

1. Testen und Testprobleme

Zunächst möchte ich erläutern, was unter dem Begriff des Testproblems zu verstehen ist. In diesem Sinne beginne ich mit dem Beispiel:

Die „Tea-testing-Lady“

Eine Lady stellt die Behauptung auf, dass sie zu erkennen vermag, ob in einer ihr gereichten Tasse zuerst Milch oder zuerst Tee eingegossen wurde, wobei jede Tasse ein Tee/Milch-Gemisch enthält.

Natürlich interessiert nun, wie man vorgeht, um die von der Lady behauptete Fähigkeit auf die Probe zu stellen. Das nahe liegendste Vorgehen scheint der praktische Test: man reicht der Lady eine Reihe von Tassen und überprüft, wie oft sie richtig liegt. Anschließend wird entschieden, ob man ihrer Behauptung Glauben schenkt, oder nicht. Dieses Vorgehen wirft allerdings einige Fragen auf: Wie viele Tassen einer festgelegten Anzahl muss die Lady erraten, wie groß ist die Chance, dass sie einfach durch „Glück“ richtig lag und welche Wahrscheinlichkeit hat sie, ihr Können auch glaubhaft zu belegen? Hier merkt man, dass es anscheinend einiger Arbeit bedarf, um an ein befriedigendes Ergebnis zu gelangen. Zunächst will ich nun einen Rahmen für unser Problem schaffen:

1.1 Definition

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum mit unbekanntem Wahrscheinlichkeitsmaß P , Θ eine Indexmenge, $(\tilde{\Omega}, \tilde{\mathcal{A}})$ ein Messraum und $X: \Omega \rightarrow \tilde{\Omega}$ eine Zufallsvariable mit unbekannter Verteilung $P_{\hat{\theta}}$.

Seien weiterhin $x \in \tilde{\Omega}$ beobachteter Wert von X und $P_{\theta} := \{P_{\vartheta}, \vartheta \in \Theta\}$ die Menge der für X in Frage kommenden Verteilungen. Sei $H \subset \Theta$. Die Frage ob $\hat{\vartheta} \in H$ aufgrund von $x \in \tilde{\Omega}$ angenommen werden kann, nennt man Testproblem.

Zurück zur Lady

Betrachten wir nun die Situation der Lady, so erkennt man das Testproblem. Reichen wir der Lady, die für uns die Funktion der Zufallsvariable X erfüllt, eine Anzahl von $n \in \Omega := \mathbb{N}$ Tassen, so wird sie über eine uns nicht bekannte Verteilung $P_{\hat{\vartheta}}$ eine Anzahl von $x \in \tilde{\Omega} := \{1, \dots, n\}$ Tassen richtig klassifizieren. Wir können außerdem P_{θ} näher bestimmen: Betrachten wir das durchgeführte Experiment, so fällt auf, dass die Anzahl x binomialverteilt ist und somit gilt:

$$P_{\theta} = \{B_{n, \vartheta} : \vartheta \in \Theta, P_{\vartheta}(X=x) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x}\}$$

Wir definieren $\Theta = [0.5, 1]$, da die Lady ja mindestens raten und maximal alles (er)schmecken kann.

Für das Testproblem fehlt uns nun noch die Beschreibung der „Menge von Verteilungen“ H für die wir prüfen wollen mit welcher Wahrscheinlichkeit $P_{\hat{\vartheta}}$ darin liegt.

1.2 Definition

Sei ein Testproblem wie oben beschrieben gegeben. Dann bezeichnet man H als Hypothese.

Aus der Definition der Hypothese resultiert die Menge $K := \Theta \setminus H$, die als Alternative bezeichnet wird.

Die Hypothese wird als Nullhypothese bezeichnet, wenn der Test darauf abzielt diese zu verwerfen, um das gewünschte Ergebnis zu bringen. Das heißt als Nullhypothese wird die Kontraposition der zu testenden Behauptung verwendet.

