Statistik I — bis jetzt

K 1 Einführung

K 2 Beschreibende Statistik

K 3 Graphiken

und jetzt

K 4 Schätzen

4.1 Eigenschaften von Schätzern

Erwartungstreue

Konsistenz

Mittlere quadratische Fehler

4.2 Herleitung von Schätzern

K 4 Schätzen

Was ist der Durschnittseinkommen einer Familie?

Welcher Prozentsatz der Bevölkerung lebt unter der Armutsgrenze?

Welchen Marktanteil hat der X-Box?

Welcher Prozentsatz der Wähler unterstützt die SPD?

Was war die Inflationsrate für April?

Wieviel schneller fährt der neue Ferrari?

Wie zuverläßig sind Volkswagen Autos?

Wieviele Leute wohnen in Deutschland?

Wieviel Bier trinkt ein Bayer?

Wir wissen nicht, wir müssen schätzen.

Theorie

Von einer Stichprobe zu einer Verteilung

Die Verteilung hat Parameter $\underline{\theta}$

θ_j wird durch die Funktion

$$\widehat{\theta}_j(X_1, X_2, \ldots, X_n)$$

geschätzt.

d.h. Aufgrund der Beobachtungen $(x_1, x_2, \ldots, x_n)$ von Zufallsgrößen $(X_1, X_2, \ldots, X_n)$ soll eine "möglichst gute" Schätzung eines unbekannten Parameters θ_j der Verteilungen der X_i angegeben werden.

Wir werden meistens o.B.d.A. nur einen Parameter schätzen.

4.1 Eigenschaften von Schätzern: Was heißt "gut"?

(a) Erwartungstreue

Beispiel 1

$$E[\bar{X}] = \frac{1}{n} E\left[\sum X_i\right] = \mu$$

Beispiel 2

$$X \sim E(\lambda)$$

$$f(x) = \lambda e^{-\lambda x}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ist ein erwartungstreuer Schätzer für} \frac{1}{\lambda}.$$

Ist $\frac{n}{\sum X_i}$ ein erwartungstreuer Schätzer für λ?

$$E\left[\frac{n}{\sum X_i}\right] =$$

Sogar für $n = 1$ geht es nicht:

$$E\left[\frac{1}{X}\right] = \int_0^\infty \frac{1}{x} \lambda e^{-\lambda x} \mathrm{d}x$$

(b) Konsistenz (im statistischen Sinne)

θ wird assymptotisch richtig geschätzt.

Die Bedingungen (a) und (b) sagen wenig über die Güte eines Schätzers aus, insbesondere bei kleineren Stichproben.

(c) Mittlere quadratische Fehler
(Je kleiner, desto besser)

(Mean Square Error)

$$= E[\widehat{\theta}^2] - 2\theta E[\widehat{\theta}] + \theta^2$$

$$= E[\widehat{\theta}^2] - E[\widehat{\theta}]^2 + E[\widehat{\theta}]^2 - 2\theta E[\widehat{\theta}] + \theta^2$$

$$MSE =$$

Für einen erwartungstreuen Schätzer gilt

$$MSE = Varianz$$

4.2 Herleitung von Schätzern

4.2.1 Methode der kleinsten Quadrate

Ein Schätzungsziel wäre die Minimierung von

wo θ uns unbekannt ist und wir das Minimum über alle möglichen Funktionen suchen. Das wird zu schwierig. Deshalb diese Methode.

Beobachtungen $(x_1, \ldots, x_n)$ und

$$E[X_i] = g_i(\theta)$$

$g_i(\theta)$ ist eine Prädiktorfunktion für X_i und wir wählen $\widehat{\theta}$, um

$$Q = \sum_{i=1}^{n} (x_i - g_i(\theta))^2$$

zu minimieren.

$$\frac{\partial Q}{\partial \theta} = 0$$

$$\Rightarrow$$

Beispiele

(1) Exponentialverteilung

$$X \sim E(\lambda)$$

$$\theta = \lambda$$

$$E[X_i] = \frac{1}{\lambda}$$

$$Q = \sum (x_i - \frac{1}{\lambda})^2$$

$$\frac{\partial Q}{\partial \lambda} = 0 \quad \Rightarrow$$

(2) Ein statistisches Modell

$$X_i = g_i(\theta) + \epsilon$$

wobei ϵ ein Meßfehler ist und $\epsilon \sim N(0, \sigma^2)$. Um $\sum \epsilon_i^2$ zu minimieren, wurden wir

minimieren.

(3) Erraten des Bereichs von Zufallszahlen

Zufallszahlen werden im Intervall $[0, \theta]$ erzeugt. Was ist der beste Schätzer für θ aus einer Stichprobe der Größe n, $\{u_1, \ldots, u_n\}$?

$$U_i \sim U[0, \theta]$$

(a) Nach dem GGZ könnte man

$$\widehat{\theta}_a =$$

nehmen. $\widehat{\theta}_a$ ist erwartungstreu und konsistent.

(b) Man könnte auch

$$\widehat{\theta}_b =$$

nehmen. $\widehat{\theta}_b$ ist konsistent, aber nicht erwartungstreu.

$$E[\widehat{\theta}_b] =$$

N.B. $\quad P(\widehat{\theta}_b \leq c) =$

Vergleich der Varianzen

$$V[\widehat{\theta}_a] =$$

$$= \frac{4}{n} V[U_i] = \frac{\theta^2}{3n}$$

$$V[\widehat{\theta}_b] = E[\widehat{\theta}_b^2] - E[\widehat{\theta}_b]^2$$

$$= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2$$

$$=$$

MSE von $\widehat{\theta}_b$ ist

$$\frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{n}{n+1}\theta - \theta\right)^2 =$$

Besser wird es, den Schätzer

$$\widehat{\theta}_c =$$

zu nehmen. $\widehat{\theta}_c$ ist erwartungstreu, konsistent und hat die Varianz

Für n= 10, sind die MSEs

$$\frac{\theta^2}{30}, \quad \frac{\theta^2}{66}, \quad \frac{\theta^2}{120}$$

Ist $\widehat{\theta}_c$ der beste Schätzer überhaupt?