# 5 Statistische Auswertungsmethode

Für die Auswertung der Ergebnisse des ABX-Tests ist zunächst die Wahl einer statistischen Methode notwendig, um anschließend mathematisch fundierte und interpretierbare Aussagen treffen zu können. Von der ABX-Company wird hier das statistische Modell der Binomialverteilung als sehr nützlich empfohlen.

## 5.1 Die Binomialverteilung

Bei einem Zufallsexperiment mit lediglich zwei möglichen Ausgängen gilt, dass das Ereignis A mit der Wahrscheinlichkeit p $(0 < p < 1)$ und das zu A komplementäre Ereignis B mit der Wahrscheinlichkeit $q = 1 - p$ eintritt. Wird dieses Zufallsexperiment mehrmals hintereinander ausgeführt, so spricht man vom sogenannten Bernoulli-Prozeß. Voraussetzung hierfür ist, dass die aufeinanderfolgenden Experimente sich nicht gegenseitig beeinflussen, d.h., dass der Ausgang der i-ten Durchführung unabhängig vom Ausgang der j-ten Durchführung ist.
Die Wahrscheinlichkeiten p für das Ereignis A und q für das Ereignis B gelten auch für jede Wiederholung des Experiments. Das jeweilige Ereignis tritt also grundsätzlich mit der gleichen Wahrscheinlichkeit ein.
Finden bei einem Bernoulli-Prozeß n Versuche statt, so kann das Ereignis A unter den n beobachteten Ergebnissen 0- bis n-mal auftreten. Ist nun X die Anzahl der Ergebnisse A unter den n beobachteten Ergebnissen, dann liefert die Binomialverteilung die Wahrscheinlichkeit für das Ereignis (X = x).
Sei BBA ... AB eine Folge der Länge n, in der x-mal der Ausgang A und (n-x)-mal der Ausgang B auftritt, so lässt sich aufgrund der Unabhängigkeit der einzelnen Versuche die Wahrscheinlichkeit P(BBA ... AB) wie folgt berechnen:

$$P(BBA \ldots AB) = P(B)P(B)P(A) \ldots P(A)P(B) = p^x (1-p)^{n-x}$$

Zwar sind bei diesem Versuch die Anzahl der Ereignisse A interessant, nicht aber deren konkrete Platzierung innerhalb der Abfolge, d.h., im wievielten Versuch ein Ereignis A aufgetreten ist. Unter den n möglichen Positionen sind genau x Positionen mit dem Ereignis A zu besetzen, was auf $\binom{n}{x}$ verschiedene Möglichkeiten erfolgen kann. Hieraus ergibt sich die gesuchte Wahrscheinlichkeit P(X=x) der Binomialverteilung:

$$P(X=x) = p_x = \binom{n}{x} p^x q^{n-x} \qquad \text{mit } x = 0,1, \dots, n$$

$$= \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{und} \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

p: Wahrscheinlichkeit für das Eintreten des Ereignisses A (beim Einzelversuch) mit $(0 < p < 1)$

q: Wahrscheinlichkeit für das Eintreten des zum Ereignis A komplementären Ereignisses B (beim Einzelversuch) mit $q = 1 - p$

n: Anzahl der durchgeführten Experimente/ Einzelversuche

x: Anzahl der Versuche, bei denen das gewünschte Ereignis eintritt

Bei solch einem Auswertungsverfahren soll geprüft werden, ob eine zuvor aufgestellte Hypothese H aufgrund der Ergebnisse des Tests beibehalten werden kann oder ob man sie ablehnt und sich damit für die sogenannte Gegenhypothese G entscheidet.
Bei Entscheidungen, die auf der Grundlage von Tests getroffen werden, können grundsätzlich jedoch auch Fehlentscheidungen bezüglich der Interpretation der Ergebnisse auftreten:

- Entscheidung für G, obwohl H richtig ist – dieser Fehler wird als Fehler erster Art bezeichnet
- Entscheidung für H, obwohl G richtig ist – dieser Fehler wird als Fehler zweiter Art bezeichnet

Die Möglichkeiten, sich richtig oder falsch zu entscheiden, lassen sich aus untenstehender Tabelle ersehen:

| | H trifft zu (Realität) | G trifft zu (Realität) |
|---|---|---|
| Entscheidung für H | richtige Entscheidung | Fehler zweiter Art |
| Entscheidung für G | Fehler erster Art | richtige Entscheidung |

