

Script generated by TTT

Title: Esparza: DWT (03.07.2012)

Date: Tue Jul 03 14:19:10 CEST 2012

Duration: 88:52 min

Pages: 36

4. Testen von Hypothesen

4.1 Einführung

Bislang haben wir versucht, Parameter von Verteilungen zu schätzen. In der Praxis ist man jedoch oft an der eigentlichen Kenntnis dieser Parameter gar nicht interessiert, sondern man möchte gewisse, damit zusammenhängende Behauptungen überprüfen.

Im Folgenden stellen wir die Bestandteile eines statistischen Tests anhand eines abstrakten Beispiels vor. Wir betrachten dazu eine Zufallsvariable X mit $\Pr[X = 1] = p$ und $\Pr[X = 0] = 1 - p$. Durch einen Test soll überprüft werden, ob $p < 1/3$ oder $p \geq 1/3$ gilt.

Definition eines Tests

Wir betrachten eine Stichprobe von n unabhängigen Stichprobenvariablen X_1, \dots, X_n , die dieselbe Verteilung wie die Zufallsvariable X besitzen. Zu einem zugehörigen Stichprobenvektor \vec{x} müssen wir nun die Frage beantworten, ob wir für diesen Versuchsausgang die Hypothese „ $p \geq 1/3$ “ annehmen oder ablehnen.

Zu definieren ist die Menge

$$K := \{\vec{x} \in \mathbb{R}^n; \vec{x} \text{ führt zur Ablehnung der Hypothese}\}.$$

K nennen wir den **Ablehnungsbereich** oder den **kritischen Bereich** des Tests.

Gewöhnlich wird K konstruiert, indem man die Zufallsvariablen X_1, \dots, X_n zu einer neuen Variablen T , der so genannten **Testgröße**, zusammenfasst. Dann unterteilt man den Wertebereich \mathbb{R} von T in mehrere Bereiche, die entweder zur Ablehnung der Hypothese führen sollen oder nicht. Dabei betrachtet man meist ein einzelnes halboffenes oder abgeschlossenes Intervall und spricht dann von einem **einseitigen** bzw. von einem **zweiseitigen** Test.

Die Menge $\tilde{K} \subseteq \mathbb{R}$ enthalte die Werte von T , die zur Ablehnung der Hypothese führen sollen. Da wir Tests immer über eine Testgröße definieren, werden wir der Einfachheit halber auch \tilde{K} als Ablehnungsbereich bezeichnen. $\tilde{K} \subseteq \mathbb{R}$ entspricht direkt dem Ablehnungsbereich $K = T^{-1}(\tilde{K}) \subseteq \mathbb{R}^n$, wie wir ihn oben festgelegt haben.

Die zu überprüfende Hypothese bezeichnen wir mit H_0 und sprechen deshalb auch von der **Nullhypothese**. Bei manchen Tests formuliert man noch eine zweite Hypothese H_1 , die so genannte **Alternative**. Im Beispiel können wir

$$H_0 : p \geq 1/3 \text{ und } H_1 : p < 1/3$$

setzen.

Manchmal verzichtet man darauf, H_1 anzugeben. Dann besteht die Alternative wie oben einfach darin, dass H_0 nicht gilt. In diesem Fall nennen wir H_1 **triviale Alternative**.

Fehler bei statistischen Tests

Bei jedem statistischen Test können mit einer gewissen Wahrscheinlichkeit falsche Schlüsse gezogen werden. Dieser Fall tritt beispielsweise ein, wenn H_0 gilt, aber das Ergebnis \bar{x} der Stichprobe im Ablehnungsbereich K liegt.

Dann spricht man von einem **Fehler 1. Art**.

Analog erhalten wir einen **Fehler 2. Art**, wenn H_0 nicht gilt und \bar{x} nicht im Ablehnungsbereich liegt.

Fehler 1. Art : H_0 gilt, wird aber abgelehnt.

Fehler 2. Art : H_0 gilt nicht, wird aber angenommen.

Die zu überprüfende Hypothese bezeichnen wir mit H_0 und sprechen deshalb auch von der **Nullhypothese**. Bei manchen Tests formuliert man noch eine zweite Hypothese H_1 , die so genannte **Alternative**. Im Beispiel können wir

$$H_0 : p \geq 1/3 \text{ und } H_1 : p < 1/3$$

setzen.

