Formale Statistische Modelle

Zusammenstellung wichtiger Formeln und Sätze

zusammengestellt von Hannes Restel,
basierend auf einem Script von Dr. Timo von Oertzen

Freie Universität Berlin, Institut für Informatik
Version 1.0 (vom 06.02.2008)

Inhaltsverzeichnis

Formale Statistische Modelle....1
Zusammenstellung wichtiger Formeln und Sätze....1
Allgemeine Rechenregeln....1
Allgemeine Formeln, Definitionen und Tricks....1
Logarithmus – Gesetze und Tricks....2
Stochastik....2
Einführung in die Stochastik....2
Zufallsvariablen....4
Mittelwerte, Varianz, Ungleichungen....4
Diskrete Zufallsverteilungen....5
Kontinuierliche Zufallsverteilungen....6
Generierende, Charakteristische, Momentgenerierende Funktionen....7
Statistik....8
Einführung in die Statistik....8
Schätzer und Maximum-Likelihood-Schätzer....8
Markov-Ketten und Markov-Prozesse....8
Periodizität:....9
Versteckte Markov Ketten:....9
Strukturgleichungsmodelle (SEMs)....9

Allgemeine Rechenregeln

Allgemeine Formeln, Definitionen und Tricks

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

$$\binom{n}{k_{1,} k_{2,\ldots}, k_r} := \frac{n!}{k_1! \cdot k_1! \cdot k_r!}$$

$$\sum_{i=0}^{n} \left(\binom{n}{i} \cdot i^2\right) = \sum_{i=0}^{n} \left(\binom{n-1}{i-1} \cdot i \cdot n\right)$$

Matrizen

Transponierte Matrix A^T*: Zeilen und Spalten vertauscht. Aus* (m, n) *wird* (n, m)

Inverse Matrix $A^{-1}: A \cdot A^{-1} = A^{-1} \cdot A = E$, *E Einheitsmatrix*, $A^{-1} = \frac{A}{det(A)}$

Stochastische Matrix *: jede Spaltensumme addiert sich jeweils zu* 1

Eigenwert λ: $A \cdot x = \lambda \cdot x$, *Eigenwert ist Streckungsfaktor jedes Spaltenvektors x*

Limes: $(\lim_{i\to\infty} a_i = a) \Rightarrow \forall \epsilon > 0 : \exists N : \forall n > N : |a_n - a| < \epsilon$

Der Limes einer Folge (a_n) existiert und kovergiert gegen einen Wert a, wenn ab einem bestimmten Element der Folge gilt, dass sich die Werte der Folge um den Wert a gruppieren und nicht mehr weiter wachsen; die a_i also maximal ϵ-weit entfernt von a sind, wobei ϵ seeehr klein ist.

Konvergenz überall

Konvergenz fast überall

Konvergenz in Verteilung: Wahrscheinlichkeit sehr vieler Zufallsvariablen X_i konvergiert punktweise gegen die Wahrscheinlichkeit einer Zufallsvariablen X

Stochastische Konvergenz: „Die Wahrscheinlichkeit, dass der Abstand zwischen der beobachteten und der theoretischen Größe größer ist als jedes positive ε (insbesondere: jedes beliebig nahe an null liegende positive) ε, geht gegen null, wenn die Anzahl der Experimente gegen unendlich geht." ()

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0,$$

Logarithmus – Gesetze und Tricks

Logarithmusgesetze :

$$\log_a(u \cdot v) = \log_a u + \log_a v$$

$$\log_a \frac{u}{v} = \log_a u - \log_a v$$

$$\log_a u^r = r \cdot \log_a u$$

$$\log_a \sqrt[n]{u} = \frac{1}{n} \log_a u \quad , n \in \mathbb{N}$$

$$\prod X = \ln(\sum X)$$

Basiswechsel :

$$\log_a b \cdot \log_b a = 1$$

$$\log_c b = \frac{\log_a b}{\log_a c} = \frac{\ln b}{\ln c}$$

Allgemeines :

$$\log_a b = c \Leftrightarrow a^c = b$$

$$a^{\log_a b} = b$$

$$\log_a 1 = 0$$

$$\log_a a = 1$$

...

