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:

$$f_{\underline{X}}(\underline{x}; \underline{\theta}) =$$

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

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

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.

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

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

$$\Rightarrow$$

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

(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 $T_i > a_i$

$$L(\underline{t}, \lambda) =$$

$$\log L =$$

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

$$\Rightarrow \quad \hat{\lambda}_1 =$$

Falls die Fälle mit unbekannten Endpunkten vernachläßigt werden, bekommen wir

$$\widehat{\lambda}_2 =$$

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

$$\widehat{\lambda}_3 =$$

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.

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

$$\log L =$$

$$\frac{\partial l}{\partial \mu} =$$

$$\frac{\partial l}{\partial \sigma} =$$

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

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

Es kann gezeigt werden, dass ML Schätzer konsistent sind.

4.2.4 Weitere ML Beispiele

(1) Gleichverteilung

$$f(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

$$L(\underline{x}; a, b) = \left(\frac{1}{b-a}\right)^n$$

$$\text{u.d.B.} \quad b \geq \max x_i \quad \text{und} \quad a \leq \min x_i$$

Das Maximum kann man mit Lagrange Methoden finden:

$$\hat{a} = \qquad\qquad \text{und} \quad \hat{b} =$$

(2) Negativ Binomial Verteilung

$$p(n) = \binom{n-1}{r-1} p^r q^{n-r}$$

Daten $n_1, n_2, \ldots, n_m$

(a) r bekannt

$$L = \prod \binom{n_i - 1}{r-1} p^r q^{n_i - r}$$

$$\log L = \sum \log \binom{n_i - 1}{r-1} + r \sum \log p + \sum (n_i - r) \log(1-p)$$

$$\frac{\partial \log L}{\partial p}$$

$$\Rightarrow \hat{p} =$$

(b) r unbekannt
r muß ganzzählig sein und $1 \le r \le \min n_i$

Wir berechnen max L für $r = 1, \ldots, \min n_i$ und nehmen den größten.

4.2.5 Funktionen von Zufallsvariablen und ML Schätzer

$$X \sim f_X(x)$$

$$Y = g(X) \qquad 1-1$$

mit Dichte

$$h_Y(y) =$$

Aus einer Stichprobe haben wir $(y_1, \ldots, y_n)$

$$L(\underline{y}, \underline{\theta}) = \prod h_Y(y_i, \underline{\theta})$$

$$= \prod f_X(g^{-1}(y_i)) \frac{\partial g^{-1}}{\partial y_i}$$

$$\log L(\underline{y}, \underline{\theta}) = \sum \log f_X(g^{-1}(y_i)) + \sum \log \frac{\partial g^{-1}}{\partial y_i}$$

$$\frac{\partial log L(\underline{y}, \underline{\theta})}{\partial \theta} = 0 \text{ genau dann wenn}$$

$$\frac{\partial log L(\underline{x}, \underline{\theta})}{\partial \theta} = 0 \quad \text{mit} \quad \{x_i\} = \{(g^{-1}(y_i))\}$$

$\Rightarrow$ ML Schätzer für θ bleiben durch eine Transformation unverändert.

4.2.6 ML Intervallschätzer

Konfidenz Intervalle (KIs)

Bis jetzt haben wir nur Punktschätzer besprochen (d.h. es wird nur einen Schätzwert angegeben. Unsere Unsicherheit um den Wert wird nicht berichtet.) Jetzt führen wir Konfidenzintervalle ein — Intervalle, die mit einer bestimmten Wahrscheinlichkeit den wahren, unbekannten Wert enthalten.

Falls wir die Verteilung unseres Schätzers kennen, können wir vor der Ziehung Vorhersageintervalle berechnen:

$$P_{\widehat{\theta}}(a < \widehat{\theta} < b) = 1 - \alpha$$

$$\text{z.B.} \quad \bar{X} \sim N(\mu, \sigma^2/n)$$

$$P_{\bar{X}}(\qquad\qquad < \bar{X} < \qquad\qquad) = 0.95$$

Diese Wahrscheinlichkeit kann umgeschrieben werden:

$$P_{\bar{X}}(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

aber es handelt sich als Wahrscheinlichkeitsintervall immer noch um $\bar{X}$ und nicht um μ.

