

Slide 1

Grundlagen der Maximum-Likelihood Schätzung

Ulrich Kohler, WZB

29. April 2005

Slide 2

Probability Function (PF)

Jede mögliche Ausprägung einer diskreten Zufallsvariable hat eine ihr zugeordnete Auftrittswahrscheinlichkeit. Die Funktion, mit der die Auftrittswahrscheinlichkeit berechnet wird heißt „Probability Function“

▷ Beispiel: Die PF der dichotomen Zufallsvariable X mit $X = 1$ für Arbeitslose und $X = 0$ für Erwerbstätige lautet:

$$P(x_i|p) = p^{x_i}(1-p)^{1-x_i} \quad (1)$$

▷ Beispiel: Die PF der Zufallsvariable $Y = \sum_{i=1}^n x_i$ lautet:

$$P(y_j|n, p) = \binom{n}{y_j} p^{y_j} (1-p)^{n-y_j} \quad (2)$$

Slide 3

Probability Density Function (PDF)

Bei kontinuierlichen Zufallsvariablen hat ein spezifischer Wert für sich stets eine Auftrittswahrscheinlichkeit von 0. Aus diesem Grund werden die „Auftrittswahrscheinlichkeiten“ für möglichst kleine Intervalle um bestimmte Werte angegeben (Dichten). Die Funktion, mit der diese Dichte berechnet wird heißt „Probability Density Function“

▷ Beispiel: Eine oft verwendete PDF einer kontinuierlichen Zufallsvariable ist die Normalverteilung:

$$f(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - \mu)^2}{2\sigma^2}\right) \quad (3)$$

Slide 4

Multiplikationstheorem

Bei mehreren voneinander unabhängigen Experimenten ergibt sich die Wahrscheinlichkeit, dass alle Ergebnisse gemeinsam eintreten aus der Multiplikation der Wahrscheinlichkeiten der Einzelergebnisse.

$$PF(Y) = \prod_{i=1}^n PF(y_i) \quad (4)$$

$$PDF(Y) = \prod_{i=1}^n PDF(y_i) \quad (5)$$

▷ Beispiel: Münzwurf mit $p = .5$. Die Wahrscheinlichkeit zuerst Kopf, dann Zahl zu werfen beträgt $.5 \times .5 = .25$

PF, PDF und Likelihood-Funktion

In (1)–(3) wird die „Auftrittswahrscheinlichkeit“ für ein bestimmtes Datum \mathbf{y}_i in Abhängigkeit gegebener Parameter θ durch die Angabe einer Funktion $f(\mathbf{y}_i|\theta)$ bestimmt.

Slide 5

▷ Beispiel: In (1) wird $P(X = x)$ für $\theta = (p)$ bestimmt, in (2) wird $P(Y = y)$ für $\theta = (p, n)$ bestimmt und in (3) wird $f(y)$ für $\theta = (\mu, \sigma^2)$ bestimmt.

In der Likelihood-Funktion $L(\theta)$ gelten dagegen die Daten als bekannt:

$$L(\theta|\mathbf{y}_i) = f(\mathbf{y}_i|\theta) \quad (6)$$

Übungsbeispiel 1

Slide 6

Angenommen Y sei eine kontinuierliche Zufallsvariable für die die Normal-PDF gilt. Wie groß ist die *Likelihood* folgender Werte für θ , gegeben dass die einzige vorhandene Beobachtung $y_i = 22$ aufweist.

- $\theta_1 = (\mu, \sigma^2) = (13, 1)$
- $\theta_2 = (\mu, \sigma^2) = (25, 1)$

Für welchen Ergebnisraum von θ ist es wahrscheinlicher, dass wir das Datum $y_i = 22$ erhalten?

Slide 7

Das Maximum-Likelihood Prinzip

Das Maximum-Likelihood Prinzip besagt, dass aus allen möglichen Werten für θ derjenige ausgewählt wird, der die gegebenen Daten am wahrscheinlichsten macht.

Technisch wird derjenige Wert für θ ausgewählt, welcher die Likelihood-Funktion maximiert.

Slide 8

Übungsbeispiel 2

Gegeben die Daten $Y = (0, 1, 2)$. Unter Annahme einer Normal-PDF mit $\sigma^2 = 1$, welcher Wert für μ macht diese Daten am wahrscheinlichsten.