Stochastik

Einführung in die Stochastik

Die Stochastik befasst sich mit Häufigkeiten und Wahrscheinlichkeiten, wobei stets alle Parameter, Variablen und Verteilungen bekannt sind.

Definition : Kolmogorov – Komplexität :

Ist die Länge des kürzesten Programms, welches die binäre Folge $(a_n)_{1..K}$ *der Länge K erzeugt*

Definition : Zufälligkeit nach Per Martin− Löf :
Eine beliebig lange binäre Folge $(a_n)_n$ heisst zufällig, wenn ein c existiert, so dass für undendlich viele Präfixe der Länge n die Kolmogorov-Komplexität mindestens $(n-c)$ beträgt.

Wahrscheinlichkeit: relative Häufigkeit eines Ereignisses im Verhältnis zu allen Ereignissen.

Wahrscheinlichkeitsmaß ***P*** ordnet jeder Menge ihre Wahrscheinlichkeit von 0 bis 1 zu.

Verteilungsfunktion : $f(x)=P(X\in(-\infty,x])$

Verteilungsdichte : $f'(x)$ (die 1. Ableitung)

Lemma : $[P(A\cup B)=P(A-A\cup B)+P(B-A\cup B)+P(A\cap B)]\leqslant P(A)+P(B)$

bedingte Wahrscheinlichkeit : $P(A|B)=\dfrac{P(A\cap B)}{P(B)}$ $[auch: P(A\cap B)=P(A|B)\cdot P(B)]$

Satz der totalen Wahrscheinlichkeit :
Sei $A_{1,} A_{2,}...$ eine abzählbare Partition von Ω und B ein Ergeignis. Dann ist

$$P(B)=\sum_i (P(B|A_i)\cdot P(A_i)).$$

Satz von Bayes : $P(A|B)=P(B|A)\cdot\dfrac{P(A)}{P(B)}$

Unabhängigkeit : *Zwei Ereignisse A, B heißen unabhängig, wenn gilt* $P(A\cap B)=P(A)\cdot P(B)$.

Es ist meist nicht notwendig, die Wahrscheinlichkeitsfunktion P auf der gesamten Potenzmenge 2^{Ω} des Wahrscheinlichkeitsraums Ω zu definieren, da wir nur an bestimmten Ereignissen interessiert sind. Wir bilden deshalb die $\sigma-Algebra\,(Sigma-Algebra)$. Diese besteht aus allen Elementarereignissen aus Ω (also Ω selbst), sowie aus Negationen und Vereinigungen der gewählten Menge von Teilmengen $A, B, C, ..$ aus Ω, welche unsere gewünschten Ereignisse darstellen.

Tatsächlich verwenden wir anstatt der $\sigma-Algebra$ jedoch die $Borelklasse$, welche als σ -Algebra der Menge aller abgeschlossenen Teilmengen definiert ist, also
$Borelklasse=\sigma(I)$*, mit I ist Menge aller abgeschlossenen Teilmengen aus* Ω. Die *Borelklasse* erweitert die $\sigma-Algebra$ also um zusätzliche Elemente.

Dichtefunktion der Wahrscheinlichkeitsverteilung : $\int_{-\infty}^{\infty} f(x)dx=1$ *, mit* $f:\mathbb{R}\rightarrow\mathbb{R}$

Wahrscheinlichleitsfunktion auf allen Intervallen[a , b] :

$$P(X\in[a,b])=\int_a^b f(x)\,dx \quad , mit\ a,b\in\mathbb{R}$$

Verteilungsfunktion : $\int_{-\infty}^{a} f(a)da$ *oder* $f(a)=P((-\infty,a])$

Die *Verteilungsfunktion* gibt also die Wahrscheinlichkeit an, in einem bestimmten Intervall bis zu einem gegebenen rechten Rand (hier das a) zu sein.

