Beispiele von Binomial KIs

(i) (Die Zeit 13. Februar, 2003 s30)
Paare sind auf Flughäfen und Bahnhöfen beobachtet worden (vermutlich in Deutschland, da ein Bochumer Wissenschaftler die Daten gesammelt hat). Aus 124 küssenden Paaren haben 80 den Kopf beim Küssen nach rechts und 44 den Kopf nach links gedreht.

$$\widehat{p} =$$

95% KIs

$$(a) \quad \text{Approx } p = 0.5 \quad (0.557, 0.733)$$

$$(b) \quad \text{"Plug-in" } p = \widehat{p} \quad (0.561, 0.729)$$

$$(c) \quad \text{Normal genau} \quad (0.558, 0.724)$$

(ii) Sonntagsfrage in der politischen Meinungsforschung z.B. 1000 Befragte CSU 55%, SPD 25%, FDP 3%

$$(a) \quad (0.519, 0.581) \quad (0.219, 0.281) \quad (-0.001, 0.061)$$

$$(b) \quad (0.519, 0.581) \quad (0.223, 0.277) \quad (0.019, 0.041)$$

$$(c) \quad (0.519, 0.581) \quad (0.224, 0.278) \quad (0.021, 0.043)$$

Browser Statistiken

Schätzungen von HitsLink, netapplications.com (April 2006):
Internet Explorer 83,88%
Mozilla Firefox 10,68%

Schätzungen von OneStat.com (Mai 2006):
Internet Explorer 85,17%
Mozilla Firefox 11,79%

“Wie bei anderen Erhebungen erklären sich die Unterschiede aber nicht nur aus dem Zufallsfehler der Stichprobe, sondern auch durch unterschiedliche Erhebungsmethoden. Während Net Applications die Statistiken von 40.000 besuchten Webseiten auswertet, gibt OneStat an, zwei Millionen Pageimpressions in 20.000 Webseiten-Besuchen aufzudröseln.” (macnews.de 12.5.06)

Welche KIs würden Sie berechnen?

4.2.7 Definition eines KIs

Sei X eine Statistik mit Verteilungsfunktion

$$F(x, \theta), \quad \theta \in \Theta$$

Eine Abbildung, C, die jedem möglichen Beobachtungsergebnis x eine Menge $C(x) \in \Theta$ zuordnet, heißt ein Konfidenzbereich für θ zum Niveau $100(1-\alpha)\%$ wenn

$$\inf_{\theta \in \Theta} P_\theta(x : c(x) \ni \theta) \geq 1 - \alpha$$

Wenn es ein Pivot (Drehpunkt) gibt, wie bei der Normalverteilung, sind KIs leicht zu berechnen, im allgemeinen nicht.

4.2.8 Assymptotische ML Konfidenz Intervalle

ML Schätzer sind assymptotisch normalverteilt.

Bedingungen

1. Der Parameterraum Ω hat endliche Dimension, ist geschlossen und kompakt und der wahre Wert θ liegt innerhalb Ω.

2. Die Wahrscheinlichkeitsverteilungen für zwei verschiedene Werte von θ sind verschieden.

3. $\frac{\partial}{\partial\theta}$, $\frac{\partial^2}{\partial\theta^2}$, $\frac{\partial^3}{\partial\theta^3}$, von $l(x;\theta)$ existieren fast sicher in der Nachbarschaft des wahren Wertes. Weiterhin in solch einer Nachbarschaft ist $\frac{1}{n}\left|\frac{\partial^3 l}{\partial\theta^3}\right| \leq$ eine Funktion von X deren Erwartungswert existiert.

4. $I(\theta) =$ ist endlich und > 0 in der Nachbarschaft des wahren Wertes.

Sei θ der für uns interessante Parameter der (glatten) Dichte f. Die Loglikelihoodfunktion ist

$$l(\theta) =$$

für eine u.i.v. Stichprobe der Größe n.
Der ML Schätzer, $\widehat{\theta}$ muß ein stationärer Punkt von $l(\theta)$ sein:

$$l'(\widehat{\theta}) = 0$$

Aus einer Taylor Entwicklung von $l'(\widehat{\theta})$ um den unbekannten wahren Wert θ_0 haben wir

$$l'(\widehat{\theta}) \approx$$

$$\Rightarrow (\widehat{\theta} - \theta_0) \approx$$

Betrachten wir

$$E[l'(\theta)] = \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta} \log f(X_i|\theta)\right]$$

$$= \sum_{i=1}^{n} \int \left[\frac{\partial}{\partial\theta} \log f(x|\theta)\right] f(x|\theta) \mathrm{d}x$$