Es lässt sich zeigen, dass die Wahrscheinlichkeit für einen Fehler erster Art umso kleiner ist, je größer die Wahrscheinlichkeit für einen Fehler zweiter Art ist. Man kann also nicht bei einer vorgegebenen Anzahl von Einzelversuchen die Wahrscheinlichkeiten für beide Fehlermöglichkeiten klein halten! In der Praxis geht man daher so vor, dass man die Wahrscheinlichkeit für den Fehler erster Art vorgibt, um zumindest eine Fehlerart unter Kontrolle halten zu können.

Die Wahrscheinlichkeit für den Fehler erster Art bezeichnet man auch als Signifikanzniveau oder Irrtumswahrscheinlichkeit, da durch ihn – durch Konvention – bestimmt wird, ab wann ein Ergebnis eine signifikante Abweichung zu dem unter der Hypothese erwarteten Ergebnis zeigt. Das Signifikanzniveau gibt somit an, ab welcher Abweichung man nicht mehr bereit ist, diese als zufällig zu bezeichnen. Die kritische Wahrscheinlichkeit P(K) wird in den meisten naturwissenschaftlichen Gebieten mit 5 % angegeben, d.h., $P(K) \leq 0,05$.[74]

Da lediglich die Entscheidung für G aus statistischer Sicht „ziemlich sicher" ist[75], versucht man, den Fehler zweiter Art zu vermeiden. Dies ist nur gewährleistet, wenn man sich bei der Auswertung der Ergebnisse für die Gegenhypothese entscheidet. Somit sollte die eigentliche Hypothese als Gegenhypothese ausgesprochen werden. Man nimmt also an, dass die beobachtete zufällige Größe dem Modell der Hypothese H entspricht – liegt dann der beobachtete Wert im Bereich der kritischen Wahrscheinlichkeit P(K), sieht man dies als statistischen Beweis dafür an, dass die Hypothese H ungültig, die Gegenhypothese G dagegen gültig ist.

74 Stahel 1995: 175.
75 Autorenkollektiv k.A.: 421.

## 5.2 Die Binomialverteilung in Bezug auf den ABX-Hörvergleich

Wie bereits zuvor ausführlich beschrieben, wird beim vorliegenden ABX-Hörvergleich getestet, ob ein Proband ein Signal X, welches mittels eines Zufallsgenerators aus den Quellen A und B bei jedem durchgeführten Experiment erneut bestimmt wird, der entsprechend richtigen Quelle A oder B zuordnen kann (also: entweder X=A oder X=B). Mit X (hier nicht das Signal X!) wird die Anzahl der korrekten Entscheidungen des Probanden gezählt. Als Ereignis wird hier also das Auftreten einer korrekten Zuordnung gewertet.

Ist für den Probanden kein Unterschied hörbar, so entscheidet er sich willkürlich für eine der beiden Quellen A und B. Er entscheidet sich demzufolge mit einer Wahrscheinlichkeit von $p = q = \frac{1}{2}$ für die richtige Quelle (oder die falsche).

Entschließt sich der Testhörer bei sämtlichen Durchführungen des Experiments rein zufällig, so lässt sich hier eine Wahrscheinlichkeitsverteilung für $p = q = \frac{1}{2}$ angeben durch:

$$P(X{=}x) = p_x = \binom{n}{x} * \left(\frac{1}{2}\right)^n \qquad \text{mit } x = 0, 1, \ldots, n$$

Wegen der oft propagierten und häufig beschriebenen klanglichen Unterschiede zwischen DSD und High-Resolution-PCM wurde diese Annahme übernommen, aber – bezogen auf diesen Hörvergleich – aufgrund der oben erwähnten Vorgehensweise aus statistischen Gründen die Hypothese formuliert, dass die Probanden zwischen den Quellen A und B <u>keinen</u> Unterschied wahrnehmen, sich also rein willkürlich oder eben zufällig entscheiden. Demzufolge wurden folgende Aussagen aufgestellt:

- Hypothese H: Es existieren <u>keine</u> wahrnehmbaren Unterschiede zwischen der Quelle A und der Quelle B
- Gegenhypothese G: Es existieren wahrnehmbare Unterschiede zwischen A und B

