

# Sequentielles Testen

Raphael Diepgen

Sequentielles Testen von Hypothesen ist ein Testen im Rahmen der Neyman-Pearson-Theorie statistischen Testens (vgl. Willmes, in diesem Band). Es geht also um vor der Datenerhebung definierte regelhafte Verfahren zur Entscheidung zwischen miteinander konkurrierenden statistischen Hypothesen – eigentlich: zwischen verschiedenen *courses of action* – auf der Basis von Daten aus Zufallsstichproben, wobei die hypothetischen Wahrscheinlichkeiten für die Entscheidungen zugunsten einer Hypothese bei angenommener Geltung einer anderen Hypothese Maßstab für die Entwicklung und Beurteilung der Entscheidungsverfahren sind. Während die klassischen Neyman-Pearson-Tests eine Entscheidung grundsätzlich erst nach Erhebung einer vollständigen Zufallsstichprobe vorab definierter Größe vorsehen, wird bei sequentiellen Tests – gelegentlich auch Folgetestpläne genannt – nach jedem sukzessiv erhobenen Stichprobendatum erneut eine Entscheidung getroffen, und zwar entweder für eine der konkurrierenden Hypothesen – und damit für die Beendigung der Datenerhebung – oder aber stattdessen für die Erhebung eines weiteren Datums. Im Vergleich zu den herkömmlichen nichtsequentiellen Tests mit fester Stichprobengröße führen sequentielle Tests bei vergleichbaren Fehlerwahrscheinlichkeiten im Schnitt schon bei wesentlich geringeren Stichprobengrößen zu Entscheidungen zwischen den konkurrierenden Hypothesen. Sequentielle Tests leisten also im Prinzip das gleiche wie herkömmliche Neyman-Pearson-Tests, dies aber in der Regel bei erheblich kleineren Stichprobenumfängen. Sequentielle Tests versprechen also insbesondere eine Verringerung des empirischen Aufwandes. Genau mit diesem Ziel wurde die Entwicklung sequentieller Tests in den vierziger Jahren von dem deutschstämmigen Mathematiker Abraham Wald im Dienste der amerikanischen Rüstungsindustrie initiiert. Die Verringerung des empirischen Aufwandes ist dann besonders wichtig, wenn die Datenerhebung teuer oder ethisch problematisch ist. Neben der Verringerung des empirischen Aufwandes versprechen sequentielle Verfahren freilich noch eine Reihe weiterer methodischer Vorteile.

Nicht nur das Hypothesentesten läßt sich sequentiell organisieren, sondern etwa auch die Bestimmung von Konfidenzintervallen (vgl. Wald, 1966). Dies soll hier aber ebensowenig Thema sein wie die selbstverständliche und überaus naheliegende Möglichkeit, im konzeptionellen Rahmen der Bayes-Statistik (vgl. Molenaar & Lewis, in diesem Band) sequentiell zu arbeiten.

## 1 Ein konkretes Beispiel

Betrachten wir zur Illustration als fiktives Beispiel das Problem, experimentell die Wirksamkeit einer neuen Therapie im Vergleich zu einer herkömmlichen Therapie

mit einer Erfolgswahrscheinlichkeit von 50% zu beurteilen. Entschieden werden soll also zwischen den beiden Hypothesen  $H_0 : p \leq 0.5$  und  $H_1 : p > 0.5$  über die bislang unbekannte Erfolgswahrscheinlichkeit  $p$  der neuen Therapie. (Letztendlich soll selbstverständlich zwischen zwei Handlungsoptionen entschieden werden, nämlich zwischen der Einführung der neuen Therapie und der Beibehaltung der herkömmlichen Therapie.) Das Entscheidungsverfahren soll folgende Eigenschaften haben: Gilt tatsächlich  $H_0$ , ist also die neue Therapie nicht besser als die alte, so soll eine (Fehl-)Entscheidung für  $H_1$  lediglich mit einer Wahrscheinlichkeit von höchstens  $\alpha = 0.05$  fallen. Beträgt aber die Erfolgswahrscheinlichkeit der neuen Therapie tatsächlich mindestens 10 Prozentpunkte mehr als die der alten, gilt also  $p \geq 0.6$ , so soll eine (Fehl-)Entscheidung zugunsten  $H_0$  lediglich mit einer Wahrscheinlichkeit von höchstens  $\beta = 0.10$  fallen.

