



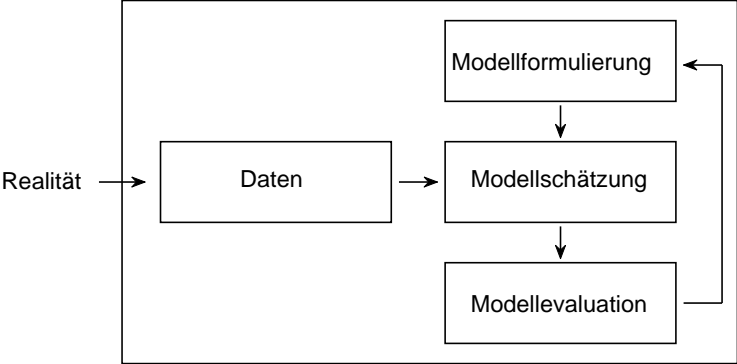
Allgemeines Lineares Modell

BSc Psychologie, SoSe 2024

Joram Soch

(6) Parameterschätzung

Naturwissenschaft



Modellformulierung

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (1)$$

Modellschätzung

$$\hat{\beta} = (X^T X)^{-1} X^T y, \quad \hat{\sigma}^2 = \frac{1}{n-p} (y - X\hat{\beta})^T (y - X\hat{\beta}) \quad (2)$$

Modellevaluation

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}, \quad F = \frac{(\hat{\varepsilon}_0^T \hat{\varepsilon}_0 - \hat{\varepsilon}^T \hat{\varepsilon})/p_1}{\hat{\varepsilon}^T \hat{\varepsilon}/(n-p)} \quad (3)$$

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für wahre, aber unbekannte, Parameterwerte oder Funktionen dieser abzugeben, typischerweise mithilfe von Daten.

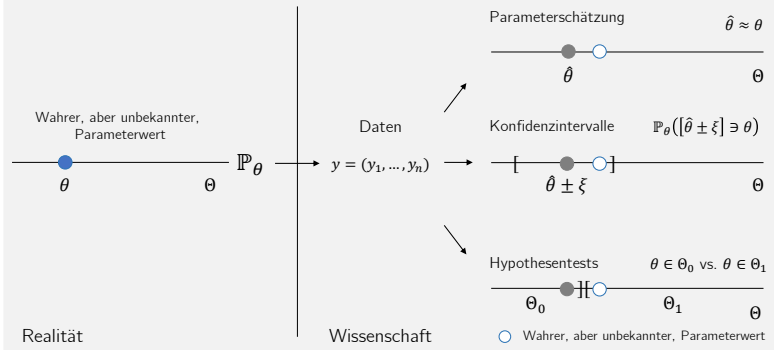
(2) Konfidenzintervalle

Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der angenommenen Verteilung der Daten eine quantitative Aussage über die mit Schätzwerten assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Ziel des Hypothesentestens ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst zuverlässigen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes liegt.

Modell und Standardprobleme Frequentistischer Inferenz



$$\theta := (\beta, \sigma^2), \quad \Theta := \mathbb{R}^p \times \mathbb{R}_{>0}, \quad \mathbb{P}_\theta(y) := \mathbb{P}_{\beta, \sigma^2}(y) \quad \text{mit WDF} \quad p_{\beta, \sigma^2}(y) := N(y; X\beta, \sigma^2 I_n)$$

Standardannahmen Frequentistischer Inferenz

Gegeben sei das Allgemeine Lineare Modell. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen der Daten des Modells ist. Aus Frequentistischer Sicht kann man unendlich oft Datensätze aus einem Modell generieren und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. den Betaparameterschätzer:

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right)^T \text{ mit } \hat{\beta}^{(1)} = (X^T X)^{-1} X^T y^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right)^T \text{ mit } \hat{\beta}^{(2)} = (X^T X)^{-1} X^T y^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right)^T \text{ mit } \hat{\beta}^{(3)} = (X^T X)^{-1} X^T y^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right)^T \text{ mit } \hat{\beta}^{(4)} = (X^T X)^{-1} X^T y^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme der Datenverteilung. Was zum Beispiel ist die Verteilung von $\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \hat{\beta}^{(3)}, \hat{\beta}^{(4)}, \dots$, also die Verteilung der Zufallsvariable $\hat{\beta} := (X^T X)^{-1} X^T y$? Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Chi-Quadrat-Zufallsvariablen

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Chi-Quadrat-Zufallsvariablen

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Theorem (Betaparameterschätzer)

Gegeben seien das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (4)$$

und der *Betaparameterschätzer*

$$\hat{\beta} := (X^T X)^{-1} X^T y. \quad (5)$$

Dann gilt, dass $\hat{\beta}$ die Summe der Abweichungsquadrate minimiert,

$$\hat{\beta} = \arg \min_{\tilde{\beta}} (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (6)$$

und dass $\hat{\beta}$ ein unverzerrter Maximum-Likelihood-Schätzer von $\beta \in \mathbb{R}^p$ ist.

Bemerkungen

- Das Theorem gibt eine Formel an, um β anhand von Designmatrix und Daten zu schätzen.
- Da $\hat{\beta}$ die Summe der Abweichungsquadrate minimiert, heißt $\hat{\beta}$ auch Kleinste-Quadrate-Schätzer (KQ-Schätzer).
- Die $\tilde{\beta}$ Notation des Maximierungarguments dient lediglich zur Abgrenzung vom w.a.u. Parameter β .
- Als ML-Schätzer ist $\hat{\beta}$ weiterhin konsistent, asymptotisch normalverteilt und asymptotisch effizient.
- Wir werden später sehen, dass $\hat{\beta}$ sogar normalverteilt ist.
- Außerdem hat $\hat{\beta}$ die "kleinste Varianz" in der Klasse der linearen unverzerrten Schätzer von β .
- Letztere Eigenschaft ist Kernaussage des *Gauss-Markov-Theorems*, auf das wir hier nicht näher eingehen wollen.
- Für eine Diskussion und einen Beweis des Gauss-Markov-Theorems, siehe z.B. Searle (1971), Kapitel 3.

