einen Parameter der Verteilung, Toleranzintervalle legen hingegen einen Bereich der Zufallsvariablen fest.

2) Zacks, S. (1970), J. Amer. Statist. Ass., 65, 307

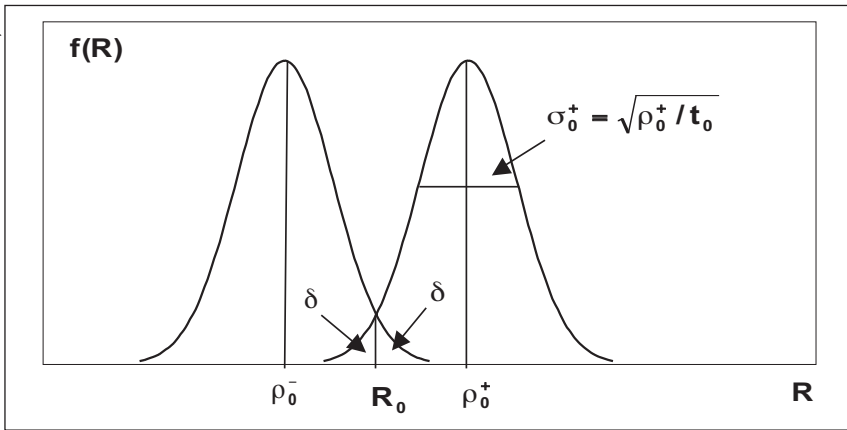


Abb. 6: Festlegung eines zweiseitigen Vertrauensintervalls für ρ_0 , Niveau $(1-2\delta)$

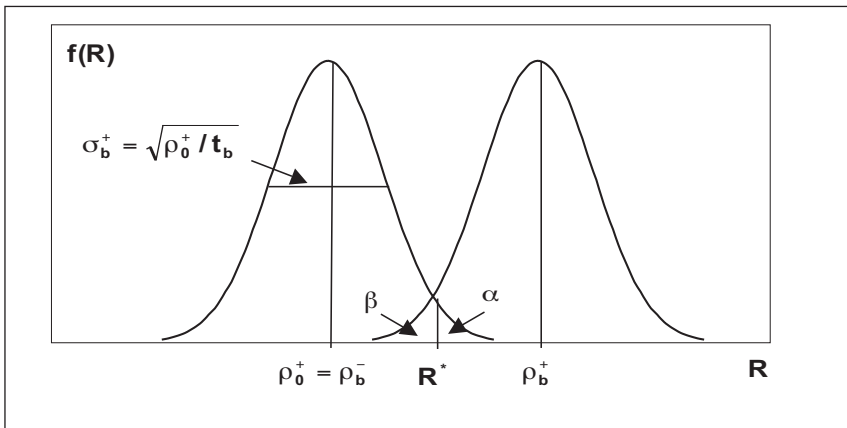


Abb. 7: Festlegung der Erkennungsgrenze R^* als $1-\alpha$ -Quantil der Verteilung um ρ_0^+ , Festlegung einer oberen ρ_b^+ - und unteren ρ_b^- -Vertrauensgrenze für ρ_b bei einem Brutto-Messwert $R_b = R^*$

ρ_0 kann das $\rho_0(\delta)$ -Quantil anstelle von ρ_0^+ verwendet werden).

Schritt 2, Abb. 6: Das $1-\alpha = \alpha'$ -Quantil der Verteilung um ρ_0^+ mit der Standardabweichung $\sigma_b^+ = \sqrt{\rho_0^+ / t_b}$ legt unsere Erkennungsgrenze R^* für eine Messzeit t_b fest. Der Messwert R_b wird direkt mit R^* verglichen (R^* ist keine Netto-Zählrate!)

Abb. 3 (d) zeigt das Verhältnis $R^* - R_0$ zum Wert (1) für die Parameter $\delta = 0,025$ und $\alpha = 0,05$. Wir sehen, dass die 5%-Marke für $R_{0,red.} < 1,96$ überschritten wird. Diese $R_{0,red.}$ -Werte entsprechen, wie vorgegeben, gerade einem Anteil von $1-\delta = 97,5\%$.

