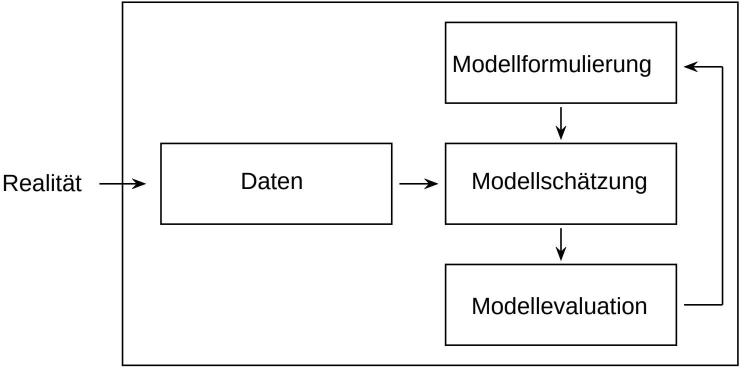
Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

Prof. Dr. Dirk Ostwald

(7) Modellevaluation

Naturwissenschaft



Überblick

Modellformulierung

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

Modellschätzung

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

Modellevaluation

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}, F = \frac{(\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon})/p_2}{\hat{\varepsilon}^T \hat{\varepsilon}/(n - p)} \tag{3}$$

Überblick

Standardprobleme Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für die wahren, aber unbekannten, Parameterwerte (oder eine Funktion derer) abzugeben, typischerweise basierend auf der Beobachtung einer Datenrealisierung.

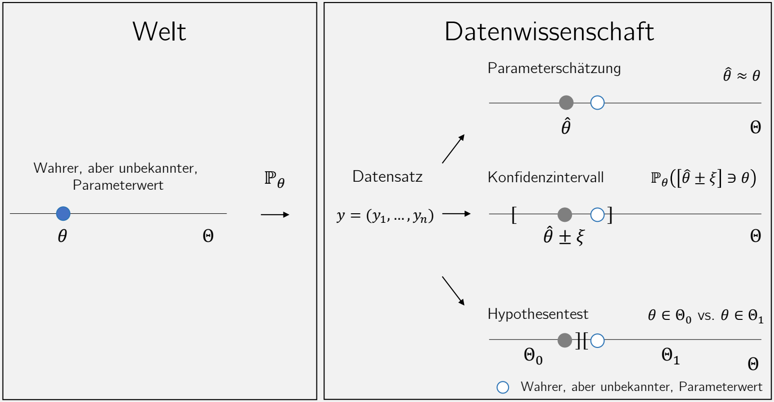
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst sinnvollen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert, sich in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Modell und Standardprobleme der Frequentistischen Inferenz



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

Überblick

Standardannahmen Frequentistischer Inferenz

Gegeben sei ein statistisches 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 basierend auf einem Modell generieren und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. den Betaparameterschätzer

$$
\begin{aligned}
&\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)^T \text{ mit } \hat{\beta}^{(1)} = (X^TX)^{-1}X^Ty^{(1)} \\
&\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)^T \text{ mit } \hat{\beta}^{(2)} = (X^TX)^{-1}X^Ty^{(2)} \\
&\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)^T \text{ mit } \hat{\beta}^{(3)} = (X^TX)^{-1}X^Ty^{(3)} \\
&\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)^T \text{ mit } \hat{\beta}^{(4)} = (X^TX)^{-1}X^Ty^{(4)} \\
&\text{Datensatz (5)} : y^{(5)} = ...
\end{aligned}
$$

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 $\hat{\beta} := (X^TX)^{-1}X^Ty$? 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.

Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Überblick

- Unter den Standadannahmen der Frequentistischen Inferenz kann man fragen, wie $\hat{\beta}$ und $\hat{\sigma}^2$ verteilt sind.
- Das erste Theorem dieses Abschnitts besagt, dass $\hat{\beta}$ multivariat normalverteilt ist
- Das zweite Theorem dieses Abschnitts besagt, dass eine skalierte Version von $\hat{\sigma}^2$ χ^2-verteilt ist.
- Diese Ergebnisse sind sehr ähnlich zu den Ergebnissen aus (7) Transformationen der Normalverteilung in WTFI.
- T- und F-Statistiken sind Funktionen von $\hat{\beta}$ und $\hat{\sigma}^2$.
- Die Verteilungen von $\hat{\beta}$ und $\hat{\sigma}^2$ bilden die Grundlage der Verteilungen der T- und F-Statistiken.
- In der konkreten Datenanalyse eines Datensatzes y berechnet man genau ein $\hat{\beta}$ und ein $\hat{\sigma}^2$.
- Probabilistische Aussagen über die aus diesen $\hat{\beta}$ und $\hat{\sigma}^2$ berechneten T- und F-Statistiken beziehen sich auf die Standardannahmen Frequentistischer Inferenz und weren vor diesem Hintergrund verstanden.

Theorem (Frequentistische Verteilung des Betaparameterschätzers)

Es sei

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

das ALM in generativer Form. Weiterhin sei

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

der Betaparameterschätzer. Dann gilt

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \tag{6}$$

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

$$\mathbb{V}(\hat{\beta}_i) = (\sigma^2 (X^T X)^{-1})_{ii} \text{ für } i = 1, ...p. \tag{7}$$

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

Schätzerverteilungen

Beweis

Das Theorem folgt direkt mit dem Theorem zur linearen Transformation von multivariaten Normalverteilung aus Einheit (4) Normalverteilungen. Speziell gilt hier:

$$\hat{\beta} \sim N\left((X^TX)^{-1}X^TX\beta, (X^TX)^{-1}X^T(\sigma^2 I_n)((X^TX)^{-1}X^T)^T\right). \tag{8}$$

Dabei können der Erwartungswertparameter wie folgt

$$(X^TX)^{-1}X^TX\beta = \beta, \tag{9}$$

und der Kovarianzmatrixparamter wie folgt

$$\begin{aligned}(X^TX)^{-1}X^T(\sigma^2 I_n)((X^TX)^{-1}X^T)^T &= (X^TX)^{-1}X^T(\sigma^2 I_n)X(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1}.\end{aligned} \tag{10}$$

Dabei hier die erste Gleichung aus der Tatsache, dass sowohl X^TX als auch ihre Inverse $(X^TX)^{-1}$ symmetrische Matrizen sind. Damit folgt dann aber sofort

$$\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1}). \tag{11}$$

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \quad (12)$$

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

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (13)$$

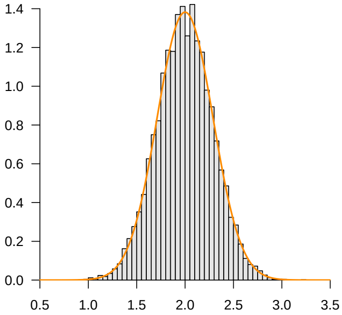
Das Stichprobenmittel von n unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter μ und Varianzparameter σ^2 ist also normalverteilt mit Erwartungswertparameter μ und Varianzparameter σ^2/n. Wir haben diese Tatsache bereits in Einheit (7) Transformationen der Normalverteilungen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

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

```
# Multivariate Normalverteilung
library(MASS)

# Modellformulierung
n        = 12                                     # Anzahl von Datenpunkten
p        = 1                                      # Anzahl von Betparametern
X        = matrix(rep(1,n), nrow = n)             # 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 des n-dimensionalen ZVs
beta_hat = rep(NaN,nsim)                          # \hat{\beta} Realisierungsarray
for(i in 1:nsim){
  y           = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
  beta_hat[i] = solve(t(X) %*% X) %*% t(X) %*% y   # \hat{\beta} = (X^T)X^{-1}X^Ty
}
```