Diese beiden Forderungen definieren im Rahmen der Testtheorie von Neyman und Pearson Signifikanzniveau und Testgüte – und damit letztlich die Stichprobengröße  $n = 200$  – für einen herkömmlichen Binomialtest. Dieser Test lautet als Entscheidungsregel:

Man unterziehe eine Zufallsstichprobe von 200 Patienten der neuen Therapie. Falls die neue Therapie bei mehr als 111 der behandelten 200 Patienten Erfolg hat, entscheide man sich zugunsten  $H_1$ . Andernfalls entscheide man sich zugunsten  $H_0$ .

Dieser klassische Test erfordert also die Erhebung einer Stichprobe des Umfangs  $n = 200$  – und damit einen erheblichen empirischen Aufwand. Der Logik dieses Verfahrens gemäß wäre man allenfalls dann zum vorzeitigen Abbruch der Datenerhebung berechtigt, wenn man entweder schon mehr als 111 Patienten erfolgreich oder schon mindestens 89 Patienten erfolglos behandelt hätte, wenn also die Entscheidung durch die bis zur Gesamtzahl 200 noch fehlenden Patienten nicht mehr beeinflusst werden könnte.

Die obigen Forderungen werden aber auch von folgendem – als rekursive Regel formulierten – sequentiellen Test erfüllt:

Man unterziehe nach und nach zufällig ausgesuchte Patienten der neuen Therapie. Es sei  $n$  die Anzahl der auf diese Art bislang behandelten Patienten und  $m$  die Anzahl derjenigen davon, bei denen die Behandlung Erfolg hatte. Falls  $m \geq 0.5503397 \cdot n + 7.128534$  ist, so beende man die Versuchsserie und entscheide sich für  $H_1$ . Ist  $m \leq 0.5503397 \cdot n - 5.552368$ , so beende man die Versuchsserie und entscheide sich für  $H_0$ . Gilt schließlich  $0.5503397 \cdot n - 5.552368 < m < 0.5503397 \cdot n + 7.128534$ , so setze man die Versuchsserie fort, unterziehe man also einen weiteren Patienten der neuen Therapie und entscheide erneut, jetzt aber mit einem um 1 erhöhten  $n$ .

Dieser sequentielle Test läßt sich sehr bequem auch graphisch organisieren, denn die auftretenden Ungleichungen definieren in einem  $(n, m)$ -Koordinatensystem jeweils

durch Geraden mit der einheitlichen Steigung 0.5503397 begrenzte Gebiete. Dieser sequentielle Test führt zu einer Entscheidung für  $H_1$  beispielsweise schon nach folgendem Verlauf von 22 Erfolgen (E) und 5 Mißerfolgen (M): E, M, E, E, E, E, M, E, E, E, E, M, E, E, E, E, E, E, M, E, E, E, M, E, E. Die Entscheidungsfindung in diesem Beispiel nach nur 27 statt 200 Probanden deutet die enormen Vorteile der sequentiellen gegenüber den klassischen Tests an.

## 2 Der sequentielle Quotiententest von Wald

### 2.1 Die Form des Tests

Der hier exemplarisch vorgestellte Test ist ein Spezialfall des sequentiellen Quotiententests von Wald (1945). Dieser für die Sequentialstatistik grundlegende Test hat folgende allgemeine Struktur:

Entschieden werden soll zwischen der Nullhypothese  $H_0 : p = p_0$  und der Alternativhypothese  $H_1 : p = p_1$ , wobei  $p$  die unbekannte Wahrscheinlichkeit eines Merkmals in einer Population ist. Das Entscheidungsverfahren auf der Basis sukzessiver zufälliger Ziehungen aus der Population soll zu vorab gewählten Größen  $\alpha$  und  $\beta$  folgenden Forderungen genügen:

- ( $F_1$ ) Die Wahrscheinlichkeit für eine Entscheidung zugunsten  $H_1$  unter der Annahme, daß  $H_0$  gilt, betrage höchstens  $\alpha$ . Formal:  $p(\text{Entscheidung für } H_1 | H_0) \leq \alpha$ . (Fehlerwahrscheinlichkeit erster Art).
- ( $F_2$ ) Die Wahrscheinlichkeit für eine Entscheidung zugunsten  $H_0$  unter der Annahme, daß  $H_1$  gilt, betrage höchstens  $\beta$ . Formal:  $p(\text{Entscheidung für } H_0 | H_1) \leq \beta$ . (Fehlerwahrscheinlichkeit zweiter Art).

