Kapitel 10

Tests für beliebige Zufallsvariable

10.1 Der Chi-Quadrat-Anpassungstest

Sei x eine ganz beliebige Zufallsvariable, deren Dichtefunktion nicht oder nicht genau bekannt ist.

Beispiel: Es seien z.B. mittels einer Messreihe mit n, $\bar{x}$ und s erste Schätzwerte für den Erwartungswert $\hat{x}$ zum Experiment und für die Streuung σ zur Messmethode bekannt. Nun soll hiermit die Dichtefunktion aufgestellt werden, aber es sei noch unklar, ob die Schätzwerte wirklich gut genug dafür sind, eine passende Formel zu liefern.
Oder es sei sogar noch unklar, ob x normalverteilt oder poissonverteilt oder binomialverteilt ist oder keins von diesen.

In solchen Fällen stellt man probehalber, also als Hypothese, eine Formel oder eine Wertetabelle für die Dichtefunktion auf und überprüft dann anhand einer Messreihe zu x mit n Messungen $x_1, \ldots, x_n$ und dem nachfolgenden Test, ob und mit welcher Sicherheit diese Hypothese zu verwerfen oder anzunehmen ist.

Der χ^2-Anpassungstest:

1. Schritt: Zunächst wird zur Auswertung der gegebenen Messreihe eine Strichliste vorbereitet wie folgt:

   - Der Wertevorrat von x wird in endlich viele Abteilungen $E_1, \ldots, E_r$ eingeteilt derart, dass jedes überhaupt nur denkbare Messergebnis zu x in genau eine Abteilung E_i fällt (in Symbolen: $x \in E_i$).
   - Anhand der hypothetischen Dichtefunktion werden die Wahrscheinlichkeiten

     $$P(x \in E_i) = p_i \;\; (i = 1, \ldots, r)$$

     vorausberechnet. (Zur Kontrolle überprüft man, dass sich tatsächlich $\sum p_i = 1$ ergibt.)

- Es muss die Bedingung

$$n \cdot p_i \geq 5 \quad \text{für alle } i$$

  erfüllt sein. Trifft dies auf einzelnde Abteilungen E_i nicht zu, so sind sie mit anderen zu wenigeren größeren Abteilungen geeignet zusammenzufassen, oder die Anzahl n der Messungen ist so weit zu erhöhen, bis die Bedingung für alle i erfüllt ist.

2. Schritt: Man trägt die Messergebnisse $x_1, \ldots, x_n$ in die so vorbereitete Strichliste ein und ermittelt die absoluten Häufigkeiten

$$H_i = \text{ Anzahl der Striche in der Abteilung } E_i \ (i = 1, \ldots, r).$$

3. Schritt: Berechne die Prüfgröße

$$\chi^2 = \sum_{i=1}^{r} \frac{(H_i - n \cdot p_i)^2}{n \cdot p_i} = \sum_{i=1}^{r} \frac{H_i^2}{n \cdot p_i} - n$$

4. Schritt: Berechne den "Freiheitsgrad" f wie folgt:

- Besagt die Hypothese, dass x normalverteilt ist mit einer Glockenfunktion als Dichtefunktion, hat man Schätzwerte für μ und σ benutzt und die p_i mit Hilfe der Φ-Tabelle berechnet, so wähle

$$f = r - 3,$$

- besagt die Hypothese, dass x poissonverteilt ist, hat man einen Schätzwert für λ benutzt und die p_i mit Hilfe der Funktion $p_\lambda(k)$ berechnet, so wähle

$$f = r - 2,$$

- besagt die Hypothese, dass x binomialverteilt ist, hat man einen Schätzwert für p benutzt und die p_i mit Hilfe der Funktion $p_{n,p}(k)$ berechnet, so wähle

$$f = r - 2,$$

- trifft keiner dieser Fälle zu und hat man die p_i-Werte frei geschätzt mit der einzigen Zusatzbedingung, dass $\sum p_i = 1$ ergibt, so wähle

$$f = r - 1.$$

5. Schritt: Schlage für dieses f in der χ^2-Tabelle[1] nach und vergleiche χ^2 mit tab-χ^2.

Auswertung:

- Ist $\chi^2 \leq$ tab-$\chi^2(95\%)$, so ist durch den Test die Falschheit der Nullhypothese nicht feststellbar. Es darf also angenommen werden, dass die benutzte Dichtefunktion zutreffend ist (ohne dass der Test die Richtigkeit dieser Annahme bestätigen kann!).

