Maximum-Likelihood-Schätzungen

für Verteilungsparameter eines ausgewählten stochastischen Prozesses

Maximum Likelihood Estimation

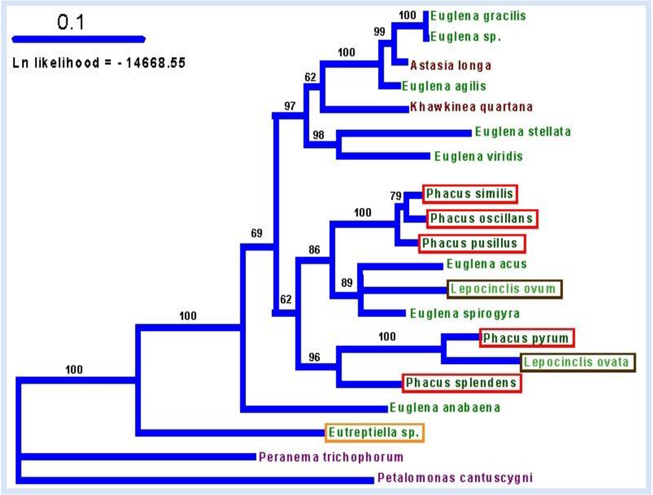
(MLE)

U. Menzel 10.3.2007

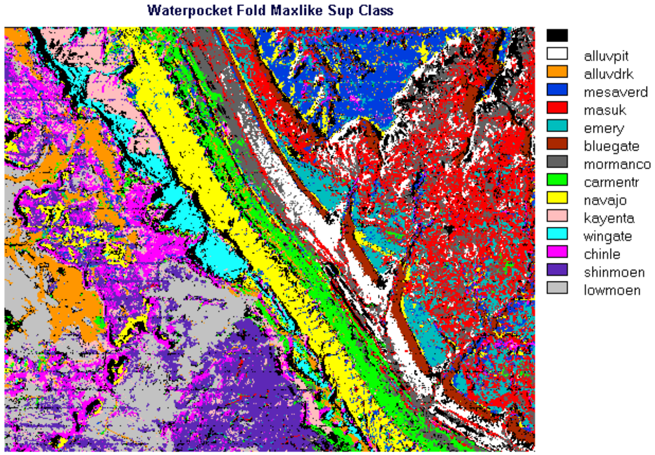
Maximum - Likelihood - Methode ist aktuell !

- R. A. Fisher (1890 - 1962)
- C. F. Gauβ (”Methode der kleinsten Quadrate”)
- Anzahl der Publikationen, die sich mit Likelihood befassen (”Likelihood” in Titel oder Abstract):
  - *Biostatistics*: 15 Publikationen im Jahr 2006
  - *Bioinformatics*: 25 Publikationen im Jahr 2006

Anwendungen der ML-Methode



- Phylogenetische Analysen auf der Grundlage der DNS-Sequenz (Stammbäume)



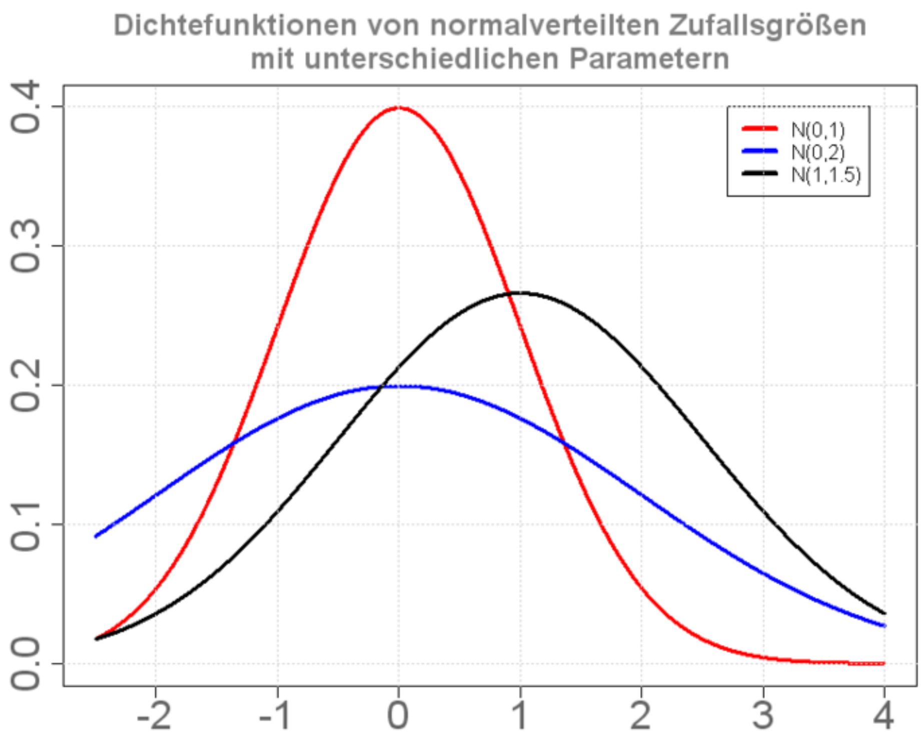
- Bilderkennung (Satellitendaten)