Der sequentielle Test regelt, wie nach jedem der sukzessiv erhobenen Stichprobendaten zwischen den drei Möglichkeiten – Beendigung der Datenerhebung mit einer Entscheidung für  $H_0$ , Beendigung der Datenerhebung mit einer Entscheidung für  $H_1$  oder schließlich Fortsetzung des Experimentes – zu entscheiden ist. Nehmen wir also an, wir haben nach und nach  $n$  Stichprobendaten erhoben und dabei in einer bestimmten Reihenfolge  $m$ -mal das fragliche Merkmal vorgefunden – und damit  $(n - m)$ -mal nicht. Dieses Ereignis – diesen Experimentalverlauf – wollen wir mit  $E$  abkürzen. Die Wahrscheinlichkeit für dieses Ereignis beträgt  $p^m \cdot (1 - p)^{n-m}$ . Es sei  $l_0(E)$  die Wahrscheinlichkeit für dieses Ereignis  $E$ , falls  $H_0$  gilt, und dementsprechend  $l_1(E)$  die Wahrscheinlichkeit für  $E$ , falls  $H_1$  gilt. Diese beiden hypothetischen Wahrscheinlichkeiten für das eingetretene Ereignis  $E$ , also diese beiden *likelihoods*, lassen sich offensichtlich leicht berechnen, denn es gilt

$$l_0(E) = p_0^m \cdot (1 - p_0)^{n-m} \quad \text{und} \quad l_1(E) = p_1^m \cdot (1 - p_1)^{n-m}. \quad (1)$$

Walds sequentieller Quotiententest, gelegentlich auch Wahrscheinlichkeitsverhältnistest genannt, besteht nun aus folgender Entscheidungsregel:

- (R<sub>1</sub>) Falls  $l_1(E)/l_0(E) \leq \beta/(1 - \alpha)$ , beende man die Datenerhebung und entscheide sich für  $H_0$ .
- (R<sub>2</sub>) Falls  $l_1(E)/l_0(E) \geq (1 - \beta)/\alpha$ , beende man die Datenerhebung und entscheide sich für  $H_1$ .
- (R<sub>3</sub>) Andernfalls setze man die Datenerhebung fort und entscheide dann erneut.

Für den mathematisch Interessierten folgt im nächsten Abschnitt der Beweis, daß diese Entscheidungsregel tatsächlich den Forderungen (F<sub>1</sub>) und (F<sub>2</sub>) genügt.

## 2.2 Die formale Begründung der Entscheidungsregel

Einen Fehler erster Art begehen wir, wenn wir uns für  $H_1$  entscheiden, obwohl  $H_0$  gilt.  $M_1$  sei die Menge aller möglichen Experimentalverläufe  $E$ , die im sequentiellen Quotiententest mit einer Entscheidung für  $H_1$  enden. Für alle diese Experimentalverläufe  $E$  aus  $M_1$  gilt nach obiger Entscheidungsregel (R<sub>2</sub>):

$$\frac{l_1(E)}{l_0(E)} \geq \frac{1 - \beta}{\alpha} \quad \text{oder dazu äquivalent} \quad l_0(E) \leq l_1(E) \cdot \frac{\alpha}{1 - \beta}. \quad (2)$$