[1]siehe Tabellen zur Statistik, S.228

- Ist $\chi^2 >$ tab-$\chi^2(95\%)$, so ist die Hypothese **wahrscheinlich** falsch, also die benutzte Dichtefunktion unzutreffend.
- Ist $\chi^2 >$ tab-$\chi^2(99\%)$, so ist die Hypothese **signifikant** falsch, also die benutzte Dichtefunktion unzutreffend.
- Ist $\chi^2 >$ tab-$\chi^2(99,9\%)$, so ist die Hypothese **hochsignifikant** falsch, also die benutzte Dichtefunktion unzutreffend.

Erläuterung: Ist der Zahlwert von p_i zutreffend angesetzt worden, so strebt für $n \to \infty$ die relative Häufigkeit $\frac{H_i}{n}$ gegen die Wahrscheinlichkeit p_i, die absolute Häufigkeit H_i also gegen den Wert $n \cdot p_i$. Das Größenverhältnis $\frac{|H_i - n \cdot p_i|}{n \cdot p_i}$ ist somit der relative Abstand zwischen H_i und $n \cdot p_i$, auch der relative Fehler, den man macht, wenn man den einen durch den anderen Wert ersetzt, und muss umso kleiner werden, je größer n ist. Die Größe

$$\chi = \sqrt{\sum_{i=1}^{r} \frac{(H_i - n \cdot p_i)^2}{n \cdot p_i}}$$

ist also ein Fehlermaß: Sie misst die Summe der relativen Fehler ($i = 1, \ldots, r$).
Führt der Test zu einer Ablehnung der Nullhypothese, so gehört der größte der r Summanden zu demjenigen p_i, das am schlechtesten zur Sachlage passt. Genauer gilt:

Merke: Stimmt die Nullhypothese im Anpassungstest, so strebt jeder Summand und damit auch χ^2 für $n \to \infty$ gegen Null. Ist ein p_i-Wert falsch, so wächst der zu H_i gehörige Summand ungefähr proportional zu n.

10.2 Der Chi-Quadrat-Unabhängigkeitstest

Zwei beliebige Zufallsvariable x und y heißen **unabhängig**[2], wenn für alle a, b, c, d stets gilt

$$P(a \leq x \leq b \text{ und zugleich } c \leq y \leq d) = P(a \leq x \leq b) \cdot P(c \leq y \leq d).$$

Es ist wissenschaftlich üblich, je zwei beliebige Zufallsvariable x und y a priori zunächst einmal als unabhängig anzusehen (= sog. "Nullhypothese"), und zwar so lange, bis mit der gewünschten Sicherheit (i.a. 99%) das Gegenteil erwiesen ist.

Der Chi-Quadrat-Unabhängigkeitstest prüft anhand einer Messreihe mit n Messungen, ob die Nullhypothese der Unabhängigkeit für zwei Zufallsvariable abgelehnt werden kann. Theoretisch gesehen ist er ein Abkömmling des Anpassungstests, in der praktischen Handhabung jedoch eigenständig.

1. Schritt: Wie beim Chi-Quadrat-Anpassungstest wird der Wertevorrat von x in r Klassen $E_1, \ldots, E_r$ eingeteilt, analog der Wertevorrat von y in m Klassen $E'_1, \ldots, E'_m$. Misst man nun bei jeder Einzelmessung sowohl den Wert von x wie den von y,

[2] vergleiche die *Unabhängigkeitsregel (Regel 66)*, 8.1, S.160

so tritt genau einer der $r \cdot m$ folgenden Fallkombinationen ein: Der x-Wert liegt in genau einem E_k ($k \in \{1, \ldots, r\}$), gleichzeitig liegt der y-Wert in genau einem E'_i ($i \in \{1, \ldots, m\}$).
Für die Strichliste bereite nun eine zweidimensionale Tabelle vor mit r Spalten für die Klassen $E_1, \ldots, E_r$ und m Zeilen für die Klassen $E'_1, \ldots, E'_m$. Zusätzlich eine letzte Spalte, um Zeilensummen zu berechnen, sowie eine letzte Zeile, um Spaltensummen zu berechnen.