B) „Robuste“ Vertrauensgrenzen für den Netto-Effekt:

Die Bestimmung exakter Vertrauensgrenzen für die Differenz von Erwartungswerten ist eine Wissenschaft für

sich! Wir wollen unsere Vertrauensgrenzen für ρ_n so konstruieren, dass sie konservativ die Wahrscheinlichkeitsvorgaben einhalten, d.h., dass $P(\rho_n^- \leq \rho_n \leq \rho_n^+) \geq 1-\gamma$. Analog dem Verfahren von Abb. 6 bilden wir zunächst Vertrauensintervalle für ρ_0 bzw. ρ_b mit $P(\rho_0^- \leq \rho_0 \leq \rho_0^+) = 1-2\delta$ bzw. $P(\rho_b^- \leq \rho_b \leq \rho_b^+) = 1-2\phi$. Hieraus folgt unmittelbar eine Abschätzung für $\rho_n = \rho_b - \rho_0$

$$(5) \quad P(\rho_b^- - \rho_0^+ \leq \rho_b - \rho_0 \leq \rho_b^+ - \rho_0^-) \geq 1-\gamma \text{ mit } \gamma/2 = 1-(1-\phi) \cdot (1-\delta)$$

Das Intervall (5) ist somit sicher nicht das kleinstmögliche Vertrauensintervall für ρ_n , aber es „funktioniert“ sozusagen immer, und das Ergebnis (5) gilt für alle Verteilungen von ρ_b und ρ_0 . Abb. 5(c, d) zeigt für (5) vergleichsweise Ergebnisse zu Aufgabe 3 mit $1-\delta = 0,99$ bzw. $1-\phi = 0,985$ und damit für das einseitige Niveau $\gamma/2 =$

2,5%. Die Vorgabe $\gamma/2 = 2,5\%$ wird aufgrund (5) nicht überschritten.

C) Nachweisgrenze ρ_n^* als obere Vertrauensgrenze ρ_n^+ zum Messwert $R_b = R^*$

Die Nachweisgrenze ρ_n^* gibt an, welche kleinste Netto-Aktivität mit unserem Verfahren noch nachweisbar ist. Nachweisbar heißt, der kleinste mögliche Probeneffekt-Messwert R_b muss noch zu einem Vertrauensbereich führen, der den wahren Netto-Wert ρ_n mit Wahrscheinlichkeit $1-\gamma$ überdeckt. Aus der Konstruktion Abb. 6 erkennen wir, dass dort ein wahrer Brutto-Wert $\rho_b = \rho_b^+$ gerade mit Fehlerwahrscheinlichkeit β den Messwert $R_b = R^*$ noch zulässt. $R_b = R^*$ führt somit bei vorgegebenem Nulleffekt nach B) zu einem kleinsten Netto-Wert $\rho_n = \rho_b^+ - \rho_0^-$, d.h., $\rho_b^+ - \rho_0^-$ ist unsere Netto-Nachweisgrenze ρ_n^* !

$$(6) \quad \rho_n^* = \rho_b^+ - \rho_0^-, \text{ mit } \rho_b^+ \text{ als oberer Vertrauensgrenze für } R_b = R^*$$

Der wahre Wert ρ_b muss natürlich immer größer als ρ_0 sein! Dies hat zur Folge, dass sich die Vertrauensbereiche von ρ_b und ρ_0 nicht überlappen dürfen, d.h., dass $\rho_b^- \geq \rho_0^+$ sein muss. Es wäre sonst immer mit einer bestimmten Wahrscheinlichkeit auch ein Ergebnis $\rho_b < \rho_0$ möglich.