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

$$N\left(\hat{\mu}; \mu, \frac{\sigma^2}{n}\right)$$



Beispiel (2) Einfache lineare Regression

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ with } \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2, \sigma^2 > 0. \tag{14}$$

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

$$\sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{s_{xx}} \begin{pmatrix} \frac{s_{xx}}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \text{ mit } s_{xx} := \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{15}$$

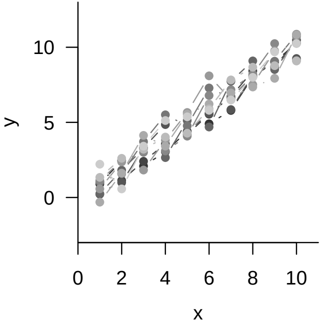
Die Varianz des Offsetparameterschätzers hängt also sowohl von der Summe der quadrierten Differenzen und dem Stichprobenmittel der unabhängigen Variablen $x_1, ..., x_n$ ab, wohingegen die Varianz des Steigungsparameterschätzers nur von der Summe der quadrierten Differenzen der $x_1, ..., x_n$ abhängt.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

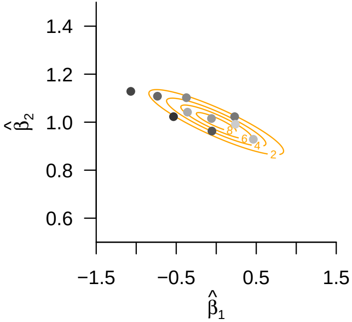
Beispiel (2) Einfache lineare Regression

```
# Multivariate Normalverteilung und Matrizenrechnung
library(MASS)
library(matlib)

# Modellformulierung
n        = 10                                          # Anzahl von Datenpunkten
p        = 2                                           # Anzahl von Betparametern
x        = 1:n                                         # Prädiktorwerte
X        = matrix(c(rep(1,n),x), nrow = n)             # Designmatrix
I_n      = diag(n)                                     # n x n Einheitsmatrix
beta     = matrix(c(0,1), nrow = p)                    # wahrer,aber unbekannter,Betaparameter
sigsqr   = .5                                          # wahrer,aber unbekannter,Varianzparameter

# Frequentistische Simulation
nsim     = 10                                          # Anzahl Realisierungen n-dimensionaler ZV
y        = matrix(rep(NaN,n*nsim), nrow = n)           # y Realisierungsarray
beta_hat = matrix(rep(NaN,p*nsim), nrow = p)           # \hat{\beta} Realisierungsarray
for(i in 1:nsim){
  y[,i]        = mvrnorm(1, X %*% beta, sigsqr*I_n)    # eine Realisierung n-dimensionaler ZV
  beta_hat[,i] = inv(t(X) %*% X) %*% t(X) %*% y[,i]    # \hat{\beta} = (X^T)X^{-1}X^Ty
}
```

Beispiel (2) Einfache lineare Regression

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





Theorem (Frequentistische Verteilung des Varianzparameterschätzers)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{16}$$

das ALM in generative Form. Weiterhin sei

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \tag{17}$$

der Varianzparameterschätzer. Dann gilt

$$\frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-p) \tag{18}$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Da es sich bei $(y - X\hat{\beta})^T (y - X\hat{\beta})$ um eine Summe normalverteilter Zufallsvariablen handelt, liegt die χ^2-Verteilung im Lichte der χ^2 Transformation aus Einheit (7) Transformationen der Normalverteilung in Wahrscheinlichkeitstheorie und Frequentistische Inferenz nahe.

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \tag{19}$$

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.

In Einheit (11) Konfidenzintervalle von Wahrscheinlichkeitstheorie und Frequentistische Inferenz hatten wir für den Fall von n unabhängig und identisch normalverteilten Zufallsvariablen die Statistik

$$U := \frac{n-1}{\sigma^2} s^2 \tag{20}$$

definiert und festgehalten, dass

$$U \sim \chi^2(n-1). \tag{21}$$

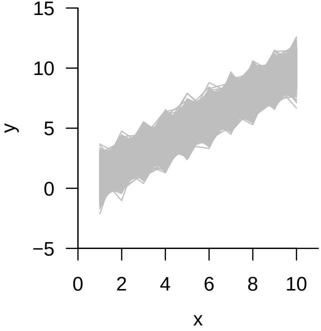
Offenbar ist U für $p = 1$ mit der im obigen Theorem betrachten Zufallsvariable $\frac{n-p}{\sigma^2}\hat{\sigma}^2$ identisch.

Beispiel (2) Einfache lineare Regression

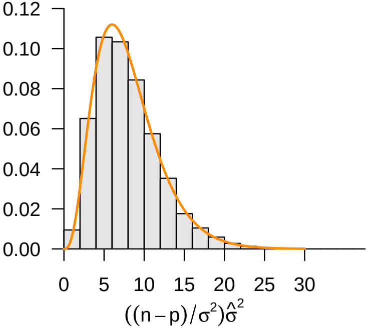
```
# Libraries
library(MASS)                                       # multivariate Normalverteilung
library(matlib)                                     # Matrizenrechnung

# Modellformulierung
n          = 10                                     # Anzahl von Datenpunkten
p          = 2                                      # Anzahl von Betparametern
x          = 1:n                                    # Prädiktorwerte
X          = matrix(c(rep(1,n),x), nrow = n)        # Designmatrix
I_n        = diag(n)                                # n x n Einheitsmatrix
beta       = matrix(c(0,1), nrow = p)               # wahrer,aber unbekannter,Betaparameter
sigsqr     = .5                                     # wahrer,aber unbekannter,Varianzparameter

# Frequentistische Simulation
nsim       = 1e3                                     # Anzahl Realisierungen n-dimensionaler ZV
y          = matrix(rep(NaN,n*nsim), nrow = n)      # y Realisierungsarray
beta_hat   = matrix(rep(NaN,p*nsim), nrow = p)      # \hat{\beta} Realisierungsarray
sigsqr_hat = rep(NaN, nsim)                         # \hat{\sigma}^2 Realisierungsarray
for(i in 1:nsim){
  y[,i]         = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung n-dimensionaler ZV
  beta_hat[,i]  = inv(t(X) %*% X) %*% t(X) %*% y[,i] # \hat{\beta}    = (X^T)X^{-1}X^Ty
  eps_hat       = y[,i] - X %*% beta_hat[,i]        # \hat{\eps}     = y-X\hat{\beta}
  sigsqr_hat[i] = (t(eps_hat) %*% eps_hat)/(n-p)    # \hat{\sigma}^2 =\hat{\eps}^T\hat{\eps}/(n-p)
}
U = ((n-p)/sigsqr)*sigsqr_hat                       # \chi^2 verteilte Zufallsvariable
```

Beispiel (2) Einfache lineare Regression

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



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



Überblick

- T-Statistiken quantifizieren die geschätzten Effekte in $\hat{\beta}$ in bezug auf die Residualvariabilität $\hat{\sigma}^2$. T-Statistiken sind also als Signal-zu-Rauschen Verhältnisse (signal-to-noise ratios) zu verstehen.

