Prof. Dr. Timm Grams
Hochschule Fulda
Fachbereich Elektrotechnik und Informationstechnik
Marquardstraße 35
36039 Fulda

Tel (0661) 78157
(0661) 9640-550
Fax (0661) 9640-559
mailto:Timm.Grams@et.hs-fulda.de
www.hs-fulda.de/~grams

Montag, 2. Juni 2008

Des Ingenieurs Schwierigkeiten mit der schließenden Statistik

Für Niklas Luhmann (1991) ist Technik das Gebiet der *Kausalsimplifikationen*, also ein in der Komplexität reduzierter und simplifizierter und von externen Einflüssen weitgehend isolierter Bereich mit festen, funktionierenden Kopplungen, so „dass (1) Abläufe kontrollierbar, (2) Ressourcen planbar und (3) Fehler (einschließlich Verschleiß) erkennbar und zurechenbar werden“.

In dieser – gemessen an der Komplexität des Lebens – doch sehr übersichtlichen Welt fällt dem Ingenieur gar nicht auf, dass jede mathematische Formel in ein Begriffsumfeld eingebettet ist, das den Interpretationsrahmen vorgibt. Für ihn sind die in den Formeln vorkommenden Größen immer klar. Er weiß, was Kraft, Masse, Geschwindigkeit, Spannung, Strom, Feldstärke bedeuten. Die Interpretation der Formeln und der Ergebnisse bereiten ihm keine Schwierigkeiten. Tatsächlich hat der Ingenieur sich eine Welt geschaffen, in der die mathematischen Formeln eine deutlich erkennbare und eindeutige Rolle spielen.

Mit dieser Denkweise ausgestattet, bekommt der Ingenieur naturgemäß Probleme in dem Moment, in dem der *Bezugsrahmen* einer Formel nicht mehr so ohne Weiteres ersichtlich ist. In der *schließenden Statistik* ist das leider so. Zwar kommen in den Formeln der schließenden Statistik immer Wahrscheinlichkeiten vor. Aber die Interpretation der Wahrscheinlichkeitswerte hängt davon ab, ob man es mit statistischer Testtheorie, mit Konfidenzintervallen, Bayes-Schätzungen oder Fiduzialintervallen zu tun hat.

Dass Schwierigkeiten mit dem Bezugsrahmen sogar den Experten auf dem Gebiet der Statistik zu schaffen machen, ist gut belegt. Beispielsweise bei der Verwechslung der Konfidenzintervalle mit den Fiduzialintervallen. Bei den Konfidenzintervallen wird der zu schätzende Parameter als konstant vorausgesetzt, bei den Fiduzialintervallen hingegen als zufällig. Gladitz (1994) stellt diese beiden Ansätze an einem einfachen Beispiel, nämlich der Schätzung einer Wahrscheinlichkeit, einander gegenüber.

Székely (1990, S. 109) schreibt: „Eine Zeitlang dachte man, dass diese beiden Intervalle praktisch identisch sind und dass die Debatte über Konfidenz- oder Fiduzialintervalle rein theoretischer Natur sei… Jedoch zeigten sich ziemlich schnell Paradoxa von Bedeutung... Im Jahre 1959 wies *C. Stein* auf einen äußerst paradoxen Fall hin.“

Ein anderes, ein aktuelles Beispiel von falscher Rahmensetzung ist im Buch von Ehrenberger (2002) auf Seite 116 zu finden. In der folgenden Kopie der Seite sind die Rahmenbezüge zur statistische Testtheorie handschriftlich nachgetragen bzw. korrigiert. Auf der Web-Page DENKFALLEN UND PARADOXA wird dieses Beispiel unter dem Stichwort „Software-Verifikation und -Test“ analysiert:

http://www2.hs-fulda.de/~grams/dnkfln.htm#_Software-Verifikation_und_-Test

Beispiel 8.2:

Die Forderung, p sei mit mindestens 99 %iger Wahrscheinlichkeit kleiner oder gleich einem bestimmtem Wert $\tilde{p}$, ist gleich der Forderung, p sei mit höchstens 1 %iger Wahrscheinlichkeit größer als eben dieser Wert. Anschaulicher: die erste Formulierung sagt in frequentistischer[1] Ausdrucksweise, dass von 100 gedachten Experimenten 99 auf ein p unter oder höchstens gleich dem Grenzwert $\tilde{p}$ geführt hätten; die zweite sagt, dass von 100 gedachten Experimenten höchstens eines auf ein p geführt hätte, das größer ist als $\tilde{p}$.