1.2.1 Anmerkung

Ein Beispiel für die Wahl einer Nullhypothese wäre der Test eines neuartigen Medikaments: das (gebräuchliche) alte Medikament wird in Form der Nullhypothese als besser wirksam angenommen und die Wirksamkeit des neuen Medikaments belegt die Rolle der Alternative. Um eine bessere Wirksamkeit der neuen Medikaments anzunehmen, muss dann die Hypothese verworfen werden.

Bei der Modellierung eines Tests wird für die Hypothese meistens die Nullhypothese aufgestellt, da man dadurch mehr Kontrolle über die Situation erlangt, weil die „plausibelsten“ Verteilungen für ein Testproblem oft gut eingrenzbar sind. Die eigentliche „Behauptung“ zeigt sich also dann in Form der Alternative, wie man zunächst nicht vermuten mag. Des weiteren hat diese Wahl der Hypothese noch Vorteile, welche die möglichen Fehler bei einem Test betreffen, aber darauf wird später eingegangen.

Was bedeutet das für die Lady?

Wir wollen nun die Hypothese und die Alternative für unser Beispielproblem festklopfen. In unserem Fall ist die Wahl einer Nullhypothese sinnvoll: wir nehmen an, dass die Lady rein zufällig wählt und über keinen sonderlich gesegneten Geschmackssinn verfügt. Dem entsprechend ist ihre Wahl aus den n gereichten Tassen binomialverteilt mit einer (Erfolgs-)Wahrscheinlichkeit von 0.5, was also die Verteilung von X im Falle der Hypothese festlegt:

$$P_{0.5}(X=x) = \binom{n}{x} 0.5^x (1-0.5)^{n-x}$$

Zur Erinnerung: x ist die Anzahl der korrekt klassifizierten Tassen und unsere Hypothese ist nun $H = \{0.5\}$. Dem entsprechend ergibt sich die Alternative zu $K = \Theta \setminus H =]0.5, 1]$.

1.3 Definition

Sei ein Testproblem wie oben beschrieben gegeben. Sei $T: \tilde{\Omega} \rightarrow \mathbb{R}$, $\tilde{\omega} \rightarrow t$ eine messbare Funktion. Sei $\hat{t} \in \mathbb{R}$ (beliebig aber fest) gewählt mit:

$T(\tilde{\omega}) \geq \hat{t}$: Die Hypothese wird für $\tilde{\omega}$ verworfen.

$T(\tilde{\omega}) < \hat{t}$: Die Hypothese wird für $\tilde{\omega}$ angenommen.

Dann heißt T Teststatistik, \hat{t} kritischer Wert und $R := \{\tilde{\omega} \text{ mit } T(\tilde{\omega}) \geq \hat{t}\}$ kritischer Bereich oder Verwerfungsbereich des Tests.

1.3.1 Anmerkung

In R , dem Verwerfungsbereich, sind also die $\tilde{\omega}$ enthalten, für die die Hypothese verworfen wird und offenbar bestimmt die Wahl des Parameters $\hat{\theta} \in \mathbb{R}$ die Größe dieses Bereichs. Das bedeutet, dass für große $\hat{\theta}$ die Wahrscheinlichkeit, dass die Hypothese angenommen wird, größer ist und gleichzeitig die Chance sinkt, dass die Alternative glaubwürdig erscheint.

Außerdem ist zu beachten, dass eine Annahme der Hypothese keinesfalls den Beweis der Hypothese bedeutet, sondern die gegebenen Daten lediglich nicht für eine Verwerfung ausreichen.

Wieder bei der Lady...

Im Fall unseres Beispiels gestaltet sich die Konstruktion einer Teststatistik relativ einfach: wir können die Identität verwenden, d.h. die Funktion liefert uns eben genau die Anzahl der richtigen Tassen. Allerdings bleibt noch die Frage nach dem kritischen Wert zu klären. Reichen wir der Lady zum Beispiel zehn Tassen, wie viele muss sie korrekt tippen, damit wir ihr glauben schenken? Welcher kritische Wert ist sinnvoll?