Wir entscheiden uns für  $H_1$ , wenn irgendein Experimentalverlauf  $E$  aus  $M_1$  beobachtet wird. Das Ereignis „Entscheidung für  $H_1$ “ entspricht also einer Oder-Verknüpfung bzw. Vereinigung aller Experimentalverläufe  $E$  aus  $M_1$ . Die Wahrscheinlichkeit für dieses Ereignis ergibt sich somit als Summe der Wahrscheinlichkeiten für alle Experimentalverläufe  $E$  aus  $M_1$ ; denn diese Experimentalverläufe „überschneiden“ sich nicht, sondern sind unvereinbare, disjunkte Elementarereignisse. Damit ergibt sich für die Fehlerwahrscheinlichkeit erster Art:

$$p(\text{Entscheidung für } H_1 | H_0) = p\left(\bigcup_{E \in M_1} E | H_0\right) \quad (\text{nach } (R_2)) \quad (3)$$

$$= \sum_{E \in M_1} p(E | H_0) \quad (\text{denn alle } E \in M_1 \text{ sind unvereinbar}) \quad (4)$$

$$= \sum_{E \in M_1} l_0(E) \quad (\text{nach Definition von } l_0(E) = p(E | H_0)) \quad (5)$$

$$\leq \sum_{E \in M_1} \left( l_1(E) \cdot \frac{\alpha}{1 - \beta} \right) \quad (\text{denn } (2) \text{ gilt für alle Summanden}) \quad (6)$$

$$= \left( \sum_{E \in M_1} l_1(E) \right) \cdot \frac{\alpha}{1 - \beta} \quad (\text{Distributivgesetz}) \quad (7)$$

$$= \left( \sum_{E \in M_1} p(E|H_1) \right) \cdot \frac{\alpha}{1 - \beta} \quad (\text{Definition von } l_1(E) = p(E|H_1)) \quad (8)$$

$$= p \left( \bigcup_{E \in M_1} E|H_1 \right) \cdot \frac{\alpha}{1 - \beta} \quad (\text{Additivität von Wahrscheinlichkeit}) \quad (9)$$

$$= p(\text{Entscheidung für } H_1|H_1) \cdot \frac{\alpha}{1 - \beta} \quad (\text{nach } (R_2)) \quad (10)$$

$$\leq \frac{\alpha}{1 - \beta} \quad (\text{denn } p(\text{Entscheidung für } H_1|H_1) \leq 1) \quad (11)$$

$$\approx \alpha, \text{ sofern } \beta \text{ klein gewählt wurde.} \quad (12)$$

Die Fehlerwahrscheinlichkeit erster Art beträgt also, da  $\beta$  üblicherweise klein gewählt wird, höchstens ungefähr  $\alpha$ . Die Forderung ( $F_1$ ) wird also vom sequentiellen Quotiententest bis auf kleine Ungenauigkeiten erfüllt. Völlig analog läßt sich zeigen, daß dieser Test auch der Forderung ( $F_2$ ) bezüglich der Fehlerwahrscheinlichkeit zweiter Art genügt. Selbstverständlich läßt sich bei Bedarf durch leichte Modifikation der Schranken  $\beta/(1 - \alpha)$  bzw.  $(1 - \beta)/\alpha$  sicherstellen, daß die maximal zugestandenen Fehlerwahrscheinlichkeiten exakt eingehalten werden.

### 2.3 Die praktische Durchführung des Tests

Für die praktische Anwendung dieses Tests formt man zweckmäßigerweise die Ungleichungen aus den Regeln ( $R_1$ ) bis ( $R_3$ ) durch beiderseitiges Logarithmieren in äquivalente lineare Ungleichungen um. Für  $p_0 = 0.5$ ,  $p_1 = 0.6$ ,  $\alpha = 0.05$  und  $\beta = 0.10$  ergeben sich dann genau die Ungleichungen, die oben den eingangs vorgestellten Sequentialtest definierten. Es dürfte schließlich unmittelbar plausibel sein, daß dieser Test auch zur Entscheidung zwischen den zusammengesetzten Hypothesen  $H_0 : p \leq 0.5$  und  $H_1 : p \geq 0.6$  geeignet ist: Je geringer die Erfolgswahrscheinlichkeit  $p$ , desto unwahrscheinlicher ist es auf jeder Stufe  $n$  des Experimentes, daß mindestens

$$0.5503397 \cdot n + 7.128534$$

Patienten erfolgreich behandelt werden; und je größer  $p$  ist, desto unwahrscheinlicher ist es wiederum, daß höchstens

$$0.5503397 \cdot n - 5.552368$$

Patienten geheilt werden.  $\alpha$  und  $\beta$  limitieren hier gleichsam die Fehlerwahrscheinlichkeiten jeweils für den ungünstigsten Fall  $p = 0.5$  bzw.  $p = 0.6$ .

