Dynamische Systeme und Zeitreihenanalyse

Multivariate Normalverteilung und ML Schätzung

Kapitel 11

Statistik und Mathematik – WU Wien

Michael Hauser

Lernziele

- Erweiterung der 2-dimensionalen Normalverteilung auf die n-dimensionale
- Rechenregeln für die multivariate Normalverteilung, MVN
- iid: identically, independently distributed
- Maximum Likelihood Schätzung, ML
- ML mit Normalverteilung
- Eigenschaften der Schätzer:
  Erwartungstreue, Effizienz, Konsistenz

Die n-dimensionale Normalverteilung

Die 1-dimensionale Normalverteilung

Sei die ZV X **normalverteilt** mit Erwartung μ und Varianz σ^2, so schreiben wir

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

Die **Dichte** der Normalverteilung, $f_N(x)$ ist mit $-\infty < x < \infty$

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$$

Sie hat die bekannte Glockenform.

Zum Vergleich mit der n-dimensionalen lesen wir sie

$$f_N(x) = (2\pi)^{-1/2}\,(\sigma^2)^{-1/2}\,\exp[-\frac{1}{2}(x-\mu)\,(\sigma^2)^{-1}\,(x-\mu)]$$

Die 1-dimensionale Normalverteilung (Fs)

Eine **standardnormalverteilte** ZV, Z,

$$Z \sim N(0,1)$$

hat $E(Z) = 0$ und $V(Z) = 1$.

Sie erhält man durch **Standardisierung** der ZV X.

$$Z = \frac{X-\mu}{\sigma}$$

Umgekehrt kann man jede beliebige normalverteilte ZV X, $X \sim N(\mu, \sigma^2)$, aus einer standardnormalverteilten erzeugen.

$$X = \mu + \sigma Z$$

Die 2-dimensionale Normalverteilung

Sei $\mathbf{X} = (X_1, X_2)'$ ein 2-dimensionaler Vektor von ZVen mit dem Vektor der Erwartungen $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ und der **Kovarianzmatrix**, Σ,

$$\Sigma = E[(\mathbf{X}-\boldsymbol{\mu})\,(\mathbf{X}-\boldsymbol{\mu})']$$

$$E(\mathbf{X}) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Die Kovarianzmatrix ist **symmetrisch**, $\Sigma = \Sigma'$, da $\sigma_{12} = Cov(X_1, X_2) = Cov(X_2, X_1) = \sigma_{21}$ ist.

$$Cov(X_1, X_2) = E[(X_1-\mu_1)(X_2-\mu_2)]$$

Die 2-dimensionale Normalverteilung (Fs)

Sind die ZVen X_1 und X_2 gemeinsam normalverteilt, $\mathbf{X} = (X_1, X_2)'$, so schreibt man

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$$

Wir setzen voraus, dass $\det(\Sigma) \neq 0$ (ie rg$(\Sigma) = 2$, Σ positiv definit) ist.

- Es gilt immer $\det(\Sigma) \geq 0$.
- $\det(\Sigma) = 0$ gilt dann, wenn lineare Abhängigkeit zwischen den X_i besteht. In diesem Fall beschreibt der Vektor $\mathbf{X}$ nur einen (hier) eindimensionalen Unterraum.

Die n-dimensionale Normalverteilung

Sei $\mathbf{X} = (X_1, \ldots, X_n)'$ ein n-dimensionaler Vektor von ZVen mit dem Vektor der Erwartungen $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ und der **Kovarianzmatrix**, Σ,

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})\,(\mathbf{X} - \boldsymbol{\mu})']$$

$$E(\mathbf{X}) = \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_n \end{pmatrix}' \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1n} \\ \cdots & & \cdots \\ \sigma_{n1} & \ldots & \sigma_{nn} \end{bmatrix}$$

Sind die X_i gemeinsam normalverteilt, schreibt man

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$$

Die Dichte der n-dimensionalen N-Verteilung

Die **Dichte der n-dimensionalen Verteilung** lautet

$$f_N(\mathbf{x}) = (2\,\pi)^{-n/2} |\Sigma|^{-1/2} \exp[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})]$$

Wir verlangen auch hier, dass rg$(\Sigma) = n$ ist.

