Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

Heinrich-Heine-Universität Düsseldorf

9. Dezember 2010

Idee

- Parameterschätzungen geben keine Auskunft über ihre Genauigkeit
- Jeder Schätzer ist selbst eine Zufallsvariable
- Alternative: Intervallschätzung
- Es gibt in der Regel kein Intervall, in welchem der zu schätzende Parameter mit Sicherheit liegt
- Eine Irrtumswahrscheinlichkeit α muss vorgegeben werden
- $1 - \alpha$ heißt dann *Konfidenzniveau*
- Bei der Schätzung eines Konfidenzintervalls sind die Randpunkte selber Zufallsvariable

Intervalle

- Für Zahlen $a \leq b$ ist

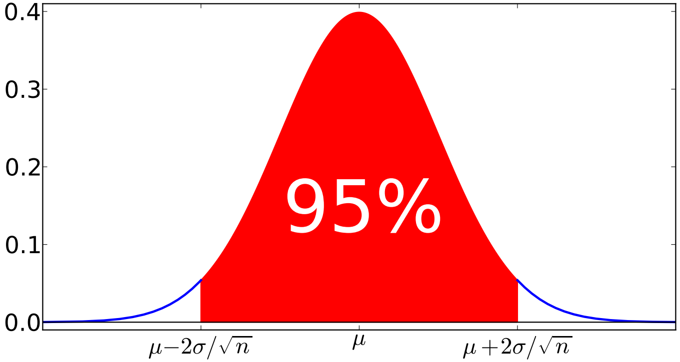
$$[a, b] = \{x \mid a \leq x \leq b\}$$

das (abgeschlossene) Intervall mit den Randpunkten a und b

- Speziallfälle

$$(-\infty, b] = \{x \mid x \leq b\}$$
$$[a, \infty) = \{x \mid x \geq a\}$$

3σ-Regel



Schätzung eines Konfidenzintervalls für den Erwartungswert bei bekannter Varianz mittels der 3σ-Regel

- $X_1, \ldots, X_n$ seien normalverteilt gemäß $N(\mu, \sigma^2)$ für bekanntes σ und unbekanntes μ
- Für μ soll ein Konfidenzintervall zum Konfidenzniveau 95% geschätzt werden.
- Verwende dazu das arithmetische Mittel

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

- 3σ-Regel

$$P\left(\mu - \frac{2\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + \frac{2\sigma}{\sqrt{n}}\right) = 0.95$$

- Beachte die Gleichwertigkeit der Ungleichungen

$$\overline{X} \leq \mu + \frac{2\sigma}{\sqrt{n}} \quad \text{und} \quad \mu \geq \overline{X} - \frac{2\sigma}{\sqrt{n}}$$

Schätzung eines Konfidenzintervalls für den Erwartungswert bei bekannter Varianz, Fortsetzung

- Damit sieht die 3σ-Regel so aus

$$P\left(\overline{X} - \frac{2\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{2\sigma}{\sqrt{n}}\right) = 0.95$$

- Also ist folgendes Intervall ein Konfidenzintervall zum Konfidenzniveau 95%

$$\left[\overline{X} - \frac{2\sigma}{\sqrt{n}}, \quad \overline{X} + \frac{2\sigma}{\sqrt{n}}\right]$$

Beispiel: Roggenpflanzen

- Gesunde Roggenpflanzen einer bestimmten Art sind im Mittel 102.5 *cm* lang, wobei die Länge um 7 *cm* streut. Die Länge sei normalverteilt
- Durch Umwelteinflüsse änderte sich die mittlere Halmlänge, die Streuung aber nicht
- Die folgenden Längen wurden gemessen

96.62	94.91	85.05	101.61	109.55
93.05	97.86	96.66	95.08	98.87

- Arithmetisches Mittel der Daten

$$\overline{x} = 96.93$$

Roggenpflanzen, Fortsetzung

- $\frac{2\sigma}{\sqrt{n}} = \frac{14}{\sqrt{10}} = 4.43$
- Also ergibt sich das folgende Konfidenzintervall zum Konfidenzniveau 95%

$$[92.50, 101.36]$$

Interpretation

- Die Frage:

  Mit welcher Wahrscheinlichkeit liegt μ zwischen 92.50 und 101.36?

  macht keinen Sinn
- μ ist der wahre Wert, also nicht zufallsbehaftet
- "Konfidenzniveau $1 - \alpha = 0.95$" bedeutet

  Das verwendete Verfahren wird in 95 von 100 Fällen das richtige Ergebnis liefern

Definition

- Es sei Θ eine Menge von Parameterwerten. Zu jedem Parameterwert $\theta \in \Theta$ gebe es eine Verteilung P_θ
- Von der Zufallsvariablen X sei bekannt, dass ihre Verteilung gleich einem der P_θ ist. Für dieses θ soll ein Konfidenzintervall geschätzt werden
- Ein Intervall $[G_u, G_o]$ mit der Eigenschaft

  $$P_\theta(G_u \leq \theta \leq G_o) = 1 - \alpha$$

  heißt *Konfidenzintervall* für den Parameter θ zum Konfidenzniveau $1 - \alpha$
- Übliche Konfidenzniveaus sind 90%, 95% und 99%
- G_u und G_o sind Zufallsvariable. Man bezeichnet sie als untere und obere *Vertrauensgrenze*