## 3 Allgemeine Ergänzungen zu sequentiellen Tests

In dem halben Jahrhundert seit Ableitung dieses sequentiellen Quotiententests durch Wald wurden unüberschaubar viele weitere sequentielle Tests für die unterschiedlichsten Fragestellungen entwickelt (vgl. etwa Bauer, Scheiber & Wohlzogen, 1986). Insbesondere gibt es neben vielen nonparametrischen Sequentialtests auch parametrische Sequentialtests, etwa sequentielle  $t$ - und  $F$ -Tests – und damit die Möglichkeit,

auch im beliebten Rahmen Allgemeiner Linearer Modelle (vgl. Andres, in diesem Band) sequentiell zu testen.

Sequentielle Tests lassen sich nicht nur – wie die klassischen Neyman-Pearson-Tests – durch die Operationscharakteristik oder OC-Funktion charakterisieren, die jeweils die Wahrscheinlichkeit einer Entscheidung für  $H_1$  in Abhängigkeit von den hypothetischen wahren Verhältnissen angibt, sondern auch durch die sogenannte ASN-Funktion, die jeweils in Abhängigkeit von den hypothetischen wahren Verhältnissen den Erwartungswert für die bis zur Entscheidung zwischen den Hypothesen benötigte Stichprobengröße, also die *average-sample-number*, angibt. Bei sequentiellen Verfahren ist die Stichprobengröße schließlich eine Zufallsvariable, die sich beispielsweise durch Erwartungswert und Varianz kennzeichnen läßt. Die OC-Funktion beschreibt gleichsam, wie gut der Test sein Ziel erreicht, korrekte Entscheidungen zu treffen, und die ASN-Funktion benennt den Preis, den wir dafür zu zahlen haben.

Sequentielle Tests erfordern nicht unbedingt eine Entscheidung nach jedem einzelnen Stichprobendatum, sondern lassen die Entscheidung auch jeweils nach Gruppen von Stichprobendaten zu (z.B. nach jedem 10. Datum, sog. Gruppierungspläne). Die möglichen versuchs- oder auswertungstechnischen Vorteile dieser gruppierten Pläne werden allerdings erkaufte durch eine leichte Erhöhung der ASN-Funktion, also durch einen gewissen empirischen Mehraufwand (vgl. Wald, 1966).

Sequentielle Tests lassen sich durch einfache Stopp- oder Abbruchregeln begrenzen, was die maximale Stichprobengröße angeht; aus offenen sequentiellen Testplänen werden so geschlossene (oder beschränkte) sequentielle Testpläne. Nach Erreichen einer hinreichend großen (und damit unter beiden Hypothesen sehr unwahrscheinlichen) Stichprobengröße läßt sich schließlich die Entscheidung zwischen den beiden Hypothesen notfalls auch über einen Münzwurf organisieren, ohne die Fehlerwahrscheinlichkeiten erster und zweiter Art nennenswert zu vergrößern.

In der Regel liegt auch noch der höchste Punkt der ASN-Funktion bei sequentiellen Testverfahren deutlich unter der festen Stichprobengröße analoger nichtsequentieller Neyman-Pearson-Tests mit gleichem Signifikanzniveau und gleicher Teststärke (vgl. etwa Wald & Wolfowitz, 1950). Alles in allem lassen sequentielle Tests bei gleicher Leistungsfähigkeit viel geringere Stichprobengrößen als ihre nichtsequentiellen Analoga erwarten. Einsparungen in der Größenordnung von 30%, 50% oder gar 70% sind keine Seltenheit; als durchschnittliche Einsparung wird in der Literatur immer wieder zur groben Orientierung ein Drittel genannt.