$N(\mathbf{0}, I_n)$

Eine standardnormalverteilte n-dimensionale ZV $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ besitzt folgende Erwartung und Kovarianzmatrix

$$E(\mathbf{Z}) = \mathbf{0} \qquad E[\mathbf{Z}\,\mathbf{Z}'] = I_n$$

Ihre Dichte ist mit

$$f_N(\mathbf{x}) = (2\,\pi)^{-n/2}|\Sigma|^{-1/2}\exp[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})]$$

$$f_N(\mathbf{z}) = (2\,\pi)^{-n/2}\exp[-\frac{1}{2}\mathbf{z}'\,\mathbf{z}]$$

Die Varianzen der einzelnen Z_i sind eins: $V(Z_i) = 1$,
die Kovarianzen sind null: $Cov(Z_i, Z_j) = 0$ für $i \neq j$.

$N(\boldsymbol{\mu}, \Sigma)$

Jede normalverteilte n-dimensionale ZV $\mathbf{X}$ kann man aus der standardnormalverteilten, $\mathbf{Z}$, durch

$$\mathbf{X} = \boldsymbol{\mu}_{n\times 1} + A_{n\times n}\,\mathbf{Z}$$

erzeugen. Es gilt

$$E[\mathbf{X}] = \boldsymbol{\mu} \qquad E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'] = A\,A' = \Sigma$$

Da

$\Sigma = E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'] = E[(A\,\mathbf{Z})(A\,\mathbf{Z})'] =$
$\ldots (A\,B)' = B'\,A' \ldots$
$= E[(A\,\mathbf{Z})\,(\mathbf{Z}'\,A')] = A\,E[\mathbf{Z}\,\mathbf{Z}']\,A' = A\,I_n\,A' = A\,A'$

Die Kovarianzmatrix $E[\mathbf{Z}\,\mathbf{Z}'] = I_n$

$Cov(Z_i\,, Z_j) = 0$ für $i \neq j$, bedeutet, dass Z_i und Z_j unkorreliert sind.

Die Dichte von Z lässt sich umformen zu

$$f_N(\mathbf{z}) = (2\,\pi)^{-n/2}\exp[-\frac{1}{2}\sum_{i=1}^{n} z_i^2] = (\frac{1}{\sqrt{2\,\pi}})^n\exp[\sum_{i=1}^{n} -\frac{1}{2}z_i^2]$$

$$= \prod_{i=1}^{n}\frac{1}{\sqrt{2\,\pi}}\exp[-\frac{1}{2}z_i^2] = \prod_{i=1}^{n} f_N(z_i)$$

Die n-dimensionale Dichte ist gleich dem Produkt der 1-dimensionalen Dichten.

Das gilt auch, wenn Σ eine Diagonalmatrix ist.

Die Unabhängigkeit von ZVen

Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B sind unabhängig voneinander, wenn gilt

$$P(A \cap B) = P(A) \cdot P(B)$$

Auf diskrete ZVen, X und Y, übertragen bedeutet dies mit $P[(X = x) \cap (Y = y)] = P[(X = x), (Y = y)]$

$$P[(X = x), (Y = y)] = P(X = x) \cdot P(Y = y)$$

Die Ws des gemeinsamen Ereignisses ist das Produkt der einzelnen Ws.

Unabhängigkeit von ZVen

Sind zwei diskrete ZVen **unabhängig**, so gilt für alle Werte x, y

$$P[(X = x), (Y = y)] = P(X = x) \cdot P(Y = y)$$

Sind zwei stetige ZVen **unabhängig**, so gilt für alle Werte x, y

$$f_{xy}(x, y) = f_x(x)\, f_y(y)$$

Die gemeinsame Dichte ist das Produkt der einzelnen Dichten.

Die gemeinsame Verteilung ist das Produkt der Randverteilungen.

Bsp: Unabhängigkeit von ZVen

- Die Komponenten der standardnormalverteilten ZV $\mathbf{Z}$, $Z_1, \ldots, Z_n$ sind unabhängig, da die gemeinsame Dichte das Produkt der einzelnen Dichten ist.
- Sind die ZVen $Z_1, \ldots, Z_n$ gemeinsam normalverteilt und unkorreliert, so sind sie auch unabhängig.
- Aus Unabhängigkeit folgt immer Unkorreliertheit, aber nicht umgekehrt. Unabhängigkeit ist die stärkere Eigenschaft.