Die Maximum-Likelihood-Methode führt eine Schätzung durch

- Grundlegende Aufgabe der Statistik
- Beispiele für Schätzungen:
  - Wahlen: Hochrechnungen (500 Personen → Gesamtbevölkerung)
  - Lebenserwartung einer Insektenart (50 Individuen → 50 Mrd. Individuen)

⇨ Zufällige Stichprobe (”*Sample*”) → ”wahre” (erzeugende) Wahrscheinlichkeitsverteilung

Das Spezifische der ML-Methode: schätzt die *Parameter* einer Verteilung



μ: Erwartungswert (Mittelwert)
σ: Standardabweichung

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2} \right]$$

- Stichprobe → Parameter, z. B. μ, σ (oder beide)
- Art der erzeugenden Verteilung muss bekannt sein

  (*Normalverteilung?, Exponentialvert.?, Weibullvert.? …*)

Beispiel: Von der Stichprobe zum Schätzwert für Verteilungs - Parameter

- suche mittleres Gewicht der Frösche in einem Teich
- Zufällige Stichprobe von 10 Fröschen (in g):
  - $y_1 = 110$; $y_2 = 115$; $y_3 = 95$; $y_4 = 101$; $y_5 = 121$; $y_6 = 130$; $y_7 = 98$; $y_8 = 99$; $y_9 = 104$; $y_{10} = 111$
- Annahme: Gewicht ist *normalverteilt*