So gesehen, ist die Aussagesicherheit ein Maß dafür, wie sehr man sich auf die in einem Experiment, in unserem Fall einem Test getroffene Auswahl verlassen kann.

Es interessiert der Zusammenhang zwischen notwendiger Testanzahl n, Aussagesicherheit β und Grenzwert $\tilde{p}$. Um darauf zu kommen, sei daran erinnert, wie Hypothesen getestet werden:

NEIN!

NEIN: Es muss heißen „Alternativ-hypothese"

Der *Test von Hypothesen* geht über die Falsifizierung ihres Komplements.

Dies verwendend, konzipieren wir ein Experiment, hier einen Test, wie folgt:

- Über die Versagenswahrscheinlichkeit p der Software wird die Hypothese „$p > \tilde{p}$" aufgestellt.
- Der Test wird mit einer so großen Zahl n von Einzelfällen konzipiert, dass die ihm unterworfene Software, sollte die Hypothese richtig sein, mit der Wahrscheinlichkeit β versagt, d. h. mindestens einen Versagensfall zeigt.
- Dann wird der Test durchgeführt.
- Versagt die Software in seinem Verlauf nicht, wird geschlossen: Die ursprüngliche Hypothese ist falsch.
- Dann gilt deren Komplement, und zwar wieder mit der Wahrscheinlichkeit β.

Dieser Denkweise folgend setzen wir nach (8.2), mindestens einmaliges Versagen in n Testfällen zugrunde legend:

$$1-(1-p)^n \le \beta.$$

$H_0 = (p < \tilde{p})$; E = kein Versagen;
$H_1 = (p = \tilde{p})$

Wenn wir statt p den Grenzwert $\tilde{p}$ nehmen, ist stattdessen zu schreiben:

$$1-(1-\tilde{p})^n = \beta = 1 - p(E|H_1) \text{ oder } p(E|H_1) = 1-\beta \quad (8.6)$$

Dies bedeutet: Wenn die *Hypothese* „$p > \tilde{p}$" wahr ist, versagt die Software im Verlauf der n Testfälle mit der Wahrscheinlichkeit β mindestens ein Mal.[2]

Versagt die Software im Verlauf dieser n Testfälle nicht, war die Hypothese „$p > \tilde{p}$" falsch. Es gilt dann deren Komplement. Dies ist im vorliegenden Fall ihr Gegenteil, „$p \le \tilde{p}$". Also gilt dann:

[1] häufigkeitsbezogener, den Quotienten zweier Anzahlen bildender

[2] Man führe sich vor Augen: Wenn p wächst, sinkt der Wert von $(1-p)$ und damit sinkt auch der Wert von $(1-p)^n$, weil stets $n >> 0$, und der Wert von $1-(1-p)^n$ steigt.

Das sind die zu den Formeln passenden Definitionen von Nullhypothese, Alternativhypothese und beobachtetem Ergebnis. Anders geht es nicht!

Literaturhinweise

Ehrenberger, W.: Software-Verifikation. Verfahren für den Zuverlässigkeitsnachweis von Software. Hanser, München, Wien 2002

Gladitz, Johannes: Fiduzialintervalle für den Parameter der Binomialverteilung mit SPSS 6.0 für Windows. RZ-Mitteilungen der Humboldt-Universität zu Berlin Nr. 9 (Dezember 1994), S. 21-26 (Erreichbar über den edoc-Server der Humboldt-Universtät zu Berlin. Artikel aus dem cms-journal)

Luhmann, N., 1991: Soziologie des Risikos, Berlin/New York: de Gruyter

Székely, Gábor J.: Paradoxa. Klassische und neue Überraschungen aus Wahrscheinlichkeitsrechnung und mathematischer Statistik. Harri Deutsch, Thun, Frankfurt am Main, 1990