1 Die Minimum Chiquadrat Methode

Wir betrachten im folgenden diskret-verteilte oder gruppierte Stichproben der Form

Klasse	1	2	...	K
Anzahl	y_1	y_2	...	y_K

Hier bezeichnet y_k, $k = 1, \ldots, K$, die beobachtete Häufigkeit in der k-ten Klasse. Sei $n = \sum_k y_k$ der Stichprobenumfang und π_k die Wahrscheinlichkeit für die Klasse k unter einem hypothetischen Verteilungsmodell. Damit resultiert $n\pi_k$ als erwartete Häufigkeit. Die Güte der Modellanpassung kann nun durch Pearson-Statistik

$$X^2 = \sum_{k=1}^{K} \frac{(y_k - n\pi_k)^2}{n\pi_k} \tag{1}$$

beschrieben werden. Durch einfache Umformung erhält man mit den empirischen Wahrscheinlichkeiten (relativen Häufigkeiten) $p_k = y_k/n$ dafür

$$\begin{aligned} X^2 &= \sum_k \frac{y_k^2 - 2y_k n\pi_k + n^2\pi_k^2}{n\pi_k} \\ &= \sum_k \left(\frac{y_k}{n}\right)^2 \frac{n}{\pi_k} - 2\underbrace{\sum_k y_k}_{=n} + n\underbrace{\sum_k \pi_k}_{=1} \\ &= n\left(\sum_k \frac{p_k^2}{\pi_k} - 1\right). \end{aligned}$$

Oft hängen die hypothetischen Wahrscheinlichkeiten π_k selbst von unbekannten Parametern λ ab. Eine Möglichkeit diese Parameter zu schätzen stellt die *Minimum Chisquare Method* dar. Hierbei wird jener Wert von λ als Schätzer verwendet, der (1) für $\pi_k = \pi_k(\lambda)$ minimiert. Dazu sucht man die Nullstelle $\hat{\lambda}$ von

$$\begin{aligned} \frac{\partial}{\partial\lambda} X^2(\lambda) &= n\frac{\partial}{\partial\lambda} \sum_k \frac{p_k^2}{\pi_k(\lambda)} \\ &= -n\sum_k \frac{p_k^2}{\pi_k^2(\lambda)} \frac{\partial\pi_k(\lambda)}{\partial\lambda}. \end{aligned} \tag{2}$$

Diese ist oft eine nichtlineare Funktion in λ und $\hat{\lambda}$ kann nur numerisch (iterativ) gefunden werden. Als zweite Ableitung resultiert

$$\frac{\partial^2}{\partial\lambda^2} X^2(\lambda) = -n\sum_k \left(\frac{-2p_k^2}{\pi_k^3(\lambda)}\left(\frac{\partial\pi_k(\lambda)}{\partial\lambda}\right)^2 + \frac{p_k^2}{\pi_k^2(\lambda)} \frac{\partial^2\pi_k(\lambda)}{\partial\lambda^2}\right). \tag{3}$$

Die Iteration lautet somit für den $(t+1)$-ten Schritt

$$\begin{aligned} \lambda^{t+1} &= \lambda^t - \frac{\partial}{\partial\lambda} X^2(\lambda)\Big|_{\lambda=\lambda^t} \left(\frac{\partial^2}{\partial\lambda^2} X^2(\lambda)\right)^{-1}\Bigg|_{\lambda=\lambda^t} \\ &= \lambda^t + \left(n\sum_k \frac{p_k^2}{\pi_k^2(\lambda)} \frac{\partial\pi_k(\lambda)}{\partial\lambda}\right)\left(n\sum_k \left(\frac{2p_k^2}{\pi_k^3(\lambda)}\left(\frac{\partial\pi_k(\lambda)}{\partial\lambda}\right)^2 - \frac{p_k^2}{\pi_k^2(\lambda)} \frac{\partial^2\pi_k(\lambda)}{\partial\lambda^2}\right)\right)^{-1} \end{aligned} \tag{4}$$

Oft kann auch die Inverse in (4) durch ihren Erwartungswert ersetzt (approximiert) werden. Das Verfahren wird dann *Fisher Iteration* genannt.

Für den Fall der Poisson-Verteilung mit Parameter λ ergibt sich für $k = 0, 1, \ldots$

$$\begin{aligned} \pi_k(\lambda) &= \frac{\lambda^k}{k!}\, e^{-\lambda} \\ \frac{\partial}{\partial\lambda}\, \pi_k(\lambda) &= \left(\frac{k}{\lambda} - 1\right) \pi_k(\lambda) \\ \frac{\partial^2}{\partial\lambda^2}\, \pi_k(\lambda) &= \left[\left(\frac{k}{\lambda} - 1\right)^2 - \frac{k}{\lambda^2}\right] \pi_k(\lambda)\,. \end{aligned} \tag{5}$$

Die Iteration wird mit einem Anfangswert λ^0 gestartet, welcher häufig durch die Momentenmethode berechnen werden kann. Für die Poisson-Verteilung gilt $\lambda^0 = \bar{y} = \sum_k k y_k / n$.

1.1 Offene Klassen

Speziell für Poisson-verteilte Größen hat man immer eine endliche letzte Klasse, z.B.