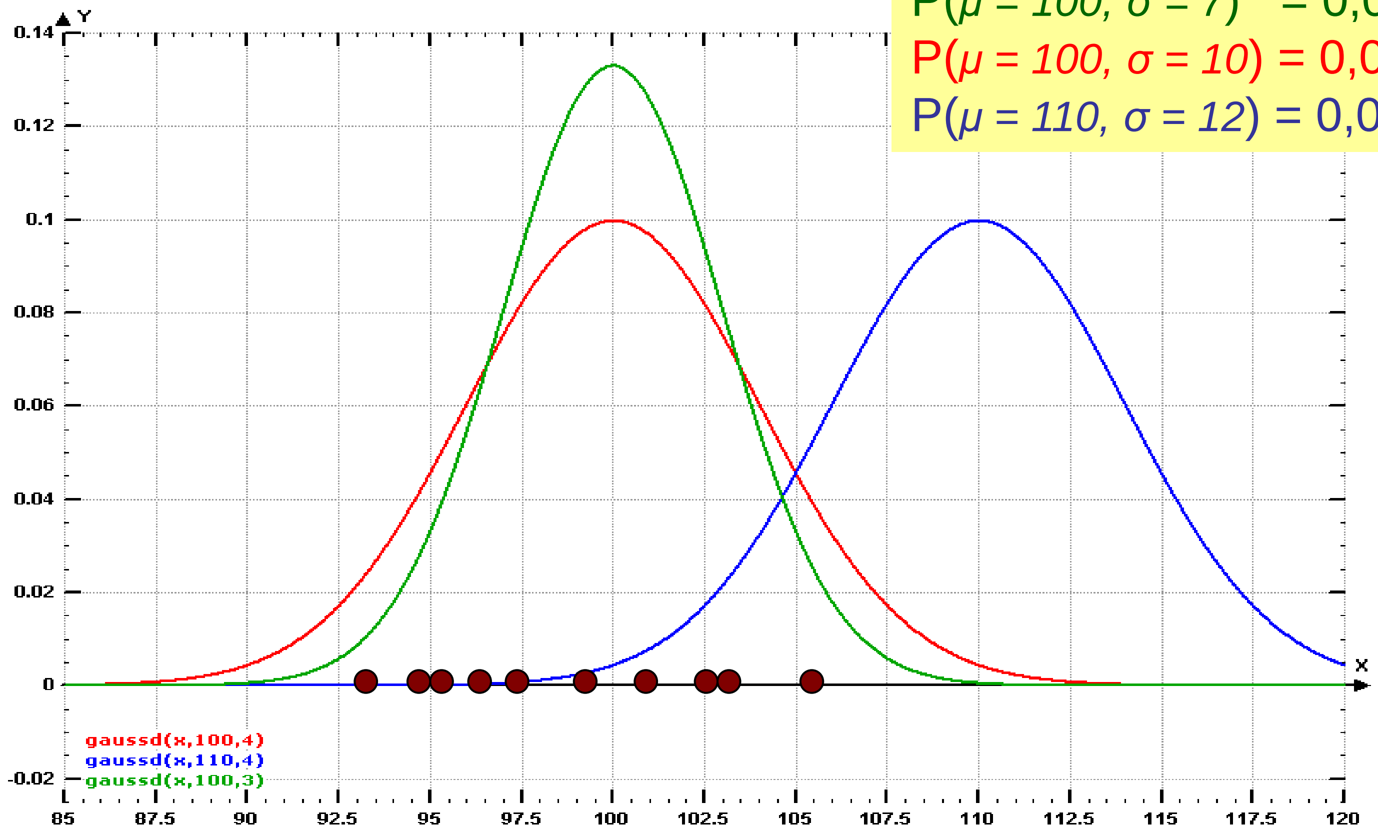
$$f(y) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left[-\frac{(y-\mu)^2}{2\sigma^2} \right] \qquad -\infty < y < \infty$$

Maximum-Likelihood: $y_1, y_2, \ldots y_{10} \rightarrow \mu, \sigma$

Ist n = 10 ausreichend ?

Grundlegende Idee der MLE

P(μ = *100*, σ = *7*) = 0,007
P(μ = *100*, σ = *10*) = 0,0104
P(μ = *110*, σ = *12*) = 0,000021



<u>Idee:</u> bestimme Parameter (μ,σ) derart, dass die beobachteten Daten plausibel (wahrscheinlich) erscheinen → **MLE.**

Die ”Wahrscheinlichkeit der Stichprobe”

θ - ein oder mehrere Verteilungsparameter

Diskrete Verteilung:

$$P = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid \theta) \quad = \quad p(y_1 \mid \theta) \quad \times \quad p(y_2 \mid \theta) \quad \times \ldots \times \quad p(y_n \mid \theta)$$

Kontinuierliche Verteilung: :

$$\frac{P}{\Delta y} \cong f(y_1, y_2, \ldots, y_n \mid \theta) \quad = \quad f(y_1 \mid \theta) \quad \times \quad f(y_2 \mid \theta) \quad \times \ldots \times \quad f(y_n \mid \theta)$$

Als Funktion des Parameters θ betrachtet

→ Likelihood - Funktion L(θ)

→ θ so wählen, dass L maximal wird

Bestimmung der plausibelsten Werte für die Verteilungsparameter aus der Likelihood

- Wir suchen *den* Wert von θ, der die Wahrscheinlichkeit der Stichprobe maximiert

⇨ Finde das θ, für welches L(θ) maximal wird !

$$\frac{\partial L}{\partial \theta} = 0 \quad \rightarrow \quad \hat{\theta}$$

$\hat{\theta}$ - Schätzer für θ

$$\frac{\partial^2 L}{\partial \theta^2} < 0 \quad an\ der\ Stelle\ \theta = \hat{\theta}$$

Der Schätzwert ist auch eine Zufallsvariable

- Entnimmt man 5 Stichproben (je 10 Frösche) und ermittelt aus jeder Stichprobe einen Schätzwert, so erhält man 5 *verschiedene* Schätzwerte.