1.4 Definition

Sei ein Testproblem wie oben beschrieben gegeben mit einer Teststatistik T und kritischem Wert \hat{t} . Sei wie oben P_{ϑ} die unbekannte Verteilung von X . Weiterhin sei $x \in \tilde{\Omega}$ ein beobachteter Ausgang. Ist nun

- a) $\hat{\theta} \in H$ und $T(x) \geq \hat{t}$ so bezeichnet man dies als Fehler erster Art.
- b) $\hat{\theta} \notin H$ und $T(x) < \hat{t}$ so bezeichnet man dies als Fehler zweiter Art.

1.4.1 Anmerkung

Ein Fehler erster Art ist also ein Verwerfen der Hypothese trotz ihrer Gültigkeit. Durch diese Art von Fehler können sehr negative Folgen eintreten (man stelle sich den Test eines neuartigen Medikaments vor).

Ein Fehler zweiter Art bedeutet eine Annahme der Hypothese, obwohl eigentlich die Alternative gültig ist. Die Auswirkung eines solchen Fehlers ist meistens nicht so gravierend, aber generell gilt, dass die Folgen der zwei Fehlertypen für jedes Testproblem genau betrachtet werden müssen.

Im Folgenden wollen wir nun das Werkzeug zur Bestimmung eines sinnvollen, kritischen Wertes erschließen.

1.5 Definition

Sei ein Testproblem wie oben beschrieben gegeben mit einer Teststatistik T und kritischem Wert \hat{t} . Unter der Gütefunktion $\beta: \Theta \rightarrow [0,1]$ versteht man die Funktion, die jedem Index $\vartheta \in \Theta$ unter der zugehörigen Verteilung P_{ϑ} seine Verwerfungswahrscheinlichkeit zuordnet:

$$\beta(\vartheta) := P_{\vartheta}(X \in R)$$

1.5.1 Anmerkung

Das bedeutet die Gütefunktion liefert die Wahrscheinlichkeit dafür, dass die zugeordnete Zufallsvariable Werte annimmt, die über die Teststatistik zum Verwerfen der Hypothese führen.

1.6 Definition

Ist β eine Gütefunktion wie oben beschrieben und existiert ein $\alpha \in [0,1]$ mit

$$\forall \vartheta \in H: \beta(\vartheta) \leq \alpha$$

so sagt man: der Test hat das Niveau α .

1.6.1 Anmerkung

Das Niveau eines Tests ist so zu interpretieren, dass für jede Verteilung der Hypothese deren Wahrscheinlichkeit verworfen zu werden kleiner oder gleich der vorgegebenen Schranke ist. Somit ist auch die Wahrscheinlichkeit eines Fehlers erster Art (die fälschliche Verwerfung der Hypothese) durch diese Schranke begrenzt.

1.7 Definition

Ist β eine Gütefunktion wie oben beschrieben, so heisst für ein $\vartheta \in K$:

$$\beta(\vartheta) \in [0,1]$$

die Macht des Tests in ϑ .

1.7.1 Anmerkung

Die Macht des Tests in einer Verteilung der Alternative ist die Wahrscheinlichkeit, dass die Hypothese für diese Verteilung abgelehnt wird, d.h. die Alternative wird angenommen. Die Wahrscheinlichkeit, dass die Alternative für ϑ verworfen wird, ist $1 - \beta(\vartheta)$. Diese Wahrscheinlichkeit begrenzt den Fehler zweiter Art (nämlich die fälschliche Annahme der Hypothese).

1.8 Finale am Teetisch

Bisher hatten wir das Testproblem für die Lady formuliert und Hypothese bzw. Alternative aufgestellt. Stehen geblieben waren wir bei der Frage nach einem sinnvollen, kritischen Wert \hat{t} zur Festlegung des Verwerfungsbereiches. Als ersten Schritt wollen wir nun die Gütefunktion für unser Testproblem aufstellen:

$$\beta_{\hat{t},n}(\vartheta) = P_{\vartheta}(X \geq \hat{t})$$

Als Erinnerung sei noch einmal gesagt, dass $\vartheta \in \Theta$ und $n \in \mathbb{N}$ die Anzahl der gereichten Tassen ist. Unter der Berücksichtigung unserer Wahl von P_{ϑ} ergibt sich

$$\beta_{\hat{t},n}(\vartheta) = P_{\vartheta}(X \geq \hat{t}) = \sum_{x=\hat{t}}^n B_{n,\vartheta}(x)$$

