Statistik

Parameterschätzung

Einführung

Beobachtete Daten $x_1, \ldots, x_n$ werden aufgefasst als **Realisierungen von i.i.d. Zufallsvariablen** $X_1, \ldots, X_n$.

Für die X_i's nehmen wir eine **Verteilungsfamilie** mit Parameter(n) θ an, z.B. Normalverteilung: $\theta = (\mu, \sigma^2)$.
Grafische Tools zur Überprüfung sind z.B. QQ-Plots.

Grundfrage: Gegeben einer Stichprobe $x_1, \ldots, x_n$, welches ist der "plausibelste" Wert eines unbekannten Parameters θ? Wir versuchen also, θ zu **schätzen** (Punktschätzung).

Die geschätzten Parameter bezeichnen wir mit $\widehat{\theta}$.

Merke: Die geschätzten Parameter $\widehat{\theta}$ ändern von Stichprobe zu Stichprobe ("the data could have been different") und entsprechen in der Regel *nicht* den wahren Parameter θ.

Allgemeines

Solange die Daten nicht realisiert sind, ist ein Schätzer eine Zufallsvariable.

Der sogenannte **Standardfehler** von $\widehat{\theta}$ ist definiert als $\sqrt{\mathrm{Var}(\widehat{\theta})}$.

Ein Schätzer heisst **erwartungstreu** (unbiased), falls

$$\mathbb{E}(\widehat{\theta}) = \theta.$$

Für erwartungstreue Schätzer ist die Varianz gerade ein Mass für die Genauigkeit, denn

$$\mathrm{Var}(\widehat{\theta}) = \mathbb{E}\left((\widehat{\theta} - \theta)^2\right)$$

(erwartete quadratische Abweichung vom wahren Wert θ).

Wir können den Erwartungswert bzw. die Varianz einer Verteilung als Modellparameter auffassen und haben dann die entsprechenden erwartungstreuen Schätzer

$$\widehat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \widehat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu}_X)^2$$

(egal, was die Verteilungsfamilie ist).

Methoden zur Parameterschätzung

Momentenmethode

Idee: Wähle die geschätzten Parameter $\widehat{\theta}$ so, dass gewisse empirische Größen mit den entsprechenden Größen aus dem Modell (der Verteilung) übereinstimmen.

k-tes **Moment einer Zufallsvariable** X ("aus Modell"):

$$\mu_k = \mathbb{E}(X^k),\ k \geq 1.$$

k-tes **empirisches Moment** ("aus Daten"):

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k,\ k \geq 1.$$

Das Moment einer Zufallsvariable hängt von den Parametern ab, d.h.

$$\mu_k = \mu_k(\theta).$$

Wähle nun $\widehat{\theta}$ so, dass folgendes Gleichungssystem erfüllt ist:

$$\begin{aligned}
\mu_1(\widehat{\theta}) &= m_1 \\
\mu_2(\widehat{\theta}) &= m_2 \\
&\vdots \\
\mu_r(\widehat{\theta}) &= m_r
\end{aligned}$$

Die Anzahl Gleichungen entspricht der Anzahl Parameter (z.B. zwei bei der Normalverteilung).

Bsp. Normalverteilung:

$$\begin{aligned}
\mu_1 &= \mu = m_1 \\
\mu_2 &= \sigma^2 + \mu^2 = m_2
\end{aligned}$$

Dies gibt die Lösung:

$$\begin{aligned}
\widehat{\mu} &= m_1 \\
\widehat{\sigma}^2 &= m_2 - m_1^2
\end{aligned}$$

Maximum-Likelihood-Methode

Idee: Wähle die geschätzten Parameter $\widehat{\theta}$ so, dass die beobachteten Daten unter diesem Modell (Verteilung) möglichst plausibel (wahrscheinlich) erscheinen.