Verteilungsdichte : *ist die* 1. *Ableitung der Verteilungsfunktion* $f'(a)$

(Dies ist die Erweiterung der diskreten Wahrscheinlichkeitsfunktion $F(x)=P(X\leqslant x)=\sum_i p_i$ in den kontinuierlichen Bereich)

Funktionen zwischen Wkt. − Räumen :
Eine Funktion, welche Ereignisse eines Ereignisraumes (Ω, F) in einen anderen Ereignisraum (Ω', F') abbildet, ist eine Fkt. zwischen Wahrscheinlichkeitsräumen.
F ist der Borelabschluss einer Menge von (Nichtelementar-)Ereignissen.

Zufallsvariablen

Definition ***Originäre Zufallsvariable X*** :
$X = (\Omega, F, P)$ *heißt originäre Zufallsvariable und wird vereinfachend* $X \in (\Omega, F)$ *geschrieben.*
F ist die Menge von Ereignissen (incl. dem σ-Abschluss aus Ω . P ist ein Wahrscheinlichkeitsmaß auf F und heißt die Verteilung auf X.

Eine originäre Zufallsvariable ist also ein 3-Tupel.

Allgemein gilt: Eine Zufallsvariable ist eine Funktion von Wahrscheinlichkeitsräumen. Diese Funktionen zwischen zwei Wahrscheinlichkeitsräumen erhalten als Argument eine Zufallsvariable und liefern eine Zufallsvariable zurück. Der (gekürzte) Satz dazu:

Sei $X = (\Omega, F, P)$ *eine Zufallsvariable, und* f *eine Funktion von* (Ω, F) *nach* (Ω', F').
Dann ist $f(X)$ *eine Zufallsvariable.*

Da das Produkt von Wkt.-Räumen existiert, existiert auch das *Produkt von Zufallsvariablen*:

Seien ZV $X_1 = (\Omega_1, F_1, P_1)$, $X_2 = (\Omega_2, F_2, P_2)$.
Dann ist die Produkt − Zufallsvariable $X = X_1 \times X_2 = (\Omega_1 \times \Omega_2, F_1 \times F_2, P)$ *mit*
$F_1, F_2, (F_1 \times F_2)$ *sind* $\sigma -$ *Algebren und* $P(A \times B) = P_1(A \in F_1) \cdot P_2(B \in F_2)$

Definition ***Vorgänger*** : Sei $X = ('X, f)$ mit $X = f('X)$ eine Zufallsvariable.
Dann heißt $'X$ ein Vorgänger von X und alle Vorgänger von $'X$ ebenfalls.

Auflistung der *Abhängigkeiten von Zufallsvariablen*:

Unabhängig − Sei Z gemeinsamer (auch direkter) Vorgänger von X und Y.
Dann heißen X und Y unabhängig, wenn ......

Determiniert − Wenn Abbildung $f(Y) = X$ existiert, so heißt *X von Y determiniert.*

Gegenseitig determiniert − Wenn Abbildungen $f(Y) = X$ und $g(X) = Y$ existieren, so heißen *X und Y gegenseitig determiniert.*

Identisch − Wenn identische Abbildung $f(Y) = X$ und $f(X) = Y$ existiert, so sind *X und Y identisch.*

identisch verteilt − Sind die Räume der Wkt-Maße von X und Y gleich, so heißen *X und Y identisch verteilt.*

Mittelwerte, Varianz, Ungleichungen

Mittelwert : $\mu = \frac{1}{N} \sum_{i=1}^{N} (x_i)$ ***Varianz*** : $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$

Covarianz : $Cov(X, Y) := E((X - E(X)) \cdot (Y - E(Y))) = \boldsymbol{E(X \cdot Y) - E(X) \cdot E(Y)}$

$\boldsymbol{Korrelation}: Cor(X,Y) := \frac{Cov(X,Y)}{\sqrt{V(X) \cdot V(Y)}}$, $-1 < Cor(X,Y) < 1$

Achtung: Selbst wenn $Cor(X,Y)=0$ können die ZV abhängig sein!

$\boldsymbol{Tschebyscheff-Ungleichung}: Wenn\ E(X)=0 \rightarrow P(|X| \geqslant \epsilon) \leqslant \frac{V(X)}{\epsilon^2}$

(Meist exakter als Markov-Ungleichung)

$\boldsymbol{Markov-Ungleichung}: P(|X| \geqslant \epsilon) \leqslant \frac{E(X)}{\epsilon}$

Gilt auch für ϵ die nicht nahe der 0 sind.

