5.7.4 Warum die Chi-Quadrat Verteilung?

(1) Binomial und Chi-Quadrat Tests
für 2 x 2 Kontingenztafeln

Als Beispiel nehmen wir die Erkältungsdaten, die Pauling berühmt gemacht hat:
279 Skifahrer in einer Doppelblindstudie (Ritzel [1961])

	Placebo	Ascorbic Acid	
Erkältet	31	17	48
Nein	109	122	231
	140	139	279

$$X^2 = 4.81$$

also, signifikant ($p < 0.05$), aber nicht sehr ($p > 0.01$).

Wir betrachten den allgemeinen Fall:

	G1	G2	
Ja	x_1	x_2	$x_1 + x_2$
Nein	$n_1 - x_1$	$n_2 - x_2$	$n_1 + n_2 - x_1 - x_2$
	n_1	n_2	$n_1 + n_2$

(a) Binomial Test

Wir vergleichen die Erfolgsquoten, p_1, p_2 für die zwei Gruppen, G1 und G2. Für groß n_1,n_2 sind die Schätzer $\widehat{p}_j = \frac{x_j}{n_j}$ approximativ normalverteilt.

Sei $\{X_{ij}, i = 1, \ldots, n \quad j = 1, 2\}$ u.i.v. Bernoulli ZV, d.h.

$$P(X_{ij} = 1) = p_j \quad P(X_{ij} = 0) = 1 - p_j$$

$$E[X_{ij}] = p_j \quad V[X_{ij}] = p_j(1 - p_j)$$

$$\Rightarrow$$

und für n_j groß gilt

$$\sum_{i=1}^{n_j} X_{ij} \sim$$

bzw.

$$\widehat{p}_j = \bar{X}_j = \sum_{i=1}^{n_j} X_{ij}/n_j \sim N\left(p_j, \frac{p_j(1 - p_j)}{n_j}\right)$$

$$\Rightarrow \widehat{p}_1 - \widehat{p}_2 \sim$$

Unser H_0 wird $p_1 = p_2 = p$ sein. Für die Varianz setzen wir den Plugin-Schätzer für p ein

$$\widehat{p} =$$

und betrachten die Teststatistik

$$T = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.2214 - 0.1223}{0.0452} = 2.193$$

$$T \sim N(0,1) \Rightarrow T^2 \sim \chi_1^2$$

und wir können zeigen, dass

$$T^2 = \frac{(n_2 x_1 - n_1 x_2)^2 (n_1 + n_2)}{n_1 n_2 (x_1 + x_2)(n_1 + n_2 - x_1 - x_2)}$$