Beweis

(1) Wir zeigen in einem ersten Schritt, dass $\hat{\beta}$ die Summe der Abweichungsquadrate

$$(\mathbf{y} - X\tilde{\beta})^T (\mathbf{y} - X\tilde{\beta}) \quad (7)$$

minimiert. Dazu halten wir zunächst fest, dass

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y} \Leftrightarrow X^T X \hat{\beta} = X^T \mathbf{y} \Leftrightarrow X^T \mathbf{y} - X^T X \hat{\beta} = \mathbf{0}_p \Leftrightarrow X^T (\mathbf{y} - X \hat{\beta}) = \mathbf{0}_p. \quad (8)$$

Weiterhin gilt dann auch, dass

$$X^T (\mathbf{y} - X \hat{\beta}) = \mathbf{0}_p \Leftrightarrow (X^T (\mathbf{y} - X \hat{\beta}))^T = \mathbf{0}_p^T \Leftrightarrow (\mathbf{y} - X \hat{\beta})^T X = \mathbf{0}_p^T. \quad (9)$$

Weiterhin halten wir ohne Beweis fest, dass für jede Matrix $X \in \mathbb{R}^{n \times p}$ gilt, dass

$$\mathbf{z}^T X^T X \mathbf{z} \geq 0 \quad \text{für alle } \mathbf{z} \in \mathbb{R}^p. \quad (10)$$

Wir betrachten nun die Summe der Abweichungsquadrate

$$(\mathbf{y} - X\tilde{\beta})^T (\mathbf{y} - X\tilde{\beta}). \quad (11)$$

Beweis (fortgeführt)

Es ergibt sich dann

$$\begin{aligned} & (\mathbf{y} - \mathbf{X}\tilde{\beta})^T (\mathbf{y} - \mathbf{X}\tilde{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta} + \mathbf{X}\hat{\beta} - \mathbf{X}\tilde{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta} + \mathbf{X}\hat{\beta} - \mathbf{X}\tilde{\beta}) \\ &= ((\mathbf{y} - \mathbf{X}\hat{\beta}) + \mathbf{X}(\hat{\beta} - \tilde{\beta}))^T ((\mathbf{y} - \mathbf{X}\hat{\beta}) + \mathbf{X}(\hat{\beta} - \tilde{\beta})) \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta}) + (\mathbf{y} - \mathbf{X}\hat{\beta})^T \mathbf{X}(\hat{\beta} - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})^T \mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T \mathbf{X}^T \mathbf{X}(\hat{\beta} - \tilde{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta}) + \mathbf{0}_p^T (\hat{\beta} - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})^T \mathbf{0}_p + (\hat{\beta} - \tilde{\beta})^T \mathbf{X}^T \mathbf{X}(\hat{\beta} - \tilde{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \tilde{\beta})^T \mathbf{X}^T \mathbf{X}(\hat{\beta} - \tilde{\beta}). \end{aligned}$$

Auf der rechten Seite obiger Gleichung ist nur der zweite Term von $\tilde{\beta}$ abhängig. Da für diesen Term gilt, dass

$$(\hat{\beta} - \tilde{\beta})^T \mathbf{X}^T \mathbf{X}(\hat{\beta} - \tilde{\beta}) \geq 0, \quad (12)$$

nimmt dieser Term genau dann seinen Minimalwert 0 an, wenn

$$(\hat{\beta} - \tilde{\beta}) = \mathbf{0}_p \Leftrightarrow \tilde{\beta} = \hat{\beta}. \quad (13)$$

Also gilt

$$\hat{\beta} = \underset{\tilde{\beta}}{\arg \min} (\mathbf{y} - \mathbf{X}\tilde{\beta})^T (\mathbf{y} - \mathbf{X}\tilde{\beta}). \quad (14)$$

Beweis (fortgeführt)

(2) Um zu zeigen, dass $\hat{\beta}$ ein Maximum-Likelihood-Schätzer ist, betrachten wir für festes $y \in \mathbb{R}^n$ und festes $\sigma^2 > 0$ die Log-Likelihood-Funktion

$$\ell : \mathbb{R}^p \rightarrow \mathbb{R}, \tilde{\beta} \mapsto \ln p_{\tilde{\beta}}(y) = \ln N(y; X\tilde{\beta}, \sigma^2 I_n), \quad (15)$$

wobei gilt, dass

$$\begin{aligned} \ln N(y; X\tilde{\beta}, \sigma^2 I_n) &= \ln \left((2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \right) \right) \\ &= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 I_n| - \frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta}). \end{aligned} \quad (16)$$

Dabei hängt allein der Term $-\frac{1}{2\sigma^2} (y - X\tilde{\beta})^T (y - X\tilde{\beta})$ von $\tilde{\beta}$ ab. Weil aber $(y - X\tilde{\beta})^T (y - X\tilde{\beta}) \geq 0$, gilt aufgrund des negativen Vorzeichens, dass dieser Term maximal wird, wenn $(y - X\tilde{\beta})^T (y - X\tilde{\beta})$ minimal wird. Dies ist aber wie oben gezeigt genau für $\tilde{\beta} = \hat{\beta}$ der Fall.

(3) Die Unverzerrtheit von $\hat{\beta}$ schließlich ergibt sich aus

$$\mathbb{E}(\hat{\beta}) = \mathbb{E} \left((X^T X)^{-1} X^T y \right) = (X^T X)^{-1} X^T \mathbb{E}(y) = (X^T X)^{-1} X^T X \beta = \beta. \quad (17)$$

□

Definition (Erklärte Daten, Residuenvektor, Residuen)

Gegeben seien das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (18)$$

und der Betaparameterschätzer

$$\hat{\beta} := (X^T X)^{-1} X^T y. \quad (19)$$

Dann heißt der Zufallsvektor

$$\hat{y} := X\hat{\beta} \quad (20)$$

die *erklärten Daten*, der Zufallsvektor

$$\hat{\varepsilon} := y - \hat{y} \quad (21)$$

heißt *Residuenvektor* und für $i = 1, \dots, n$ heißen die Komponenten dieses Zufallsvektors

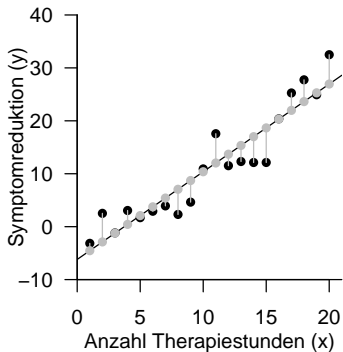
$$\hat{\varepsilon}_i := y_i - \hat{y}_i \quad (22)$$

die *Residuen*.

Bemerkungen

- Wir haben diese Begriffe bereits für den Fall einfacher linearer Regression mit Ausgleichsgerade eingeführt (vgl. Einheit (2) in *Allgemeines Lineares Modell*).