⇨ $\hat{\theta}$ ist selbst eine Zufallsvariable (eine ”Statistik”)

⇨ Man kann den Erwartungswert und die Varianz des Schätzwertes berechnen:

$$E(Y) = \int_{-\infty}^{+\infty} y \cdot f(y) \cdot dy$$

$$V(Y) = E\left[\left(Y - E(Y)\right)^2\right]$$

Kontinuierliche Zufallsvariable, Normalverteilung

$$f(y_i \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left[-\frac{(y_i-\mu)^2}{2\sigma^2}\right]$$

$$L(\mu,\sigma) = f(y_1, y_2, \ldots, y_n \mid \mu,\sigma) = f\left(y_1 \mid \mu,\sigma\right) \;\times\; \ldots \;\times\; f\left(y_n \mid \mu,\sigma\right)$$

$$= \left\{\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left[-\frac{(y_1-\mu)^2}{2\sigma^2}\right]\right\} \;\times\; \ldots \;\times\; \left\{\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left[-\frac{(y_n-\mu)^2}{2\sigma^2}\right]\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\cdot\exp\left[\frac{-1}{2\sigma^2}\cdot\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

$$\ln(L) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

Normalverteilung, Schätzung für μ

$$\ln(L) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\frac{\partial \ln(L)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \mu = n \cdot \mu$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Ein Schätzer für den Mittelwert der Normalverteilung
ist das arithmetische Mittel der Beobachtungswerte.

Erwartungswert des Schätzers für μ

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}\cdot n\cdot \mu = \mu$$

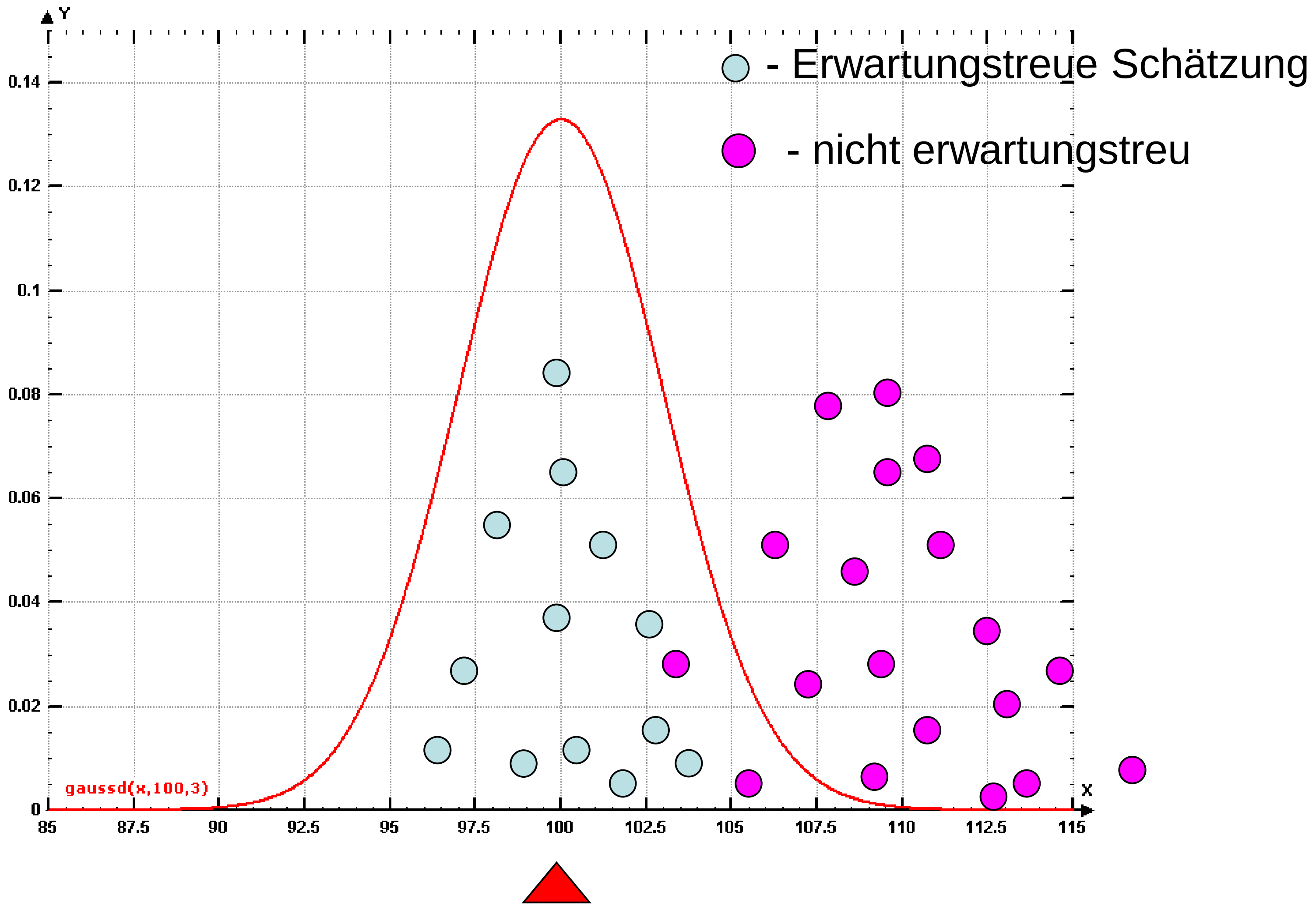
verwendet wurde:

$$E(y) = \int y \cdot \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] dy = \mu$$

Der *Erwartungswert des Schätzers für den Parameter* ist wieder der Parameter selbst.

⇨ Der Schätzer für μ ist erwartungstreu ("unbiased")

Erwartungstreue



Varianz des Schätzers für μ

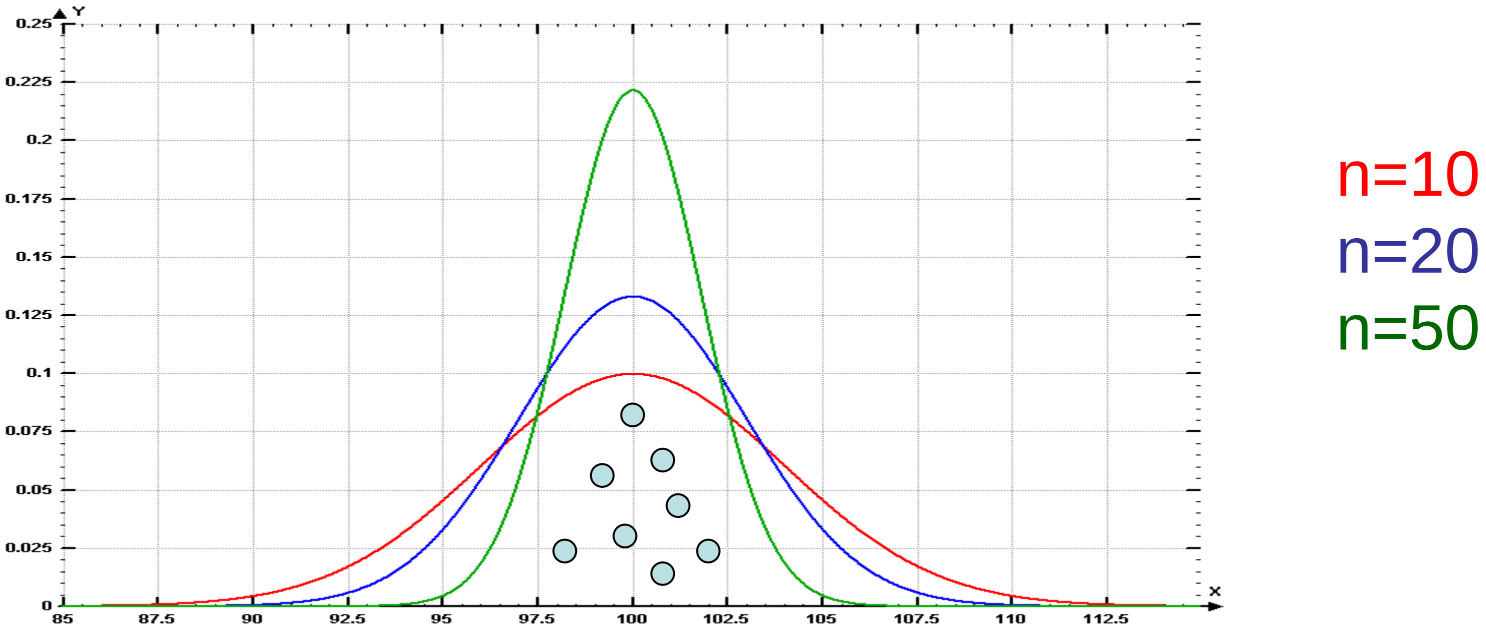
$$V(\hat{\mu}) = V\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(y_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{1}{n}\cdot\sigma^2$$