Manchmal verzichtet man darauf, H_1 anzugeben. Dann besteht die Alternative wie oben einfach darin, dass H_0 nicht gilt. In diesem Fall nennen wir H_1 **triviale Alternative**.

Ein echter, also nicht-trivialer Alternativtest läge beispielsweise vor, wenn wir ansetzen

$$H'_0 : p \geq 1/3 \text{ und } H'_1 : p \leq 1/6.$$

Beispiel 100

Wir untersuchen eine Festplatte, von der bekannt ist, dass sie zu einer von zwei Baureihen gehört. Die mittleren Zugriffszeiten dieser Baureihen betragen 9ms bzw. 12ms. Wir möchten nun herausfinden, zu welchem Typ die betrachtete Festplatte gehört, indem wir die Zugriffszeit bei n Zugriffen bestimmen. Hier würde man dann ansetzen: $H_0 : \mu \leq 9$ und $H_1 := \mu \geq 12$, wobei μ die mittlere Zugriffszeit bezeichnet.

Für die Beurteilung eines Tests ist es wesentlich, mit welcher Wahrscheinlichkeit diese beiden Fehler eintreten können. Ziel ist es natürlich, diese Wahrscheinlichkeiten möglichst klein zu halten. Allerdings sind die Minimierung des Fehlers 1. Art und des Fehlers 2. Art gegenläufige Ziele, so dass ein vernünftiger Ausgleich zwischen beiden Fehlern gefunden werden muss. Wenn man beispielsweise $K = \emptyset$ setzt, so erhält man Wahrscheinlichkeit Null für den Fehler 1. Art, da H_0 immer angenommen wird. Allerdings tritt der Fehler 2. Art dann mit Wahrscheinlichkeit Eins ein, wenn H_0 nicht gilt.

Die Wahrscheinlichkeit für den Fehler 1. Art wird mit α bezeichnet, und man spricht deshalb gelegentlich vom α -Fehler. α heißt auch **Signifikanzniveau** des Tests.

In der Praxis ist es üblich, sich ein Signifikanzniveau α vorzugeben (übliche Werte hierfür sind 0,05, 0,01 oder 0,001) und dann den Test so auszulegen (also den Ablehnungsbereich K so zu bestimmen), dass die Wahrscheinlichkeit für den Fehler 1. Art den Wert α besitzt.

2 (SPSS14, deu).pdf (application/pdf-Objekt) - Mozilla Firefox

Biometrie_L2 (SPSS14, deu).pdf (appli... x | W Statistischer Test - Wikipedia

Biometrie_L2 (SPSS14, deu).pdf (application/pdf-Objekt) | e_L2 (SPSS14, deu).pdf

Inhaltsverzeichnis

| | |
|---|-----------|
| Kapitel 3 Wahrscheinlichkeitsrechnung | 4 |
| 3.1. Einführung | 4 |
| 3.2. Wahrscheinlichkeitsrechnung | 7 |
| 3.3. Übungen | 28 |
| Kapitel 4 Statistische Test I | 31 |
| 4.1. Das Prinzip von statistischen Tests | 31 |
| 4.2. t-Test | 35 |
| 4.3. Wilcoxon Rangsummentest | 41 |
| 4.4. Übungen | 45 |
| Kapitel 5 Statistische Tests II | 46 |
| 5.1. Mehr zum t-Test bei unabhängigen Stichproben | 46 |
| 5.2. Chi-Quadrat Test | 52 |
| 5.3. Gepaarte Tests | 57 |
| 5.4. Einseitige versus zweiseitige Tests | 66 |
| 5.5. Übungen | 68 |
| Anhänge | 72 |
| D. Exakte Tests | 72 |
| E. Äquivalenzstudien | 79 |
| F. Beschreibung von statistischen Methoden in medizinischen | |

DE 14:56 03.07.2012

idf* - PDF Annotator

2012-dwt.pdf | Biometrie_L2 (SPSS14, deu).pdf (application/pdf-Objekt)

Stift

p-Wert

Die Wahrscheinlichkeit für den Fehler 1. Art wird mit α bezeichnet, und man spricht deshalb gelegentlich vom α -Fehler. α heißt auch **Signifikanzniveau** des Tests.