Erklärte Daten und Residuenvektor bei einfacher linearer Regression



$$\bullet y_i \quad \bullet \hat{y}_i = (X\hat{\beta})_i = \begin{pmatrix} 1 & x_i \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad \text{---} \quad \hat{\varepsilon}_i = \hat{y}_i - y_i \quad i = 1, \dots, n$$

Theorem (Erklärte Daten und Residuenvektor als Matrixprodukte)

Gegeben seien das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (23)$$

sowie die erklärten Daten \hat{y} und der Residuenvektor $\hat{\varepsilon}$. Dann ergeben sich die erklärten Daten und der Residuenvektor als Matrixmultiplikationen des Datenvektors y von links

$$\begin{aligned} \hat{y} &= Py \\ \hat{\varepsilon} &= Ry, \end{aligned} \quad (24)$$

wobei die *Projektionsmatrix* P und die *Residuen-bildende Matrix* R wie folgt gegeben sind:

$$\begin{aligned} P &:= X(X^T X)^{-1} X^T \\ R &:= (I_n - P). \end{aligned} \quad (25)$$

Bemerkungen

- P wird auch als *projection matrix* bezeichnet.
- R wird auch als *residual-forming matrix* bezeichnet.

Beweis

Die erklärten Daten ergeben sich zu

$$\begin{aligned}\hat{y} &:= X\hat{\beta} \\ &= X(X^T X)^{-1} X^T y \\ &= P y .\end{aligned}\tag{26}$$

Der Residuenvektor ergibt sich zu

$$\begin{aligned}\hat{\varepsilon} &:= y - \hat{y} \\ &= y - X\hat{\beta} \\ &= y - X(X^T X)^{-1} X^T y \\ &= (I_n - X(X^T X)^{-1} X^T) y \\ &= R y .\end{aligned}\tag{27}$$

□

Theorem (Varianzparameterschätzer)

Gegeben sei das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (28)$$

Dann ist

$$\hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \quad (29)$$

ein unverzerrter Schätzer von $\sigma^2 > 0$.

Bemerkungen

- Für einen Beweis, siehe Searle (1971), Kapitel 3 oder Rencher and Schaalje (2008), Kapitel 7.
- Mit Definition des Residuenvektors und der Residuen bieten sich für $\hat{\sigma}^2$ auch folgende Schreibweisen an:

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - p} = \frac{1}{n - p} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n - p} \sum_{i=1}^n (y_i - (X\hat{\beta})_i)^2 \quad (30)$$

- σ^2 wird also durch eine skalierte residuelle Quadratsumme geschätzt.
- Es handelt sich bei $\hat{\sigma}^2$ *nicht* um einen Maximum-Likelihood-Schätzer von σ^2 .
- Der Maximum-Likelihood-Schätzer des Varianzparameters ist $\hat{\sigma}_{ML}^2 := \frac{1}{n} \hat{\varepsilon}^T \hat{\varepsilon}$.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Chi-Quadrat-Zufallsvariablen

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Unabhängige und identisch normalverteilte Zufallsvariablen

Wir betrachten das Szenario von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter $\mu \in \mathbb{R}$ und Varianzparameter σ^2 :

$$y_i \sim N(\mu, \sigma^2) \quad \text{für } i = 1, \dots, n. \quad (31)$$

Dann gilt, wie unten gezeigt wird, dass

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y} \quad \text{und} \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 =: s_y^2. \quad (32)$$

In diesem Fall ist also der Betaparameterschätzer mit dem Stichprobenmittel \bar{y} der y_1, \dots, y_n und der Varianzparameterschätzer mit der Stichprobenvarianz s_y^2 der y_1, \dots, y_n identisch.

Für $\hat{\beta}$ ergibt sich

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= (1_n^T 1_n)^{-1} 1_n^T y \\ &= \left((1 \quad \dots \quad 1) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)^{-1} (1 \quad \dots \quad 1) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= n^{-1} \sum_{i=1}^n y_i \\ &= \frac{1}{n} \sum_{i=1}^n y_i \\ &=: \bar{y}.\end{aligned}$$

Für $\hat{\sigma}^2$ ergibt sich

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-1} (y - X\hat{\beta})^T (y - X\hat{\beta}) \\ &= \frac{1}{n-1} (y - \mathbf{1}_n \bar{y})^T (y - \mathbf{1}_n \bar{y}) \\ &= \frac{1}{n-1} \left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{y} \right)^T \left(\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{y} \right) \\ &= \frac{1}{n-1} (y_1 - \bar{y} \quad \cdots \quad y_n - \bar{y}) \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \\ &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &=: s_y^2.\end{aligned}$$

Unabhängige und identisch normalverteilte Zufallsvariablen

Simulation einer Parameterschätzung

```
# Modellformulierung
library(MASS)
n      = 12
p      = 1
X      = matrix(rep(1,n), nrow = n)
I_n    = diag(n)
beta   = 2
sigsqr = 1

# multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betaparameter
# n x p Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Datenrealisierung
y      = mvrnorm(1, X %>% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs y

# Parameterschätzung
beta_hat = solve(t(X) %>% X) %>% t(X) %>% y # Betaparameterschätzer
eps_hat  = y - X %>% beta_hat             # Residuenvektor
sigsqr_hat = t(eps_hat) %>% eps_hat) / (n-p) # Varianzparameterschätzer

# Ausgabe
cat( "beta      : ", beta,
     "\nhat{beta} : ", beta_hat,
     "\nsigsqr     : ", sigsqr,
     "\nhat{sigsqr} : ", sigsqr_hat)

> beta      : 2
> hat{beta} : 1.64
> sigsqr    : 1
> hat{sigsqr} : 1.12
```

Unabhängige und identisch normalverteilte Zufallsvariablen

Simulation der Schätzerunverzerrtheit

```
# Modellformulierung
library(MASS)
n = 12
p = 1
X = matrix(rep(1,n), nrow = n)
I_n = diag(n)
beta = 2
sigsqr = 1

# multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betaparameter
# n x p Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Frequentistische Simulation
nsim = 1e4
beta_hat = rep(NA,n)
sigsqr_hat = rep(NA,n)
for(i in 1:nsim){
  y = mvrnorm(1, X %*% beta, sigsqr*I_n)
  beta_hat[i] = solve(t(X) %*% X) %*% t(X) %*% y
  eps_hat = y - X %*% beta_hat[i]
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat) / (n-p)
}

# Ausgabe
cat( "wahrer, aber unbekannter Betaparameter : ", beta,
     "\ngeschätzter Erwartungswert des Betaparameterschätzers : ", mean(beta_hat),
     "\nwahrer, aber unbekannter Varianzparameter : ", sigsqr,
     "\ngeschätzter Erwartungswert des Varianzparameterschätzers : ", mean(sigsqr_hat))

> wahrer, aber unbekannter Betaparameter : 2
> geschätzter Erwartungswert des Betaparameterschätzers : 2
> wahrer, aber unbekannter Varianzparameter : 1
> geschätzter Erwartungswert des Varianzparameterschätzers : 0.998
```