Herleitung:
Seinen X und Y zwei ua ZVen der Einfachheit halber mit $E(X) = E(Y) = 0$.

$$Cov(X,Y) = E[X\,Y] = \int x\,y\,f(x,y)\,dx\,dy$$
$$= \int x\,y\,f_x(x)\,f_y(y)\,dx\,dy = (\int x\,f_x(x)\,dx)(\int y\,f_y(y)\,dy) = 0$$

iid

Ein Vektor von ZVen $\mathbf{X} = (X_1, \ldots, X_n)'$ ist **id** (independently distributed) **unabhängig verteilt**, wenn die ZVen X_i unabhängig sind,

$$f_{1\ldots n}(x_1, \ldots, x_n) = f_1(x_1) \cdot \ldots \cdot f_n(x_n)$$

wobei $f_{1\ldots n}$ die gemeinsame Dichte der ZVen $(X_1, \ldots, X_n)$ und f_i die Dichte für die ZV X_i bezeichnet.

Ein Vektor von ZVen $\mathbf{X} = (X_1, \ldots, X_n)'$ ist **iid** (identically, independently distributed), wenn die X_i unabhängig sind und dieselbe Verteilung besitzen. Mit $f_i = f_j = f$

$$f_{1\ldots n}(x_1, \ldots, x_n) = \Pi_i f(x_i)$$

Beispiele - iid

Beispiel 1:

Seien $X_1, \ldots, X_n$ iid, normalverteilt mit $E(X_i) = \mu = const$ und $V(X_i) = \sigma^2 = const$, dh $X_i \sim N(\mu, \sigma^2)$, folgt

- X_i iid $N(\mu, \sigma^2)$
- $Cov(X_i, X_j) = 0$ für $i \neq j$, $\quad Corr(X_i, X_j) = 0$ für $i \neq j$
- $f_{1\ldots n}(x_1, \ldots, x_n) = \prod f_N(x_i; \mu, \sigma^2) = \prod \frac{1}{\sqrt{2\,\pi}} \exp[-\frac{1}{2}(\frac{x_i - \mu}{\sigma})^2]$
- $X \sim N(\boldsymbol{\mu}, \sigma^2\, I)$
  $f_{1\ldots n}(x_1, \ldots, x_n) = (\frac{1}{\sqrt{2\,\pi}})^n \exp[-\frac{1}{2} \sum (\frac{x_i - \mu}{\sigma})^2]$

Beispiel 2:

In der linearen Regression wird üblicherweise angenommen, dass die Residuen (gemeinsam) normalverteilt und unkorreliert sind, also iid.

Die Maximum Likelihood Schätzung, ML

Schätzen

Schätzen eines Modells bedeutet die Parameter des Modells so zu wählen, dass die Eigenschaften des geschätzten Modells möglichst gut mit den Eigenschaften der beobachteten Reihe übereinstimmen.

Maximum Likelihood kann dort angewendet werden, wo die Verteilung der Fehler (als) bekannt (angenommen) ist.

Wir betrachten eine Familie von Grundgesamtheiten. Aus jeder dieser Grundgesamtheiten könnte unsere Stichprobe stammen. Die Frage ist nun aus welcher.

ML wählt jene Grundgesamtheit, welche die größte Plausibilität aufweist. Sie wird durch das Maximum der Likelihood Funktion bestimmt.

Das Maximum Likelihood Prinzip - Bsp 1

Angenommen es liegt (nur) eine Beobachtung, sagen wir mit dem Wert 1.5, vor.

Die Grundgesamtheiten, aus der unsere (einzige) Beobachtung stammen kann, seien entweder durch eine

- N$(-1,4)$- oder durch eine
- N$(1,4)$-verteilte Zufallsvariable

beschrieben. Wir wollen wissen, aus welcher Grundgesamtheit gezogen wurde, woher also die Beobachtung stammt.