- T-Statistiken erlauben die Frequentistische Evaluation von linearen Kombinationen der $\hat{\beta}$ Komponenten. Eine einfache Linearkombination von $\hat{\beta}$ ist das Skalarprodukt meinem Einheitsvektor. Solch eine Linearkombination "extrahiert" eine Komponente von $\hat{\beta}$ und erlaubt damit ihre Frequentistische Evaluation.

- T-Teststatistiken erlauben weiterhin die Evaluation von Linearkombinationen von $\hat{\beta}$ im Sinne Frequentistischer Hypothesentests. Wir betrachten hier zunächst die nur die funktionale Form von T-Teststatistiken und ihre Frequentistische Verteilung. Der Einsatz von T-Teststatistiken im Rahmen von Einstich- und Zweistichproben T-Tests ist Inhalt von (9) T-Tests.

Definition (t-Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

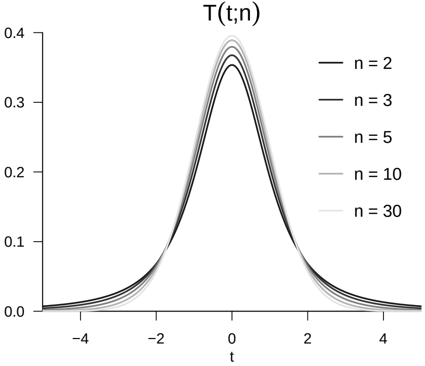
$$p : \mathbb{R} \to \mathbb{R}_{>0}, t \mapsto p(t) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \tag{22}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass T einer t-Verteilung mit n Freiheitsgraden unterliegt und nennen T eine t-Zufallsvariable mit n Freiheitsgraden. Wir kürzen dies mit $T \sim t(n)$ ab. Die WDF einer t-Zufallsvariable bezeichnen wir mit

$$T(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \tag{23}$$

Bemerkungen

- Die Verteilung ist um 0 symmetrisch
- Steigendes n verschiebt Wahrscheinlichkeitsmasse aus den Ausläufen zum Zentrum.
- Ab $n = 30$ gilt $T(t; n) \approx N(0, 1)$.


T(t;n)
0.4
0.3
0.2
0.1
0.0
−4
−2
0
2
4
t
n = 2
n = 3
n = 5
n = 10
n = 30

Theorem (T-Transformation)

$Z \sim N(0,1)$ sei eine Z-Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2-Zufallsvariable mit n Freiheitsgraden, und Z und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

$$T := \frac{Z}{\sqrt{U/n}} \tag{24}$$

eine t-verteilte Zufallsvariable mit n Freiheitsgraden, es gilt also $T \sim t(n)$.

Bemerkungen

- Das Theorem geht auf Student (1908) zurück.
- Das Theorem ist das zentrale Resultat der Frequentistischen Statistik.
- Zabell (2008) gibt hierzu einen historischen Überblick.
- Die T-Statistik und T-Tests sind wichtige Anwendungsfälle.

Definition (Nichtzentrale t-Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

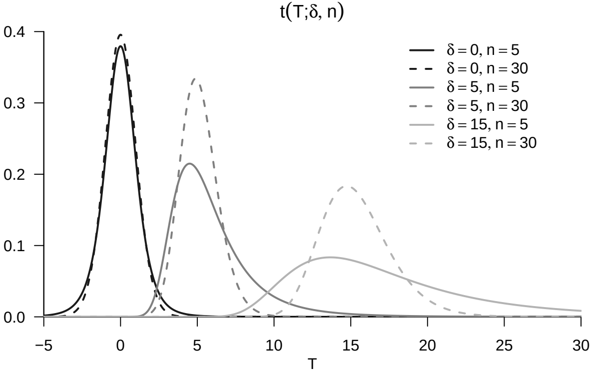
$$p : \mathbb{R} \to \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n}{2}\right)(n\pi)^{\frac{1}{2}}} \times \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2}\left(t\left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right) d\tau. \tag{25}$$

Dann sagen wir, dass T einer nichtzentralen t-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n unterliegt und nennen T eine *nichtzentrale t-Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n*. Wir kürzen dies mit $t(\delta, n)$ ab. Die WDF einer nichtzentralen t-Zufallsvariable bezeichnen wir mit $t(T; \delta, n)$. Die KVF und inverse KVF einer nichtzentralen t-Zufallsvariable bezeichnen wir mit $\psi(\cdot; \delta, n)$ und $\psi^{-1}(\cdot; \delta, n)$, respektive.

Bemerkungen

- Eine nichtzentrale t-Zufallsvariable mit $\delta = 0$ ist eine t-Zufallsvariable.
- Es gilt also $t(T; 0, n) = t(T; n)$.
- Die funktionale Form der WDF findet sich zum Beispiel in Lehmann (1986), Seite 254, Gl. (80).

Wahrscheinlichkeitsdichtefunktionen nichtzentraler t-Verteilungen



Theorem (Nichtzentrale T-Transformation)

$X \sim N(\mu, 1)$ sei eine normalverteilte Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 Zufallsvariable mit Freiheitsgradparameter n, und X und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

$$T := \frac{X}{\sqrt{U/n}} \tag{26}$$

eine nichtzentrale t-Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter n, es gilt also $T \sim t(\mu, n)$

Bemerkung

- Wir verzichten auf einen Beweis.

Definition (T-Statistik)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{27}$$

das ALM in generativer Form. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \tag{28}$$

die Betaparameter- und Varianzparameterschätzer, respektive. Dann ist für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ die *T-Statistik* definiert als

$$T := \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \tag{29}$$

Bemerkungen

- Die T-Statistik hängt via $\hat{\beta}$ und $\hat{\sigma}^2$ von den Daten y ab.
- Der Kontrastgewichtsvektor projiziert $\hat{\beta}$ auf einen Skalar $c^T \hat{\beta}$.
- Die folgende Intuition zur T-Statistik ist im Allgemeinen hilfreich

$$T = \frac{\text{Geschätzte Effektstärke}}{\text{Geschätzte stichprobenumfangskalierte Datenvariabilität}} \tag{30}$$

- Intuitiv repräsentiert eine T-Statistik also ein Signal-zu-Rauschen Verhältnis.

T-Statistiken

Beispiel (1) Unabhängige und identisch normalverteilte Zufallsvariablen

Es sei

$$y \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R} \text{ und } \sigma^2 > 0. \tag{31}$$

das ALM Szenario unabhängiger und identisch normalverteilter Zufallsvariablen und es sei $c := 1$. Dann gilt für die T-Statistik

$$T = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{y}}{\sqrt{s_y^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \frac{\bar{y}}{s_y} \tag{32}$$

was der Einstichproben-T-Teststatistik für den Fall $\mu_0 = 0$ entspricht (vgl. Einheit (13) Einstichproben-T-Tests in Wahrscheinlichkeitstheorie und Frequentistische Inferenz und Einheit (9) T-Tests in Allgemeines Lineares Modell). Die hier betrachtete T-Statistik nimmt hohe Werte für hohe Werte von $\bar{y}$ (Effekt), kleine Werte von s_y^2 (Datenvariabilität) und hohe Werte von n (Stichprobenumfang) an.

Eine beliebte Definition in diesem Zusammenhang ist *Cohen's d* als *Effektstärkenmaß*. Es gilt

$$d := \frac{\bar{y}}{s_y}, \tag{33}$$

so dass

$$T = \sqrt{n} d \text{ bzw. } d = \frac{1}{\sqrt{n}} T. \tag{34}$$

Cohen's d ist also ein stichprobenumfangunabhängiges Signal-zu-Rauschen Verhältnis.