Diskrete Zufallsverteilungen

$\boldsymbol{Vorbedingungen}: \omega \in \Omega\ ist\ Ereignis,\ p\ ist\ Wkt.\ für\ ein\ Ereignis\ und\ 1-p\ die\ Gegen-Wkt.\ dazu$

$\mathbf{1-}\boldsymbol{Bernoulli} - Werfe\ eine\ Münze\ ein\ einziges\ Mal.\ P(\omega)\ ist\ p,\ wenn\ \omega=1, und\ 1-p\ sonst$

$\boldsymbol{konstante\ Verteilung} - eine\ konstante\ ZV\ ist\ eine\ Variable\ \text{im}\ herkömmlichen\ Sinn.$

$\boldsymbol{Bernoulli} - P(\omega) = p^{\sum_1^n \omega_i} \cdot (1-p)^{\sum_1^n 1-\omega_i}$

Bsp: Ziehen ganz oft Kugeln aus einer Urne mit 2 Farben mit Zurücklegen;
Wkt. dafür, dass ein ganz bestimmter Pfad auftritt, d.h.
eine ganz bestimmte Sequenz auftritt (z.B. 'rot,rot,blau')

$\boldsymbol{Gleichverteilung} - \quad P(\omega) = \frac{1}{|\Omega|}\ mit\ |\Omega| < \infty$

$\boldsymbol{Binomialverteilung} - \quad P(\omega) = \binom{n}{\omega} \cdot p^{\omega} \cdot (1-p^{(n-\omega)}),$ $n\ Anzahl\ gezogener\ Kugeln,$ $\omega\ Anzahl\ gezogener\ Kugeln\ der\ gewünschten\ Farbe,$ $p\ Prozentsatz\ der\ gewünschten\ Farbe\ an\ Gesamtanzahl$

Ziehen n Kugeln aus Urne mit 2 Farben mit Zurücklegen.
Die Anzahl ω erfolgreicher Versuche nach Durchführung von insgesamt n Versuchen;
Ist also nicht wie bei der $n-Bernoulli$ die Wkt. eines Pfades, sondern die Wkt. aller Pfade!

Multinomialverteilung −
Ziehe ' n ' mal aus Urne mit Zurücklegen. Urne enthält Kugeln ' m ' verschiedener Farben.
Gibt Wkt. an , eine bestimmte Farbkombination zu ziehen (*z.B.* 2mal *blau* , 3mal *rot* , 2mal *grün*).

$$\omega_m := n - \sum_{i=1}^{m-1} \omega_i$$

$$P(X=(\omega_1, \omega_2, \dots, \omega_m)) = \underbrace{\binom{n}{\omega_1 \omega_2 \dots \omega_m}}_{\text{ist Multinomialkoeffizient}} \cdot \prod_{i=1}^{m} \left(\frac{N_i}{\sum_n N_i} \right)^{\omega_i}$$

N_i = Gesamtanzahl Kugeln der Farbe i
, ω_i = Anzahl gezogener Kugeln der Farbe i
$n = \sum(\omega_i)$ = alle gezogenen Kugeln

2 − *Hypergeometrisch* :− $P(X=\omega) = \dfrac{\binom{M}{\omega} \cdot \binom{N-M}{n-\omega}}{\binom{N}{n}}$,

N Gesamtanzahl Kugeln ,
M Anzahl Kugeln der gewünschten Farbe ,
n Anzahl Ziehungen ,
ω Anzahl gezogener Kugeln der gewünschten Farbe

Ziehen aus Urne mit 2 Farben ohne Zurücklegen. Wkt. dafür,
dass ω viele Kugeln der gewünschten Farbe gezogen wurden.