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Chi-Quadrat-Zufallsvariablen

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Einfache lineare Regression

Wir betrachten das Modell der einfachen linearen Regression:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad \text{für } i = 1, \dots, n. \quad (33)$$

Dann gilt, wie unten gezeigt wird, dass

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2} \bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix} \quad \text{und} \quad \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2, \quad (34)$$

wobei

- \bar{x} und \bar{y} die Stichprobenmittel der x_1, \dots, x_n bzw. der y_1, \dots, y_n bezeichnen,
- c_{xy} die Stichprobenkovarianz der x_1, \dots, x_n und y_1, \dots, y_n bezeichnet und
- s_x^2 die Stichprobenvarianz der x_1, \dots, x_n bezeichnet.

Wie bereits gesehen (siehe Einheit (1) in *Allgemeines Lineares Modell*) sind die Bezeichnungen "Stichproben"kovarianz und "Stichproben"varianz bezüglich der x_1, \dots, x_n hier lediglich formal gemeint, da keine Annahme zugrundeliegt, dass die x_1, \dots, x_n Realisierungen von Zufallsvariablen sind. Die x_1, \dots, x_n sind vorgegebene Werte.

Einfache lineare Regression

Wir halten fest, dass für eine parametrische Designmatrixspalte sich der entsprechende Betaparameterschätzer aus der Stichprobenkovarianz der respektiven Spalte mit den Daten geteilt durch die Stichprobenvarianz der entsprechenden Spalte ergibt und somit einer "standardisierten" Stichprobenkovarianz entspricht.

Ein Vergleich mit den Parametern der Ausgleichsgerade in (1) Regression zeigt weiterhin die Identität der Betaparameterschätzerkomponenten $\hat{\beta}_0$ und $\hat{\beta}_1$ mit den dort unter dem Kriterium der Minimierung der quadrierten vertikalen Abweichungen hergeleiteten Parametern. Dies ist nicht überraschend, da sowohl $\hat{\beta}$ als auch die Parameter der Ausgleichsgerade die Funktion

$$q(\tilde{\beta}) = \sum_{i=1}^n (y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2 = (y - X\tilde{\beta})^T (y - X\tilde{\beta}) \quad (35)$$

hinsichtlich $\tilde{\beta}$ minimieren.

Um die Form des Betaparameterschätzers herzuleiten, halten wir zunächst fest, dass

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \bar{y} - \sum_{i=1}^n \bar{x} y_i + \sum_{i=1}^n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - \bar{y} n \bar{x} - \bar{x} n \bar{y} + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} - n \bar{x} \bar{y} + n \bar{x} \bar{y} \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}.\end{aligned}\tag{36}$$

Weiterhin halten wir fest, dass

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2x_i\bar{x} + \sum_{i=1}^n \bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2.\end{aligned}\tag{37}$$

Einfache lineare Regression

Aus der Definition von $\hat{\beta}$ ergibt sich

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \left(\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} \\ &= \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.\end{aligned}\tag{38}$$

Mit der Formel für die Inverse einer 2×2 Matrix

$$A := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}\tag{39}$$

(siehe Einheit (3) in *Allgemeines Lineares Modell*) und unter Gebrauch der oben hergeleiteten Identität $s_x^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$ ergibt sich die Inverse von $X^T X$ wie folgt:

$$\begin{aligned}(X^T X)^{-1} &= \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} = \frac{1}{n \sum_{i=1}^n x_i^2 - (n\bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \\ &= \frac{1}{n (\sum_{i=1}^n x_i^2 - n\bar{x}^2)} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \\ &= \frac{1}{ns_x^2} \begin{pmatrix} s_x^2 + n\bar{x}^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \\ &= \frac{1}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.\end{aligned}\tag{40}$$

Einfache lineare Regression

Mit der oben hergeleiteten Identität $c_{xy} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$ ergibt sich also

$$\begin{aligned}\hat{\beta} &= \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{s_x^2} & -\frac{\bar{x}}{s_x^2} \\ -\frac{\bar{x}}{s_x^2} & \frac{1}{s_x^2} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right)n\bar{y} - \frac{\bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i}{s_x^2} - \frac{n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{n\bar{y}}{n} + \frac{\bar{x}^2 n\bar{y}}{s_x^2} - \frac{\bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} + \frac{\bar{x}n\bar{y} - \bar{x} \sum_{i=1}^n x_i y_i}{s_x^2} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} - \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \bar{x} \\ \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{s_x^2} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} - \frac{c_{xy}}{s_x^2} \bar{x} \\ \frac{c_{xy}}{s_x^2} \end{pmatrix} .\end{aligned}\tag{41}$$

Simulation einer Parameterschätzung

```
# Modellformulierung
library(MASS)
n = 10 # multivariate Normalverteilung
p = 2 # Anzahl Datenpunkte
x = 1:n # Anzahl Betaparameter
X = matrix(c(rep(1,n),x), nrow = n) # Prädiktorwerte
I_n = diag(n) # n x p Designmatrix
beta = matrix(c(0,1), nrow = p) # n x n Einheitsmatrix
sigsqr = 1 # wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs y

# Parameterschätzung
beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
eps_hat = y - X %*% beta_hat # Residuenvektor
sigsqr_hat = t(eps_hat) %*% eps_hat / (n-p) # Varianzparameterschätzer

# Ausgabe
cat( "beta      : ", beta,
     "\nhat{beta} : ", beta_hat,
     "\nsigsqr     : ", sigsqr,
     "\nhat{sigsqr} : ", sigsqr_hat)

> beta      : 0 1
> hat{beta} : -1.04 1.07
> sigsqr    : 1
> hat{sigsqr} : 0.337
```

