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 iPhone?

Welcher Prozentsatz der Wähler unterstützt die SPD?

Was war die Inflationsrate für April?

Wieviel schneller fährt der neue Mercedes?

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

$$E\left[\widehat{\theta}\right] = \theta$$

Beispiel 1

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

Beispiel 2

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

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

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

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

$$E\left[\frac{n}{\sum X_i}\right] = \int_{X_1} \ldots \int_{X_n} \frac{n}{\sum X_i} \lambda^n e^{-\lambda \sum x_i} \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

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)

$$n \to \infty \Rightarrow P(|\widehat{\theta}_n - \theta| > \epsilon) \to 0$$

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)

$$MSE(\theta, \widehat{\theta}) = E[(\widehat{\theta} - \theta)^2]$$

(Mean Square Error)

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

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

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

$$MSE = Varianz + Bias^2$$

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

$$MSE = E[(\widehat{\theta} - \theta)^2]$$

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.

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

$$\Rightarrow \sum_{i=1}^{n} (x_i - g_i(\theta)) \frac{\partial g_i}{\partial \theta} = 0$$

Beispiele
(1) Exponentialverteilung

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

$$\theta = \lambda$$

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

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

$$\frac{\partial Q}{\partial \lambda} = 0 \quad \Rightarrow \quad -\frac{2}{\lambda^2} \sum \left(x_i - \frac{1}{\lambda} \right) = 0$$

$$\frac{n}{\lambda} = \sum x_i \quad \Rightarrow \quad \hat{\lambda} = \frac{n}{\sum x_i}$$

(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

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

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

$$\widehat{\theta}_a = \frac{2}{n} \sum_{i=1}^{n} u_i$$

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

(b) Man könnte auch

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

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

$$E[\widehat{\theta}_b] = \frac{n}{n+1}\theta$$

$$\text{N.B.} \quad P(\widehat{\theta}_b \leq c) = \left(\frac{c}{\theta}\right)^n$$

Vergleich der Varianzen

$$V[\widehat{\theta}_a] = \left(\frac{2}{n}\right)^2 V\left[\sum_{i=1}^{n} U_i\right]$$

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

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

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

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

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

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

Besser wird es, den Schätzer

$$\widehat{\theta}_c = \frac{n+1}{n} \max\{u_i\}$$

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

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

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?

Millionen Zugvögel sterben jedes Jahr in Nordamerika (SZ 26.4.20012)

"Diese Zahl haben Wissenschaftler der University of California ermittelt indem sie verendete Vögel an 38 Masten zählten und auf alle 84000 Einrichtungen in Kanada und den USA hochrechneten."

"Besonders tückisch sind die höchsten zwei Prozent der Masten. An ihnen kommt es zu mehr als zwei Drittel aller Kollisionen."

Mieses Image belastet die Bistümer in Bayern (AZ 6.5.2003)

”Bei der größten gesellschaftspolitischen Online-Umfrage Perspective Deutschland der Unternehmensberatung McKinsey...”

”Dramatisch klingen McKinseys demoskopische Werte aus 350,000 Rückmeldungen...”

”Die Diözese Augsburg landete tief unten auf der Skala: 27 Prozent der Teilnehmer aus ihrer Region haben in sie wenig oder überhaupt kein Vertrauen.”

”Der Sprecher des Augsburger Bischofs Dammertz zweifelt zwar daran, ob eine Online-Umfrage für die deutsche Bevölkerung wirklich repräsentativ sein kann... aber sehr beunruhigend findet er die Daten schon.”

4.2.2 Maximum Likelihood Schätzer (ML)

Das Likelihood einer vorliegenden Stichprobe

$$\underline{x} = (x_1, x_2, \ldots, x_n)$$

unter einem bestimmten Modell mit multivariaten Dichte f und Parametervektor $\underline{\theta}$ ist

$$L(\underline{x}; \underline{\theta}) = f_{X_1,\ldots,X_n}(\underline{x}; \underline{\theta})$$

Likelihood ist eine Funktion von $\underline{\theta}$ mit "Parametern" $\underline{X}$. Die gemeinsame Dichte $f_{\underline{X}}$ ist eine Funktion von $\underline{X}$ mit Parametern $\underline{\theta}$.

Mit $f_{\underline{X}}$ geht es um die Wahrscheinlichkeit von Daten gegeben das Modell:

$$f_{\underline{X}}(\underline{x}; \underline{\theta}) = W(Daten|Modell)$$

Mit $L(\underline{x}; \underline{\theta})$ geht es um das Liklihood des Modells gegeben die Daten:

$$L(\underline{x}; \underline{\theta}) = L(Modell|Daten)$$

ML Schätzer maximieren das Likelihood.
(Vorsicht: Likelihoods sind keine Wahrscheinlichkeiten!)

Beispiele

(1) Exponentialverteilung

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

$$L(x_1, x_2, \ldots, x_n) = \lambda^n e^{-\lambda \sum x_i}$$

Ein ML Schätzer maximiert L oder, der Einfachheit halber,

$$\log L$$

was auch mathematisch gleich ist.

$$\log L = n \log \lambda - \lambda \sum x_i$$

$$\frac{\partial \log L}{\partial \lambda} = 0$$

$$\Rightarrow \frac{n}{\lambda} - \sum x_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum x_i}$$

(2) Krankheitsdauer

Angenommen

$$T \sim E(\lambda)$$

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

Aus einer Studie haben wir die Beobachtungen $t_1, \ldots, t_j$ und $(n-j)$ Fälle mit $T_i > a_i$

$$L(\underline{t}, \lambda) = \lambda^j e^{-\lambda \sum_{i=1}^{j} t_i} \prod_{i=j+1}^{n} e^{-\lambda a_i}$$

$$\log L = j \log \lambda - \lambda \sum_{i=1}^{j} t_i - \lambda \sum_{i=j+1}^{n} a_i$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{j}{\lambda} - \sum_{i=1}^{j} t_i - \sum_{i=j+1}^{n} a_i = 0$$

$$\Rightarrow \quad \hat{\lambda}_1 = \frac{j}{\sum_{i=1}^{j} t_i + \sum_{i=j+1}^{n} a_i}$$

Falls die Fälle mit unbekannten Endpunkten vernachläßigt werden, bekommen wir

$$\widehat{\lambda}_2 = \frac{j}{\sum_{i=1}^{j} t_i}$$

Oder wir nehmen an, dass $t_i = a_i$:

$$\widehat{\lambda}_3 = \frac{n}{\sum_{i=1}^{j} t_i + \sum_{i=j+1}^{n} a_i}$$

Beide Schätzer sind offensichtlich schlechter als $\widehat{\lambda}_1$, weil sie die verfügbare Information nicht voll ausnutzen.

4.2.3 Eigenschaften von ML Schätzern

ML Schätzer sind nicht immer erwartungstreu. z.B.

$$X \sim N(\mu, \sigma^2)$$

$$\log L = l(\mu, \sigma) = -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (X_i - \mu)^2$$

Die ML Schätzer sind $\widehat{\mu} = \bar{X}$ und

$$\widehat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Es kann gezeigt werden, dass ML Schätzer konsistent sind.