Es gilt

$$\int \left[\frac{\partial}{\partial\theta} \log f(x_i|\theta)\right] f(x|\theta)\mathrm{d}x = \int \frac{\partial}{\partial\theta} f(x|\theta)\mathrm{d}x \qquad (2)$$

und wenn wir sagen dürfen, dass

$$\int \frac{\partial}{\partial\theta} f(x|\theta)\mathrm{d}x = \frac{\partial}{\partial\theta} \int f(x|\theta)\mathrm{d}x$$

ist

$$\int [\frac{\partial}{\partial\theta} \log f(x|\theta)] f(x|\theta)\mathrm{d}x = 0$$

und

$$E[l'(\theta)] = 0$$

$$V[l'(\theta)] = \sum_{i=1}^{n} E[(\frac{\partial}{\partial\theta} \log f(X_i|\theta))^2]$$

$$= nI(\theta) \quad \text{per Definition}$$

$I(\theta)$ wird die Fisher-Information genannt. (Man findet auch den Ausdruck ”Scorefunktion” für $U_\theta = \frac{\partial}{\partial\theta} \log f(x|\theta)$).

Jetzt betrachten wir den Nenner von (1)

$$l''(\theta) = \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} \log f(X_i|\theta)$$

Wir zeigen, dass

$$E[l''(\theta)] = -nI(\theta)$$

$$E[l''(\theta)] = \frac{\partial}{\partial\theta} E[l'(\theta)]$$

$$E[l'(\theta)] = \int \left[\frac{\partial}{\partial\theta} \log f(x|\theta)\right] f(x|\theta)\mathrm{d}x = 0$$

$$\Rightarrow \text{ (hier wird (2) wieder benutzt)}$$

$$\int \left[\frac{\partial^2 \log f(x|\theta)}{\partial\theta^2}\right] f(x|\theta)\mathrm{d}x + \int \left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right]^2 f(x|\theta)\mathrm{d}x$$

$$= 0$$

$$\Rightarrow E[l''(\theta)] = -V[l'(\theta)]$$

$$= -nI(\theta)$$

Für n groß, nach dem GGZ

$$\frac{1}{n} l''(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \log f(x_i, \theta_0) \to -I(\theta_0)$$

und daher aus (1)

$$\sqrt{n}(\widehat{\theta} - \theta_0) \approx \frac{n^{-1/2} l'(\theta_0)}{I(\theta_0)}$$

und

$$E[\sqrt{n}(\widehat{\theta} - \theta_0)] = 0$$

$$V[\sqrt{n}(\widehat{\theta} - \theta_0)] \approx \frac{1}{I(\theta_0)}$$

$$V[(\widehat{\theta} - \theta_0)] \approx \frac{1}{nI(\theta_0)}$$

$$l'(\theta_0) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(x_i|\theta)|_{\theta=\theta_0}$$

ist eine Summe von u.i.v. Zufallsvariablen. Nach dem ZGS

$$\sqrt{nI(\theta_0)}(\widehat{\theta} - \theta_0)$$

Beispiel (1)

X hat die Dichte

$$f(x) = (\theta + 1)x^{\theta} \qquad 0 \leq x \leq 1$$

eine Beta Verteilung mit $a = \theta + 1$ und $b = 1$

$$E[X] =$$

$$l(x;\theta) = n \log(\theta + 1) + \theta \sum \log x_i$$

$$\widehat{\theta} =$$

$$\frac{\partial}{\partial\theta} \log f = \frac{1}{\theta + 1} + \log x$$

$$\Rightarrow I(\theta) = E[(\frac{1}{\theta + 1} + \log X)^2]$$

$$\frac{\partial^2}{\partial\theta^2} \log f = -\frac{1}{(\theta + 1)^2}$$

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

$$\Rightarrow V[\widehat{\theta}] = \frac{(\theta+1)^2}{n}$$

Assymptotisch

$$\widehat{\theta} = (-1 - \frac{n}{\sum \log x_i}) \sim$$

Im allgemeinen gilt

$$P(-1.96 < \sqrt{nI(\theta)}(\widehat{\theta} - \theta) < 1.96) = 0.95$$

$$P(|\widehat{\theta} - \theta| < \frac{1.96}{\sqrt{n}}(\theta + 1)) = 0.95$$

$$P_{\widehat{\theta}}(\qquad\qquad\qquad < \theta < \qquad\qquad\qquad) = 0.95$$

Ein 95% KI für θ wenn wir $\widehat{\theta}$ mit dem beobachteten Wert ersetzen.