Das **Maximum Likelihood Prinzip** besagt:
Ich wähle die Grundgesamtheit, die mir die größte "Wahrscheinlichkeit" für das Auftreten meiner Beobachtung liefert.

Das Maximum Likelihood Prinzip - Bsp 1 (Fs)

Man vergleicht nun den Wert der Dichte an der Stelle 1.5

- für den Fall, daß eine Grundgesamtheit N$(-1, 4)$ vorliegt,

$$f_N(1.5 \,|\, \mu = -1, \sigma^2 = 4)$$

mit dem Wert der Dichte an der Stelle 1.5,

- im Falle der anderen Grundgesamtheit, N$(1, 4)$,

$$f_N(1.5 \,|\, \mu = 1, \sigma^2 = 4)$$

Ist

$$f_N(1.5 \,|\, \mu = -1, \sigma^2 = 4) \;<\; f_N(1.5 \,|\, \mu = 1, \sigma^2 = 4)\;,$$

so ist die N(1,4) die plausiblere Verteilung.
Gilt das Größer-Zeichen, so ist die N(-1,4) die plausiblere.

Das Maximum Likelihood Prinzip - Bsp 1 (Fs)

Die **Likelihood Funktion**, L, ist unter Weglassen von σ^2 (da fix)

$$L(\mu \,|\, x) = f_N(x \,|\, \mu, 4)$$

eine Funktion in μ.

Die Likelihood wird nun bez μ maximiert

$$\max_{\mu \in \{-1,+1\}} L(\mu \,|\, x)$$

wobei in diesem Beispiel nur zwischen $\mu = -1$, oder $\mu = 1$ gewählt werden kann.

Das Maximum Likelihood Prinzip - Bsp 1 (Fs)

(1) Setzen wir die Daten ein, erhalten wir

$$L(\mu \,|\, x = 1.5) = f(1.5 \,|\, \mu,\, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{1.5-\mu}{\sigma})^2]$$

mit $\sigma = 2$ (fix) also

$$L(\mu \,|\, x = 1.5) = f(1.5 \,|\, \mu,\, 4) = \frac{1}{2\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{1.5-\mu}{2})^2]$$

Das Maximum Likelihood Prinzip - Bsp 1 (Fs)

(2a) Nun untersuchen wir den Wert der Likelihood Funktion an den Stellen $\mu = -1$ und $\mu = +1$.

$$L(\mu = 1 \,|\, 1.5) = \frac{1}{2\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{1.5-1}{2})^2] = 0.193,$$

(2b) und

$$L(\mu = -1 \,|\, 1.5) = \frac{1}{2\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{1.5-(-1)}{2})^2] = 0.091$$

(3) Also vergleichen wir:

$$L(1 \,|\, 1.5) > L(-1 \,|\, 1.5)$$

Der Maximum Likelihood Schätzer für μ, $\widehat{\mu}$, ist 1. $\widehat{\mu} = 1$.

Das Maximum Likelihood Prinzip - Bsp 2

Angenommen es liegt eine normalverteilte ZV X vor mit $X \sim N(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt sind.

Wir ziehen n mal und erhalten durch unabhängige Ziehungen (Ziehungen mit Zurücklegen, oder unendliche Grundgesamtheit) die Werte $x_1, \ldots, x_n$. Die Ziehungen sind selbst Realisationen von ZVen $X_1, \ldots, X_n$, die so verteilt sind wie X, also

$$X_i \text{ iid } N(\mu, \sigma^2)$$

Wir wollen nun basierend auf unserer Stichprobe Schätzer für μ und σ^2 bestimmen.

Das Maximum Likelihood Prinzip - Bsp 2 (Fs)

Im Bsp 1 haben wir die Verteilung der Stichprobe nach dem unbekannten Parameter maximiert.

