Qualitätsmanagement

Übung 8

15. Juni 2007

Maximum-Likelihood-Methode:

- Wir möchten eine Stichprobe untersuchen.
- In dieser Stichprobe sind uns aber verschiedene Kennwerte, die wir für die Schätzung der Grundgesamtheit benötigen, nicht bekannt (Erwartungswert, Standardabweichung).
- Anhand von Stichprobenwerten gelingt es uns die Kennwerte möglichst genau zu schätzen.
- Die Maximum-Likelihood-Methode besagt, daß zu einem festen Stichprobenergebnis $(x_1, x_2, ..., x_n)$ derjenige Schätzwert für den unbekannten zu schätzenden Parameter ϑ zu wählen ist, unter dem **von vornherein** die Wahrscheinlichkeit für das Eintreten dieses Stichprobenergebnisses am größten ist.

$$L[\vartheta|(x_1, ..., x_n)] = \begin{cases} \prod_{i=1}^{n} P(X_i = x_i|\vartheta)(diskret) \\ \prod_{i=1}^{n} f(x_i|\vartheta)(stetig) \end{cases}$$

- Die Funktion $L[\vartheta|(x_1, ..., x_n)]$ gibt zu jedem ϑ-Wert an, wie wahrscheinlich es **war** , die Stichprobenwerte $(x_1, ..., x_n)$ zu beobachten, wenn ϑ der wahre Parameter ist.
- Einem Parameterwert ϑ, für den die Beobachtung $(x_1, ..., x_n)$ **von vornherein** wahrscheinlicher war als für einen anderen Parameterwert wird **im nachhinein** die höhere Glaubwürdigkeit (Plausibilität) zugesprochen, der wahre Parameterwert der Grundgesamtheit zu sein.
- Ein Parameterwert ϑ_1 ist aufgrund der Beobachtung $(x_1, ..., x_n)$ **plausibler** als ein Parameterwert ϑ_2, wenn gilt $L[\vartheta_1|(x_1, ..., x_n)] > L[\vartheta_2|(x_1, ..., x_n)]$

Beispiel:

Wir befinden uns auf See und möchten die Qualität der Fanggewässer untersuchen. Dazu fangen wir n=15 Fische und untersuchen, ob diese Fische krank sind oder nicht.

- Sei X:"Anzahl der kranken Fische beim i-ten Zug" mit i = 1,...,15
- $x_i \approx$ bernoulli-verteilt mit π
- Y= $\sum x_i$ "Anzahl der kranken Fische bei n = 15 Zügen; y $\approx$ B(n;π) $\approx$ B(15;$\hat{\pi}$)
- $(x_1, ..., x_{15}) = 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0$

Beispiel:

	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0											
.											
.											
.											
4	0.04	0.19	0.23	0.22	0.13	0.04	0	0	0	0	0
.											
.											
.											
15											

Beispiel:

	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0											
.											
.											
.											
4	0.04	0.19	**0.23**	0.22	0.13	0.04	0	0	0	0	0
.											
.											
.											
15											

Aufgabe 1:

Sie interessieren sich für die Unfallhäufigkeit an einem Verkehrsknotenpunkt. Sowohl die Polizei als auch die Versicherungsgesellschaften haben diesen Sachverhalt bereits in der Vergangenheit untesucht, so daß Sie auf diese Unterlagen zurückgreifen können:

Versicherung:

x	0	1	2	$x \neq 0{,}1{,}2$
P(X=x)	0,7	0,1	0,2	0

Polizei:

x	0	1	2	$x \neq 0{,}1{,}2$
P(X=x)	0,1	0,4	0,5	0

Aufgabe 1:

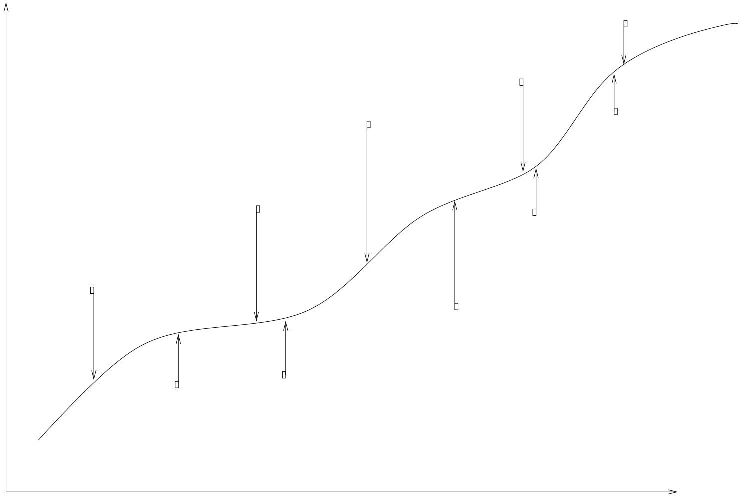
- Wie nennt man ein solches "Zurückgreifen" auf vorhandene Daten ?

Leider unterscheiden sich beide Wahrscheinlichkeitsverteilungen doch recht erheblich voneinander. Um zwischen beiden Verteilungen entscheiden zu können, stellen Sie sich selbst an 5 zufällig ausgewählten Stichtagen an diese Straßenkreuzung und registrieren die Unfälle. Die Stichprobe ergab folgende Ergebnisse: (0,2,0,2,1)

- Wie lautet Ihre Entscheidung nach dem Maximum-Likelihood-Prinzip?

Methode der kleinsten Quadrate

- Methode wird auch Least Squares Method genannt.
- Wiederum möchten wir eine Stichprobe untersuchen.
- Wiederum sind uns aber in dieser Stichprobe verschiedene Kennwerte, die wir für die Schätzung der Grundgesamtheit benötigen, nicht bekannt (Erwartungswert, Standardabweichung).
- Vorgehen: Wir versuchen innerhalb einer Punktwolke (= die beobachteten Werte) eine Funktion zu legen, so daß die Abweichungen zwischen der Funktion und den einzelnen Messwerten minimal wird.



Es gilt:

$$Q(\mu_i(\vartheta)/(x_1, ..., x_n) = \sum(x_i - \mu_i(\vartheta))^2 \Rightarrow Min$$

- Ein Parameterwert $Q(\mu_1(\vartheta)/(x_1, ..., x_n)$ ist aufgrund der Beobachtung $(x_1, ..., x_n)$ plausibler als ein Parameterwert $(\mu_2(\vartheta)/(x_1, ..., x_n)$, wenn gilt $Q(\mu_1(\vartheta)/(x_1, ..., x_n) < Q(\mu_2(\vartheta)/(x_1, ..., x_n)$

Aufgabe 2:

Benutzen Sie die Angaben aus Aufgabe 1: Wie lautet Ihre Entscheidung nach der Methode der kleinsten Quadrate?