Theorem (T-Teststatistik)

Es sei

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{35}$$

das ALM in generativer Form. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T y \text{ und } \hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n-p} \tag{36}$$

die Betaparameter- und Varianzparameterschätzer, respektive. Schließlich sei für einen *Kontrastgewichtsvektor* $c \in \mathbb{R}^p$ und einen *Nullhypothesenbetaparameter* $\beta_0 \in \mathbb{R}^p$ die *T-Teststatistik* definiert als

$$T := \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \tag{37}$$

Dann gilt

$$T \sim t(\delta, n-p) \text{ mit } \delta := \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} \tag{38}$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- T ist eine Funktion der Parameterschätzer, δ ist eine Funktion der wahren, aber unbekannten, Parameter
- Für $c^T \beta = c^T \beta_0$, also bei Zutreffen der Nullhypothese, gilt $\delta = 0$ und damit $T \sim t(n-p)$.
- Für $c^T \beta \neq c^T \beta_0$ kann die Verteilung von T zur Herleitung von Powerfunktionen benutzt werden..

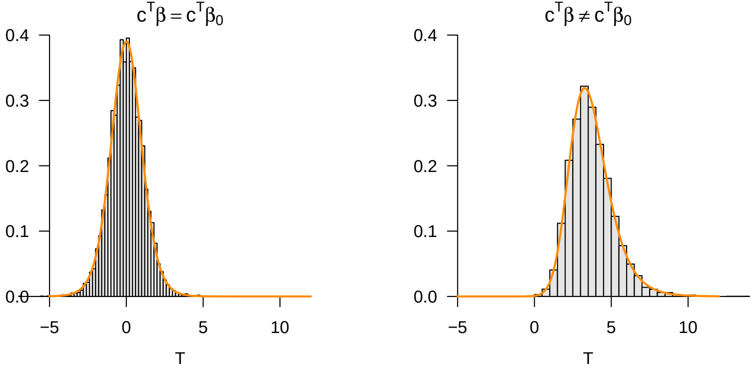
Simulation (1) Unabhängig und identische normalverteilte Zufallsvariablen

Wahre, aber unbekannte, Hypothesenszenarien $c^T\beta = c^T\beta_0$ und $c^T\beta \neq c^T\beta_0$

```
# Libraries
library(MASS)                                                   # multivariate Normalverteilung

# Modellformulierung
n          = 12                                                  # Anzahl von Datenpunkten
p          = 1                                                   # Anzahl von Betaparametern
X          = matrix(c(rep(1,n)), nrow = n)                       # Designmatrix
I_n        = diag(n)                                             # Einheitsmatrix
beta       = c(0,1)                                              # wahre , aber unbekannte , Betaparameter
nscn       = length(beta)                                        # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr     = 1                                                   # wahrer, aber unbekannter, Varianzparameter
c          = 1                                                   # Kontrastvektor von Interessse
beta_0     = 0                                                   # Nullhypothesenbetaparameter

# Frequentistische Simulation
nsim       = 1e4                                                 # Anzahl Simulationen
delta      = rep(NaN, nscn)                                      # Anzahl Nichtzentralitätsparameter
Tee        = matrix(rep(NaN, nscn*nsim), ncol = nscn)            # T-Teststatistik Realisierungsarray
for(s in 1:nscn){                                                # Hypothesenszenarien
  delta[s]    = ((t(c) %*% beta[s] - t(c) %*% beta_0)/           # Nichtzentralitätsparameter
                sqrt(sigsqr*t(c)%*%solve(t(X)%*%X)%*%c))
  for(i in 1:nsim){                                              # Simulationsiterationen
    y          = mvrnorm(1, X %*% beta[s], sigsqr*I_n)           # y
    beta_hat   = solve(t(X) %*% X) %*% t(X) %*% y                # \hat{\beta}
    eps_hat    = y - X %*% beta_hat                              # \hat{\eps}
    sigsqr_hat = (t(eps_hat) %*% eps_hat)/(n-p)                  # \hat{\sigma}^2
    Tee[i,s]   = ((t(c) %*% beta_hat - t(c) %*% beta_0)/         # T
                  sqrt(sigsqr_hat*t(c)%*%solve(t(X)%*%X)%*%c))
  }
}
```

Simulation (1) Unabhängig und identische normalverteilte Zufallsvariablen

Wahre, aber unbekannte, Hypothesenszenarien $c^T\beta = c^T\beta_0$ und $c^T\beta \neq c^T\beta_0$



T-Statistiken

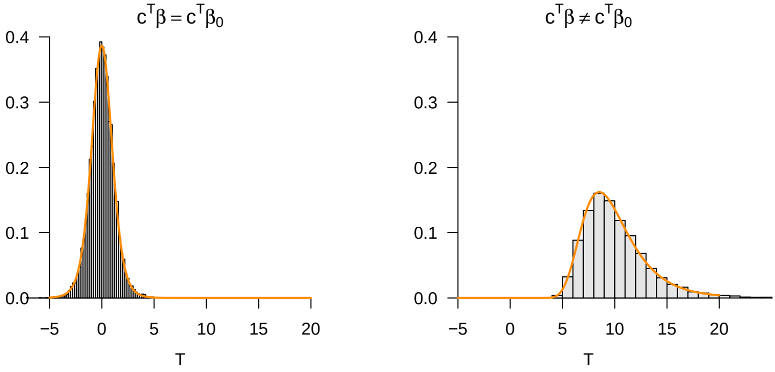
Simulation (1) Einfache lineare Regression

Wahre, aber unbekannte, Hypothesenszenarien $c^T\beta = c^T\beta_0$ und $c^T\beta \neq c^T\beta_0$

```r
# Libraries
library(MASS)                                               # multivariate Normalverteilung
library(matlib)                                             # Matrizenrechnung

# Modellformulierung
n           = 10                                            # Anzahl von Datenpunkten
p           = 2                                             # Anzahl von Betaparametern
x           = 1:n                                           # Prädiktorwerte
X           = matrix(c(rep(1,n),x), ncol = p)               # Designmatrix
I_n         = diag(n)                                       # Einheitsmatrix
beta        = matrix(c(1,0,1,1), nrow = 2)                  # wahre , aber unbekannte , Betaparameter
nscn        = ncol(beta)                                    # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr      = 1                                             # wahrer, aber unbekannter, Varianzparameter
c           = matrix(c(0,1), nrow = 2)                      # Kontrastvektor von Interessse
beta_0      = matrix(c(0,0), nrow = 2)                      # Nullhypothesenbetaparameter

# Frequentistische Simulation
nsim        = 1e4                                           # Anzahl Simulationen
delta       = rep(NaN, nscn)                                # Anzahl Nichtzentralitätsparameter
Tee         = matrix(rep(NaN, nscn*nsim), ncol = nscn)      # T-Teststatistik Realisierungsarray
for(s in 1:nscn){                                           # Hypothesenszenarien
  delta[s]     = ((t(c) %*% beta[,s] - t(c) %*% beta_0)/    # Nichtzentralitätsparameter
                 sqrt(sigsqr*t(c)%*%inv(t(X)%*%X)%*%c))
  for(i in 1:nsim){                                         # Simulationsiterationen
    y          = mvrnorm(1, X %*% beta[,s], sigsqr*I_n)     # y
    beta_hat   = inv(t(X) %*% X) %*% t(X) %*% y             # \hat{\beta}
    eps_hat    = y - X %*% beta_hat                         # \hat{\eps}
    sigsqr_hat = (t(eps_hat) %*% eps_hat)/(n-p)             # \hat{\sigma}^2
    Tee[i,s]   = ((t(c) %*% beta_hat - t(c) %*% beta_0)/    # T
                  sqrt(sigsqr_hat*t(c)%*%inv(t(X)%*%X)%*%c))
  }
}
```