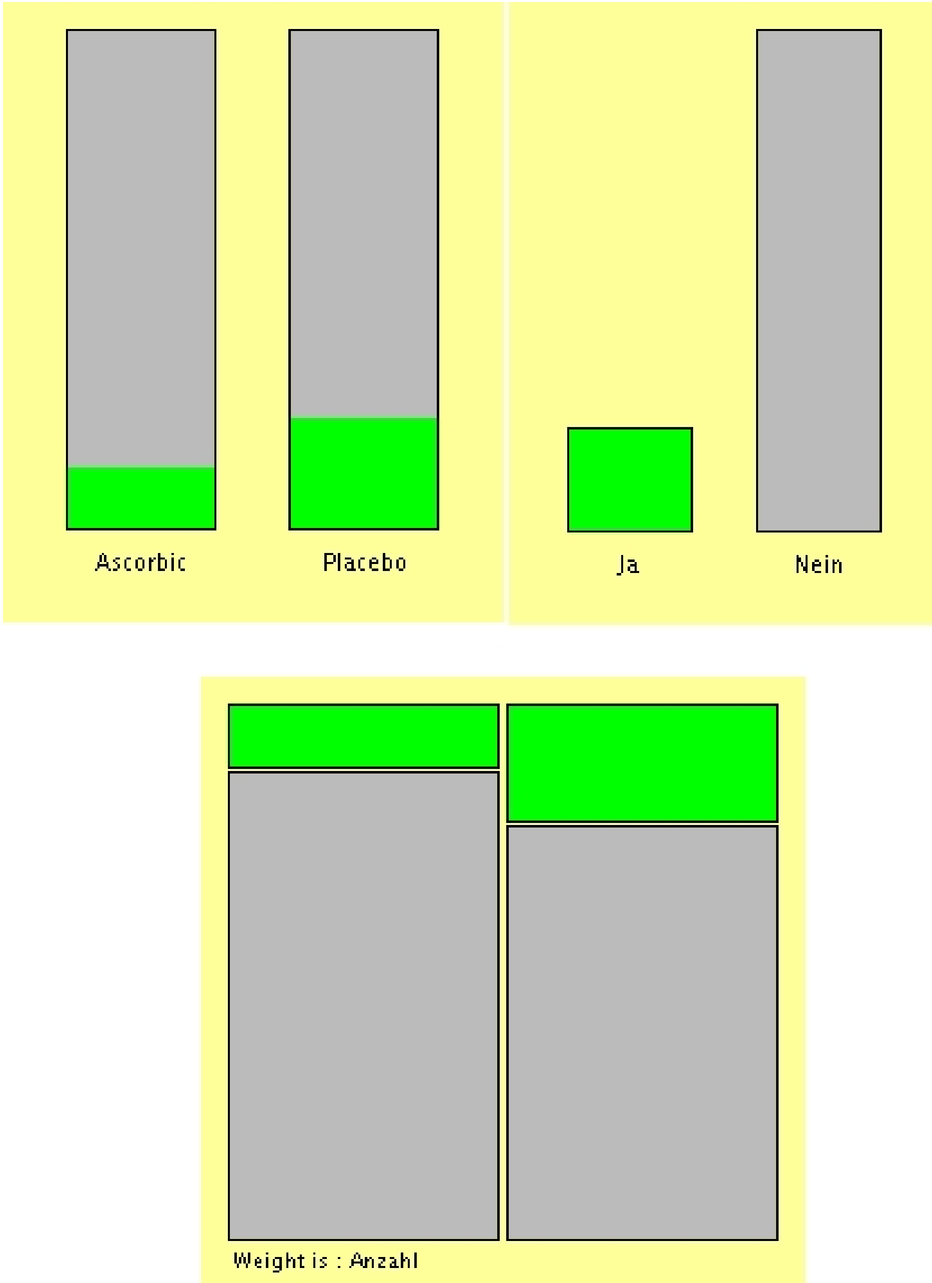
(b) Chi-Quadrat Test

Für die angegebene 2x2 Kontingenztafel gilt

$$X^2 = \left(x_1 - \frac{(x_1 + x_2) n_1}{n_1 + n_2}\right)^2 \left(\frac{1}{e_{11}} + \frac{1}{e_{21}} + \frac{1}{e_{12}} + \frac{1}{e_{22}}\right)$$

$$= \frac{(n_2 x_1 - n_1 x_2)^2 (n_1 + n_2)}{n_1 n_2 (x_1 + x_2)(n_1 + n_2 - x_1 - x_2)}$$

$$= T^2$$



Erkältung gegen Ascorbic acid/Placebo

Oben: Spineplot und Säulendiagramm mit Erkältungen selektiert

Unten: Mosaicplot mit Erkältungen selektiert

(2) Chi-Quadrat Tests für u. Poisson Verteilungen

Sei die Anzahl Beobachtungen in der Zelle i

$$X_i \quad \text{u.} \sim P(\lambda_i)$$

Für λ_i groß gilt approximativ

$$X_i \sim$$

$$\Rightarrow \frac{X_i - \lambda_i}{\sqrt{\lambda_i}} \sim N(0, 1)$$

$$\Rightarrow \frac{(X_i - \lambda_i)^2}{\lambda_i} \sim$$

$$\Rightarrow X^2 = \sum_{i=1}^{m} \frac{(X_i - \lambda_i)^2}{\lambda_i} \sim$$

Aber im allgemeinen sind die Zellen nicht unabhängig und wir müssen die λ_i's schätzen. Für jeden geschätzten Parameter nimmt man einen Freiheitsgrad weg. Der Beweis für die asymptotische Richtigkeit dieser Vorgehensweise ist aufwändig.

5.7.5 Beispiel — Tests von π

Gibt es Muster in der Dezimaldarstellung von π?

1) Die Anzahl von Nullen testen

$$H_0 : P(0) = 0.1$$

$$H_A : P(0) \neq 0.1$$

Sei X_{0n} die Anzahl Nullen in den ersten n Stellen

$$X_{0n} \sim B(n, 0.1)$$

$$n \text{ groß } \Rightarrow$$

Ein Test des Niveaus $\alpha = 0.05$ hat in diesem Falle den Annahmebereich

z.B. $n = 1000000 \Rightarrow [99412, 100588]$ und in der Tat gibt es 99959 Nullen in den ersten Million Stellen.
Aber was ist mit den Zahlen $1, \ldots, 9$?