Nun lässt sich, nachdem wir uns für eine Anzahl zu reichender Tassen n und ein gewünschtes Niveau α entschieden haben, der entsprechende kritische Wert \hat{t} mit

$$\beta_{\hat{t},n}(0.5) = P_{0.5}(X \geq \hat{t}) = \sum_{x=\hat{t}}^n B_{n,0.5}(x) \leq \alpha$$

berechnen. Wählt man zum Beispiel als Niveau $\alpha = 0.05$ und als Tassenanzahl $n = 5$, so erhält man durch Einsetzen den kritischen Wert $\hat{t} = 5$ als Lösung, denn für $\hat{t} = 4$ ist bereits

$$\beta_{4,5}(0.5) = P_{0.5}(X \geq 4) = \binom{5}{4} (0.5)^5 + (0.5)^5 \approx 0.187 > \alpha$$

Somit muss die Lady bei fünf gereichten Tassen auch alle korrekt bestimmen. Um nun zu bewerten wie sinnvoll ein derartiger Test ist, stellen wir uns die Frage welche Chance die

Lady hat ihr Können als glaubhaft zu belegen. Dazu betrachten wir die Macht des Tests für verschiedene $\vartheta \in K$. Der Wert Gütefunktion, also in diesem Fall die Macht, ist nun also von ϑ abhängig:

$$\beta_{4,5}(\vartheta) = P_{\vartheta}(X \geq 5) = \sum_{x=5}^5 B_{5,\vartheta}(x) = \vartheta^5$$

Setzen wir einige Werte ein:

$$\beta_{4,5}(0.6) = 0,07776$$

$$\beta_{4,5}(0.75) \approx 0,23730$$

$$\beta_{4,5}(0.9) = 0,59049$$

Wir erkennen, dass die Lady, falls sie mit einer Chance von 60% die richtige Wahl treffen würde, in unserem Test nur eine Wahrscheinlichkeit von weniger als acht Prozent hätte, diese Fähigkeit zu belegen. Selbst bei einer Trefferwahrscheinlichkeit von 90% gelingt ihr der Nachweis nur mit einer Wahrscheinlichkeit von rund 60%.

Aufgrund dieser Erkenntnisse ist fraglich, ob der aktuelle Versuchsaufbau in dieser Art sinnvoll ist. Über die Betrachtung der Macht der Alternative(n) ist es also möglich den praktischen Aussagegehalt eines Tests näher zu bewerten.

Um die Situation für die Lady zu verbessern, können wir unser Experiment sensibler machen, so dass auch geringe Abweichungen von der Hypothese stärker ins Gewicht fallen. Dazu erhöhen wir die Anzahl der gereichten Tassen. Setzen wir zum Beispiel $n=42$, so können wir zum Niveau $\alpha=0.05$ den kritischen Wert $\hat{t}=27$ bestimmen. Die Macht für eine Verteilung mit einer Erfolgchance von 60% lässt sich dann durch Einsetzen bestimmen als $\beta_{27,42}(0.6) \approx 0,33$.

Auch hier muss abgewogen werden, wie sensibel der Test sein sollte und in wie weit der Aufwand einer größeren Datenerhebung (z.B. in Bezug auf Zeit oder Kosten) die genaueren Ergebnisse rechtfertigt.

Abschließend wollen wir noch erkunden in wie weit der Test anfällig für Fehler zweiter Art ist, das heißt die Hypothese würde angenommen, obwohl eigentlich die Alternative gültig ist. Für ein $\vartheta \in K$ ist die Wahrscheinlichkeit für einen solchen Fehler durch $1 - \beta_{\hat{t},n}(\vartheta)$ begrenzt, d.h. die Wahrscheinlichkeit für einen solchen Fehler hängt direkt von der Wahl von ϑ ab. Der Fehler zweiter Art wird wahrscheinlicher, je mehr wir uns der Hypothese annähern, was logisch erscheint, da kleinere Abweichungen schwieriger signifikant feststellbar sind und damit die Annahmewahrscheinlichkeit der Hypothese steigt. Dies kann durch folgende Grenzwertbildung veranschaulicht werden, wobei wir ein beliebig gewähltes Niveau α vorgeben:

$$\lim_{\vartheta \rightarrow 0.5} (1 - \beta_{\hat{t},n}(\vartheta)) = 1 - \beta_{\hat{t},n}(0.5) \geq 1 - \alpha$$