Zurück zum Beispiel

Das Konfidenzintervall zum Konfidenzniveau 95% für den Erwartungswert einer normalverteilten Zufallsvariable mit bekannter Varianz σ^2 ist

$$[G_u, G_o]$$

mit

$$G_u = \overline{X} - \frac{2\sigma}{\sqrt{n}} \qquad G_o = \overline{X} + \frac{2\sigma}{\sqrt{n}}$$

Dabei ist n der Stichprobenumfang

Quantile

- Φ die Verteilungsfunktion der Standardnormalverteilung
- Die Zahl q_α mit $\Phi(q_\alpha) = \alpha$ heißt α-*Quantil* der Standardnormalverteilung. Die wichtigsten Quantile der Standardnormalverteilung sind tabelliert

$\Phi(u)$	70%	80%	90%	95%	97.5%	99%	99.5%
u	0.524	0.842	1.282	1.645	1.960	2.326	2.576

- Z. B. ist das 0.975-Quantil gleich 1.960, also näherungsweise gleich 2
- Umrechnungsformel

$$q_\alpha = -q_{1-\alpha}$$

- Beispiel

$$q_{0.05} = -q_{0.95} = -1.645$$

Einseitige Konfidenzintervalle

- Bei einigen Fragestellungen ist statt eines Intervalls nur eine obere oder eine untere Vertrauensgrenze für den Erwartungswert von Interesse
- Variation des Beispiels der Roggenpflanzen: Schätze Intervall der Form

$$(-\infty, G_o]$$

- Bei gleichem Konfidenzniveau $1-\alpha$ ist G_o bei der Schätzung eines einseitigen Konfidenzintervalls etwas kleiner als bei einem zweiseitigen

Schätzung eines einseitigen Konfidenzintervalls für den Erwartungswert bei bekannter Varianz

- $X_1, \ldots, X_n$ seien normalverteilt gemäß $N(\mu, \sigma^2)$ für bekanntes σ und unbekanntes μ
- Für μ soll ein einseitiges Konfidenzintervall $(-\infty, G_o]$ zum Konfidenzniveau $1 - \alpha$ geschätzt werden
- Verwende dazu das arithmetische Mittel

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

- Seine Standardisierung ist

$$Y = \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu)$$

Schätzung eines einseitigen Konfidenzintervalls für den Erwartungswert bei bekannter Varianz, Fortsetzung

- $q_{1-\alpha}$ sei das $(1-\alpha)$-Quantil der Standardnormalverteilung
- Y ist standardnormalverteilt, also

$$P(Y \geq -q_{1-\alpha}) = 1 - \Phi(-q_{1-\alpha}) = \Phi(q_{1-\alpha}) = 1 - \alpha$$

- Für $\overline{X}$ heißt das

$$P\left(\frac{\sqrt{n}}{\sigma}(\overline{X} - \mu) \geq -q_{1-\alpha}\right) = 1 - \alpha$$

- Man stellt μ frei

$$P\left(\mu \leq \overline{X} + \frac{\sigma \cdot q_{1-\alpha}}{\sqrt{n}}\right) = 1 - \alpha$$

- Das Konfidenzintervall ist $\left(-\infty, \overline{X} + \frac{\sigma \cdot q_{1-\alpha}}{\sqrt{n}}\right]$

Beispiel: Roggenpflanzen

- Zurück zu den Roggenpflanzen: Es soll nur noch untersucht werden, ob die Umwelteinflüsse die Halmlänge *verringern*
- Die Daten waren

$$\overline{x} = 96.93 \qquad \sigma = 7$$
$$n = 10 \qquad 1 - \alpha = 0.95$$

- Das Quantil ist

$$q_{0.95} = 1.645$$

- Damit erhält man als obere Vertrauensgrenze für das einseitige Konfidenzintervall

$$\overline{x} + \frac{\sigma \cdot q_{1-\alpha}}{\sqrt{n}} = 96.93 + \frac{7 \cdot 1.645}{\sqrt{10}} = 100.57$$

- Zum Vergleich: Die obere Vertrauensgrenze des zweiseitigen Konfidenzintervalls zum selben Konfidenzniveau war 101.36

Tabelle: Konfidenzintervall für den Erwartungswert bei bekannter Varianz

- zweiseitig zum Konfidenzniveau $1-\alpha$

$$\overline{X} - \frac{\sigma \cdot q_{1-\alpha/2}}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{\sigma \cdot q_{1-\alpha/2}}{\sqrt{n}}$$

- einseitig, obere Vertrauensgrenze zum Konfidenzniveau $1-\alpha$

$$\mu \leq \overline{X} + \frac{\sigma \cdot q_{1-\alpha}}{\sqrt{n}}$$

- einseitig, untere Vertrauensgrenze zum Konfidenzniveau $1-\alpha$

$$\overline{X} - \frac{\sigma \cdot q_{1-\alpha}}{\sqrt{n}} \leq \mu$$

Wichtige Werte der Quantile

Beim zweiseitigen Konfidenzintervall wird $q_{1-\alpha/2}$ verwendet

$$1-\alpha = 0.66 \qquad 1-\frac{\alpha}{2} = 0.83 \qquad q_{0.83} = 0.95 \cong 1$$

$$1-\alpha = 0.95 \qquad 1-\frac{\alpha}{2} = 0.975 \qquad q_{0.975} = 1.96 \cong 2$$

$$1-\alpha = 0.99 \qquad 1-\frac{\alpha}{2} = 0.995 \qquad q_{0.995} = 2.56 \cong 3$$