Simulation (1) Einfache lineare Regression

Wahre, aber unbekannte, Hypothesenszenarien $c^T\beta = c^T\beta_0$ und $c^T\beta \neq c^T\beta_0$



Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Überblick

- F-Statistiken können als Modellvergleichskriterien zwischen ALMs verstanden werden. Im Gegensatz zu T-Statistiken kann das Ziel der Berechnung von F-Statistiken sein, nicht nur Linearkombinationen von Betaparameterschätzwerten probabilistisch zu evaluieren, sondern die Modellanpassung an einen Datensatz insgesamt zu evaluieren.
- Wir motivieren die funktionale Form der F-Statistik als Residualquadratsummendifferenz hier vor dem Hintergrund eine Likelihood-Quotienten zum Modellvergleich. Die Modellvergleichskapazität von F-Statistiken ist allerdings etwas beschränkt, da sich die F-Statistik auf ALMs und insbesondere geschachtelte ALMs bezieht, in denen ein Modell Bestandteil eines anderen Modells ist.
- F-Statistiken bilden die Grundlage für Hypothesentests im Rahmen der Varianzanalyse (Einheiten (10) Einfaktorielle Varianzanalyse, (11) Zweifaktorielle Varianzanalyse, (13) Kovarianzanalyse). Der Einsatz von F-Statistiken ist aber *perse* nicht auf Varianzanalysen beschränkt.

Definition (f-Zufallsvariable)

X sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

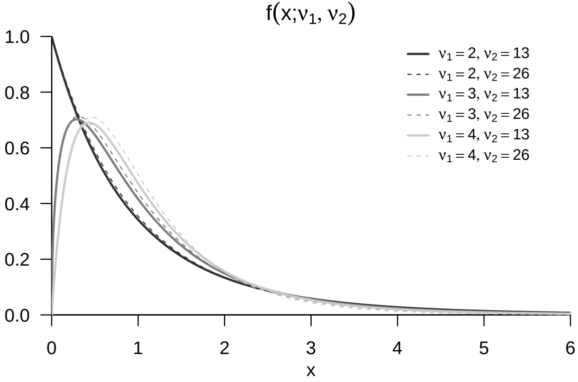
$$p_X : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto p_X(x) := \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1}{2}-1}}{(\nu_1 x + \nu_2)^{\frac{\nu_1+\nu_2}{2}}}, \tag{39}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer f-Verteilung mit Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen X eine f-Zufallsvariable mit Freiheitsgradparametern ν_1 und ν_2. Wir kürzen dies mit $X \sim f(\nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f-Zufallsvariable bezeichnen wir mit $f(x; \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer f-Zufallsvariable bezeichnen wir mit $\varphi(x; \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer f-Zufallsvariable bezeichnen wir mit $\varphi^{-1}(x; \nu_1, \nu_2)$.

Bemerkungen

- f-Zufallsvariablen sind nach Ronald A. Fisher benannt.
- George W. Snedecor hat die KVF der F-Verteilung wohl 1934 basierend auf Arbeiten von Fisher tabuliert.

Wahrscheinlichkeitsdichtefunktionen von f-Verteilungen



Theorem (F-Transformation)

$V \sim \chi^2(\nu_1)$ und $W \sim \chi^2(\nu_2)$ seien zwei unabhängige χ^2-Zufallfsvariablen mit ν_1 und ν_2 Freiheitsgraden, respektive. Dann ist die Zufallsvariable

$$F := \frac{V/\nu_1}{W/\nu_2} \tag{40}$$

eine f-verteilte Zufallsvariable mit ν_1, ν_2 Freiheitsgraden, es gilt also $F \sim f(\nu_1, \nu_2)$.

Bemerkungen

- Wir verzichten auf einen Beweis.
- Das Theorem kann bewiesen werden, in dem man zunächst ein Transformationstheorem für Quotienten von Zufallsvariablen mithilfe des multivariaten Transformationstheorems und Marginalisierung herleitet und dieses Theorem dann auf die WDF von χ^2-verteilten ZVen anwendet. Dabei ist die Regel zur Integration durch Substitution von zentraler Bedeutung.

Definition (Nichtzentrale f-Zufallsvariable)

X sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und Wahrscheinlichkeitsdichtefunktion

$$p_X : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto$$

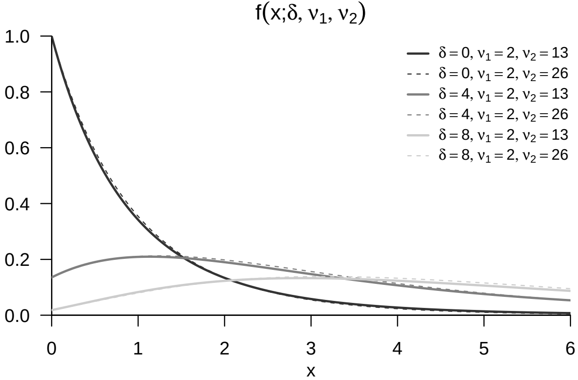
$$p_X(x) := \sum_{k=0}^{\infty} \frac{e^{-\delta/2}(\delta/2)^k}{\frac{\Gamma(\nu_2/2)\Gamma(\nu_1/2+k)}{\Gamma(\nu_2/2+\nu_1/2+k)} k!} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2+k} \left(\frac{\nu_2}{\nu_2+\nu_1 x}\right)^{(\nu_1+\nu_2)/2+k} x^{\nu_1/2-1+k} \tag{41}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer nichtzentralen f-Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen X eine nichtzentrale f-Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2. Wir kürzen dies mit $X \sim f(\delta, \nu_1, \nu_2)$ ab. Die Wahrscheinlichkeitsdichtefunktion (WDF) einer f-Zufallsvariable bezeichnen wir mit $f(x; \delta, \nu_1, \nu_2)$, die kumulative Verteilungsfunktion (KVF) einer nichtzentralen f-Zufallsvariable bezeichnen wir mit $\varphi(x; \delta, \nu_1, \nu_2)$, und die inverse kumulative Verteilungsfunktion einer nichtzentralen f-Zufallsvariable bezeichnen wir mit $\varphi^{-1}(x; \delta, \nu_1, \nu_2)$.

Bemerkungen

- Es gilt $f(0, \nu_1, \nu_2) = f(\nu_1, \nu_2)$.

WDFen von nichtzentralen f-Verteilungen



Theorem (Nichtzentrale F-Transformation)

$V \sim \chi^2(\delta, \nu_1)$ und $W \sim \chi^2(\nu_2)$ seien eine nichtzentrale χ^2 Zufallsvariable und eine χ^2-Zufallfsvariablen mit ν_1 und ν_2 Freiheitsgraden, respektive, und V und W seien unabhängig. Dann ist die Zufallsvariable

$$F := \frac{V/\nu_1}{W/\nu_2} \tag{42}$$

eine nichtzentral f-verteilte Zufallsvariable mit Nichtzentralitätsparameter δ und ν_1, ν_2 Freiheitsgraden, es gilt also $F \sim f(\delta, \nu_1, \nu_2)$.