Einfache lineare Regression

Simulation der Schätzerunverzerrtheit

```
# Modellformulierung
library(MASS)
n      = 10
p      = 2
x      = 1:n
X      = matrix(c(rep(1,n),x), nrow = n)
I_n    = diag(n)
beta   = matrix(c(0,1), nrow = p)
sigsqr = 1

# multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betaparameter
# Prädiktorwerte
# n x p Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Frequentistische Simulation
nsim   = 1e4
beta_hat = matrix(rep(NaN,p*nsim), nrow = p)
sigsqr_hat = rep(NaN,nsim)
for(i in 1:nsim){
  y      = mvrnorm(1, X %*% beta, sigsqr*I_n)
  beta_hat[,i] = solve(t(X) %*% X) %*% t(X) %*% y
  eps_hat = y - X %*% beta_hat[,i]
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat) / (n-p)
}

# Ausgabe
cat( "wahrer, aber unbekannter Betaparameter      : ", beta,
     "\ngeschätzter Erwartungswert des Betaparameterschätzers : ", rowMeans(beta_hat),
     "\nwahrer, aber unbekannter Varianzparameter      : ", sigsqr,
     "\ngeschätzter Erwartungswert des Varianzparameterschätzers : ", mean(sigsqr_hat))

> wahrer, aber unbekannter Betaparameter      : 0 1
> geschätzter Erwartungswert des Betaparameterschätzers : -0.00191 1
> wahrer, aber unbekannter Varianzparameter      : 1
> geschätzter Erwartungswert des Varianzparameterschätzers : 1.01
```

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Chi-Quadrat-Zufallsvariablen

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Definition (χ^2 -Zufallsvariable)

$\{Z_1, \dots, Z_n\}$ mit $Z_i \sim N(0, 1)$ für $i = 1, \dots, n$ sei eine Menge von n unabhängigen standardnormalverteilten Zufallsvariablen. Dann nennen wir die Zufallsvariable

$$X := \sum_{i=1}^n Z_i^2 \quad (42)$$

eine χ^2 -verteilte Zufallsvariable mit Freiheitsgradparameter n . Wir schreiben $X \sim \chi^2(n)$. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer χ^2 -Zufallsvariable bezeichnen wir mit $\chi^2(x; n)$.

Bemerkungen

- Die Summe von n quadrierten standardnormal-verteilten Zufallsvariablen ist eine χ^2 -verteilte Zufallsvariable.
- Chi-Quadrat-Zufallsvariablen sind zentral für die Definition von T- und F-Statistiken.

Theorem (WDF einer χ^2 -Zufallsvariable)

X sei eine χ^2 -Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Freiheitsgradparameter n . Dann ist die Wahrscheinlichkeitsdichtefunktion von X gegeben durch

$$\chi^2(\cdot; n) : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, x \mapsto \chi^2(x; n) = \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left(-\frac{x}{2}\right), \quad (43)$$

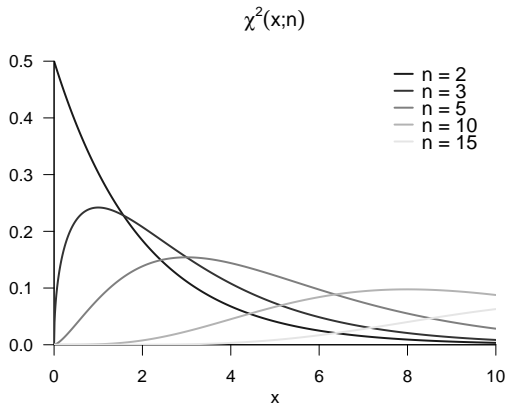
wobei \exp die Exponentialfunktion und Γ die Gammafunktion bezeichne.

Bemerkungen

- Wir verzichten auf einen Beweis, der [sich hier findet](#) (englisch).
- Die χ^2 -Verteilung ist auf die positiven reellen Zahlen beschränkt. Steigendes n verschiebt die Wahrscheinlichkeitsmasse in den höheren positiven Bereich.

Chi-Quadrat-Zufallsvariablen

Wahrscheinlichkeitsdichtefunktionen von χ^2 -Zufallsvariablen



Definition (Nichtzentrale χ^2 -Zufallsvariable)

$\{Y_1, \dots, Y_n\}$ mit $Y_i \sim N(\mu_i, 1)$ für $i = 1, \dots, n$ sei eine Menge von n unabhängigen und normalverteilten Zufallsvariablen mit Erwartungswertparameter μ_i und Varianzparameter 1. Dann nennen wir die Zufallsvariable

$$X := \sum_{i=1}^n Y_i^2 \quad (44)$$

eine nichtzentral χ^2 -verteilte Zufallsvariable mit Nichtzentralitätsparameter $\delta := \sum_{i=1}^n \mu_i^2$ und Freiheitsgradparameter n . Wir schreiben $X \sim \chi^2(\delta, n)$. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer nichtzentralen χ^2 -Zufallsvariable bezeichnen wir mit $\chi^2(x; \delta, n)$.

Bemerkungen

- Die Summe von n quadrierten normalverteilten Zufallsvariablen mit nicht notwendigerweise gleichen Erwartungswertparametern, aber gleichem Varianzparameter ist eine nichtzentral χ^2 -verteilte Zufallsvariable.
- Eine nichtzentrale χ^2 -Zufallsvariable mit $\delta = 0$ ist eine χ^2 -Zufallsvariable. Es gilt also $\chi^2(x; 0, n) = \chi^2(x; n)$.

Theorem (WDF einer nichtzentralen χ^2 -Zufallsvariable)

X sei eine nichtzentrale χ^2 -Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$, Nichtzentralitätsparameter δ und Freiheitsgradparameter n . Dann ist die WDF von X gegeben durch

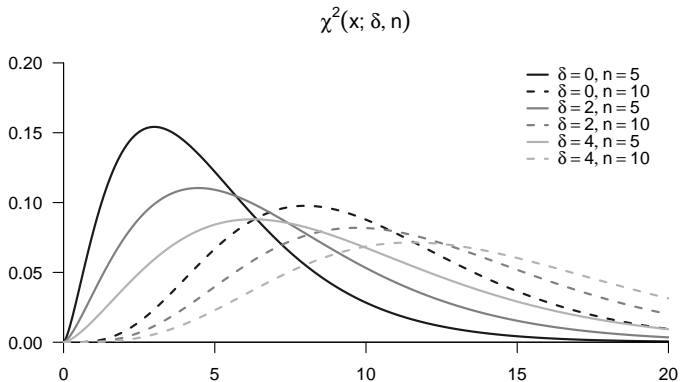
$$\begin{aligned} \chi^2(\cdot; \delta, n) : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, x \mapsto \chi^2(x; \delta, n) := & \frac{1}{2} \left(\frac{x}{\delta}\right)^{\frac{n}{4} - \frac{1}{2}} \exp\left(-\frac{x + \delta}{2}\right) \\ & \times \left(\frac{\sqrt{\delta x}}{2}\right)^{\frac{n}{2} - 1} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta x}{4}\right)^j}{j! \Gamma\left(\frac{n}{2} + j\right)}, \end{aligned} \quad (45)$$

wobei \exp die Exponentialfunktion und Γ die Gammafunktion bezeichne.