Den Grenzwert dürfen wir aufgrund der Stetigkeit von 1 bzw. $\beta_{\hat{t},n}(\vartheta)$ auf $]0.5, 1]$ bilden und $\beta_{\hat{t},n}(0.5)$ ist aufgrund des gesetzten Niveaus maximal α .

Wir erkennen also, dass es nicht möglich ist für alle $\vartheta \in K$ eine gemeinsame, obere Schranke für den Fehler zweiter Art, wie etwa $\alpha'=0.2$ anzugeben. Uns bleibt als einzige Möglichkeit für die Begrenzung dieses Fehlers die Erhöhung der Versuche n . Dazu wählen wir ein $\vartheta \in K$, zum Beispiel $\vartheta=0.7$ und vergrößern n so weit, dass die Macht $\beta_{\hat{t},n}(0.7) \geq 1 - \alpha'$ ist. Somit hätten wir für $\vartheta \in [0.7, 1] \subset K$ den Fehler zweiter Art auf maximal α' begrenzt.

2. Ein verfeinertes Modell für die Tea-Testing-Lady

In unserem Modell für die Situation der Lady sind wir bisher implizit davon ausgegangen, dass die Lady bei jeder ihr gereichten Tasse völlig unabhängig von den vorhergehenden Tassen entscheidet. Betrachtet man die Lage kritisch, so könnte man nun einwenden, dass wir uns über die Unabhängigkeit der Entscheidungen der Lady nicht sicher sein können. Das hieße allerdings, dass wir die Möglichkeit der Abhängigkeit in unser Testproblem mit einfließen lassen müssten, also unsere Wahl von P_Θ modifizieren sollten.

2.1 Definition

Sei ein Testproblem wie oben definiert gegeben. Sei die Wahl aller in Frage kommenden Verteilungen unzureichend realistisch. Man spricht dann von einem Fehler dritter Art.

2.2 Verfeinerung des Testproblems

Wir gehen nun im Folgenden davon aus, dass die Lady möglicherweise nicht unabhängig entscheidet. Außerdem konkretisieren wir noch: die Lady erhält nach einer Klassifikation keine Information darüber, ob sie richtig oder falsch lag. Dem entsprechend überarbeiten wir unseren Test Stück für Stück.

Als erstes betrachten wir $\tilde{\omega} \in \tilde{\Omega}$ und definieren $\tilde{\omega} := (y_1, \dots, y_n, z_1, \dots, z_n)$. Dies entspricht der Folge der gereichten Tassentypen y_i und der von der Lady klassifizierten Tassentypen z_i , wobei $i \in \{1, \dots, n\}$ ist. Wir bezeichnen die Tassentypen mit 1 und 2, also gilt $y_i, z_i \in \{1, 2\}$. Der Index i bezeichnet die i -te Tasse. Ist also für ein $i \in \{1, \dots, n\}$ $y_i = z_i$, so hat die Lady diese Tasse richtig klassifiziert.

Die Testdurchführung wird über die im folgende definierte Zufallsvariable $X: \Omega \rightarrow \{1, 2\}^{2n}$ beschrieben:

Um über $\tilde{\omega} \in \{1, 2\}^{2n}$ bessere Aussagen machen zu können, werden wir zunächst zwei weitere Zufallsvariablen konstruieren. Wir definieren $Y_i(\tilde{\omega}) := y_i$ und $Z_i(\tilde{\omega}) := z_i$, sowie $Y := (Y_1, \dots, Y_n)$ und $Z := (Z_1, \dots, Z_n)$ und setzen $X := (Y, Z)$.