In der Praxis ist es üblich, sich ein Signifikanzniveau α vorzugeben (übliche Werte hierfür sind 0,05, 0,01 oder 0,001) und dann den Test so auszulegen (also den Ablehnungsbereich K so zu bestimmen), dass die Wahrscheinlichkeit für den Fehler 1. Art den Wert α besitzt.

Geändert | Gehe zu nächster Seite mit Anmerkungen | 378 von 463 | C:\Users\esparza\De...\2012-dwt.pdf (463 Seiten) | DE 14:56 03.07.2012

Konstruktion eines einfachen Tests

Wir konstruieren einen Test für den Parameter p einer Bernoulli-verteilten Zufallsvariablen X . Wir setzen

$$H_0 : p \geq p_0, \quad H_1 : p < p_0.$$

Als Testgröße verwenden wir

$$\rightarrow p = p_0$$

$$p \neq p_0 \quad \text{— gesuchte M.H.}$$

$$T := X_1 + \dots + X_n.$$

Für größere Wahrscheinlichkeiten p erwarten wir auch größere Werte für T . Deshalb ist es sinnvoll, einen Ablehnungsbereich der Art $K := [0, k]$ für T zu wählen, wobei $k \in \mathbb{R}$ geeignet festzulegen ist. Wir konstruieren hier also einen einseitigen Test, während für eine Nullhypothese $H_0 : p = p_0$ sowohl zu kleine als auch zu große Werte von T zur Ablehnung von H_0 führen sollten und somit ein zweiseitiger Test vorzuziehen wäre.

$$[np_0 - \delta, np_0 + \delta]$$

$$n, \epsilon, p_0, \alpha$$

T ist binomialverteilt. Da wir von einem großen Stichprobenumfang n ausgehen, bietet es sich an, die Verteilung von T nach dem Grenzwertsatz von de Moivre (siehe Korollar 82) durch die Normalverteilung zu approximieren.

Sei

$$\tilde{T} := \frac{T - np}{\sqrt{np(1-p)}}.$$

\tilde{T} ist annähernd standardnormalverteilt.

Wir berechnen für jeden Wert von k ($K = [0, k]$) das zugehörige Signifikanzniveau α des Tests.

$$\begin{aligned} \text{Fehlerwahrscheinlichkeit 1. Art} &= \max_{p \in H_0} \Pr_p[T \in K] \\ &= \max_{p \in H_0} \Pr_p[T \leq k] \end{aligned}$$

$$\begin{aligned} \text{Fehlerwahrscheinlichkeit 2. Art} &= \sup_{p \in H_1} \Pr_p[T \notin K] \\ &= \sup_{p \in H_1} \Pr_p[T > k] \end{aligned}$$

Für den Fehler 1. Art α erhalten wir

$$\begin{aligned} \alpha &= \max_{p \geq p_0} \Pr_p[T \leq k] = \Pr_{p=p_0}[T \leq k] \\ &= \Pr_{p=p_0} \left[\tilde{T} \leq \frac{k - np}{\sqrt{np(1-p)}} \right] \\ &= \Pr \left[\tilde{T} \leq \frac{k - np_0}{\sqrt{np_0(1-p_0)}} \right] \approx \Phi \left(\frac{k - np_0}{\sqrt{np_0(1-p_0)}} \right). \end{aligned}$$

Für den Fehler 1. Art α erhalten wir

$$\begin{aligned} \alpha &= \max_{p \geq p_0} \Pr_p[T \leq k] = \Pr_{p=p_0}[T \leq k] \\ &= \Pr_{p=p_0} \left[\tilde{T} \leq \frac{k - np}{\sqrt{np(1-p)}} \right] \\ &= \Pr \left[\tilde{T} \leq \frac{k - np_0}{\sqrt{np_0(1-p_0)}} \right] \approx \Phi \left(\frac{k - np_0}{\sqrt{np_0(1-p_0)}} \right). \end{aligned}$$

Unter Verwendung der Quantile der Standardnormalverteilung ergibt sich damit:

- Ist k so gewählt, dass $(k - np_0)/\sqrt{np_0(1-p_0)} = z_\alpha$, so ist das Signifikanzniveau gleich α .
- Ist das gewünschte Signifikanzniveau α des Tests vorgegeben, so erhält man den Wert $k = k(n)$ in Abhängigkeit vom Umfang n der Stichprobe durch

$$k = z_\alpha \cdot \sqrt{np_0(1-p_0)} + np_0. \quad (7)$$

Kleinere Werte für k verkleinern zwar den Fehler 1. Art, vergrößern jedoch den Annahmehereich und damit die Wahrscheinlichkeit für einen Fehler 2. Art.