Wenn wir den Stichprobenwert $\bar{x}$ für $\bar{X}$ einsetzen, bekommen wir ein sogenanntes Konfidenzintervall, und wir sagen, dass

$$(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$$

ein 95% Konfidenzintervall für μ ist.

Wenn wir viele Stichproben ziehen und Konfidenzintervalle aus jeder berechnen, werden wir finden, dass ungefähr 95% der Intervalle den wahren Wert μ enthalten werden.

Falls σ unbekannt ist aber wir $X \sim N(\mu, \sigma^2)$ annehmen dürfen, dann können wir mit $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ arbeiten.

Weitere Beispiele von KIs

(1) Gleichverteilung

$$X \sim G(0, b)$$

Wir beobachten $x = 7$, was ist eine 95% KI für b?

$$P_X(0 < X < 0.95b) = 0.95 \quad \Rightarrow \quad \left(\frac{x}{0.95}, \infty\right)$$

$$P_X(0.05b < X < b) = 0.95 \quad \Rightarrow \quad (x, 20x)$$

$$P_X(0.025b < X < 0.975b) = 0.95 \quad \Rightarrow \quad \left(\frac{x}{0.975}, 40x\right)$$

Meistens versuchen wir, das kurzeste Intervall zu finden. In diesem Fall haben wir die Intervalle

$$(7.37, \infty) \qquad (7, 140) \qquad (7.18, 280)$$

und wir würden das zweite nehmen.

(2) Proportionen — Binomial Verteilung

$$X \sim B(N, p)$$

Für N groß gilt

$$Y = \frac{X}{N} \sim N\left(p, \frac{p(1-p)}{N}\right)$$

und ein 95% Vorhersageintervall für Y wäre

$$P_Y\left(p - 1.96\sqrt{\frac{p(1-p)}{N}} < Y < p + 1.96\sqrt{\frac{p(1-p)}{N}}\right)$$

$$P_Y\left((Y-p)^2 < 1.96^2 \quad \frac{p(1-p)}{N}\right) = 0.95$$

Y und N sind bekannt. Für welche Werte von p gilt

$$(Y-p)^2 < 1.96^2 \quad \frac{p(1-p)}{N} \quad ?$$

Drei Möglichkeiten
(a) Als Approximation für die Varianz setzen wir $p = 0.5$ und

$$\frac{p(1-p)}{N} \leq \frac{1}{4N} \qquad \forall \quad p \in [0,1]$$

$$\Rightarrow P_Y(Y - 1.96\frac{1}{\sqrt{4N}} < p < Y + 1.96\frac{1}{\sqrt{4N}}) \geq 0.95$$

(b) “Plug-in Schätzer”

Wir benutzen $\frac{y(1-y)}{N}$ als Approximation für die Varianz

$$P_Y\left(Y - 1.96\sqrt{\frac{y(1-y)}{N}} < p < Y + 1.96\sqrt{\frac{y(1-y)}{N}}\right)$$

$$\approx 0.95$$

Dieser Schätzer ist besser als (a). Er ist am Rand (p nahe 0 oder 1) schlecht und die Coverage (Überdeckung) ist wegen der Normalapproximation unregelmäßig.

(c) Eine genauere Methode

Lösen wir die quadratische Gleichung

$$(y-p)^2 = 1.96^2 \quad \frac{p(1-p)}{N}$$

um Wurzeln p_1 und p_2 zu finden

$$\Rightarrow P_Y(p_1 < p < p_2) = 0.95$$

(Für eine eingehende Studie der vielen Alternativen, siehe Brown L, Cal, T.T., DasGupta, A. (2001) Interval Estimation for a Binomial Proportion. Statistical Science 16: 101-133)