verwendet wurde:

$$V\left(y_i\right) = \sigma^2$$

Für n gegen ∞ geht die Varianz des Schätzers gegen Null.
⇨ Der Schätzer für μ ist konsistent.

Konsistenz eines Schätzers



- Ein erwartungstreuer Schätzer heißt konsistent, wenn die Varianz des Schätzers gegen Null geht, sobald die Größe der Stichprobe (n) gegen Unendlich geht.
- Der Schätzer konvergiert dann ”in Wahrscheinlichkeit” gegen den wahren Wert.

Normalverteilung, Schätzung für σ

$$\ln(L) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \hat{\mu}\right)^2$$

$$\frac{\partial \ln(L)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\left(y_i - \hat{\mu}\right)^2 = 0$$

$$\hat{\sigma}_0{}^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}\left(y_i - \hat{\mu}\right)^2$$

$$aber(!): \quad E(\hat{\sigma}_0{}^2) = \frac{n-1}{n}\cdot\sigma^2 \neq \sigma^2$$

$$Korrektur: \quad \hat{\sigma}^2 = \frac{n}{n-1}\cdot\sigma_0^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(y_i - \hat{\mu}\right)^2$$

Schätzer für σ^2 ist konsistent.

Normalverteilung, Schätzer für Mittelwert und Varianz

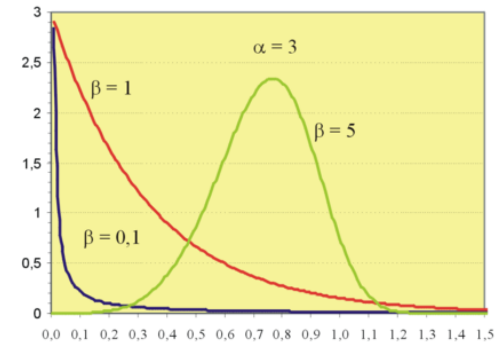
y_i – beobachtete Werte(Stichprobe,”Sample”); i=1,2,…,n
n – Anzahl der Werte in der Stichprobe

Erwartungstreue, konsistente Schätzer sind:

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(y_i - \hat{\mu}\right)^2$$

Weibull-Verteilung



$$f(y \mid \theta) = \left(\frac{2y}{\theta}\right) \cdot \exp\left[-\frac{y^2}{\theta}\right] \quad y > 0$$

$$L(y_1, y_2, \ldots y_n \mid \theta) = f(y_1 \mid \theta) \quad \times \quad f(y_2 \mid \theta) \quad \times \quad \ldots \quad \times \quad f(y_n \mid \theta)$$

$$= \left(\frac{2y_1}{\theta}\right) \cdot \exp\left[-\frac{{y_1}^2}{\theta}\right] \quad \times \quad \left(\frac{2y_2}{\theta}\right) \cdot \exp\left[-\frac{{y_2}^2}{\theta}\right] \quad \times \quad \ldots \quad \times \quad \left(\frac{2y_n}{\theta}\right) \cdot \exp\left[-\frac{{y_n}^2}{\theta}\right]$$

$$= \left(\frac{2}{\theta}\right)^n \cdot \exp\left[-\frac{1}{\theta} \cdot \sum_{i=1}^{n} y_i^2\right] \cdot (y_1 \cdot y_2 \cdot y_3 \cdot \ldots \cdot y_n) \quad setze: \quad u = \sum_{i=1}^{n} y_i^2 \quad "Statistik"$$

