

## 1 Definition

Gegeben sind  $X_1, \dots, X_n$  unabhängige, jeweils  $N(0, 1)$  verteilte Zufallsvariablen.

Dann gilt:

$$Z := \sum_{i=1}^n X_i^2 \quad Z \sim \chi^2(n)$$

$\alpha$ - Fraktile der  $\chi^2(n)$ - Verteilung näherungsweise:

$$x_\alpha = \frac{1}{2}(\tilde{x}_\alpha + \sqrt{2n-1})^2 \text{ wobei } \tilde{x}_\alpha \text{ das } \alpha\text{- Fraktile der } N(0, 1)$$

Weiter gilt:  $E[Z] = n$   $V[Z] = 2n$

(aus Bamberg/ Bauer: Statistik, 11. Auflage, Oldenburg 2001)

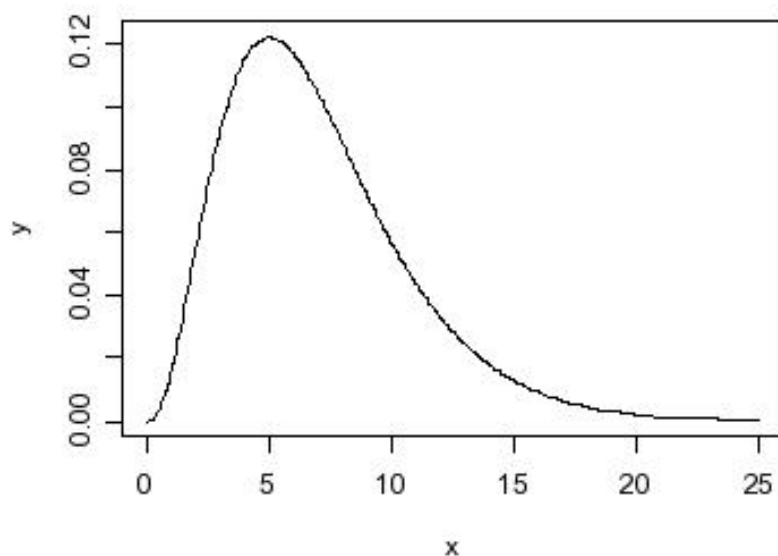
## 2 Definition

Wahrscheinlichkeitsverteilung, die zur Signifikanzprüfung eingesetzt werden kann.

(aus: ILMES - Internet-Lexikon der Methoden der empirischen Sozialforschung,  
<http://www.lrz-muenchen.de/wlm/ilmes.htm>)

## Plot von $\chi^2(7)$ mit R

```
x <- seq(0, 25, length = 1000), y <- dchisq(x, df = 7), plot(x, y, type = "l")
```



## Test auf Überzufälligkeit von Zusammenhängen in Kreuztabellen (Contingency- Table)

Kontingenztafel mit  $r$  Zeilen und  $c$  Spalten

$$\text{Teststatistik } X^2 = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \text{ mit } e_{ij} = \frac{(ZS_i)(SS_j)}{n}$$

$$X^2 \sim \chi^2((r-1)(c-1))$$

Problem bei großen Kontingenztafeln (>2x2):

$\chi^2$ - Test ist ein globaler Test;

Teststatistik gibt nicht an welche Zusammenhänge es gibt, sondern nur, ob es welche gibt.

Deshalb Modellierung mit komplexeren Verfahren, z.B.: einem log-linearen Modell.

## $\chi^2$ Anpassungstest (Goodness-of-fit-Test)

$$H_0 : P(X = i) = \pi_i, i = 1, \dots, k$$

$$H_1 : P(X = i) \neq \pi_i, \text{ für ein } i$$

$$\chi^2 = \sum_{i=1}^k \frac{(h_i - n\pi_i)^2}{n\pi_i}$$

Bei voll spezifizierter Verteilung: approximativ  $\chi^2(k-1)$  verteilt.

Bei  $l$  geschätzten Parametern: approximativ  $\chi^2(k-1-l)$  verteilt.

Ablehnungsbereich:  $\chi^2 > \chi_{1-\alpha}^2(k-1)$  bzw.  $\chi^2 > \chi_{1-\alpha}^2(k-1-l)$

(aus Fahrmeir, Künstler, Pigeot, Tutz: **Statistik**, Springer 2003)

## Beispiel für einen $\chi^2$ - Unabhängigkeitstest

Gegeben einer Erhebung zur Parteipräferenz an einem Sonntag, wollen wir überprüfen ob die unterschiedlichen Häufigkeiten von Frauen und Männern, rein zufällig sind.

Dazu folgende Daten:

	CDU/CSU	SPD	FDP	Grüne	Rest	
Männer	144	153	17	26	95	435
Frauen	200	145	30	50	71	496
gesmat	344	298	47	76	166	931

hieraus ergibt sich folgende "Unabhängigkeitstabelle":

	CDU/CSU	SPD	FDP	Grüne	Rest	
Männer	160.73	139.24	21.96	35.51	77.56	435
Frauen	183.27	158.76	25.04	40.49	88.44	496
gesmat	344	298	47	76	166	931

Nullhypothese:  $H_0 : P(X = i, Y = j) = P(X = i)P(Y = j), i = 1, 2, j = 1, \dots, 5$

Alternativhypothese:  $H_1 : P(X = i, Y = j) \neq P(X = i)P(Y = j)$  für mindestens ein Paar  $(i,j)$ .

Aus den gegebenen Daten erhalten wir folgende Teststatistik:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^5 \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}} = 20.065 > 9.488 = \chi_{0.95}^2(4)$$

Es stellt sich heraus, dass auf einem Signifikanzniveau von  $\alpha = 0.05$  ein Zusammenhang zwischen der Wahl der Partei und dem Geschlecht besteht.

## Beispiel für einen $\chi^2$ - Anpassungstest

Wir wollen zu einem Niveau von  $\alpha = 0.05$  überprüfen ob die Anzahl der Anrufe bei einer Hotline im Verlauf von je 15min. einer Poissonverteilung mit  $\lambda = 3$  unterliegt.

Hierzu haben wir 50 Zeitintervalle gegeben:

Anzahl der Anrufe k	0	1	2	3	4	5	6	7	8	9	10	>10
Anzahl der Zeitintervalle mit k-Anrufen	1	6	8	10	6	6	7	1	1	2	2	0

Dazu teilen wir die Daten in 6 Intervalle

	$h_i$	$n\pi_i$
$(-\infty; 1]$	7	$9.95 = 50 * 1.999 = 50 * (0.050 + 0.149) = 50 * (\frac{e^{-3}\lambda^0}{0!} + \frac{e^{-3}\lambda^1}{1!})$
$(1; 2]$	8	$11.2 = \dots$
$(2; 3]$	10	$11.2 = \dots$
$(3; 4]$	6	$8.4 = \dots$
$(4; 5]$	6	$5 = \dots$
$(5; \infty]$	13	$4.55 = \dots$

Nullhypothese:

Die Beobachtungen stimmen mit einer Poissonverteilung mit Parameter  $\lambda = 3$  überein.

Alternativhypothese:

Sie unterliegen einer anderen Verteilung oder  $\lambda$  müsste anders gewählt werden.

Aus den gegebenen Daten erhalten wir folgende Teststatistik:

$$\chi^2 = \sum_{i=1}^6 \frac{(h_i - n\pi_i)^2}{n\pi_i} = 18.5 > 11.07 = \chi_{0.95}^2(5)$$

Die Anrufe unterliegen nicht der gewählten Poissonverteilung.