Likelihood-Funktion $L(\cdot)$:

Parameter $\longmapsto$ W'keit der beob. Daten unter diesem Parameter

Diskrete Verteilungen: $L(\cdot)$ ist das Produkt der Einzelw'keiten

$$L(\theta) = \prod_{i=1}^{n} p_X(x_i \mid \theta) = p_X(x_1 \mid \theta) \cdots p_X(x_n \mid \theta).$$

Stetige Verteilungen: $L(\cdot)$ ist das Produkt der Dichten

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i \mid \theta) = f_X(x_1 \mid \theta) \cdots f_X(x_n \mid \theta).$$

Merke: Daten $x_1, \ldots, x_n$ sind nach Beobachtung *fix*, man variiert in der Likelihood-Funktion die Parameter!

Der **Maximum-Likelihood-Schätzer** von θ ist

$$\widehat{\theta} = \mathrm{argmax}_\theta L(\theta).$$

Oft ist es einfacher, die **log-Likelihood-Funktion**

$$\ell(\theta) = \log(L(\theta))$$

zu maximieren, denn das Maximum von $L(\cdot)$ und $\ell(\cdot)$ ist an der gleichen Stelle. Also

$$\widehat{\theta} = \mathrm{argmax}_\theta \ell(\theta).$$

Lukas Meier, meier@stat.math.ethz.ch

Statistische Tests

Allgemein

Grundfrage: Ist ein bestimmter Parameterwert θ_0 (z.B. Sollwert) mit den beobachteten Daten "verträglich" oder nicht?

Basierend auf einer Stichprobe soll also ein Entscheid gefällt werden. Spezifiziere hierzu die **Nullhypothese** H_0

$$H_0 : \theta = \theta_0$$

und die **Alternativhypothese** H_A (je nach Fragestellung)

$$\begin{aligned} H_A : \theta &\neq \theta_0 \quad (\text{"zweiseitig"}) \\ \theta &> \theta_0 \quad (\text{"einseitig nach oben"}) \\ \theta &< \theta_0 \quad (\text{"einseitig nach unten"}). \end{aligned}$$

Folgende Fehlerarten sind möglich:

	Belasse H_0	Verwerfe H_0
H_0 wahr	Kein Fehler	Fehler 1. Art
H_0 falsch	Fehler 2. Art	Kein Fehler

Ein statistischer Test kontrolliert gemäß Konstruktion die Wahrscheinlichkeit für einen Fehler 1. Art durch das **Signifikanzniveau** α, d.h.

$$\mathbb{P}(\text{Verwerfe } H_0 \text{ fälschlicherweise}) \leq \alpha.$$

Oft verwendet man $\alpha = 0.05$.

Wenn wir die Nullhypothese nicht verwerfen können, ist sie damit *nicht* bewiesen ("absence of evidence is not evidence of absence").

Aufbau eines statistischen Tests

1. Wähle ein geeignetes Modell für die Daten.
2. Lege die Nullhypothese $H_0 : \ \theta = \theta_0$ fest.
   Spezifiziere die Alternative H_A anhand der Problemstellung.
3. Wähle das Signifikanzniveau α, z.B. $\alpha = 0.05$ oder 0.01.
4. Konstruiere den **Verwerfungsbereich** für H_0, so dass gilt

   $$\mathbb{P}_{\theta_0}(\text{Fehler 1. Art}) \leq \alpha.$$

   Die Form des Verwerfungsbereichs hängt von der Alternative H_A ab.
5. Betrachte, ob die Daten, bzw. eine Funktion davon, die sg. **Teststatistik** T, in den Verwerfungsbereich fällt. Falls ja, so verwerfe H_0 zugunsten von H_A. Man sagt dann auch, dass ein statistisch signifikantes Resultat vorliegt.

p-Wert

Frage: Wie "extrem" liegt der beobachtete Wert der Teststatistik in der Verteilung unter der Nullhypothese? Die "Richtung" ist durch die Alternativhypothese vorgegeben.

Der beobachtete Wert der Teststatistik sei t (fix). Der p-Wert ist dann definiert als die Wahrscheinlichkeit für ein mindestens so extremes Ereignis, wenn wir annehmen, dass H_0 stimmt.
Bei einseitigen Alternativen betrachten wir die Ereignisse $\{T \geq t\}$ bzw. $\{T \leq t\}$. Bei einer zweiseitigen Alternative nimmt man die mindestens so extremen Ereignisse auf *beiden* Seiten.