#### 4 Vermutliche Gründe für die geringe Verbreitung sequentieller Tests

Angesichts dieser enormen ökonomischen Vorteile sequentieller Tests gegenüber herkömmlichen Verfahren bedarf es der Erklärung, warum sequentielles Testen im psychologischen und sozialwissenschaftlichen Forschungsalltag kaum einmal vorkommt – anders als etwa in der industriellen Qualitätskontrolle. Als erste Erklärung bietet sich hier der Hinweis auf das geringe Kostenbewußtsein im bürokratisierten und öffentlich finanzierten Wissenschaftsbetrieb an, das bekanntlich bis zur aggressiven Ablehnung jedweder Kosten-Nutzen-Erwägung für den Bereich „freier“ Forschung und Wissenschaft geht. Plausibler erscheinen indessen andere Erklärungen:

Sequentielle Tests zwingen Forscher, tatsächlich vor der Datenerhebung die zu testenden statistischen Hypothesen explizit zu formulieren – insbesondere auch die Mindestgröße relevanter Unterschiede oder Effekte – und die Schranken für die Fehlerwahrscheinlichkeiten erster und zweiter Art festzulegen. Ohne diese Festlegungen „funktionieren“ sequentielle Tests einfach nicht. Von ihrer Theorie her erfordern zwar auch die Neyman-Pearson-Tests notwendig solche Festlegungen vor der Datenerhebung; ohne solche Festlegung sind sie sinnlos. Sie ermöglichen es aber von ihrer äußeren Durchführungspraxis her, diese Festlegungen erst im Nachhinein zu treffen. Die äußere Form klassischer Neyman-Pearson-Tests eröffnet damit ein weites Feld der „Als-ob-Statistik“: Der Forscher, der tatsächlich – wie auch durch die weitverbreiteten Statistikprogramm Pakete nahegelegt – erst nach der Datenerhebung „Tests“ festlegt und diese „Tests“ dann als Neyman-Pearson-Tests verkauft, tut so, als ob er diese Festlegungen vor der Datenerhebung getroffen hätte. Den Signifikanzen („ $p$ -Werten“) in den veröffentlichten Forschungsberichten läßt sich schließlich nicht ansehen, ob sie vorab festgelegte Fehlerwahrscheinlichkeiten im Sinne von Neyman und Pearson sind oder aber lediglich im Nachhinein berechnete Überschreitungswahrscheinlichkeiten im Sinne der problematischen und zu Recht häufig kritisierten Ideen von Ronald Fisher (1966, 1972), sofern es sich hierbei überhaupt noch um bedeutungshaltige Angaben handelt und nicht nur um mehr oder minder sinnfreie Ergebnisse irgendwelcher unverstandener Computeralgorithmen. Genau diese Form der „Als-ob-Statistik“ ist bei sequentiellen Tests nicht mehr möglich. Sequentielle Tests bedrohen daher den gegenwärtigen Zustand, in dem die Testtheorie von Neyman und Pearson als ideologischer Überbau für eine ganz anders orientierte Testpraxis dient. Sequentielle Tests würden den unüberbrückbaren Konflikt zwischen dem Neyman-Pearson’schen „Über-Ich“ und dem Fisher’schen (wenn nicht gar rein deskriptivstatistischen) „Ego“ aufbrechen lassen (vgl. Gigerenzer, 1987). Indem schließlich Sequentialstatistik den probabilistischen Charakter von Inferenzstatistik unmittelbar erlebbar macht, bedroht sie die – auch psychisch entlastende – Illusion der „Mechanisierung des induktiven Schließens“ (vgl. Gigerenzer, 1987): Spätestens wenn Forscher am Ende eines geschlossenen sequentiellen Plans die Entscheidung zwischen den beiden Hypothesen mit der Münze bestimmen sollen, dürfte ihnen der probabilistische, also vom Zufall abhängige Charakter eines jeden Neyman-Pearson-Tests klar werden. Und dann könnten sie endlich erkennen, daß solche Neyman-Pearson-Tests – anders als etwa in der industriellen Qualitätskontrolle – in wissenschaftlichen Kontexten, gar in der Grundlagenforschung, eigentlich selten einen Sinn haben – ihrer täglich tausendfachen „Als-ob-Anwendung“ zum Trotz. Sich vor der Datenerhebung verbindlich darauf festzulegen, wie man sich nach der Datenerhebung zwischen lediglich zwei *courses of action* entscheiden wird – worunter Neyman und Pearson wohl kaum das schlichte „Annehmen“ oder „Verwerfen“ von Hypothesen, was auch immer das eigentlich sei, verstanden haben dürften –, dies macht in wissenschaftlichen Kontexten nur selten Sinn. Der Verlust dieser Illusion könnte wiederum viele und mächtige Interessen im Wissenschaftsbetrieb tangieren sowie mühsam erworbene und dann liebgegewonnene Gewohnheiten entwerten. Kurzum: Sequentielles Testen von Hypothesen ist vermutlich deswegen so wenig verbreitet, weil es Forscher dazu zwingen würde, mit der offiziell proklamierten Neyman-Pearson-Theorie statistischen Testens Ernst zu machen (vgl. Diepgen, 1987).