Hypergeometrisch − $P(X=(\omega_1, \omega_2, \dots, \omega_m)) = \dfrac{\prod_{i=1}^{m} \binom{N_i}{\omega_i}}{\binom{\sum_m N_i}{n}}$,

m Anzahl Farben ,
n Anzahl gezogener Kugeln ,
N_i Anzahl Kugeln der Farbe i
ω_i Anzahl gezogener Kugeln der Farbe i

Ziehen aus Urne mit m Farben ohne Zurücklegen = multinomial ohne Zurücklegen

Geometrische Verteilung − $P(X=\omega) = p^1 \cdot (1-p)^{\omega-1}$
Wkt. dafür, dass bei ω-vielen Versuchen das gewünschtes Ereignis genau einmal auftritt.
Bzw.: Wie viele Versuche brauche ich, bis gewünschtes Ergebnis auftritt. (Abbruchverteilung)
Bemerkung: Geometrische Verteilung ist gedächtnislos

Negativ Binomial (***Pascal***) − $\binom{\omega-1}{m-1} \cdot p^m \cdot (1-p)^{\omega-m}$
Anzahl Versuche, bis erwünschtes Ereignis m-Mal aufgetreten ist.
Die Anzahl ω von Versuchen, die benötigt werden,
um eine vorgegebene Anzahl von m Erfolgen zu erzielen

Poisson − Verteilung − $P(X=\omega) = \dfrac{e^{-\lambda} \cdot \lambda^{\omega}}{\omega!}$,

λ Proportionalitätsfaktor für Zeiteinheit ,
ω Anzahl gewünschter Ereignisse

Wieviele Ereignisse treten in einem Zeitraum λ auf?

Kontinuierliche Zufallsverteilungen

Def. ***Kontinuierliche Verteilungen*** *:* Verteilungen von ZV auf kontinuierlichen Ω
heißen *kontinuierliche Verteilungen.*

Gamma − Funktion *:* Die *Gamma − Funktion* ist die Fortführung der Fakultät ins Kontinuierliche
(mittels Integration), d.h. sehr umgangssprachlich geschrieben: $\Gamma(r) = r!, r \in \mathbb{R}$

Exponential − Verteilung: $\alpha \cdot e^{-\alpha \cdot x}$
Fortsetzung der Geometrischen Verteilung ins Kontinuierliche. Ist gedächtnislos.
Mißt die Zeit bis zum Eintreffen eines Ereignisses.

Erlang − Verteilung: $\frac{\alpha^m}{(m-1)!} \cdot x^{m-1} \cdot e^{-\alpha \cdot x}$
Fortsetzung der Pascal'schen-Verteilung ins Kontinuierliche.
Wir warten auf das m-malige Eintreten eines Poisson-Prozesses und brechen dann ab.
Bsp: Wie lange dauert es, bis x Anrufe eingegangen sind, wenn im Mittel in jeder Minute y Anrufe eingehen?

Normal − Verteilung: $\sqrt{\frac{1}{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \; dx$
Entsteht als Grenzverteilung der Binomialverteilung.

$$1-Bernoulli- \quad E(X)=p \;,\; V(X)=p \cdot q \;,\; Charakt.\,Fkt.=q+p\,e^{is}$$
$$Konstant- \quad E(X)=x \;,\; V(X)=0 \cdot q \;,\; Charakt.\,Fkt.=e^{isx}$$
$$Binomial- \quad E(X)=n \cdot p \;,\; V(X)=n \cdot p \cdot q \;,\; Charakt.\,Fkt.=(q+p\,e^{is})^n$$
$$Geometrisch- \quad E(X)=\frac{1}{p} \;,\; V(X)=\frac{q}{p^2} \;,\; Charakt.\,Fkt.=\left(\frac{pe^{is}}{1-qe^{is}}\right)^m$$
$$Poisson- \quad E(X)=\lambda \;,\; V(X)=\lambda^2 \;,\; Charakt.\,Fkt.=e^{\lambda \cdot e^{is}-1}$$