Bemerkungen

- Wir verzichten auf einen Beweis.

Motivation der F-Statistik | Modellpartitionierung und Likelihood Ratios

Definition (Vollständiges und reduziertes Modell)

Für $p > 1$ mit $p = p_1 + p_2$ seien

$$X := \begin{pmatrix} X_1 & X_2 \end{pmatrix} \in \mathbb{R}^{n \times p} \text{ mit } X_1 \in \mathbb{R}^{n \times p_1} \text{ und } X_2 \in \mathbb{R}^{n \times p_2}, \tag{43}$$

sowie

$$\beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^{p} \text{ mit } \beta_1 \in \mathbb{R}^{p_1} \text{ und } \beta_2 \in \mathbb{R}^{p_2} \tag{44}$$

Partitionierungen einer $n \times p$ Designmatrix und eines p-dimensionalen Betaparametervektors. Dann nennen wir

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{45}$$

das *vollständige Modell* und

$$y = X_1\beta_1 + \varepsilon_1 \text{ mit } \varepsilon_1 \sim N(0_n, \sigma^2 I_n) \tag{46}$$

das *reduzierte Modell* und sprechen von einer *Partionierung eines (vollständigen) Modells*.

Bemerkungen

- Man sagt auch, dass das reduzierte Modell im vollständigen Modell *geschachtelt (nested)* ist.

Motivation der F-Statistik | Modellpartitionierung und Likelihood Ratios

Definition (Likelihood-Quotient von vollständigem und reduzierten Modell)

Für $p = p_1 + p_2, p > 1$ sei eine Partitionierung eines vollständigen ALMs gegeben und $\sigma^2 > 0$ als bekannt vorausgesetzt. Weiterhin seien

$$L : \mathbb{R}^p \to \mathbb{R}_{>0}, \beta \mapsto L(\beta) := p_\beta(\upsilon) = N(\upsilon; X\beta, \sigma^2 I_n) \tag{47}$$

und

$$L_1 : \mathbb{R}^{p_1} \to \mathbb{R}_{>0}, \beta_1 \mapsto L(\beta_1) := p_{\beta_1}(\upsilon) = N(\upsilon; X_1\beta_1, \sigma^2 I_n) \tag{48}$$

die Likelihood-Funktionen von vollständigem und reduzierten Modell, respektive. Dann ist für einen Datenvektor $\upsilon \in \mathbb{R}^n$ der *Likelihood-Quotient von vollständigem und reduziertem Modell* gegeben als

$$:= \frac{\max_{\beta \in \mathbb{R}^p} p_\beta(\upsilon)}{\max_{\beta_1 \in \mathbb{R}^{p_1}} p_{\beta_1}(\upsilon)} = \frac{\max_{\beta \in \mathbb{R}^p} L(\beta)}{\max_{\beta_1 \in \mathbb{R}^{p_1}} L_1(\beta_1)} = \frac{L(\hat{\beta})}{L_1(\hat{\beta}_1)} \tag{49}$$

F-Statistiken

Motivation der F-Statistik | Modellpartitionierung und Likelihood Ratios

Bemerkungen

- Der Likelihood-Quotient setzt die Wahrscheinlichkeitsdichten eines beobachteten Datenssatzes $\upsilon \in \mathbb{R}^n$ unter dem vollständigem und dem reduzierten Modell nach Optimierung der jeweiligen Modellbetaparameter ins Verhältnis.
- Ein hoher Wert des Likelihood-Quotienten entspricht einer höhereren Wahrscheinlichkeit(sdichte) des beobachteten Datensatzes $\upsilon \in \mathbb{R}^n$ unter dem optimierten vollständigem Modell und vice versa.
- Die Wahrscheinlichkeiten beobachtbarerer Daten nach Modellschätzung unter verschiedenen Modellen zu betrachten ist ein allgemeines Vorgehen zum Modellvergleich. Letztlich erlaubt dieses Vorgehen, verschiedene wissenschaftliche Theorien (= Modelle) über die Genese beobachtbarer Daten quantitativ zu vergleichen und die damit verbundene Unsicherheit zu quantifizieren.
- Modellvergleiche sind ein zentrales Thema in der Bayesianischen Statistik, die die Logik von Likelihood-Quotienten zum Beispiel unter den Begriffen der Bayes Factors oder der des Bayesian Information Criterions auf arbiträre probabilistische Modelle, d.h. nicht nur geschachtelte ALMs generalisiert.
- Allerdings sind, wie hier gesehen, Modellvergleiche auch im Rahmen der Frequentistischen Statistik möglich und sinnvoll, Modellvergleiche sind also kein Alleinstellungsmerkmal der Bayesianischen gegenüber der Frequentistischen Statistik.

F-Statistiken

Motivation der F-Statistik | Modellpartitionierung und Likelihood Ratios

Theorem (Likelihood-Quotient und Residualquadratsummendifferenz)

Für $p = p_1 + p_2, p > 1$ sei eine Partitionierung eines vollständigen ALMs gegeben und $\sigma^2 > 0$ als bekannt vorausgesetzt. Weiterhin seien

$$\hat{\varepsilon} := y - X\hat{\beta} \text{ und } \hat{\varepsilon}_1 := y - X_1\hat{\beta}_1 \tag{50}$$

die Residuenvektoren des vollständigen und es reduzierten Modells, respektive. Dann gilt

$$\mathrm{ln} = \frac{\hat{\varepsilon}_1^T\hat{\varepsilon}_1 - \hat{\varepsilon}^T\hat{\varepsilon}}{2\sigma^2} \tag{51}$$

Bemerkungen

- ln und die F-Statistik sind nicht identisch, aber eng verwandt.
- $\hat{\varepsilon}_1^T\hat{\varepsilon}_1$ ist die Residualquadratsumme (RQS), die bei Schätzung des reduzierten Modells von y resultiert.
- $\hat{\varepsilon}^T\hat{\varepsilon}$ ist die RQS, die bei Schätzung des vollständigen Modells anhand desselben y resultiert.
- Für $\beta_2 := 0_{p_2}$ sind vollständiges und reduziertes Modell identisch, es muss also $\hat{\varepsilon}_1^T\hat{\varepsilon}_1 \geq \hat{\varepsilon}^T\hat{\varepsilon}$ gelten.
- $\hat{\varepsilon}_1^T\hat{\varepsilon}_1 - \hat{\varepsilon}^T\hat{\varepsilon}$ misst, inwieweit für $\beta_2 \neq 0_{p_2}$ die Spalten von X_2 die RQS verringern.
- Ist $\hat{\varepsilon}_1^T\hat{\varepsilon}_1$ groß, aber $\hat{\varepsilon}^T\hat{\varepsilon}$ klein, so ist ln groß und der Beitrag der X_2-Spalten zur Datenerklärung groß.
- Sind $\hat{\varepsilon}_1^T\hat{\varepsilon}_1$ und $\hat{\varepsilon}^T\hat{\varepsilon}$ gleich groß, also ln klein, ist und der Beitrag der X_2-Spalten zur Datenerklärung klein.
- "Groß" und "klein" von $\hat{\varepsilon}_1^T\hat{\varepsilon}_1$ und $\hat{\varepsilon}^T\hat{\varepsilon}$ werden in Relation zur doppelten Fehlervektorvarianz σ^2 gemessen.