Da sowohl von der International Telecommunication Union (ITU)[76] als auch von der ABX-Company[77] ein Signifikanz-Niveau von 5% empfohlen wird, lässt sich die kritische Wahrscheinlichkeit P(K) mit $p = q = \frac{1}{2}$ nach folgender Gleichung bestimmen:

$$P(K) = \sum_{k=c}^{n} P(X = c) = \left(\frac{1}{2}\right)^{n} * \sum_{k=c}^{n} \binom{n}{k} \leq 0{,}05 \qquad \text{mit } k = 0,1, \ldots, n$$

c: Anzahl der korrekten Entscheidungen

n: Anzahl der durchgeführten Versuche

$\sum_{k=c}^{n} P(X = c)$: Eintrittswahrscheinlichkeit $P(X \geq c)$, sie gibt an, wie hoch die Wahrscheinlichkeit ist, in n Versuchen mindestens c korrekte Antworten zu geben

Die ABX-Company empfiehlt eine Mindestanzahl von 16 durchzuführenden Versuchen. Trotz der größeren Anstrengung für den Probanden wurde die Zahl der Versuche auf 20 erhöht, da die Aussagekraft der statistischen Auswertung natürlich umso höher ist, je mehr Versuche durchgeführt werden (s. Tabelle der Eintrittswahrscheinlichkeiten am Ende dieses Kapitels).

Anhand von zwei Beispielen soll nun die Auswertung und Interpretation von Testergebnissen verdeutlicht werden:

Fall 1: Berechnung der Eintrittswahrscheinlichkeit für mindestens 14 korrekte Entscheidungen bei 20 durchgeführten Versuchen:

$$P(X \geq 14) = \left(\frac{1}{2}\right)^{20} * \sum_{k=14}^{20} \binom{20}{14} = 0{,}058$$

Bei rein willkürlicher Entscheidung für A oder B beträgt die Wahrscheinlichkeit für mindestens 14 korrekte Entscheidungen bei 20 Versuchen also 5,8%. $P(X \geq 14)$ liegt also gerade eben nicht im Bereich der kritischen Wahrscheinlichkeit $P(K) \leq 0{,}05$. In diesem Fall würde man die Hypothese H beibehalten und die Gegenhypothese G verwerfen. Man könnte also die Vermutung äußern,

[76] ITU-R BS.1116-1 1997: 27.

[77] http://www.pcavtech.com/abx/ .

dass ein Unterschied zwischen den Quellen A und B vom Probanden nicht wahrgenommen wird. Es sei aber an dieser Stelle noch einmal erwähnt, dass hier ein Fehler zweiter Art (Entscheidung für H, obwohl G zutrifft) nicht auszuschließen ist.

Fall 2: Berechnung der Eintrittswahrscheinlichkeit für mindestens 15 korrekte Entscheidungen bei 20 durchgeführten Versuchen:

$$P(X \geq 15) = \left(\frac{1}{2}\right)^{20} * \sum_{k=15}^{20} \binom{20}{15} = 0{,}021$$

Bei rein willkürlicher Entscheidung für A oder B beträgt die Wahrscheinlichkeit für mindestens 15 korrekte Entscheidungen bei 20 Versuchen also 2,1%.
Damit liegt $P(X \geq 15)$ gerade im Bereich der kritischen Wahrscheinlichkeit $P(K) \leq 0{,}05$. Es ergibt sich, dass man aufgrund der Entscheidungsregel die Hypothese H verwerfen und die Gegenhypothese G annehmen würde. In diesem Fall könnte man also davon ausgehen, dass der Proband einen Unterschied zwischen den Quellen A und B wahrgenommen hat.

Sicherlich mag es ein wenig befremdlich erscheinen, dass die Differenz von nur einer korrekten Antwort bei der Auswertung eine Rolle spielen kann, aber ein Schwellwert muss festgelegt werden – dieser liegt durch Konvention bei 5%.
Aus der nachfolgenden Grafik-Übersicht Nr.41[78] können – neben anderen – die verschiedenen Eintrittswahrscheinlichkeiten bei 20 durchgeführten Versuchen entnommen werden:

[78] Die enthaltenen Werte wurden von http://www.pcavtech.com/abx/abx_bino.htm übernommen.