$$Gleichverteilt\,auf\,[0,1]- \quad E(X)=\frac{1}{2} \;,\; V(X)=\frac{1}{12} \;,\; Charakt.\,Fkt.=\frac{e^{is}-1}{is}$$
$$Exponential- \quad E(X)=\frac{1}{\alpha} \;,\; V(X)=\frac{1}{\alpha^2} \;,\; Charakt.\,Fkt.=\frac{\alpha}{\alpha-is}$$
$$Erlang- \quad E(X)=\frac{m}{\alpha} \;,\; V(X)=\frac{1}{m \cdot \alpha^2} \;,\; Charakt.\,Fkt.=\left(\frac{\alpha}{\alpha-is}\right)^m$$
$$Gamma- \quad E(X)=\frac{\beta}{\alpha} \;,\; V(X)=\frac{1}{\beta \cdot \alpha^2} \;,\; Charakt.\,Fkt.=\left(\frac{\alpha}{\alpha-is}\right)^\beta$$
$$Normalverteilung- \quad E(X)=\mu \;,\; V(X)=\sigma^2 \;,\; Charakt.\,Fkt.=e^{\mu is-\frac{\sigma^2 s^2}{2}},\left(bei\;Std-Normalvtlg: e^{-\frac{s^2}{2}}\right)$$
$$Chi^2- \quad E(X)=n \;,\; V(X)=2n$$

Chi^2 − Verteilung (eine Testverteilung)− $X_{1,\ldots}, X_n$ normalverteilte ZV mit Parametern $\mu=0, \sigma=1$.
Dann ist $(X_1)^2+\ldots+(X_n)^2\,Chi^2$-verteilt mit n Freiheitsgraden.
Die Chi-Quadrat-Verteilung ist eine so genannte Stichprobenverteilung, die bei der Schätzung von Verteilungsparametern, beispielsweise der Varianz, Anwendung findet. Man benutzt sie zur Beschreibung der Summe unabhängiger quadrierter standardnormalverteilter Zufallsvariablen.

Generierende, Charakteristische, Momentgenerierende Funktionen

Generierende Funktion: $\phi_X(T)=\sum_{i=o}^{\infty}(P(X=i)\,T^i)$, mit T^i ist der i-te konkrete Wert einer Variablen.

ist X diskret, so gilt: $\phi_X(T)=E(T^x)$, mit x ist Anzahl der Variablen von X.

***Charakteristische Funktion (der Verteilung)*:** $\Phi_X(s) = \int_{-\infty}^{\infty} (f_X(t) \cdot e^{ist}) dt$

Dies ist eine Fouriertransformation auf der Wahrscheinlichkeitsdichte f_X.

Es gilt: $\Phi_X(s) = E(e^{isX}) = \phi_X(e^{is})$

***Momentgenerierende Funktion*:** $M(s) = \Phi_X(-is) = \int_{-\infty}^{\infty} (f_X(t) \cdot e^{ts}) dt$

Setzen $-is$ anstatt s in die charakteristische Fkt Φ ein, um den Imaginärteil i los z werden.

Leiten wir nun $M(s)$ (bzw. $\Phi_X(-is)$ $r-$mal ab und setzen $s=0$, so erhalten wir das r-te Moment!!

Also: $M^{(r)}(0) = r-$tes Moment (Als Rechnung: $M^{(r)}(0) = \int_{-\infty}^{\infty} (f(t) \cdot t^r) dt$)

Das ***Gesetz der großen Zahlen*** besagt, dass sich die Häufigkeit eines Zufallsergebnisses immer weiter an dessen Wahrscheinlichkeit annähert, je häufiger das Zufallsexperiment durchgeführt wird.

Bsp: „Die Wahrscheinlichkeit, dass eine Münze beim Werfen Kopf zeigt, betrage ½. Je häufiger die Münze geworfen wird, desto näher wird der Anteil der Würfe, bei denen Kopf erscheint, beim theoretischen Wert ½ liegen. Trotzdem wird der absolute Abstand anwachsen."

Der ***Zentrale Grenzwertsatz*** besagt, dass der Durchschnitt von sehr vielen Zufallsvariablen mit beschränkter Varianz eine Normalverteilung ist.

Statistik

Die Statistik befasst sich mit Häufigkeiten und Wahrscheinlichkeiten, wobei einige Parameter, Variablen oder Verteilungen unbekannt sind und deshalb auf vielfältige Art und Weise geschätzt werden müssen. Die Statistik findet ins Besondere in *real-world* Problemen Anwendung, da Prozesse in der realen Welt meist nicht komplett erfasst, messbar und berechenbar sind.