Anhand des p-Werts kann man direkt den Testentscheid ablesen:

$$\text{Verwerfe } H_0 \iff \text{p-Wert} \leq \alpha \text{ (Signifikanzniveau).}$$

Man sieht zusätzlich, wie *stark* die Nullhypothese verworfen wird (oder nicht).

Alternative Interpretation: Der p-Wert ist das *kleinste* Signifikanzniveau, bei dem die Nullhypothese *gerade noch* verworfen wird.

Achtung: Der p-Wert ist *nicht* die Wahrscheinlichkeit, dass die Nullhypothese stimmt.

Macht

Frage: Wie wahrscheinlich ist es, dass wir die Nullhypothese verwerfen, wenn in der Tat $\theta_A \in H_A$ gilt?

Die Macht ist also gemäß Definition $1 - \mathbb{P}_{\theta_A}$ (W'keit Fehler 2. Art).

Zur Berechnung der Macht müssen wir also einen *bestimmten* Parameterwert $\theta \in H_A$ annehmen. Es gibt nicht "die" Macht.

Vertrauensintervalle

Allgemeines

Grundfrage: Welche Parameterwerte θ sind mit den beobachteten Daten "verträglich"?

Ein $(1-\alpha) \times 100\%$-Vertrauensintervall I für den Parameter θ ist ein *zufälliges* Intervall I, welches den Parameter θ mit Wahrscheinlichkeit $1-\alpha$ "einfängt", d.h.

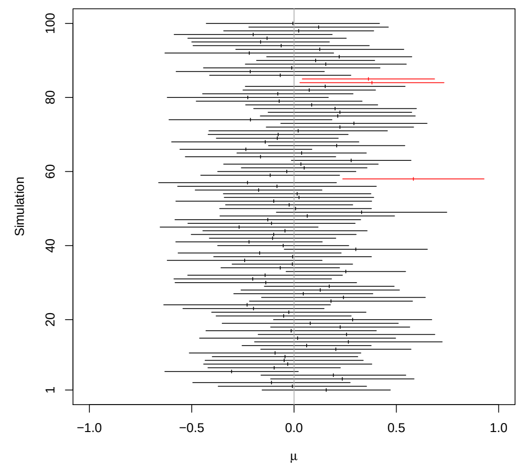
$$\mathbb{P}(I \ni \theta) = 1 - \alpha.$$

Vertrauensintervalle haben oft (aber nicht immer) die Form

$$\widehat{\theta} \pm \cdots$$

Interpretation der Überdeckungswahrscheinlichkeit: Ein einzelnes Vertrauensintervall enthält den Parameter θ oder nicht (weil der Parameter fix) ist. Wenn wir aber ein "Experiment" viele Male wiederholen könnten, dann würden im Schnitt $(1-\alpha) \times 100\%$ der Vertrauensintervalle den wahren Parameterwert enthalten.

Hilfsbild:



Dualitätssatz: Das $(1-\alpha) \times 100\%$ Vertrauensintervall besteht aus allen Parameterwerten, die im Sinne eines statistischen Tests zum Signifikanzniveau α mit den Daten verträglich sind (üblicherweise nimmt man den zweiseitigen Test). Mathematisch:

$$I = \{\theta : \text{Nullhypothese } H_0 : \theta = \theta_0 \text{ wird } \textit{nicht} \text{ verworfen}\}.$$

Anhand eines Vertrauensintervalls kann man also direkt den Testentscheid ablesen. Wenn ein Wert θ_0 im $(1-\alpha) \times 100\%$-Vertrauensintervall enthalten ist, so wird die entsprechende Nullhypothese auf dem Signifikanzniveau α *nicht* verworfen, ansonsten schon.

Informationsgehalt: Testentscheid $\prec$ p-Wert $\prec$ VI

Lukas Meier, meier@stat.math.ethz.ch