2) Multinomial Test

Sei n_i die Anzahl von 'i', $i = 0, \ldots, 9$

$$P(\{n_i\}) =$$

$$p_i = P(i)$$

$$H_0 : p_i = 0.1 \quad \forall i$$

$$H_A : p_i \neq 0.1 \text{ für einige } i$$

Annahmebereich B:
Alle Zustände $(n_0, \ldots, n_9)$ die unter H_0 eine Gesamtwahrscheinlichkeit $\geq 1 - \alpha$ haben und wo

$$P_{H_0}(\{n_i\} \in B) \geq P_{H_0}(\{n_i\} \notin B)$$

d.h. die Wahrscheinlichkeit von jedem Ereignis im Annahmebereich ist größer gleich die Wahrscheinlichkeit von jedem Ereignis im Ablehnungsbereich

3) Ein anderes Modell

Sei $n_i \sim P(np_i) \quad \forall i$ unabhängig von den anderen n_j's und zuerst ohne der Einschränkung $\sum n_i = n$.

$$P(\{n_i\}) =$$

Unter der Bedingung, daß $\sum n_i = n$, muss dann

$$P_p(\{n_i\} | \sum n_i = n) \propto$$

Aus einem Vergleich mit dem Resultat für die Multinomiale Verteilung muß

$$P_p(\{n_i\} | \sum n_i = n) =$$

Daraus schliessen wir, dass beide Ansätze zum selben Test führen. Aber der Zustandsraum ist zu groß. n identische Kugeln auf m verschiedene Zellen bedeutet

$$\binom{n+m-1}{m-1}$$

$$m = 10, n = 100 \Rightarrow \binom{109}{9} >> 12^9$$

4) χ^2 Test

Für np_i groß gilt

$$n_i \sim N(np_i, np_i)$$

$$\Rightarrow \qquad\qquad\qquad \sim \chi^2_1$$

Für $\{n_i\}$ unabhängig

$$X^2 = \qquad\qquad\qquad \sim \chi^2_{10}$$

Unter der Einschränkung, dass $\sum n_i = n$, hat X^2 eine χ^2_9 Verteilung, weil wir einen Freiheitsgrad verloren haben.

$$\chi^2_{9,0.95} = 16.92, \chi^2_{9,0.99} = 21.67$$

Bei den Tests von Kanada war der höchste X^2 Wert 9.32.

Natürlich gibt es viele andere Tests auf die Zufälligkeit von Zahlenreihen — u.a. den Poker Test und den Lücken Test.

Frage: gibt es Muster in den ersten Dezimalstellen von e?

5.7.6 Eine wichtige Eigenschaft von X^2

Für eine Kontingenztafel gilt

$$X^2 = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

und unter H_0

$$= \sum_i \sum_j \frac{(n\widehat{p}_{ij} - n\widehat{p}_i\widehat{p}_j)^2}{n\widehat{p}_i\widehat{p}_j}$$

$$=$$

d.h. der Wert von X^2 steigt direkt mit n für gegebene beobachtete Proportionen.

Wenn ein Resultat nicht signifikant ist, kann man leicht ausrechnen, wie groß die Stichprobe hätte sein müssen, um Signifikanz zu erreichen, gegeben die selben Proportionen.

5.7.7 Zusammenfassendes zum χ^2 Test

χ^2 Tests werden sehr oft verwendet. Man muss überprüfen, welche Annahmen gemacht worden sind, welche Nullhypothese getestet worden ist, und wieviele Tests gemacht worden sind. (Es gibt jede Menge Möglichkeiten, Kategorien einer Kontingenztafel zu kombinieren.)

Der p-Wert allein genügt nicht. Man sollte untersuchen, wo die Abweichungen von der Nullhypothese erscheinen. Dafür sind die X^2 Beiträge der einzelnen Zellen aufschlussreich.

χ^2 Tests berücksichtigen nicht ordinale Struktur. Deshalb sind andere Testverfahren besser, wenn es um Anpassungstests geht. Trotzdem können χ^2 Tests als erster Versuch nützlich sein.

Nur große Werte von X^2 werden als signifikant betrachtet, nicht kleine Werte, sogar wenn sie sehr unwahrscheinlich sind. (“Die beobachtete Werte sind zu gut.”)