Die Y_i sind als Zufallsvariablen für die gereichten Tassen aufgrund unseres Versuchsaufbaus auf jeden Fall unabhängig, da wir der Lady immer einen rein zufälligen Tassentyp reichen. Es gilt demnach $P(Y_i=1) = P(Y_i=2) = 0.5$, also Gleichverteilung. Da wir beliebig geartete Abhängigkeiten für die Entscheidungen der Lady zulassen wollen, können wir über die Verteilung der Z_i keine weiteren Aussagen treffen, außer, dass es sich um eine Verteilung auf $\{1, 2\}^n$ handelt.

Die Menge der generell in Frage kommenden Verteilungen für X sei also:

$$\Theta := \{ \text{Verteilungen über } \{1, 2\}^{2n} \text{ mit } Y_1, \dots, Y_n \text{ unabhängig; } P(Y_i=1) = P(Y_i=2) = 0.5 \}$$

Betrachten wir dies im Vergleich zum ersten Modell, so fällt auf, dass Θ nun kein gut zu erfassendes Intervall (von Erfolgswahrscheinlichkeiten) mehr ist, sondern wir unsere Verteilungen über sich selbst indizieren: P_ϑ mit $\vartheta \in \Theta$ ist also die Verteilung zum Parameter ϑ , wobei dieser Parameter selbst die gleiche Verteilung wie P_ϑ ist.

Die Indexmenge Θ muss so allgemein werden, da wir auch unser Modell entsprechend verallgemeinert haben.

Als nächstes müssen wir uns überlegen, wie nun unsere Hypothese zu formulieren ist. Da wir wieder eine Nullhypothese formulieren wollen, gehen wir davon aus, dass die Lady rein zufällig einen Tassentyp benennt und ihre Entscheidungen Z nicht von den gereichten Tassen Y abhängig sind, denn wir wenn die Lady keine Geschmacksunterschiede erkennen kann, so ist eine Abhängigkeit ebenfalls nicht möglich. Sei also $H \subset \Theta$ so gewählt, dass gilt:

$$\forall \vartheta \in H: P_\vartheta \text{ Verteilung mit } Y \text{ und } Z \text{ unabhängig}$$

Nun wollen wir eine Teststatistik konstruieren. Wir betrachten jedes Telexperiment (d.h. jede einzelne Tassenreichung) und definieren $T_i(\tilde{\omega}) := 1$ falls $y_i = z_i$ und sonst $T_i(\tilde{\omega}) := 0$. Es gilt also

$$T_i(\tilde{\omega}) = y_i - z_i + 1 \pmod{2}$$

Denn bei einem Nicht-Treffer ist die Differenz $y_i - z_i = 1$ und sonst $y_i - z_i = 0$. Die Anzahl der Treffer ist somit bestimmbar als

$$T(\tilde{\omega}) := T_1(\tilde{\omega}) + \dots + T_n(\tilde{\omega})$$

Für ein $\vartheta \in H$ und eine beliebige Trefferkonstellation $(u_1, \dots, u_n) \in \{0, 1\}^n$ kann die Wahrscheinlichkeit, dass diese Trefferkonstellation auftritt wie folgt bestimmt werden:

$$\begin{aligned} & P_\vartheta(T_1 = u_1, \dots, T_n = u_n) \\ &= \sum_{z \in \{1, 2\}^n} P_\vartheta(Z = z, Y_i = z_i - u_i + 1 \pmod{2} \text{ mit } i = 1, \dots, n) \end{aligned}$$

*: betrachtet man die Formel für Y_i näher, ergibt sich durch Einsetzen:

$$\begin{aligned} 1 - 0 + 1 &= 2, \\ 2 - 0 + 1 &= 3, \\ 1 - 1 + 1 &= 1, \\ 2 - 1 + 1 &= 2 \end{aligned}$$

wobei die erste Zahl $z_i \in \{1, 2\}$ und die zweite Zahl $u_i \in \{0, 1\}$ ist. Für die korrekte Modellierung eines Treffers soll $z_i = y_i$ gelten. Das $\pmod{2}$ gilt in obiger Gleichung also nur für den Fall $z_i = 2, u_i = 0$.