Beispiel (2)

$\{X_i\}$ u.i.v. von einer Rayleigh Dichte

$$f(x) = \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} \qquad x \geq 0$$

(Die Rayleigh Verteilung folgt aus Y, W u.i.v. $\sim N(0, \sigma^2)$. Nach der Transformation in Polarkoordinaten (r, ϕ) hat r eine Rayleigh Verteilung.)

$$l(x;\theta) =$$

$$\frac{\partial l}{\partial \theta} = 0 \quad \Rightarrow \quad -\frac{2}{\theta} + \frac{1}{\theta^3} \sum x^2 \quad \Rightarrow \quad \hat{\theta} = \sqrt{\frac{\sum x^2}{2n}}$$

$$\frac{\partial}{\partial \theta} \log f =$$

$$\frac{\partial^2}{\partial \theta^2} \log f =$$

$$E[X^2] = 2\theta^2 \quad \Rightarrow \quad I(\theta) = \frac{4}{\theta^2}$$

und assymptotisch

$$\hat{\theta} = \sqrt{\frac{\sum x^2}{2n}} \sim N\left(\theta, \frac{\theta^2}{4n}\right)$$

4.2.9 Beste Schätzer

Um Resultate über Optimalität zu erzielen, beschränken wir uns auf erwartungstreue Schätzer für reguläre Modelle. Ein einparametriges Modell heißt regulär wenn

1. Θ ist ein offenes Intervall in $\mathbb{R}$

2. Die Likelihoodfunktion ist auf X x Θ strikt positiv und nach θ differenzierbar. Somit existiert die Scorefunktion

$$U_\theta(x) = \frac{\partial}{\partial\theta} \log f(x|\theta)$$

3. Für jedes $\theta \in$ Θ existiert $V_\theta[U_\theta] = I(\theta)$ und ist nicht 0. Es gilt die Vertauschungsrelation

Die Cramér-Rao Ungleichung

Gegeben sei ein reguläres statistisches Modell $P_X(x;\theta)$, $\theta \in \Theta$, ein Intervall. Dann gilt für jeden erwartungstreuen Schätzer, $\tilde{\theta}$ von θ, dass

$$Var_\theta[\tilde{\theta}] \geq$$

Beweis

$$E[\tilde{\theta}] = \theta$$

und für die Scorefunktion $U_\theta = \frac{\partial}{\partial\theta} \log f(x|\theta)$ gilt

$$E[U_\theta] = 0$$

Deshalb haben wir

$$Cov[\tilde{\theta}, U_\theta] = E[\tilde{\theta}, U_\theta]$$

$$E[\tilde{\theta}, U_\theta] = \int_X \tilde{\theta} \frac{\partial}{\partial\theta} \log f(x|\theta) f(x|\theta) \mathrm{d}x$$

$$= \int_X \tilde{\theta} \frac{\partial}{\partial\theta} f(x|\theta) \mathrm{d}x$$

$$= \frac{\partial}{\partial\theta}\int_X \tilde{\theta} f(x|\theta)\mathrm{d}x = \frac{\partial}{\partial\theta} E[\tilde{\theta}] = 1$$

Betrachten wir die Varianz von der Funktion $(\tilde{\theta} - U_\theta / I(\theta))$

$$0 \le V[\tilde{\theta} - U_\theta / I(\theta)]$$

$$=$$

$$= V[\tilde{\theta}] - 1/I(\theta)$$

Q.E.D.

Cramér-Rao liefert uns eine untere Grenze für die Varianz von erwartungstreuen Schätzern für reguläre Modelle (z.B. es gilt nicht für das Beispiel mit der Gleichverteilung). Es sagt uns nicht, wie wir einen optimalen Schätzer finden. Deshalb sprechen wir von der Effizienz eines Schätzers:

Gegeben einen erwartungstreuen Schätzer, $\tilde{\theta}$, eines Parameters für ein reguläres Modell, dann ist die Effizienz von $\tilde{\theta}$

$$eff(\tilde{\theta}) =$$