Bemerkung

- Wir verzichten auf einen Beweis.
- Die nichtzentrale Chi-Quadrat-Verteilung spielt für die residuelle Quadratsumme im ALM eine große Rolle.

Wahrscheinlichkeitsdichtefunktionen nichtzentraler χ^2 -Zufallsvariablen



Theorem (Cochrans Theorem für multivariate Normalverteilungen)

ξ sei ein multivariat normalverteilter n -dimensionaler Zufallsvektor mit sphärischem Kovarianzmatrixparameter:

$$\xi \sim N(\mu, \sigma^2 I_n). \quad (46)$$

Weiterhin sei $A \in \mathbb{R}^{n \times n}$ eine quadratische Matrix und die Zufallsvariable v als quadratische Form von ξ definiert:

$$v := \xi^T A \xi / \sigma^2. \quad (47)$$

Dann gilt:

$$v \sim \chi^2(\mu^T A \mu, \text{rg}(A)). \quad (48)$$

Bemerkung

- Wir verzichten auf einen Beweis.
- Das Theorem stellt einen tiefen Zusammenhang zwischen multivariater Normalverteilung und Chi-Quadrat-Verteilung her: Die mit einer Matrix A gewichtete und durch die Varianz σ^2 normalisierte quadratische Form eines multivariat normalverteilten Zufallsvektors ist eine nichtzentral χ^2 -verteilte Zufallsvariable, wobei der Nichtzentralitätsparameter sich aus dem Erwartungswertparameter der multivariaten Normalverteilung und der Matrix A und der Freiheitsgradparameter sich als der Rang dieser Matrix ergibt.
- Namensgeber des Theorems ist William G. Cochran. Cochrans Theorem ist grundlegend für die Frequentistische Schätzerverteilung des Varianzparameterschätzers im Rahmen der ein- und mehrfaktoriellen Varianzanalyse.

Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Chi-Quadrat-Zufallsvariablen

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Theorem (Frequentistische Verteilung des Betaparameterschätzers)

Gegeben seien das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (49)$$

und der Betaparameterschätzer

$$\hat{\beta} := (X^T X)^{-1} X^T y \quad (50)$$

Dann gilt

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (51)$$

Bemerkungen

- Es gilt also wie bereits gesehen $\mathbb{E}(\hat{\beta}) = \beta$ und außerdem $\mathbb{C}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$.
- Die Varianzen der Komponenten von $\hat{\beta}$ sind die Diagonalelemente von $\mathbb{C}(\hat{\beta})$, also

$$\mathbb{V}(\hat{\beta}_i) = (\sigma^2 (X^T X)^{-1})_{ii} \quad \text{für} \quad i = 1, \dots, p. \quad (52)$$

- Die Streuung von $\hat{\beta}$ hängt von σ^2 und der Designmatrix X ab. σ^2 ist ein experimentell nicht beeinflussbarer wahrer, aber unbekannter Parameter. X dagegen kann gewählt werden, um zum Beispiel die Diagonalelemente von $\mathbb{C}(\hat{\beta})$ bei festem σ^2 zu minimieren.

Frequentistische Schätzerverteilungen

Beweis

Das Theorem folgt direkt mit dem Theorem zur linearen Transformation von multivariaten Normalverteilungen (siehe Einheit (4) in *Allgemeines Lineares Modell*). Speziell gilt hier:

$$\hat{\beta} \sim N\left((X^T X)^{-1} X^T X \beta, (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T\right). \quad (53)$$

Der Erwartungswertparameter vereinfacht sich dann zu

$$(X^T X)^{-1} X^T X \beta = \beta. \quad (54)$$

Der Kovarianzmatrixparameter vereinfacht sich wie folgt:

$$\begin{aligned} (X^T X)^{-1} X^T (\sigma^2 I_n) ((X^T X)^{-1} X^T)^T &= (X^T X)^{-1} X^T (\sigma^2 I_n) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}. \end{aligned} \quad (55)$$

Dabei folgt hier die erste Gleichung aus der Tatsache, dass sowohl $X^T X$ als auch ihre Inverse $(X^T X)^{-1}$ symmetrische Matrizen sind. Insgesamt gilt damit

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X^T X)^{-1}\right). \quad (56)$$

□

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \quad \text{mit} \quad X := \mathbf{1}_n \in \mathbb{R}^n, \quad \beta := \mu \in \mathbb{R}, \quad \sigma^2 > 0. \quad (57)$$

das ALM-Szenario unabhängiger und identisch normalverteilter Zufallsvariablen. Wir haben bereits gesehen, dass $\hat{\beta} = \bar{y}$. Das Theorem zur Frequentistischen Verteilung des Betaparameterschätzers impliziert damit

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (58)$$

Das Stichprobenmittel von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter μ und Varianzparameter σ^2 ist also normalverteilt mit Erwartungswertparameter μ und Varianzparameter σ^2/n . Wir haben diese Tatsache bereits unter dem Begriff der *Mittelwerttransformation* kennengelernt (siehe Einheit (7) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*).