Für den Fehler 1. Art α erhalten wir

$$\begin{aligned} \alpha &= \max_{p \geq p_0} \Pr_p[T \leq k] = \Pr_{p=p_0}[T \leq k] \\ &= \Pr_{p=p_0} \left[\tilde{T} \leq \frac{k - np}{\sqrt{np(1-p)}} \right] \\ &= \Pr \left[\tilde{T} \leq \frac{k - np_0}{\sqrt{np_0(1-p_0)}} \right] \approx \Phi \left(\frac{k - np_0}{\sqrt{np_0(1-p_0)}} \right). \end{aligned}$$

Unter Verwendung der Quantile der Standardnormalverteilung ergibt sich damit:

- Ist k so gewählt, dass $(k - np_0)/\sqrt{np_0(1-p_0)} = z_\alpha$, so ist das Signifikanzniveau gleich α .
- Ist das gewünschte Signifikanzniveau α des Tests vorgegeben, so erhält man den Wert $k = k(n)$ in Abhängigkeit vom Umfang n der Stichprobe durch

$$k = z_\alpha \cdot \sqrt{np_0(1-p_0)} + np_0. \quad (7)$$

Kleinere Werte für k verkleinern zwar den Fehler 1. Art, vergrößern jedoch den Annahmehereich und damit die Wahrscheinlichkeit für einen Fehler 2. Art.

Bei echten **Alternativtests** werden für hinreichend große Stichproben und einen geeignet eingestellten Ablehnungsbereich beide Testfehler klein.

Beispiel 101

Die Abbruchrate p der Transaktionen in einem Online-Datenbanksystem wurde bereits früher einmal ermittelt. Allerdings sind die entsprechenden Daten verloren gegangen und die Entwickler erinnern sich nur noch, dass das Ergebnis entweder $p = 1/3$ oder $p = 1/6$ lautete. Unter dieser Annahme würde man den Test wie folgt ansetzen:

$$H_0 : p \geq 1/3, \quad H_1' : p \leq 1/6.$$

Bei echten **Alternativtests** werden für hinreichend große Stichproben und einen geeignet eingestellten Ablehnungsbereich beide Testfehler klein.

Beispiel 101

Die Abbruchrate p der Transaktionen in einem Online-Datenbanksystem wurde bereits früher einmal ermittelt. Allerdings sind die entsprechenden Daten verloren gegangen und die Entwickler erinnern sich nur noch, dass das Ergebnis entweder $p = 1/3$ oder $p = 1/6$ lautete. Unter dieser Annahme würde man den Test wie folgt ansetzen:

$$H_0 : p \geq 1/3, \quad H_1' : p \leq 1/6.$$

Beispiel (Forts.)

Für den Fehler 2. Art erhält man nun:

$$\begin{aligned} \text{Fehlerwahrsch. 2. Art} &= \max_{p \leq 1/6} \Pr_p[T > k] \\ &\approx 1 - \Phi\left(\frac{k - (1/6) \cdot n}{\sqrt{(1/6) \cdot (5/6)n}}\right). \end{aligned}$$

Mit den obigen Werten $k = 25$ und $n = 100$ ergibt sich mit

$$\Phi\left(\frac{150 - 100}{\sqrt{5} \cdot 10}\right) = \Phi(\sqrt{5}) \approx 0,9871$$

ein Fehler 2. Art der Größe 0,0129, während sich für die **triviale Alternative** $H_1 : p \leq 1/3$ ein Wert von etwa 0,95 ergibt.

4.2 Praktische Anwendung statistischer Tests

Das im vorhergehenden Abschnitt konstruierte Testverfahren taucht in der Literatur unter dem Namen **approximativer Binomialtest** auf.

Die folgende Tabelle 1 gibt einen Überblick über die Eckdaten dieses Tests.

Beispiel (Forts.)