2. Schritt: Trage die n Messergebnisse in die so vorbereitete Strichtabelle ein: Liegt bei einer Einzelmessung der x-Wert in E_k, der y-Wert in E'_i, so wird ein Strich in der k-ten Spalte und i-ten Zeile gemacht. Anschließend berechne die absoluten Häufigkeiten und trage sie in eine gleichgebaute Tabelle ein:

$$H_{ik} = \text{ Anzahl der Striche in der } i\text{-ten Zeile und } k\text{-ten Spalte}$$

3. Schritt: Es muss die Bedingung

$$H_{ik} \geq 5 \quad \text{für alle } i \text{ und } k$$

erfüllt sein. Ist dies nicht der Fall, so muss entweder die Klasseneinteilung weniger fein gewählt werden oder die Anzahl n der Messungen so weit erhöht werden, bis die Bedingung erfüllt ist.

4. Schritt: Berechne die i-te Zeilensumme Z_i für $i = 1, \ldots, m$ und trage sie in der i-ten Zeile und letzten Spalte ein.
Berechne die k-te Spaltensumme S_k für $k = 1, \ldots, r$ und trage sie in der k-ten Spalte und letzten Zeile ein.

Bezeichnung: *Die vollständig ausgefüllte Tafel heißt* **Kontingenztafel**.

5. Schritt: Mache die Probe: Es muss gelten $\sum Z_i = n$ und $\sum S_k = n$.

6. Schritt: Berechne

$$\chi^2 = \sum_{i=1}^{r} \sum_{k=1}^{m} \frac{\left(H_{ik} - \frac{Z_i \cdot S_k}{n}\right)^2}{\frac{Z_i \cdot S_k}{n}}$$

7. Schritt: Berechne

$$f = (r-1) \cdot (m-1)$$

8. Schritt: Schlage für dieses f in der χ^2-Tabelle[3] nach und vergleiche χ^2 mit tab-χ^2.

Auswertung:

- Ist $\chi^2 \leq$ tab-$\chi^2(95\%)$, so ist durch den Test die Falschheit der Nullhypothese nicht feststellbar. Es ist also weiterhin anzunehmen, dass die Variablen x und y unabhängig sind (ohne dass der Test die Richtigkeit dieser Annahme bestätigen kann!).
- Ist $\chi^2 >$ tab-$\chi^2(95\%)$, so sind die Variablen x und y **wahrscheinlich** abhängig.
- Ist $\chi^2 >$ tab-$\chi^2(99\%)$, so sind die Variablen x und y **signifikant** abhängig.
- Ist $\chi^2 >$ tab-$\chi^2(99,9\%)$, so sind die Variablen x und y **hochsignifikant** abhängig.

[3] siehe Tabellen zur Statistik, S.228

Erläuterung: Für $n \to \infty$ strebt auf jeden Fall

$\frac{Z_i}{n}$ = relative Häufigkeit von $y \in E'_i$ gegen $P(y \in E'_i)$ und

$\frac{S_k}{n}$ = relative Häufigkeit von $x \in E_k$ gegen $P(x \in E_k)$.