Antwort: Derjenige, bei dem die *Likelihood*-Funktion

$$L(\theta|Y) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y_i - \mu)^2}{2}\right) \quad (7)$$

am höchsten ist.

Lösungswege:

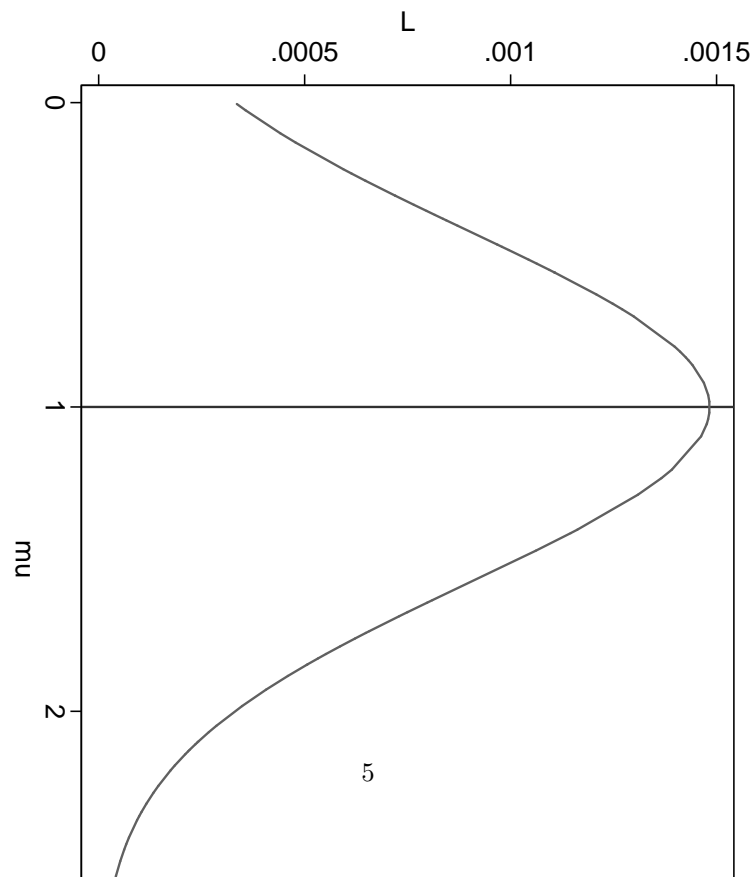
- Grafische Lösung: Ausprobieren verschiedener Werte für μ und Darstellung in einer Grafik
- Analytische Lösung: Nullsetzen der ersten Ableitung

Grafische Lösung mit Stata

Slide 9

```
. version 9
. set obs 100
obs was 0, now 100
. gen mu = uniform()*3
. gen L_0 = 1/(2*_pi)*exp(-(0-mu)^2/2)
. gen L_1 = 1/(2*_pi)*exp(-(1-mu)^2/2)
. gen L_2 = 1/(2*_pi)*exp(-(2-mu)^2/2)
. gen L = L_0 * L_1 * L_2
. line L mu, sort xline(1)
. graph export mlest05_1.eps, replace
(file mlest05_1.eps written in EPS format)
```

Slide 10



Analytische Lösung

Lösung durch Nullsetzen der 1. Ableitung der *logarithmierten* Likelihood-Funktion.

Slide 11

$$\frac{\partial -n(\ln \sqrt{2\pi}) - \frac{1}{2} \sum (y_i - \mu)^2}{\partial \mu} = \sum y_i - n\mu$$

Nullsetzen und nach μ auflösen

$$\begin{aligned} 0 &= \sum y_i - n\mu \\ \mu &= \frac{1}{n} \sum y_i = \bar{y}_i \end{aligned}$$

Modelle

In statistischen Modellen nehmen wir an, dass die Parameter in θ von so genannten „unabhängigen Variablen“ abhängen.

Zum Beispiel könnte man annehmen, dass μ ein lineares Modell von X ist, also $\mu = \mathbf{X}\mathbf{b}$.

Slide 12

$$\frac{\partial -n(\ln \sqrt{2\pi}) - \frac{1}{2} \sum (y_i - \mathbf{x}_i \mathbf{b})^2}{\partial \mu} = \frac{1}{2} \sum x_{ik} (y_i - \mathbf{x}_i \mathbf{b})$$

Nullsetzen und nach μ auflösen

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})\mathbf{X}'\mathbf{Y}$$