

# Maximum Likelihood

## Version 1.6

Uwe Ziegenhagen

15. November 2005

### Logarithmengesetze

$$\log_a(b) + \log_a(c) = \log_a(b \cdot c) \quad (1)$$

$$\log_a(b) - \log_a(c) = \log_a(b/c) \quad (2)$$

$$\log_a(b^c) = c \cdot \log_a(b) \quad (3)$$

### Ableitungsregeln

$$D(u + v) = u' + v' \quad (4)$$

$$D(u \cdot v) = u'v + v'u \quad (5)$$

$$D\left(\frac{u}{v}\right) = \frac{u'v - uv'}{v^2} \quad (6)$$

$$D(u[v]) = u'[v] \cdot v' \quad (7)$$

# 1 Einführung

*Maximum Likelihood* bedeutet soviel wie “Methode der größten Mutmaßlichkeit”, diese Methode wurde von dem australischen Statistiker Ronald Aylmer Fisher (1890-1962) entwickelt.

## 2 Beispiel

Bevor zwei Studenten per Münzwurf entscheiden, wer die Rechnung zahlt, soll sichergestellt werden, dass die Wahrscheinlichkeiten für Kopf und Zahl in etwa gleich sind.

Beim sechsmaligen Wurf erhalten sie die folgende Reihe:  $K, K, Z, Z, K, Z$

Die Zufallsvariable  $X$  “Auftreten von Kopf” ist dabei bernoulli-verteilt. Wenn die *wahre* Wahrscheinlichkeit für ‘Kopf’  $p = 0.1$  betragen würde, wäre die Wahrscheinlichkeit, genau diese Reihe zu erhalten:

$$0.1^3 \cdot 0.9^3 = 0.000729$$

Bei einer wahren Wahrscheinlichkeit von  $p = 0.5$  wäre diese Wahrscheinlichkeit:

$$0.5^3 \cdot (1 - 0.5)^3 = 0.015625$$

Der Wert  $p = 0.5$  ist damit mutmaßlicher als  $p = 0.1$ .

Man spricht bewußt von *Mutmaßlichkeit* und nicht von Wahrscheinlichkeit, denn die Wahrscheinlichkeit  $p$  kennen wir nicht und die Reihenfolge  $K, K, Z, Z, K, Z$  hätte bei *jedem* Wert von  $p$  ( $0 < p < 1$ ) erzielt werden können. Der Wert  $p = 0.5$  ist nur mutmaßlicher als der Wert  $p = 0.1$ .

Zur Bestimmung des mutmaßlichsten Wertes für  $p$  bilden wir die *Likelihoodfunktion* und leiten diese anschließend mit der Produktregel nach  $\theta$ , unserem zu schätzenden Parameter ab.

Seien  $X_1, \dots, X_n$  unabhängig identisch verteilte Stichprobenvariablen<sup>1</sup> einer Zufallsvariable  $X$  mit der Dichte  $f(x, \theta)$ . Dann gilt für die gemeinsame Dichte

---

<sup>1</sup>Die einzelnen Ausprägungen sind auch Zufallsvariablen!

$$f(x_1, \dots, x_n) = f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

Diese Funktion ist eine Funktion der  $x_i$  und  $\theta$ , die Beobachtungen  $x_i$  sind dabei zufällige Realisationen der Stichprobenvariablen  $X_1, \dots, X_n$ , während  $\theta$  als fix angesehen wird. Vertauschen wir die Rollen der Parameter, so erhalten wir die Likelihood-Funktion  $L$  mit fixen Beobachtungen und einem variablen  $\theta$ .

$$L(\theta) = L(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i, \theta)$$

Für unser Beispiel sieht die Likelihood-Funktion wie folgt aus:

$$\mathcal{L}(1, 1, 0, 0, 1, 0 | \theta) = \theta \cdot \theta \cdot (1 - \theta) \cdot (1 - \theta) \cdot \theta \cdot (1 - \theta) = \theta^3 \cdot (1 - \theta)^3$$