$$= \sum_{z \in \{1, 2\}^n} P_\vartheta(Z = z) P_\vartheta(Y_i = z_i - u_i + 1 \pmod{2} \text{ mit } i = 1, \dots, n)$$

Dieser Schritt ist zulässig, da wir uns im Bereich der Hypothese aufhalten und somit Y und Z unabhängig sind.

$$= \sum_{z \in \{1, 2\}^n} P_\vartheta(Z = z) \frac{1}{2^n} = \frac{1}{2^n}$$

Letzteres gilt, da die Y_i nach Voraussetzung mit $P(Y_i = 1) = P(Y_i = 2) = 0.5$ gleichverteilt sind und für eine konkrete Auswahl $\{1, 2\}^n$ genau 2^n Möglichkeiten bestehen. Die Summation von $P_\vartheta(Z = z)$ über ganz $\{1, 2\}^n$ ergibt nach Definition eines Wahrscheinlichkeitsmaßes genau 1.

Wir haben nun gezeigt, dass (T_1, \dots, T_n) unter der Hypothese Bernoulli-verteilt ist mit Erfolgswahrscheinlichkeit 0.5 . Somit ergibt sich - wie bei unserem ersten Modell - wieder die Binomialverteilung für die Anzahl der Treffer, nur dass diese diesmal direkt von unserer Teststatistik $T(\tilde{\omega}) = T_1(\tilde{\omega}) + \dots + T_n(\tilde{\omega})$ geliefert werden.

$$P_{\vartheta}(T=x) = \binom{n}{x} 0.5^x (1-0.5)^{n-x}, \text{ mit } \vartheta \in H, \text{ } x \text{ Anzahl der Treffer}$$

In einem nächsten Schritt können wir den Verwerfungsbereich wie gehabt als $R = \{\tilde{\omega} \in \tilde{\Omega} \text{ mit } T(\tilde{\omega}) \geq \hat{t}\}$ bestimmen, wobei der kritische Wert \hat{t} erneut so gesetzt wird, dass wir für den Test ein gewünschtes Niveau α erreichen. Das Vorgehen ist analog zum ersten Modell.

Problematisch wird die Betrachtung der Macht bestimmter $\vartheta \in K$. Da wir die Alternative möglichst wenig eingeschränkt haben, ist sie nicht mehr ausreichend konkretisierbar. Allerdings könnte man sich überlegen, dass eine Prüfung gerade der Verteilungen der Alternative interessant ist, für die eine höhere Anzahl von Treffern wahrscheinlich ist. In diesem Sinne könnte man zum Beispiel wie im vorherigen Modell die Binomialverteilung mit 0.6 als Erfolgchance heranziehen.

In den meisten Fällen ist eine solche „Vereinfachung“ der Alternative nicht weiter problematisch. Die Forderung der Verhinderung eines Fehlers „dritter Art“ nach einem vollständigen Modell bezieht sich vor allem auf die Konstruktion der Hypothese, denn wenn diese nicht korrekt konstruiert wird, würde zum Beispiel die Alternative möglicherweise durch Verteilungen, die eigentlich der Hypothese zugesprochen werden müssten, unzutreffend gestärkt.

3. Ein verwandtes Problem: Testen außersinnlicher Wahrnehmung (ASW)

Beim Testen auf ASW soll eine Person auf so genannte „Präkognition“ getestet werden. Das bedeutet, dass die Person im Stande wäre bestimmte Ereignisse vorherzusehen. Die Ausführung dieses Tests ist mit der Situation der tea-testing-lady eng verwandt:

Seien zwei Symbole 1 und 2 gegeben. Die Person gibt in jedem Versuch an, ob sie als nächstes Symbol 1 oder 2 vorhersieht. Über einen Münzwurf wird danach das Symbol bestimmt.

Für den Versuch i wird zunächst in $Z_i \in \{1,2\}$ die Aussage der Person festgehalten und danach über den Münzwurf das tatsächliche Symbol in $Y_i \in \{1,2\}$ gesetzt. Sei nun $Y := (Y_1, \dots, Y_n)$ und $Z := (Z_1, \dots, Z_n)$.