Fazit

- 1) Um den Fehler 1. Art tatsächlich zu beschränken, müssen wir bei der Festlegung der Erkennungsgrenze zusätzlich spezifizieren, welcher Bereich der Nulleffektverteilung durch unsere Nulleffektmessung R_0 berücksichtigt werden soll (Festlegung als Toleranzintervall).
- 2) Unsere Erkennungsgrenze R^* stellt einen Probeneffekt durch direkten Vergleich mit dem Brutto-Messwert R_b fest. R^* stellt den kleinstmöglichen Brutto-Messwert dar, falls Aktivität in Höhe der Nachweisgrenze ρ_n^* vorliegt.
- 3) Schätzwerte der wahren Werte

Formeln nach DIN 25482

(1)	$R_n^* = \frac{1}{2t_0} k_{1-\alpha}^2 \left(1 + \sqrt{1 + \frac{4R_0 t_0}{k_{1-\beta}^2} \left(1 + \frac{t_0}{t_b} \right)} \right)$	Teil 1, Tabelle 1
(1a)	$R_n^* = k_{1-\alpha} \sqrt{R_0 \left(\frac{1}{t_0} + \frac{1}{t_b} \right)}$	Teil 1, Tabelle 1
(1')	$R_n^* = k_{1-\alpha} \cdot \sqrt{\rho_0 / t_0}$	aus (1) bzw. (1a) für $t_0 \rightarrow \infty$
(2)	$\rho_n^* = (k_{1-\alpha} + k_{1-\beta}) \sqrt{\rho_0 \left(\frac{1}{t_0} + \frac{1}{t_b} \right)} + \frac{1}{4} (k_{1-\alpha} + k_{1-\beta})^2 \left(\frac{1}{t_0} + \frac{1}{t_b} \right)$	Teil 1, Tabelle 1
(2a)	$\rho_n^* = (k_{1-\alpha} + k_{1-\beta}) \sqrt{\rho_0 \left(\frac{1}{t_0} + \frac{1}{t_b} \right)}$	Teil 1, Tabelle 1
(2b)	$\rho_n^* = k_{1-\alpha} \sqrt{\rho_0 / t_b} + k_{1-\beta} \sqrt{\rho_n^* / t_b + \rho_0 / t_b}$	implizite Vorschrift nach Formel (3), Abschnitt 6.4 Teil 10
(2')	$\rho_n^* = 2k_{1-\alpha} \cdot \sqrt{\rho_0 / t_0} + k_{1-\alpha}^2 / t_b$	aus (2) bzw. (2b) für $t_0 \rightarrow \infty$ und $\alpha = \beta$
(3)	$R_n - k_{1-\gamma/2} \sqrt{\frac{R_0}{t_0} + \frac{R_b}{t_b}} \leq \rho_n \leq R_n + k_{1-\gamma/2} \sqrt{\frac{R_0}{t_0} + \frac{R_b}{t_b}}$	Teil 1, Tabelle 1

sind $\hat{\rho}_0 = R_0$, $\hat{\rho}_b = R_b$ bzw. $\hat{\rho}_n = R_b - R_0$. Die Vertrauensintervalle für den wahren Wert ρ_n gewährleisten, dass $\rho_n = \rho_b - \rho_0 \geq 0$. Eine Zensur negativer ρ_n -Werte entfällt.

- 4) Bei kleinen Impulszahlen lassen sich unsere Überlegungen auf Poisson-Verteilungen anwenden.

Die Bestimmung der Vertrauensgrenzen erfolgt am besten numerisch (siehe Webseite der BA-Karlsruhe www.ba-karlsruhe.de).

- 5) Grundsätzlich kann ein Bezug von Erkennungs- bzw. Nachweisgrenzen auf Konfidenz- bzw. auf Toleranzintervalle immer auch als sta-

tistischer Hypothesentest formuliert werden.

- 6) Unsere Überlegungen führen für den Idealfall $t_0 \rightarrow \infty$ natürlich zu den Ergebnissen (1) und (2).

Wolfgang Kraut, Karlsruhe
E-Mail: kraut@ba-karlsruhe.de □