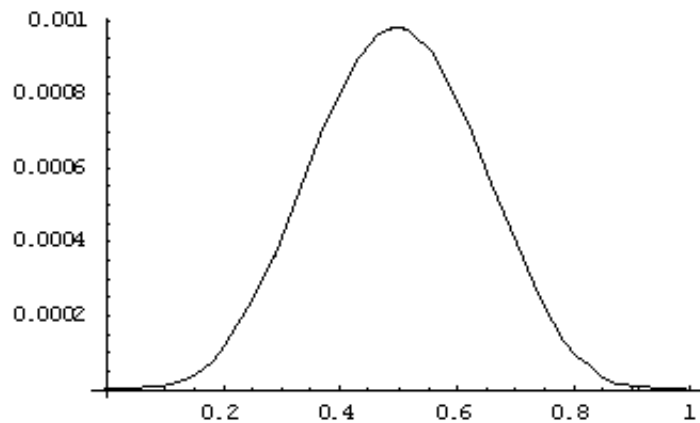


Abbildung 1: Plot der Likelihood-Funktion

Ableitung der Likelihood-Funktion:

$$\frac{\partial \mathcal{L}}{\partial \theta} = 3\theta^2 \cdot (1 - \theta)^3 + \theta^3 \cdot 3(1 - \theta)^2 \cdot (-1) \stackrel{!}{=} 0$$

$$3\theta^2 \cdot (1 - \theta)^3 - \theta^3 \cdot 3(1 - \theta)^2 = 0$$

$$3\hat{\theta}^2 \cdot (1 - \hat{\theta})^3 = \hat{\theta}^3 \cdot 3(1 - \hat{\theta})^2$$

$$3\hat{\theta}^2 \cdot (1 - \hat{\theta})(1 - \hat{\theta})^2 = \hat{\theta} \cdot \hat{\theta}^2 \cdot 3(1 - \hat{\theta})^2$$

$$1 - \hat{\theta} = \hat{\theta}$$

$$\hat{\theta} = 1/2 = x/n = \bar{x}$$

Oftmals gestaltet sich die Ableitung der Log-Likelihoodfunktion einfacher als die Ableitung der Likelihoodfunktion selbst, denn aus den Produkten werden hier durch die Logarithmen-Gesetze Summen. Für unser Beispiel bedeutet das:

$$\ln \mathcal{L} = \ln \left( \theta^3 \cdot (1 - \theta)^3 \right)$$

$$\ln \mathcal{L} = 3 \cdot \ln(\theta) + 3 \cdot \ln(1 - \theta)$$

$$\frac{\partial \ln \mathcal{L}}{\partial \theta} = \frac{3}{\theta} - \frac{3}{1 - \theta} \stackrel{!}{=} 0$$

Auflösen nach  $\theta$  ergibt auch hier  $\hat{\theta} = 3/6 = x/n = \bar{x}$ .

## 2.1 Exponentialverteilung

Wir haben  $n$  Werkstücke, die jeweils im Zeitpunkt  $t_i$  kaputt gingen. Um einen ML-Schätzer für den Parameter  $\lambda$  zu finden, bilden wir wieder die Likelihoodfunktion:

$$\mathcal{L}(t_1, t_2, \dots, t_n | \theta) = \prod_{i=1}^n \theta e^{-\theta \cdot t_i}$$

Durch Vereinfachen erhalten wir:

$$\mathcal{L}(t_1, t_2, \dots, t_n | \theta) = \theta^n \cdot e^{-\theta \cdot \sum_{i=1}^n t_i}$$

(Hinweis:  $a^{b \cdot c} \cdot a^{b \cdot d} \cdot a^{b \cdot e} = a^{b(c+d+e)}$ )

Durch Logarithmieren erhalten wir die Log-Likelihoodfunktion, die wir anschließend nach  $\theta$  ableiten:

$$\ln \mathcal{L}(t_1, t_2, \dots, t_n | \theta) = n \cdot \ln(\theta) - \theta \cdot \sum_{i=1}^n t_i$$

$$\frac{\partial \ln \mathcal{L}(t_1, t_2, \dots, t_n | \theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n t_i \stackrel{!}{=} 0$$

$$\frac{n}{\theta} = \sum_{i=1}^n t_i \Rightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^n t_i}$$

## 2.2 Poissonverteilung

Die Dichtefunktion der Poisson-Verteilung ist definiert als

$$f_{Po}(x, \lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

Als Likelihoodfunktion mit  $\theta = \lambda$  erhalten wir für  $n$  Versuche:

$$\mathcal{L}(x_1, x_2, \dots, x_n | \lambda) = \frac{e^{-\lambda} \cdot \lambda^{x_1}}{x_1!} \dots \frac{e^{-\lambda} \cdot \lambda^{x_n}}{x_n!} = \frac{e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i}}{x_1! \dots x_n!}$$

und bilden wiederum die Log-Likelihoodfunktion, die wir dann ableiten.