Wert k	0	1	$\ldots$	$K-1$	$\geq K$
Anzahl	y_0	y_1	$\ldots$	y_{K-1}	y_K
Modell	π_0	π_1	$\ldots$	π_{K-1}	π_K

Hier beschreibt π_K die Wahrscheinlichkeit in zumindest Klasse K zu realisieren, also

$$\pi_K = 1 - \sum_{k=0}^{K-1} \pi_k\,. \tag{6}$$

Durch diesen kleinen Unterschied resultieren für $k = 0, \ldots, K-1$ zwar wiederum die Ergebnisse aus (5), aber für die letzte Klasse, $k = K$, gilt jetzt speziell

$$\frac{\partial}{\partial\lambda}\, \pi_K(\lambda) = -\sum_{k=0}^{K-1} \frac{\partial}{\partial\lambda}\, \pi_k(\lambda) \tag{7}$$

$$\frac{\partial^2}{\partial\lambda^2}\, \pi_K(\lambda) = -\sum_{k=0}^{K-1} \frac{\partial^2}{\partial\lambda^2}\, \pi_k(\lambda) \tag{8}$$

mit den $\pi_k(\lambda)$ Ableitungen, $k = 0, \ldots, K-1$, so wie in (5) spezifiziert.

1.2 Beispiel

Wir passen der Stichprobe über Häufigkeiten von $n = 109$ k-silbigen Worten in einem slawischen Text

k	0	1	2	3	4	≥ 5	gesamt
Anzahl	44	35	17	6	6	1	109

eine Poisson-Verteilung mit Parameter λ an. Als Initialisierung für die Iteration (4) verwenden wir $\lambda^0 = \sum_k k y_k / 109 = 1.06$, was einen X^2-Wert von 11.3 ergibt. Weiteres Iterieren liefert

```
it= 0 lambda= 1.064220 X2= 11.27654
it= 1 lambda= 1.128125 X2= 10.44464
it= 2 lambda= 1.136973 X2= 10.43276
it= 3 lambda= 1.137104 X2= 10.43276
it= 4 lambda= 1.137104 X2= 10.43276
it= 5 lambda= 1.137104 X2= 10.43276
```

Man sieht hierbei, dass sich bereits nach zwei Iterationen der Wert des Schätzers fast nicht mehr ändert. Die X^2 Werte in Abhängigkeit vom Parameterschätzer sind in der Abbildung dargestellt.

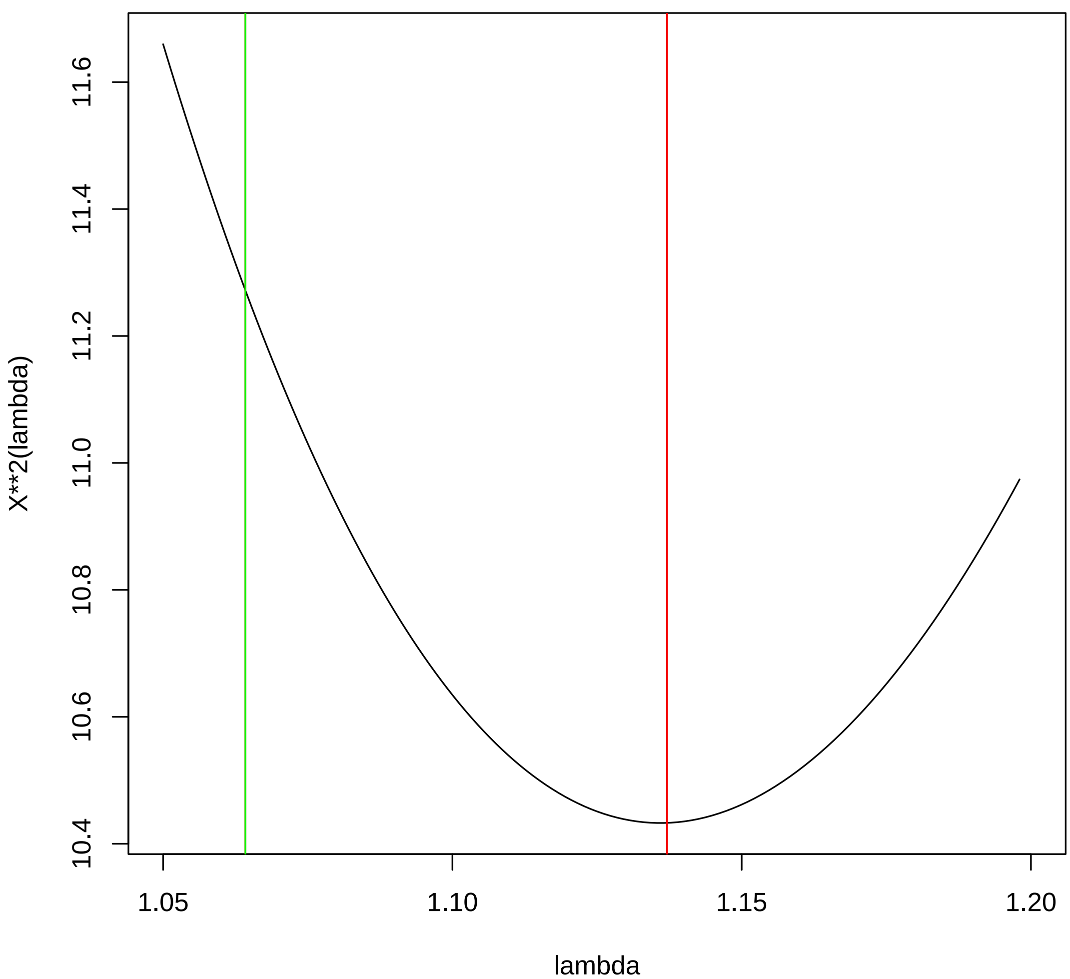

Abbildung 1: X^2 Statistiken für verschiedene Werte von λ. Weiters eingezeichnet sind die Ergebnisse für den Initialwert (links) und den Endwert (Mitte) von λ.