$$= \left(\frac{2}{\theta}\right)^n \cdot \exp\left[-\frac{u}{\theta}\right] \cdot (y_1 \cdot y_2 \cdot y_3 \cdot \ldots \cdot y_n)$$

$$= g(u, \theta) \times h(y_1, y_2, y_3, \ldots, y_n)$$

Die Zufallsvariable u ist eine minimale ”erschöpfende Statistik” für θ
(Faktorisierungskriterium der Likelihood-Funktion)

MLE für θ in Weibull-Verteilung

$$L(y_1, y_2, \dots y_n \mid \theta) = \left(\frac{2}{\theta}\right)^n \cdot \exp\left[-\frac{u}{\theta}\right] \cdot (y_1 \cdot y_2 \cdot y_3 \cdot \dots \cdot y_n) \qquad mit \quad u = \sum_{i=1}^{n} y_i^2$$

$$\ln(L) = n \cdot \ln(2) - n \cdot \ln(\theta) - \frac{u}{\theta} + \ln(y_1 \cdot y_2 \cdot y_3 \cdot \dots \cdot y_n)$$

$$\frac{d\ln(L)}{d\theta} = -\frac{n}{\theta} + \frac{u}{\theta^2} = 0$$

$$\hat{\theta} = \frac{u}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i^2 \qquad \text{Schätzer für } \theta\text{; Funktion von u}$$

$$\hat{\theta} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i^2 = \frac{u}{n}$$

$$E(\hat{\theta}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^{n} y_i^2\right) = \frac{1}{n} \sum_{i=1}^{n} E(y_i^2) \qquad Subst.\ w = y^2 \rightarrow E(y_i^2) = \theta$$

$$E(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \theta = \frac{1}{n} \cdot n \cdot \theta = \theta \qquad erwartungstreu$$

Minimalschätzung für die Weibull - Verteilung

- Der Schätzer für θ ist:
  - ein erwartungstreuer Schätzer
  - eine Funktion der minimalen ”erschöpfenden Statistik”

⇨ Damit ist er von allen möglichen erwartungs-treuen Schätzern derjenige mit der kleinsten Varianz (Minimalschätzer,”***MVUE=Minimum-Variance Unbiased Estimator***”)

- **ML – Methode führt oft automatisch zu einem Minimalschätzer !**

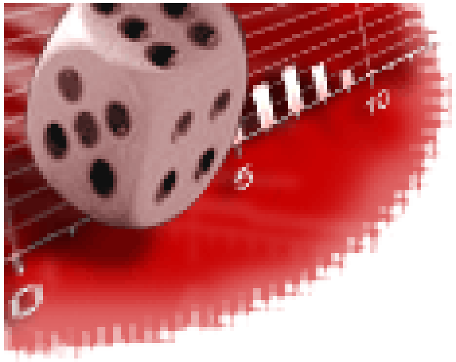
Zusammenfassung 1

- Ausgehend von einer zufälligen Stichprobe und einer Annahme über die zugrunde liegende Verteilung ermittelt die ML-Methode die plausibelsten Parameter dieser Verteilung.
- Die Likelihood-Funktion ist die gemeinsame Verteilungsfunktion (kontinuierlicher Fall) bzw. die kombinierte Wahrscheinlichkeit (diskreter Fall) der Stichprobe, aufgefasst als Funktion der Verteilungsparameter.
- Der plausibelste Wert für den Parameter der Verteilungsfunktion ist derjenige, welcher die Likelihood-Funktion maximiert. Dieser wird als Schätzer verwendet.

Zusammenfassung 2

- Ein Schätzer ist:
  - erwartungstreu, wenn sein Erwartungswert gleich dem zu schätzenden Parameter ist,
  - konsistent, wenn er erwartungstreu ist und wenn seine Varianz für unendlich große Stichproben gegen Null geht.
- Ein Minimalschätzer (MVUE) ist ein erwartungstreuer Schätzer mit kleinstmöglicher Varianz. Die ML-Methode führt oft automatisch zu einem Minimalschätzer.