$$\ln \mathcal{L}(x_1, x_2, \dots, x_n | \lambda) = -n \cdot \lambda + (\ln \lambda) \sum_{i=1}^n x_i - \ln \left( \prod_{i=1}^n x_i! \right)$$

$$\frac{\partial \ln \mathcal{L}}{\partial \theta} = -n + \frac{\sum_{i=1}^n x_i}{\lambda} \stackrel{!}{=} 0$$

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i}{n}$$

## 2.3 Normalverteilung

### 2.3.1 Schätzen von $\mu$

Die Dichtefunktion der Normalverteilung ist definiert als:

$$f(\theta, \sigma, x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\theta}{2\sigma}\right)^2}$$

Erhalten wir bei der  $n$ -fachen Durchführung eines Zufallsexperiments  $n$  Realisationen der Zufallsvariable  $X$ , können wir die Likelihoodfunktion für  $\theta = \mu$  als Produkt der Normalverteilungen dieser  $n$  Versuche bestimmen:

$$\mathcal{L}(x_1, \dots, x_n, \sigma | \theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{\sum_{i=1}^n -\left(\frac{x_i - \theta}{2\sigma}\right)^2}$$

Anders aufgeschrieben:

$$\mathcal{L}(x_1, \dots, x_n, \sigma | \theta) = (\sqrt{2\pi}\sigma)^{-n} \cdot e^{\sum_{i=1}^n -\left(\frac{x_i - \theta}{2\sigma}\right)^2}$$

Durch Logarithmieren erhalten wir die Log-Likelihoodfunktion.

$$\ln \mathcal{L}(x_1, \dots, x_n, \sigma | \theta) = \ln \left( (\sqrt{2\pi}\sigma)^{-n} \cdot e^{\frac{-\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}} \right)$$

$$\ln \mathcal{L}(x_1, \dots, x_n, \sigma | \theta) = -n \left[ \ln(\sqrt{2\pi}) + \ln(\sigma) \right] - \frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}$$

$$\frac{\partial \ln \mathcal{L}(x_1, \dots, x_n, \sigma | \theta)}{\partial \theta} = \frac{-2 \sum_{x=1}^n (x_i - \theta)(-1)}{2\sigma^2} = \frac{\sum_{x=1}^n (x_i - \theta)}{\sigma^2} \stackrel{!}{=} 0$$

$$\frac{\sum_{x=1}^n (x_i - \theta)}{\sigma^2} \stackrel{!}{=} 0$$

$$\sum_{x=1}^n (x_i - \hat{\theta}) = 0$$

$$\sum_{x=1}^n x_i - n \cdot \hat{\theta} = 0$$

$$\sum_{x=1}^n x_i = n \cdot \hat{\theta}$$

$$\frac{\sum_{x=1}^n x_i}{n} = \hat{\theta}$$

$$\hat{\theta} = \bar{x}$$

Das arithmetische Mittel  $\bar{x}$  ist also der Maximum-Likelihood Schätzer für  $\mu$ .

### 2.3.2 Schätzen von $\sigma$

Um den ML-Schätzer für die Varianz zu bestimmen, bilden wir die partielle Ableitung der Likelihoodfunktion nach  $\sigma^2$  (Unser zu schätzender Parameter ist jetzt  $\sigma^2$ ,  $\mu$  wird als bekannt angesehen).

$$\ln \mathcal{L}(x_1, \dots, x_n, \mu | \theta) = \ln \left( (\sqrt{2\pi\theta})^{-n} \cdot e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\theta^2}} \right)$$

$$\frac{\partial \ln \mathcal{L}(x_1, \dots, x_n, \mu | \theta)}{\partial \theta} = -n \cdot \frac{1}{\theta} - \frac{1}{2} \cdot (-2) \cdot \theta^{-3} \cdot \sum_{i=1}^n (x_i - \mu)^2 \stackrel{!}{=} 0$$

$$\frac{-n}{\hat{\theta}} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\hat{\theta}^3} = 0$$

$$\frac{\sum_{i=1}^n (x_i - \mu)^2}{\hat{\theta}^3} = \frac{n}{\hat{\theta}}$$

$$\hat{\theta}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

### 3 Literatur

- Becker/Genschel, Schließende Statistik, Springer-Verlag
- Schira, Statistische Methoden der VWL und BWL, Addison-Wesley Verlag
- Wikipedia, <http://de.wikipedia.org>