Die Verteilung der Stichprobe ist auf Grund der iid Eigenschaft das Produkt der einzelnen Dichten:

$$f(x_1, \ldots, x_n) = \prod_i f_N(x_i) = (\frac{1}{\sigma\sqrt{2\,\pi}})^n \exp[-\frac{1}{2}\sum_i (\frac{x_i - \mu}{\sigma})^2]$$

Die Likelihood Funktion ist daher bei gegebenen Daten

$$L(\mu, \sigma^2 | x_1, \ldots, x_n) = (\frac{1}{\sigma\sqrt{2\,\pi}})^n \exp[-\frac{1}{2}\sum_i (\frac{x_i - \mu}{\sigma})^2]$$

Das Maximum Likelihood Prinzip - Bsp 2 (Fs)

Wir maximieren $L(\mu, \sigma^2 | x_1, \ldots, x_n)$ nach μ und σ^2. Dazu bilden wir die Bedingungen 1.Ordnung, die partiellen Ableitungen nach μ und σ^2 und setzen sie Null.

$$\frac{\partial L}{\partial \mu} = 0 \qquad \frac{\partial L}{\partial \sigma^2} = 0$$

$\frac{\partial L}{\partial \mu} = const \cdot \exp[\ldots] \cdot [-\frac{1}{2} \sum_i 2 \, (\frac{x_i - \mu}{\sigma}) \, (-\frac{1}{\sigma})] = 0$

Wir vereinfachen.

$\sum_i (x_i - \mu) = 0$ dh $\sum_i x_i - n\,\mu = 0$ also $\widehat{\mu} = \frac{1}{n} \sum_i x_i$

Der Maximum Likelihood Schätzer für die Erwartung ist das arithmetische Mittel, $\widehat{\mu} = \overline{x}$.

Die Log-Likelihood Funktion

Das Gleiche führen wir *nicht* für die Varianz durch. Die Rechnungen sind zu mühsam.

Wir bedienen uns nun eines Tricks. Die Funktion $L(\mu, \sigma^2)$ besitzt an den gleichen Stellen ihre Maxima, wie $\log L(\mu, \sigma^2)$, weil der $\log(\ldots)$ eine streng monotone Transformation ist. Die Funktion

$$\log L(\mu, \sigma^2)$$

heißt **Log-Likelihood Funktion**.

$$\log L(\mu, \sigma^2 | x_1, \ldots, x_n) = -n \, \log \sigma - n \, \log \sqrt{2\,\pi} - \frac{1}{2\,\sigma^2} \sum_i (x_i - \mu)^2$$

Wir maximieren statt L die Funktion $\log(L)$.

Das Maximum Likelihood Prinzip - Bsp 2 (Fs)

Nochmals die Lösung für μ:

$$\frac{\partial \log L}{\partial \mu} = 0 : \quad \frac{1}{2\,\sigma^2} 2 \sum_i (x_i - \mu) = 0$$

Daher die gleiche Lösung wie oben: $\widehat{\mu} = \frac{1}{n} \sum_i x_i = \overline{x}$

Für σ:

$$\frac{\partial \log L}{\partial \sigma} = 0 : \quad -\frac{n}{\sigma} + \frac{\sum (x_i - \mu)^2}{\sigma^3} = 0$$

Somit ist $\sigma^2 = [\sum (x_i - \mu)^2]/n$. Setzen wir für μ unseren ML-Schätzer ein, ergibt sich

$$\widehat{\sigma^2} = \frac{1}{n} \sum_i (x_i - \widehat{\mu})^2$$

ML - Zusammenfassung

Wir sind in den Bsp von iid normalverteilten ZVen ausgegangen. Die gemeinsame Verteilung haben wir durch Multiplikation der einzelnen Dichten erreicht.

Wichtig für eine Maximum Likelihood Schätzung ist,

- dass die gemeinsame Verteilung der X_i bekannt ist.

Die Funktion hängt von den

- unbekannten Parametern und
- den (gegebenen) Beobachtungen aus der Stichprobe ab.

ML - Zusammenfassung (Fs)

Beispiel:
Die Verteilung kann theoretisch auch eine allgemeine MVN sein. Allerdings müssten die unbekannten Kovarianzen σ_{ij} geschätzt werden.

Dieser Ansatz stößt an Grenzen, da wir nur n Beobachtungen haben, aber mit (konstantem) Erwartungswert und (konstanter) Varianz noch immer

$$1 + 1 + (n^2 - n)/2 \quad (> n)$$

verschiedene Parameter berechnen müssten.