Für den Fehler 2. Art erhält man nun:

$$\begin{aligned} \text{Fehlerwahrsch. 2. Art} &= \max_{p \leq 1/6} \Pr_p[T > k] \\ &\approx 1 - \Phi\left(\frac{k - (1/6) \cdot n}{\sqrt{(1/6) \cdot (5/6)n}}\right). \end{aligned}$$

Mit den obigen Werten $k = 25$ und $n = 100$ ergibt sich mit

$$\Phi\left(\frac{150 - 100}{\sqrt{5} \cdot 10}\right) = \Phi(\sqrt{5}) \approx 0,9871$$

ein Fehler 2. Art der Größe 0,0129, während sich für die *triviale Alternative* $H_1 : p \leq 1/3$ ein Wert von etwa 0,95 ergibt.

Tabell: Approximativer Binomialtest

Annahmen:

X_1, \dots, X_n seien unabhängig und identisch verteilt mit $\Pr[X_i = 1] = p$ und $\Pr[X_i = 0] = 1 - p$, wobei p unbekannt sei. n sei hinreichend groß, so dass die Approximation aus Korollar 82 brauchbare Ergebnisse liefert.

Hypothesen:

- a) $H_0 : p = p_0$ gegen $H_1 : p \neq p_0$,
- b) $H_0 : p \geq p_0$ gegen $H_1 : p < p_0$,
- c) $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$.

Testgröße:

$$Z := \frac{h - np_0}{\sqrt{np_0(1-p_0)}},$$

wobei $h := X_1 + \dots + X_n$ die Häufigkeit bezeichnet, mit der die Ereignisse $X_i = 1$ aufgetreten sind.

Ablehnungskriterium für H_0 bei Signifikanzniveau α :

- a) $|Z| > z_{1-\alpha/2}$,
- b) $Z < z_\alpha$,
- c) $Z > z_{1-\alpha}$.

4.3 Allgemeines Vorgehen bei statistischen Tests

1. **Schritt:** Formulierung von Annahmen. Ganz ohne Annahmen kommt man meist nicht aus. Übliche Annahmen betreffen meist die Verteilung der Stichprobenvariablen und deren Unabhängigkeit.
2. **Schritt:** Formulierung der Nullhypothese.
3. **Schritt:** Auswahl des Testverfahrens.
4. **Schritt:** Durchführung des Tests und Entscheidung.

4.4 Ausgewählte statistische Tests

4.4.1 Wie findet man das richtige Testverfahren?

Statistische Tests kann man nach mehreren Kriterien in Klassen einteilen.

• Anzahl der beteiligten Zufallsgrößen

Sollen zwei Zufallsgrößen mit potentiell unterschiedlichen Verteilungen verglichen werden, für die jeweils eine Stichprobe erzeugt wird, (**Zwei-Stichproben-Test**) oder wird nur eine einzelne Zufallsgröße untersucht (**Ein-Stichproben-Test**)?

Bei der Fragestellung

Beträgt die mittlere Zugriffszeit auf einen Datenbankserver im Mittel höchstens 10ms?

hat man es mit einem Ein-Stichproben-Test zu tun, während die Untersuchung der Frage

Hat Datenbankserver A eine kürzere mittlere Zugriffszeit als Datenbankserver B?

auf einen Zwei-Stichproben-Test führt.

- **Formulierung der Nullhypothese**

Welche Größe dient zur Definition der Nullhypothese? Hierbei werden in erster Linie Tests unterschieden, die Aussagen über verschiedene so genannte **Lageparameter** treffen, wie z.B. den **Erwartungswert** oder die **Varianz** der zugrunde liegenden Verteilungen.

Im Zwei-Stichproben-Fall könnte man beispielsweise untersuchen, ob der Erwartungswert der Zufallsgröße A größer oder kleiner als bei Zufallsgröße B ist.

Gelegentlich wird zur Formulierung der Nullhypothese auch der so genannte **Median** betrachtet: Der Median einer Verteilung entspricht dem (kleinsten) Wert x mit $F(x) = 1/2$.

Neben solchen Tests auf Lageparameter gibt es z.B. auch Tests, die auf eine **vorgegebene Verteilung** oder auf ein Maß für die Abhängigkeit verschiedener Zufallsgrößen testen.

- **Annahmen über die Zufallsgrößen**

Was ist über die Verteilung der untersuchten Größe(n) bekannt? Bei entsprechenden Annahmen könnte es sich z.B. um die Art der Verteilung, den Erwartungswert oder die Varianz handeln.