Vielen Dank für Ihre Aufmerksamkeit



http://puffer.genpat.uu.se/MLE/
uwe.menzel@genpat.uu.se

Aufgabe

- $y_1, y_2, \ldots y_n$ sei eine zufällige Stichprobe von einer Poisson-Verteilung mit dem Mittelwert ⍰
  - Finden Sie einen ML-Schätzer ⍰ˆ für ⍰
  - Berechnen Sie den Erwartungswert und die Varianz dieses Schätzers
  - Ist der Schätzer erwartungstreu ?
  - Ist der Schätzer konsistent ?

ML für eine diskrete Verteilung

- n Versuche $y_1, y_2, \ldots y_n$ mit jeweils zwei möglichen Resultaten: $y_i = 0,1$
- p sei die Wahrscheinlichkeit des Erfolges (y_i=1)

$$L(y_1, y_2, \ldots, y_n \mid p) = P(y_1 \mid p) \quad \times \quad P(y_2 \mid p) \quad \times \quad \ldots \quad \times \quad P(y_n \mid p)$$

$$= p^{y_1} \cdot (1-p)^{1-y_1} \times p^{y_2} \cdot (1-p)^{1-y_2} \times \ldots \times p^{y_n} \cdot (1-p)^{1-y_n}$$

$$= p^{y} \cdot (1-p)^{n-y} \quad mit \quad y = \sum y_i$$

$$\ln(L) = y \cdot \ln(p) + (n-y) \cdot \ln(1-p)$$

$$\frac{d \ln(L)}{dp} = y \cdot \left(\frac{1}{p}\right) + (n-y) \cdot \left(\frac{-1}{1-p}\right) = 0$$

$$\frac{y}{p} = \frac{n-y}{1-p}$$

$$\hat{p} = \frac{y}{n}$$

Weitere Anwendungen

- Genauigkeit eines Messgerätes (”50 ± 0,5”))
- phylogenetische Analyse von DNA- oder Proteinsequenzen mit Maximum-Likelihood (http://abacus.gene.ucl.ac.uk/software/paml.html)
- Classifier
  http://www.eduspace.esa.int/eduspace/subdocument/default.asp?document=521
- Shoreline Mapping
  http://www.eomonline.com/EOM_Jul05/article.php?Article=feature3
- Localization by Maximum-Likelihood

  http://www.nasatech.com/Briefs/Oct98/NPO20392.html

Programme

- R: *mle*-package (im *stats4* package)
- Matlab code for the ML estimation (http://www.netlab.tkk.fi/tutkimus/com2/fbm/index.shtml)

Invarianzeigenschaft der ML-Methode

- Sei θ' ein Schätzer für θ

⇨ dann ist f(θ') ein Schätzer für f(θ)

Faktorisierung der Likelihood-Funktion

$$L\left(y_1, y_2, \ldots\, y_n \mid \theta\right) = g(u,\theta) \times h\left(y_1, y_2, y_3, \ldots, y_n\right) \quad mit \quad u = \sum_{i=1}^{n} y_i^2$$

Sei u eine von einer Stichprobe y_1,y_2,..,y_n abgeleitete Statistik. Dann ist u eine erschöpfende Statistik für den Parameter θ, wenn (und nur wenn) die Likelihood in zwei nichtnegative Faktoren $g(u,\theta) \times h(y_1,y_2,\ldots,y_n)$ zerlegt werden kann, wobei

- g ausschließlich eine Funktion von u und θ ist und
- h nicht von θ abhängt.

(*engl.: "sufficient statistics")*

(h ist oft identisch 1, z. B. bei derNormalverteilung)

Erschöpfend bedeutet, dass die in der Stichprobe enthaltene Information vollständig für die Schätzung ausgenutzt wird. Es geht keine Information verloren, wenn z. B. 10 Werte einer Stichprobe auf einen einzigen Wert (arithmetisches Mittel) reduziert werden.

Weitergehende Theorie

- Eine Schätzung, bei der Vorwissen in Form einer a priori-Wahrscheinlichkeit einfließt, wird Maximum-A-Posteriori-Schätzung (MAP) genannt.