F-Statistiken

Definition (F-Statistik)

Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{52}$$

mit der Partitionierung

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 \in \mathbb{R}^{n \times p_1}, X_2 \in \mathbb{R}^{n \times p_2}, \text{ und } \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 \in \mathbb{R}^{p_1}, \beta_2 \in \mathbb{R}^{p_2}, \tag{53}$$

mit $p = p_1 + p_2$ gegeben. Weiterhin seien mit

$$\hat{\beta}_1 := (X_1^T X_1)^{-1} X_1^T y \text{ und } \hat{\beta} := (X^T X)^{-1} X^T y \tag{54}$$

die Residuenvektoren

$$\hat{\varepsilon}_1 := y - X_1 \hat{\beta}_1 \text{ und } \hat{\varepsilon} := y - X\hat{\beta} \tag{55}$$

definiert. Dann ist die F-Statistik definiert als

$$F := \frac{\frac{\hat{\varepsilon}_1^T \hat{\varepsilon}_1 - \hat{\varepsilon}^T \hat{\varepsilon}}{p_2}}{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p}} \tag{56}$$

F-Statistiken

Bemerkungen

- Der Zähler der F-Statistik
$$\frac{\hat{\varepsilon}_1^T\hat{\varepsilon}_1 - \hat{\varepsilon}^T\hat{\varepsilon}}{p_2} \tag{57}$$
misst, inwieweit die p_2 Regressoren in X_2 die Residualquadratsumme reduzieren und zwar im Verhältnis zur Anzahl dieser Regressoren. Das heißt, dass bei gleicher Größe der Residualquadratsummenreduktion (und gleichem Nenner) ein größerer F Wert resultiert, wenn diese durch weniger zusätzliche Regressoren resultiert, also p_2 klein ist (und vice versa). Im Sinne der Anzahl der Spalten von X und der entsprechenden Komponenten von β favorisiert die F-Statistik also weniger "komplexe" Modelle.
- Für den Nenner der F-Statistik gilt
$$\frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-p} = \hat{\sigma}^2, \tag{58}$$
wobei $\hat{\sigma}^2$ hier der aufgrund des vollständigen Modells geschätzte Schätzer von σ^2 ist. Werden die Daten tatsächlich unter dem reduzierten Modell generiert, so kann das vollständige Modell dies durch $\hat{\beta}_2 \approx 0_{p_2}$ abbilden und erreicht eine ähnliche σ^2 Schätzung wie das reduzierte Modell. Werden die Daten de-facto unter dem vollständigem Modell generiert, so ist $\hat{\varepsilon}^T\hat{\varepsilon}/(n-p)$ ein besserer Schätzer von σ^2 als $\hat{\varepsilon}_1^T\hat{\varepsilon_1}/(n-p)$, da sich für diesen Datenvariabilität, die nicht durch die p_1 Regressoren in X_1 erklärt wird, in der Schätzung von σ^2 widerspiegeln würde. Der Nenner der F-Statistik ist also in beiden Fällen der sinnvollere Schätzer von σ^2.
- Zusammengenommen misst die F-Statistik also die Residualquadratsummenreduktion durch die p_2 Regressoren in X_2 gegenüber den p_1 Regressoren in X_1 pro Datenvariabilitäts (σ^2)- und Regressor (p_2)-Einheit.

F-Statistiken

Beispiel (1) Einfache lineare Regression

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 := 1_n, X_2 := (x_1, ..., x_n)^T, \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 = \text{Interzept}, \beta_2 = \text{Steigung} \quad (59)$$

```
# Libraries
library(MASS)                                            # Multivariate Normalverteilung

# Modellformulierung
nmod   = 2                                               # Anzahl Modelle
n      = 10                                              # Anzahl Datenpunkte
p      = 2                                               # Anzahl Betaparameter
p_1    = 1                                               # Anzahl Betaparameter reduziertes Modell
p_2    = 1                                               # Anzahl zusätzlicher Betaparameter vollständiges Modell
p      = p_1 + p_2                                       # Anzahl Betaparameter im vollständigem Modell
x      = 1:n                                             # Prädiktorwerte
X      = matrix(c(rep(1,n),x), nrow = n)                 # Designmatrix des vollständigen Modells
X_1    = X[,1]                                           # Designmatrix des reduzierten Modells
I_n    = diag(n)                                         # n x n Einheitsmatrix
beta   = matrix(c(1,0,1,.5), nrow = 2)                   # wahre , aber unbekannte , Betaparameter
nscn   = ncol(beta)                                      # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr = 1                                               # wahrer, aber unbekannter, Varianzparameter

# Modellsimulation und Evaluierung
Eff    = matrix(rep(NaN, nscn), nrow = nscn)             # F-Statistiik Realisierungsarray
for(s in 1:nscn){                                        # Szenarieniterationen
  y               = mvrnorm(1, X %*%beta[,s], sigsqr*I_n) # Datenrealisierung
  beta_hat_1      = solve(t(X_1)%*%X_1)%*%t(X_1)%*%y     # Betaparameterschätzer reduziertes Modell
  beta_hat        = solve(t(X)  %*%X  )%*%t(X)  %*%y     # Betaparameterschätzer vollständiges Modell
  eps_1_hat       = y-X_1%*%beta_hat_1                   # Residuenvektor reduziertes Modell
  eps_hat         = y-X%*%beta_hat                       # Residuenvektor vollständiges Modell
  eps_1_eps_1_hat = t(eps_1_hat) %*% eps_1_hat           # RQS reduziertes Modell
  eps_eps_hat     = t(eps_hat)   %*% eps_hat             # RQS vollständiges Modell
  Eff[s]          = (((eps_1_eps_1_hat-eps_eps_hat)/p_2)/ # F-Statistik
                      (eps_eps_hat/(n-p)))}
```

```
> F-Statistik für beta_2  = 0_{p_2}: 0.404
> F-Statistik für beta_2 != 0_{p_2}: 24.4
```

Theorem (F-Statistik)

Für $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$ und $\sigma^2 > 0$ sei ein ALM der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{60}$$

mit der Partitionierung

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, X_1 \in \mathbb{R}^{n \times p_1}, X_2 \in \mathbb{R}^{n \times p_2}, \text{ und } \beta := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta_1 \in \mathbb{R}^{p_1}, \beta_2 \in \mathbb{R}^{p_2}, \tag{61}$$

mit $p = p_1 + p_2$ gegeben. Schließlich sei

$$K := \begin{pmatrix} 0_{p_1} \\ 1_{p_2} \end{pmatrix} \in \mathbb{R}^p \tag{62}$$

ein Kontrastgewichtsvektor. Dann gilt

$$F \sim f(\delta, p_2, n - p) \text{ mit } \delta := \frac{K^T\beta \left(K^T(X^TX)^{-1}K\right)^{-1} K^T\beta}{\sigma^2} \tag{63}$$

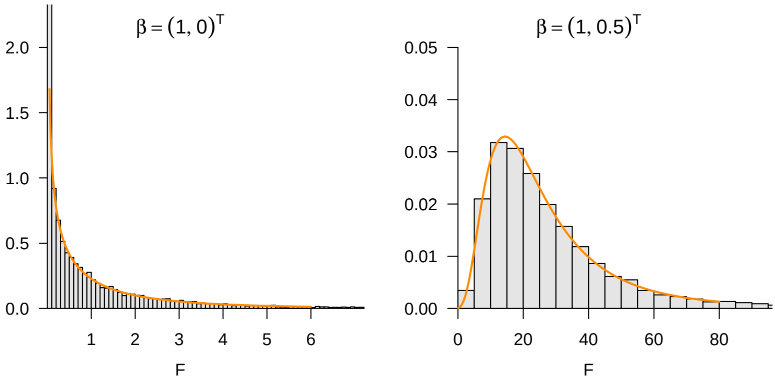
Bemerkungen

- Wir verzichten auf einen Beweis.
- F ist eine Funktion der Parameterschätzer, δ ist eine Funktion der wahren, aber unbekannten, Parameter
- Diese Verteilung von F kann zum Nullhypothesentesten und zur Powerfunktionsevaluation genutzt werden.