Als Hypothese formulieren wir die Annahme, dass keine ASW existiert. Dies kommt einer Unabhängigkeit von Y und Z gleich. Nun ist zweifelsfrei zu erkennen, dass der folgende Test mathematisch äquivalent zu dem aus Abschnitt 2 ist.

Bisher haben wir noch nicht betrachtet, was passiert, wenn wir – zum Beispiel obigen Test auf ASW – mit Feedback versehen. Das bedeutet nach jedem Versuch wird der zu testenden Person mitgeteilt, welches Symbol im letzten Versuch tatsächlich aufgetreten ist, also weiß die Person vor Versuch i auch alle Y_1, \dots, Y_{i-1} . Um dieser Änderung nachzukommen, muss natürlich die Hypothese verändert werden. Außerdem benötigen wir eine weitere Definition:

3.1 Definition

Sei (Ω, \tilde{A}_1, P) ein Wahrscheinlichkeitsraum, $(\tilde{\Omega}, \tilde{A}_2)$ Meßraum. Seien $A, B, C \in \tilde{A}$ Ereignisse. Dann heißt A bedingt unabhängig von B unter C , wenn gilt:

$$P(A \cap B | C) = P(A | C) P(B | C)$$

Seien $Y_1, \dots, Y_n, Z_1, \dots, Z_n$ Zufallsvariablen von $\Omega \rightarrow \tilde{\Omega}$ mit Verteilung P' und C eine Menge der Art $\{Y_1 = y_1, Z_1 = z_1, \dots, Y_{i-1} = y_{i-1}, Z_{i-1} = z_{i-1}\}$ für $i = 1, \dots, n$, dann heißen Y_i, Z_i bedingt unabhängig unter C , falls gilt:

$$\forall z = (z_1, \dots, z_n) \in \tilde{\Omega}, \forall y = (y_1, \dots, y_n) \in \tilde{\Omega} \forall (Z_i, Y_i) \forall I \subseteq \{1, \dots, n\} \forall J \subseteq \{1, \dots, n\}:$$

$$P((Z_i = z_i, i \in I) \wedge (Y_j = y_j, j \in J) | C) = P(Z_i = z_i, i \in I | C) P((Y_j = y_j, j \in J) | C)$$

3.2 Konstruktion der Hypothese

Wenn nun die Versuchsperson über die Y_1, \dots, Y_{i-1} informiert wurde, so soll die (Null-) Hypothese diesen Umstand als unbedeutend bestimmen; das heißt in der Hypothese soll für alle $i \leq n$ dann Z_i bedingt unabhängig von (Y_1, \dots, Y_n) unter gegebenem $(Y_1, Z_1, \dots, Y_{i-1}, Z_{i-1})$ sein.

Sei $\vartheta \in H$. Dann gilt $\forall i, 1 \leq i \leq n, \forall z = (z_1, \dots, z_n), \forall y = (y_1, \dots, y_n)$:

$$\begin{aligned} P_\vartheta(Z_i = z_i, Y_i = y_i, \dots, Y_n = y_n | C) \\ = P_\vartheta(Z_i = z_i | C) P_\vartheta(Y_i = y_i, \dots, Y_n = y_n | C) \end{aligned}$$

$$\text{mit } C = \begin{cases} \{Y_1 = y_1, Z_1 = z_1, \dots, Y_{i-1} = y_{i-1}, Z_{i-1} = z_{i-1}\} & \text{für } i \geq 2 \\ \tilde{\Omega} & \text{für } i = 1 \end{cases}$$

Der erste Schritt ist die Anwendung der Definition bedingter Unabhängigkeit und die Menge C ist so zu interpretieren, dass für $i \geq 2$ alle bisher vorhergesehenen und alle ermittelten Symbole bekannt sind. Für $i = 1$ wird nur von der Kenntnis der möglichen Symbole ausgegangen.

Auch unter dieser Hypothese erhält man die Aussage, dass die Trefferzahl binomialverteilt mit Erfolgswahrscheinlichkeit 0.5 ist.