4.4.2 Ein-Stichproben-Tests für Lageparameter

Beim approximativen Binomialtest wird ausgenutzt, dass die Binomialverteilung für große n nach dem Grenzwertsatz von de Moivre (Korollar 82) gegen die Normalverteilung konvergiert. Aus diesem Grund kann man diesen Test auch als Spezialfall eines allgemeineren Testverfahrens ansehen, nämlich des **Gaußtest**, der nun dargestellt wird.

Bei der Fragestellung

Beträgt die mittlere Zugriffszeit auf einen Datenbankserver im Mittel höchstens 10ms?

hat man es mit einem Ein-Stichproben-Test zu tun, während die Untersuchung der Frage

Hat Datenbankserver A eine kürzere mittlere Zugriffszeit als Datenbankserver B?

auf einen Zwei-Stichproben-Test führt.

4.4 Ausgewählte statistische Tests

4.4.1 Wie findet man das richtige Testverfahren?

Statistische Tests kann man nach mehreren Kriterien in Klassen einteilen.

- **Anzahl der beteiligten Zufallsgrößen**

Sollen zwei Zufallsgrößen mit potentiell unterschiedlichen Verteilungen verglichen werden für die jeweils eine Stichprobe erzeugt wird, (Zwei-Stichproben-Test) oder wird nur eine einzelne Zufallsgröße untersucht (Ein-Stichproben-Test)?

Tabelle: t-Test

Annahmen:

X_1, \dots, X_n seien unabhängig und identisch verteilt mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
Alternativ gelte $\mathbb{E}[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$, und n sei groß genug.

Hypothesen:

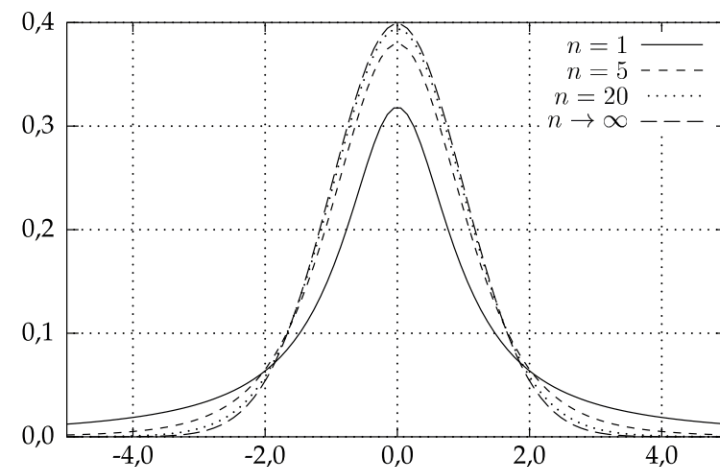
- a) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$,
- b) $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$,
- c) $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$.

Testgröße:

$$T := \frac{\bar{X} - \mu_0}{S} \sqrt{n}.$$

Ablehnungskriterium für H_0 bei Signifikanzniveau α :

- a) $|T| > t_{n-1, 1-\alpha/2}$,
- b) $T < t_{n-1, \alpha}$,
- c) $T > t_{n-1, 1-\alpha}$.



Dichte der t-Verteilung mit n Freiheitsgraden

4.4.4 Nicht an Lageparametern orientierte Tests

Wir betrachten in diesem Abschnitt exemplarisch den χ^2 -Anpassungstest. Bei einem Anpassungstest wird nicht nur der Lageparameter einer Verteilung getestet, sondern es wird die Verteilung als Ganzes untersucht.

Beim approximativen Binomialtest (siehe Tabelle 1) haben wir streng genommen bereits einen Anpassungstest durchgeführt. Bei der Nullhypothese $H_0 : p = p_0$ wird untersucht, ob es sich bei der betrachteten Zufallsgröße um eine Bernoulli-verteilte Zufallsvariable mit Parameter p_0 handelt. Beim χ^2 -Test gehen wir nun einen Schritt weiter: Wir nehmen an, dass die Zufallsgröße X genau k verschiedene Werte annimmt. Ohne Beschränkung der Allgemeinheit sei $W_X = \{1, \dots, k\}$. Die Nullhypothese lautet nun

$$H_0 : \Pr[X = i] = p_i \quad \text{für } i = 1, \dots, k.$$