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

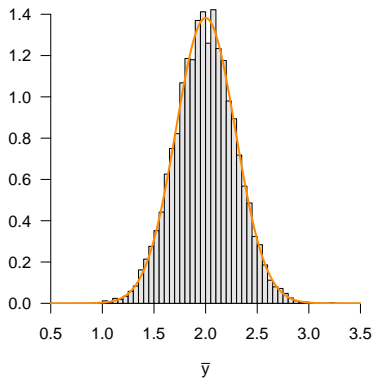
```
# Modellformulierung
library(MASS) # multivariate Normalverteilung
n = 12 # Anzahl Datenpunkte
p = 1 # Anzahl Betparameter
X = matrix(rep(1,n), nrow = n) # n x p Designmatrix
I_n = diag(n) # n x n Einheitsmatrix
beta = 2 # wahrer, aber unbekannter Betaparameter
sigsqr = 1 # wahrer, aber unbekannter Varianzparameter

# Frequentistische Simulation
nsim = 1e4 # Anzahl Realisierungen n-dimensionaler ZV
beta_hat = rep(NaN,nsim) # \hat{\beta} Realisierungsarray
for(i in 1:nsim){
  y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs
  beta_hat[i] = solve(t(X) %*% X) %*% t(X) %*% y # \hat{\beta} = (X^T X)^{-1} X^T y
}
```

Frequentistische Schätzerverteilungen

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



Beispiel (2) Einfache lineare Regression

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \quad \text{mit} \quad \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2, \sigma^2 > 0. \quad (59)$$

das ALM-Szenario der einfachen linearen Regression. Wir haben bereits gesehen, dass

$$\sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{s_x^2} \begin{pmatrix} \frac{s_x^2}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \quad \text{mit} \quad s_x^2 := \sum_{i=1}^n (x_i - \bar{x})^2. \quad (60)$$

Die Varianz des Offsetparameterschätzers hängt also sowohl von der Summe der quadrierten Differenzen und dem Stichprobenmittel der unabhängigen Variablen x_1, \dots, x_n ab, wohingegen die Varianz des Steigungsparameterschätzers nur von der Summe der quadrierten Differenzen der x_1, \dots, x_n abhängt. Die Kovarianz von Offset- und Steigungsparameterschätzern hängt vom Mittelwert der x_1, \dots, x_n ab.

Beispiel (2) Einfache lineare Regression

```
# Modellformulierung
library(MASS)
n      = 10
p      = 2
x      = 1:n
X      = matrix(c(rep(1,n),x), nrow = n)
I_n    = diag(n)
beta   = matrix(c(0,1), nrow = p)
sigsqr = 0.5

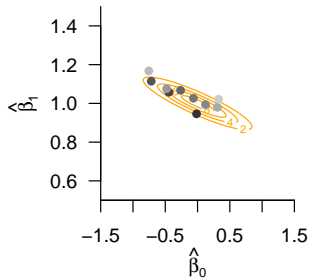
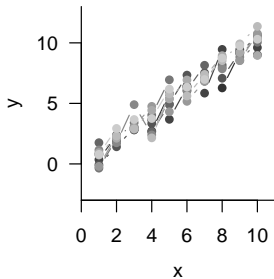
# multivariate Normalverteilung
# Anzahl Datenpunkte
# Anzahl Betparameter
# Prädiktorwerte
# n x p Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Frequentistische Simulation
nsim   = 10
y      = matrix(rep(NaN,n*nsim), nrow = n)
beta_hat = matrix(rep(NaN,p*nsim), nrow = p)
for(i in 1:nsim){
  y[,i]      = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs
  beta_hat[,i] = solve(t(X)%*%X) %*% t(X)%*%y[,i] # \hat{\beta} = (X^T X)^{-1} X^T y
}
}
```

Beispiel (2) Einfache lineare Regression

$$y \sim (X\beta, \sigma^2 I_n)$$

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$



Theorem (Frequentistische Verteilung des Varianzparameterschätzers)

Gegeben seien das ALM

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \quad (61)$$

und der Varianzparameterschätzer

$$\hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p}. \quad (62)$$

Dann gilt

$$\frac{n - p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n - p). \quad (63)$$

Bemerkungen

- Das Theorem ist für das ALM zentral, da es für die Formulierung von sowohl T-Statistiken als auch F-Statistiken eine wichtige Rolle spielt.
- Da es sich bei $(y - X\hat{\beta})^T (y - X\hat{\beta})$ um eine Summe quadrierter normalverteilter Zufallsvariablen handelt, liegt die χ^2 -Verteilung im Lichte der χ^2 -Transformation der Normalverteilung bereits nahe (vgl. Einheit (8) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*).

Frequentistische Schätzerverteilungen

Beweis

Wir zeigen zunächst, dass sich $(n-p)\hat{\sigma}^2$ als quadratische Form von y schreiben lässt:

$$\begin{aligned}(n-p)\hat{\sigma}^2 &= (n-p) \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n-p} = (y - X\hat{\beta})^T (y - X\hat{\beta}) \\ &= (y - \hat{y})^T (y - \hat{y}) = \hat{\varepsilon}^T \hat{\varepsilon} = (Ry)^T (Ry) = y^T R^T Ry \\ \text{mit } R &:= I_n - P = I_n - X(X^T X)^{-1} X^T .\end{aligned}\tag{64}$$

Im nächsten Schritt bemerken wir, dass die Residuen-bildende R symmetrisch ist

$$\begin{aligned}R^T &= (I_n - P)^T = (I_n - X(X^T X)^{-1} X^T)^T = I_n^T - (X(X^T X)^{-1} X^T)^T \\ &= I_n - X(X^T X)^{-1} X^T = I_n - P = R\end{aligned}\tag{65}$$

und dass R idempotent ist, d.h. mit sich selbst multipliziert wieder R ergibt

$$\begin{aligned}RR &= (I_n - P)(I_n - P) = I_n - P - P + PP \\ &= I_n - 2P + X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T \\ &= I_n - 2P + X(X^T X)^{-1} X^T = I_n - 2P + P = I_n - P = R ,\end{aligned}\tag{66}$$

sodass gilt:

$$(n-p)\hat{\sigma}^2 = y^T R^T Ry = y^T R R y = y^T R y .\tag{67}$$

Frequentistische Schätzerverteilungen

Beweis

Nun sind die Anwendungsbedingungen von Cochran's Theorem für multivariate Normalverteilungen gegeben:

$$\xi \sim N(\mu, \sigma^2 I_n) \Rightarrow v := \xi^T A \xi / \sigma^2 \sim \chi^2(\mu^T A \mu, \text{rg}(A)) . \quad (68)$$

Im vorliegenden Fall haben wir:

$$y \sim N(X\beta, \sigma^2 I_n) \quad \text{und} \quad \frac{n-p}{\sigma^2} \hat{\sigma}^2 = y^T R y / \sigma^2 . \quad (69)$$

Also gilt:

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2((X\beta)^T R (X\beta), \text{rg}(R)) . \quad (70)$$