| Anzahl der Versuche | Anzahl der Versuche mit korrekter Entscheidung | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 1 | 0,500 | | | | | | | | | | | | | | | | | | | |
| 2 | 0,750 | 0,250 | | | | | | | | | | | | | | | | | | |
| 3 | 0,875 | 0,500 | 0,125 | | | | | | | | | | | | | | | | | |
| 4 | 0,938 | 0,688 | 0,312 | 0,063 | | | | | | | | | | | | | | | | |
| 5 | 0,969 | 0,813 | 0,500 | 0,188 | **0,031** | | | | | | | | | | | | | | | |
| 6 | 0,984 | 0,891 | 0,656 | 0,344 | 0,109 | **0,016** | | | | | | | | | | | | | | |
| 7 | 0,992 | 0,938 | 0,773 | 0,500 | 0,227 | 0,062 | **0,008** | | | | | | | | | | | | | |
| 8 | 0,996 | 0,965 | 0,855 | 0,637 | 0,363 | 0,145 | **0,035** | **0,004** | | | | | | | | | | | | |
| 9 | 0,998 | 0,980 | 0,910 | 0,746 | 0,500 | 0,254 | 0,090 | **0,020** | **0,002** | | | | | | | | | | | |
| 10 | 0,999 | 0,989 | 0,945 | 0,828 | 0,623 | 0,377 | 0,172 | 0,055 | **0,011** | **0,001** | | | | | | | | | | |
| 11 | 1,000 | 0,994 | 0,967 | 0,887 | 0,726 | 0,500 | 0,274 | 0,113 | **0,033** | **0,006** | **0,000** | | | | | | | | | |
| 12 | 1,000 | 0,997 | 0,981 | 0,927 | 0,806 | 0,613 | 0,387 | 0,194 | 0,073 | **0,019** | **0,003** | **0,000** | | | | | | | | |
| 13 | 1,000 | 0,998 | 0,989 | 0,954 | 0,867 | 0,709 | 0,500 | 0,291 | 0,133 | **0,046** | **0,011** | **0,002** | **0,000** | | | | | | | |
| 14 | 1,000 | 0,999 | 0,994 | 0,971 | 0,910 | 0,788 | 0,605 | 0,395 | 0,212 | 0,090 | **0,029** | **0,006** | **0,001** | **0,000** | | | | | | |
| 15 | 1,000 | 1,000 | 0,996 | 0,982 | 0,941 | 0,849 | 0,696 | 0,500 | 0,304 | 0,151 | 0,059 | **0,018** | **0,004** | **0,000** | **0,000** | | | | | |
| 16 | 1,000 | 1,000 | 0,998 | 0,989 | 0,962 | 0,895 | 0,773 | 0,598 | 0,402 | 0,227 | 0,105 | **0,038** | **0,011** | **0,002** | **0,000** | **0,000** | | | | |
| 17 | 1,000 | 1,000 | 0,999 | 0,994 | 0,975 | 0,928 | 0,834 | 0,685 | 0,500 | 0,315 | 0,166 | 0,072 | **0,025** | **0,006** | **0,001** | **0,000** | **0,000** | | | |
| 18 | 1,000 | 1,000 | 0,999 | 0,996 | 0,985 | 0,952 | 0,881 | 0,760 | 0,593 | 0,407 | 0,240 | 0,119 | **0,048** | **0,015** | **0,004** | **0,001** | **0,000** | **0,000** | | |
| 19 | 1,000 | 1,000 | 1,000 | 0,998 | 0,990 | 0,968 | 0,916 | 0,820 | 0,676 | 0,500 | 0,324 | 0,180 | 0,084 | **0,032** | **0,010** | **0,002** | **0,000** | **0,000** | **0,000** | |
| 20 | 1,000 | 1,000 | 1,000 | 0,999 | 0,994 | ,0979 | 0,942 | 0,868 | 0,748 | 0,588 | 0,412 | 0,252 | 0,132 | 0,058 | **0,021** | **0,006** | **0,001** | **0,000** | **0,000** | **0,000** |

**Grafik Nr.41:** Tabelle der Eintrittswahrscheinlichkeiten. Hervorgehoben sind die Eintrittswahrscheinlichkeiten, die im Bereich des Signifikanzniveaus von 5% liegen. Die gerahmte Zeile gilt für den vorliegenden Fall bei 20 durchgeführten Versuchen.