F-Statistiken

```
# Libraries
library(MASS)                                                  # Multivariate Normalverteilung

# Modellformulierung
nmod   = 2                                                     # Anzahl Modelle
n      = 10                                                    # Anzahl Datenpunkte
p      = 2                                                     # Anzahl Betaparameter
p_1    = 1                                                     # Anzahl Betaparameter im reduzierten Modell
p_2    = 1                                                     # Anzahl zusätzlicher Betaparameter im vollständigen Modell
p      = p_1 + p_2                                             # Anzahl Betaparameter im vollständigem Modell
x      = 1:n                                                   # Prädiktorwerte
X      = matrix(c(rep(1,n),x), nrow = n)                       # Designmatrix des vollständigen Modells
X_1    = X[,1]                                                 # Designmatrix des reduzierten Modells
I_n    = diag(n)                                               # n x n Einheitsmatrix
beta   = matrix(c(1,0,1,.5), nrow = 2)                         # wahre , aber unbekannte , Betaparameter
nscn   = ncol(beta)                                            # Anzahl wahrer, aber unbekannter, Hypothesenszenarien
sigsqr = 1                                                     # wahrer, aber unbekannter, Varianzparameter
K      = matrix((c(0,1)), nrow = 2)

# Frequentistische Simulation
nsim   = 1e4                                                   # Anzahl Realisierungen des n-dimensionalen ZVs
delta  = rep(NaN,nscn)                                         # Nichtzentralitätsparameterarray
Eff    = matrix(rep(NaN, nscn*nsim), nrow = nscn)              # F-Statistiik Realisierungsarray
for(s in 1:nscn){                                              # Szenarieniterationen
  delta[s] = (t(t(K)%*%beta[,s])%*%                            # Nichtzentralitätsparameter
            solve(t(K)%*%solve(t(X)%*%X)%*%K) %*%
             (t(K)%*%beta[,s])/sigsqr)
  for(i in 1:nsim){                                            # Simulationsiterationen
    y               = mvrnorm(1, X %*%beta[,s], sigsqr*I_n)    # Datenrealisierung
    beta_hat_1      = solve(t(X_1)%*%X_1)%*%t(X_1)%*%y         # Betaparameterschätzer reduziertes Modell
    beta_hat        = solve(t(X)  %*%X  )%*%t(X)  %*%y         # Betaparameterschätzer vollständiges Modell
    eps_1_hat       = y-X_1%*%beta_hat_1                       # Residuenvektor reduziertes Modell
    eps_hat         = y-X%*%beta_hat                           # Residuenvektor vollständiges Modell
    eps_1_eps_1_hat = t(eps_1_hat) %*% eps_1_hat               # RQS reduziertes Modell
    eps_eps_hat     = t(eps_hat)   %*% eps_hat                 # RQS vollständiges Modell
    Eff[s,i]        = (((eps_1_eps_1_hat-eps_eps_hat)/p_2)/    # F-Statistik
                          (eps_eps_hat/(n-p)))
  }
}
```

Beispiel (1) Einfache lineare Regression



Schlussbemerkungen

- Die Theorie von T- und F-Statistiken kann unter dem Gesichtspunkt von linearen Transformationen von Betaparametern verallgemeinert und integriert werden. Dabei betrachtet man Hypothesen der Form

$$K\beta = \beta_0, \tag{64}$$

  wobei $K \in \mathbb{R}^{s \times p}$ eine beliebige Matrix ist. Zum Beispiel ergibt sich für $K \in \mathbb{R}^{1 \times p}$ der hier betrachtete Fall von T-Statistiken und für $K = I_p \in \mathbb{R}^{p \times p}$ der hier betrachtete Fall von F-Statistiken.

- Man kann weiterhin zeigen, dass im Wesentlichen gilt

$$F = T^2, \tag{65}$$

  dass also auch das Quadrat einer T-Statistik f-verteilt ist und T-Statistiken damit spezielle F-Statistiken sind.

- Dennoch wird in der Anwendung sehr stark zwischen T- und F-Statistiken unterschieden und es ist sinnvoll, sich der unterschiedlichen Anwenungsfälle von T- und F-Statistiken bewusst zu sein. Eine sehr gute Einführung in die allgemeine Kontrasttheorie gibt Searle (1971), Chapter 3.6.

Schätzerverteilungen

T-Statistiken

F-Statistiken

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie das Theorem zur Frequentistischen Verteilung des Betaparameterschätzers wieder.
2. Geben Sie die Verteilung des Betaparameterschätzers im Szenarion von n u.i.v. Zufallsvariablen wieder.
3. Geben Sie das Theorem zur Frequentistischen Verteilung des Varianzparameterschätzers wieder.
4. Geben Sie die Verteilung des skalierten Varianzparameterschätzers bei n u.i.v. Zufallsvariablen wieder.
5. Skizzieren Sie die WDFen von t-Zufallsvariablen mit 2, 10 und 30 Freiheitsgraden.
6. Skizzieren Sie die WDFen von nichtzentralen t-Zufallsvariablen mit Nichtzentralitätsparametern 0,5 und 15.
7. Geben Sie die Definition der T-Statistik wieder.
8. Erläutern Sie die Definition der T-Statistik.
9. Warum kann die T-Statistik als Signal-zu-Rauschen Verhältnis interpretiert werden.
10. Geben Sie die Form der T-Statistik im Szenario von n u.i.v. Zufallsvariablen wieder.
11. Erläutern Sie den Zusammenhang der T-Statistik und Cohen's d.
12. Geben Sie die Definition der T-Teststatistik wieder.

Selbstkontrollfragen

13. Erläutern Sie die Definition der T-Teststatistik.
14. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.
15. Geben Sie die Definition eines vollständigem und eines reduziertem ALMs wieder.
16. Geben Sie die Definition des Likelihood-Quotienten eines vollständigen und eines reduzierten ALMs wieder.
17. Geben Sie das Theorem zu Likelihood-Quotient und Residualquadratsummendifferenz wieder.
18. Definieren Sie die F-Statistik.
19. Erläutern Sie den Zusammenhang zwischen F-Statistik und Likelihood-Quotient.
20. Definieren Sie und erläutern Sie den Zähler der F-Statistik
21. Definieren Sie und erläutern Sie den Nenner der F-Statistik.
22. Erläutern Sie die F-Statistik.
23. Geben Sie das Theorem zur Verteilung der F-Statistik wieder.

References

Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Wiley Series in Probability and Statistics.

Searle, S. R. 1971. *Linear Models*. A Wiley Publication in Mathematical Statistics. New York: Wiley.

Student. 1908. "The Probable Error of a Mean." *Biometrika* 6 (1): 1–25.

Zabell, S. L. 2008. "On Student's 1908 Article 'The Probable Error of a Mean'." *Journal of the American Statistical Association* 103 (481): 1–7. https://doi.org/10.1198/016214508000000030.