Für den Nichtzentralitätsparameter ergibt sich:

$$\begin{aligned} (X\beta)^T R (X\beta) &= \beta^T X^T (I_n - P) X \beta \\ &= \beta^T X^T (I_n - X(X^T X)^{-1} X^T) X \beta \\ &= \beta^T (X^T X - X^T X (X^T X)^{-1} X^T X) \beta \\ &= \beta^T (X^T X - X^T X) \beta \\ &= \beta^T 0_{pp} \beta \\ &= 0 . \end{aligned} \quad (71)$$

Frequentistische Schätzerverteilungen

Beweis

Da R idempotent ist, ist ihr Rang gleich ihrer Spur und für den Freiheitsgradparameter ergibt sich:

$$\begin{aligned}\operatorname{rg}(R) &= \operatorname{sp}(R) = \operatorname{sp}(I_n - P) \\ &= \operatorname{sp}(I_n - X(X^T X)^{-1} X^T) \\ &= \operatorname{sp}(I_n) - \operatorname{sp}(X(X^T X)^{-1} X^T) \\ &= \operatorname{sp}(I_n) - \operatorname{sp}(X^T X (X^T X)^{-1}) \\ &= \operatorname{sp}(I_n) - \operatorname{sp}(I_p) \\ &= n - p .\end{aligned}\tag{72}$$

Zusammengenommen folgt daraus:

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(0, n-p) .\tag{73}$$

Da eine nichtzentrale Chi-Quadrat-Verteilung mit Nichtzentralitätsparameter 0 äquivalent zu einer Chi-Quadrat-Verteilung mit gleichem Freiheitsgradparameter ist, gilt schließlich:

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n-p) .\tag{74}$$

□

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \quad \text{mit} \quad X := \mathbf{1}_n \in \mathbb{R}^n, \quad \beta := \mu \in \mathbb{R}, \quad \sigma^2 > 0. \quad (75)$$

das ALM-Szenario unabhängiger und identisch normalverteilter Zufallsvariablen. Wir haben bereits gesehen, dass in diesem Fall $\hat{\beta}$ mit dem Stichprobenmittel \bar{y} identisch ist und dass $\hat{\sigma}^2$ mit der Stichprobenvarianz s_y^2 übereinstimmt.

Bei der Betrachtung von Konfidenzintervallen (vgl. Einheit (10) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*) hatten wir für den Fall von n unabhängigen und identisch normalverteilten Zufallsvariablen die Statistik

$$U := \frac{n-1}{\sigma^2} S^2 \quad (76)$$

definiert und festgehalten, dass

$$U \sim \chi^2(n-1). \quad (77)$$

Offenbar ist U für $p = 1$ mit der im obigen Theorem betrachteten Zufallsvariable $\frac{n-p}{\sigma^2} \hat{\sigma}^2$ identisch.

Beispiel (2) Einfache lineare Regression

```
# Modellformulierung
library(MASS)
n      = 10
p      = 2
x      = 1:n
X      = matrix(c(rep(1,n),x), nrow = n)
I_n    = diag(n)
beta   = matrix(c(0,1), nrow = p)
sigsqr = 0.5

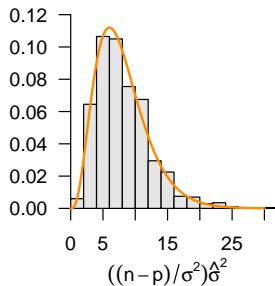
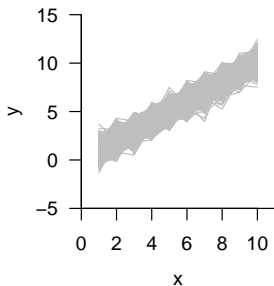
# multivariate Normalverteilung
# Anzahl von Datenpunkten
# Anzahl von Betparametern
# Prädiktorwerte
# n x p Designmatrix
# n x n Einheitsmatrix
# wahrer, aber unbekannter Betaparameter
# wahrer, aber unbekannter Varianzparameter

# Frequentistische Simulation
nsim   = 1e3
y      = matrix(rep(NaN,n*nsim), nrow = n)
beta_hat = matrix(rep(NaN,p*nsim), nrow = p)
sigsqr_hat = rep(NaN, nsim)
for(i in 1:nsim){
  y[,i] = mvrnorm(1, X %>% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs y
  beta_hat[,i] = solve(t(X)%%X) %>% t(X)%%y[,i] # \hat{\beta} = (X^T X)^{-1} X^T y
  eps_hat = y[,i] - X %>% beta_hat[,i] # \hat{\epsilon} = y - X \hat{\beta}
  sigsqr_hat[i] = (t(eps_hat) %>% eps_hat)/(n-p) # \hat{\sigma}^2 = \hat{\epsilon}^T \hat{\epsilon} / (n-p)
}
U = ((n-p)/sigsqr)*sigsqr_hat # \chi^2-verteilte Zufallsvariable
```

Beispiel (2) Einfache lineare Regression

$$y \sim (X\beta, \sigma^2 I_n)$$

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n-p)$$



Allgemeine Theorie

Unabhängige und identisch normalverteilte Zufallsvariablen

Einfache lineare Regression

Chi-Quadrat-Zufallsvariablen

Frequentistische Schätzerverteilungen

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie das Theorem zum Betaparameterschätzer wieder.
2. Warum ist der Betaparameterschätzer ein Maximum-Likelihood-Schätzer?
3. Geben Sie das Theorem zum Varianzparameterschätzer wieder.
4. Warum ist der Varianzparameterschätzer kein Maximum-Likelihood-Schätzer?
5. Geben Sie die Definition der erklärten Daten und des Residuenvektors wieder.
6. Geben Sie das Theorem zu erklärten Daten und Residuenvektor als Matrixprodukte wieder.
7. Geben Sie die Parameterschätzer bei unabhängigen und identisch normalverteilten Zufallsvariablen an.
8. Geben Sie die Parameterschätzer bei einfacher linearer Regression an.
9. Geben Sie die Definition einer χ^2 -Zufallsvariable wieder.
10. Geben Sie die Definition einer nichtzentralen χ^2 -Zufallsvariable wieder.
11. Geben Sie das Theorem zur Verteilung des Betaparameterschätzers wieder.
12. Geben Sie das Theorem zur Verteilung des Varianzparameterschätzers wieder.

Rencher, Alvin C., and G. Bruce Schaalje. 2008. *Linear Models in Statistics*. 2nd ed. Hoboken, N.J: Wiley-Interscience.

Searle, S. R. 1971. *Linear Models*. Wiley Classics Library. New York, NY: Wiley.