Eigenschaften von Schätzern

Erwartungstreue, Unverzerrtheit

Ein Schätzer $\widehat{\theta}_n$ für einen Parameter θ (n bezieht sich auf den Stichprobenumfang, der zur Verfügung steht) heißt **erwartungstreu**, bzw **unverzerrt**, wenn gilt

$$E(\widehat{\theta}_n) = \theta$$

Hier wird die Erwartung über alle möglichen Stichproben vom Umfang n gebildet.

Beispiel:
In der Regressionsanalyse sind wir bemüht erwartungstreue Schätzer zu finden.

Asymptotische Erwartungstreue

Ein Schätzer $\widehat{\theta}_n$ für einen Parameter θ heißt **asymptotisch erwartungstreu**, bzw **asymptotisch unverzerrt**, wenn gilt

$$\lim_{n\to\infty} E(\widehat{\theta}_n) = \theta$$

Beispiel:
In der Zeitreihenanalyse haben wir es in der Regel nur mit asymptotischer Erwartungstreue zu tun.

Effizienz

Ein Schätzer $\widehat{\theta}_n$ für einen Parameter θ heißt **effizient**, wenn er bezüglich einer Klasse von Modellen die kleinste Varianz aufweist.

$$\sigma_{\widehat{\theta}} \quad \text{ist minimal.}$$

Beispiel: Unter den üblichen Annahmen sind die Kleinstquadrat-Schätzer im linearen Regressionsmodell effizient.

In der Zeitreihenanalyse haben wir es in der Regel nur mit **asymptotischer Effizienz**

$$\lim_{n\to\infty} \sigma_{\widehat{\theta}_n} \quad \text{ist minimal}$$

zu tun. Die Eigenschaften in kleinen Stichproben sind analytisch schwierig zu bestimmen. ($\rightarrow$ Monte Carlo Simulationsstudien)

Normalverteilter Schätzer

Ein Schätzer $\widehat{\theta}_n$ ist normalverteilt, wenn zB gilt

$$\widehat{\theta}_n \sim N(\theta, \sigma^2_{\widehat{\theta}_n})$$

Beispiel:
Unter den üblichen Annahmen sind die Schätzer in der linearen Regression normalverteilt. Nach Standardisierung mit der (geschätzten) Standardabweichung erhalten wir die bekannte t-Statistik, t-Verteilung mit $(n-k)$ Freiheitsgraden

$$t_{\widehat{\theta}} = \frac{\widehat{\theta}_n}{\widehat{\sigma}_{\widehat{\theta}_n}} \sim t_{n-k}$$

Hypothesen sind:

$$H_0: \quad \theta = \theta_0 = 0 \quad \text{und} \quad H_A: \quad \theta \neq \theta_0 = 0$$

Normalverteilter Schätzer

Ein Schätzer $\widehat{\theta}_n$ heißt asymptotisch normalverteilt, wenn gilt

$$\frac{\widehat{\theta}_n - \theta}{\widehat{\sigma}_{\widehat{\theta}_n}} \xrightarrow{d} N(0,1)$$

Die Verteilung des standardisierten Schätzers konvergiert für größer werdende n gegen die Normalverteilung.

Für große n ist der Schätzer näherungsweise normalverteilt.

(*Konvergenz in der Verteilung, convergence in distribution*: $\xrightarrow{d}$)

Konsistenz eines Schätzers

Konsistenz eines Schätzers ist ein etwas schwächeres Konzept. Ein Folge von Schätzern $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_n$ ist **konsistent** für den Parameter θ, wenn für jedes $\epsilon > 0$

$$\lim_{n\to\infty} P(|\widehat{\theta}_n - \theta| < \epsilon) = 1$$

Man schreibt auch $\quad plim\ \widehat{\theta}_n = \theta.$

Diese Definition besagt, dass der Unterschied zwischen der Schätzung $\widehat{\theta}_n$ und θ immer kleiner wird (kleiner als jedes $\epsilon > 0$), in dem Sinn, dass die Ws eines Unterschieds zwischen beiden gegen Null geht. (*Konvergenz in der Wahrscheinlichkeit, conv. in probability*)

Bemerkung: Unerwünscht sind Schätzer, die nicht konsistent sind. Für diese ist die Ws eines Unterschieds zwischen wahrem und geschätztem Wert immer positiv.