## 5 Weiterführende Literatur

In der sozialwissenschaftlichen Publikationswelt finden sich nur wenige Beiträge zum sequentiellen Testen. Grundsätzlich ist daher dem an weiterführender Literatur Interessierten zu raten, sich eher in medizinischen, technikwissenschaftlichen oder mathematischen Bibliotheken zu orientieren.

Aus der Feder von Wald (1966) selbst stammt eine auch heute noch gut lesbare und ausdrücklich für den mathematischen Laien geschriebene englischsprachige Einführung in die grundlegende Theorie sequentieller Tests.

Eine recht umfassende deutschsprachige Darstellung haben vor einigen Jahren Bauer, Scheiber und Wohlzogen (1986) vorgelegt. Dieses Büchlein bietet dem Anwender eine nennenswerte Auswahl von für die Praxis aufbereiteten sequentiellen Verfahren einschließlich der Erörterung vieler technischer Detailprobleme. Außerdem bietet dieses Arbeitsbuch ein ausführlichst erläutertes Verzeichnis weiterführender Literatur; auf dieses Verzeichnis sei hier abkürzend verwiesen.

Als sequentialstatistisches Standardwerk innerhalb der Mathematik gilt die Darstellung von Ghosh (1970). Dieses sehr mathematisch orientierte Buch dürfte indessen für den sozialwissenschaftlichen Anwender ähnlich schwerverständlich sein wie die vergleichbaren Werke von Govindarajulu (1975) und Wetherill (1975).

Anwendungsorientierte Kapitel zu sequentiellen Tests finden sich schließlich in den bekannten und weitverbreiteten statistischen Werken von Weber (1980) und – eingeschränkt auf verteilungsfreie Verfahren – Lienert (1978).

### Literaturverzeichnis

- Bauer, P., Scheiber, V. & Wohlzogen, F. X. (1986). *Sequentielle statistische Verfahren*. Stuttgart: Gustav Fischer.
- Diepgen, R. (1987). Dropje voor dropje. Oder: Sequentialstatistik, die ignorierte Alternative. *Zeitschrift für Sozialpsychologie*, 18, 19–27.
- Fisher, R. A. (1966). *The design of experiments* (8th ed.). Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1972). *Statistical methods for research workers* (17th ed.). Edinburgh: Oliver and Boyd.
- Ghosh, B. H. (1970). *Sequential tests of statistical hypotheses*. Reading: Addison-Wesley.
- Gigerenzer, G. (1987). Probabilistic thinking and the fight against subjectivity. In L. Krüger, G. Gigerenzer & M. Morgan (Eds.), *The probabilistic revolution: Vol. 2. Ideas in the sciences* (pp. 11–33). Cambridge: MIT Press.
- Govindarajulu, Z. (1975). *Sequential statistical procedures*. New York: Wiley.
- Lienert, G. A. (1978). *Verteilungsfreie Methoden in der Biostatistik. Band II*. Meisenheim: Hain.
- Wald, A. (1945). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16, 117–186.
- Wald, A. (1966). *Sequential analysis* (8th ed.). New York: Wiley.
- Wald, A. & Wolfowitz, J. (1950). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19, 326–339.
- Weber, E. (1980). *Grundriß der biologischen Statistik*. Stuttgart: Fischer.
- Wetherill, G. B. (1975). *Sequential methods in statistics*. London: Chapman and Hall.