Nach den Grenzwertregeln folgt

$$\frac{Z_i \cdot S_k}{n} = n \cdot \frac{Z_i}{n} \cdot \frac{S_k}{n} \longrightarrow n \cdot P(y \in E'_i) \cdot P(x \in E_k)$$

Falls nun die Nullhypothese der Unabhängigkeit stimmt, so strebt außerdem jedes

$\frac{H_{ik}}{n}$ = relative Häufigkeit, dass $y \in E'_i$ und zugleich $x \in E_k$ gilt,

gegen $P(y \in E'_i) \cdot P(x \in E_k)$. Daraus würde folgen, dass

$$H_{ik} \longrightarrow n \cdot P(y \in E'_i) \cdot P(x \in E_k) \quad \text{für alle } i \text{ und } k$$

gilt. In diesem Fall streben also H_{ik} und $\frac{Z_i \cdot S_k}{n}$ gegen denselben Wert, und der Abstand zwischen beiden muss immer kleiner werden. Das Größenverhältnis

$$\frac{\left| H_{ik} - \frac{Z_i \cdot S_k}{n} \right|}{\frac{Z_i \cdot S_k}{n}}$$

ist der relative Abstand zwischen H_{ik} und $\frac{Z_i \cdot S_k}{n}$, auch der relative Fehler, den man macht, wenn man den einen Wert durch den anderen ersetzt. Die Größe

$$\chi = \sqrt{\sum_{i=1}^{r} \sum_{k=1}^{m} \frac{\left(H_{ik} - \frac{Z_i \cdot S_k}{n} \right)^2}{\frac{Z_i \cdot S_k}{n}}}$$

ist somit ein Fehlermaß: Sie misst die Summe sämtlicher $r \cdot m$ relativen Fehler.
Führt der Test zu einer Ablehnung der Nullhypothese, so gehört der größte aller $r \cdot m$ Summanden in dieser Summe zu demjenigen H_{ik}, an dem sich am stärksten die Abhängigkeit der Variablen erkennen lässt.

Merke: Stimmt die Nullhypothese im Unabhängigkeitstest, so strebt jeder Summand und damit auch χ^2 für $n \to \infty$ gegen Null. Ist für eine Kombination (i, k) die Unabhängigkeitsbedingung

$$P(x \in E_k \text{ und zugleich } y \in E'_i) = P(x \in E_k) \cdot P(y \in E'_i)$$

nicht erfüllt, so wächst der zu H_{ik} gehörige Summand ungefähr proportional zur Anzahl n der Messungen.

Spezialfall:
Der einfachste Spezialfall liegt vor, wenn x und y beide ja/nein-Variable sind. Dann ist $r = 2$, $m = 2$, also $f = 1$, und die Kontingenztafel besteht nur aus 4 Feldern plus letzter Spalte und letzter Zeile (sog. "Vierfeldertafel").
x misst dann, ob ein Ereignis A eintritt, ja oder nein; y misst, ob ein Ereignis B eintritt, ja oder nein. Die Berechnung der Prüfgröße vereinfacht sich dann sehr, man kann folgende Formel benutzen:

$$\chi^2 = \frac{n(H_{11}H_{22} - H_{12}H_{21})^2}{Z_1 Z_2 S_1 S_2}$$

Mit diesem χ^2 und mit $f = 1$ ist der χ^2-Unabhängigkeitstest durchzuführen.

Hinweis:
Aus der Schule sind eventuell Vierfeldertafeln bekannt, aber in einem anderen Zusammenhang: Es werden nicht absolute Häufigkeiten eingetragen, sondern Wahrscheinlichkeiten. Einige dieser Wahrscheinlichkeiten sind vorgegeben, die übrigen lassen sich ergänzend bestimmen, und man errechnet dann mit diesen Tafeln sogenannte "bedingte Wahrscheinlichkeiten", kann außerdem auch prüfen, ob Ereignisse A und B unabhängig sind.
In 9.4[4] haben wir gesehen, dass zum verlässlichen Schätzen einer Wahrscheinlichkeit i.a. schon etliche Tausend Messungen erforderlich sind. In der hier besprochenen Situation sind Vorkenntnisse über Wahrscheinlichkeiten nicht gegeben, es werden auch keine Wahrscheinlichkeiten errechnet: Man besitzt lediglich die auszuwertende Messreihe, die auch kürzer sein darf, solange nur die Bedingung $H_{ik} \geq 5$ für alle i und k erfüllt ist.

[4] *Schätzung von unbekanntem p*, S.191