Einführung in die Statistik

Was ist *'Confusion of the Inverse'*: Fälschlicherweise wird oft gedacht: $P(A|B) = P(B|A)$.

Es muss aber heißen: $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$. (*nach Satz der totalen Wkt.*)

***Statistisches Modell*:** Ein Modell ist ein Tupel (L, X) von latenten und beobachteten Variablen, wobei $L = (L_{1,\dots}, L_n)$ und $X = (X_{1,\dots}, X_m)$.

Die Verteilung der X hängt ab oder wird determiniert von den L.

***Stichprobe / Datensatz*:** Eine Menge von beobachteten Variablen einer oder mehrerer Beobachtungsobjekte; ist also eine Matrix.

***Modellfamilie*:** Hängt die Verteilung der Variablen von einem oder mehreren Parametern ab (welche selbst Zufallsvariablen sind), so handelt es sich um eine Modellfamilie.

***Hypothese*:** eine anhand empirischer Daten zu prüfende Annahme.

***α− Fehler (Irrtumswahrscheinlichkeit, 'false positives')*:** $P(T \neq H|H)$

Beispielsweise wird eine Person zu Unrecht als krank bezeichnet, obwohl sie tatsächlich gesund ist.

***β− Fehler ('false negatives')*:** $P(T = H|\bar{H})$

Bsp: Eine Person wird zu Unrecht als gesund bezeichnet, obwohl sie tatsächlich krank ist.

***Power (Sensitivität)*:** $1-\beta$-Fehler oder $P(H|\Theta = \theta)$ *mit festem* θ

Schätzer und Maximum-Likelihood-Schätzer

Ein *Schätzer* ist ein Kennwert (oder eine Menge von Kennwerten), welche sich einem Parameter der Modellfamilie annähern.

$$\textbf{Maximum}-\textbf{Likelihood}-\textbf{Schätzer} : \hat{\theta} = argmax_\theta\left(\prod_{i=1}^{N}(f_\theta(X))\right) \quad = argmax_\theta \sum_{i=1}^{N}(\ln(f_\theta(X)))$$

Der Maximum-Likelihood-Schätzer (oder universelle Schätzer) des Modells ist derjenige Schätzer von allen Schätzern, für den die Likelihood am größten ist

Markov-Ketten und Markov-Prozesse

Jeder Zustand ist genau vom Vorgängerzustand abhängig.

Jede Markovkette besitzt genau eine stabile Verteilung, selbst wenn diese Verteilung nicht unbedingt eingenommen werden muss.

Existieren mehrere starke Zusammenhangskomponenten, so gibt es ggf. mehrere potentielle stabile Verteilungen, welche vom Startzustand abhängig sind und sich so entscheidet. Ist die Markovkette ergodisch, so konvergiert sie gegen genau einen stabilen Zustand.

***Periodizität*:**

- ggT aller Wege eines Zustands zu sich selbst
- aller Zustände eines Graphen / einer starken Zusammenhangskomponente: *kgV*
- Ist ein Zustand aperiodisch, so ist die gesamte starke Zusammenhangskomponente aperiodisch

$$\textbf{stabile Verteilung } \boldsymbol{\beta} : M \cdot \beta = \beta \text{, } M \textit{ ist Übergangsmatrix} \text{ und } \beta \textit{ ist Verteilung auf den Zuständen}$$

Versteckte Markov Ketten:

M = (L,X) mit gemeinsamen Ereignisraum und alle Xi sind durch die Li determiniert.

Schätzung des wahrscheinlichlichsten Pfades (=Zustandsvektors) aus einem Datensatz:

(schneller durch Viterbi-Algorithmus)

Strukturgleichungsmodelle (SEMs)

$$\textbf{SEM} : M = (L, X) \textit{ mit } X = \Lambda \cdot L \quad \begin{array}{l} X \textit{ beobachtete Variablen (oder Merkmale)}, \\ L \textit{ Latente Variablen (oder Faktoren)} \end{array}$$

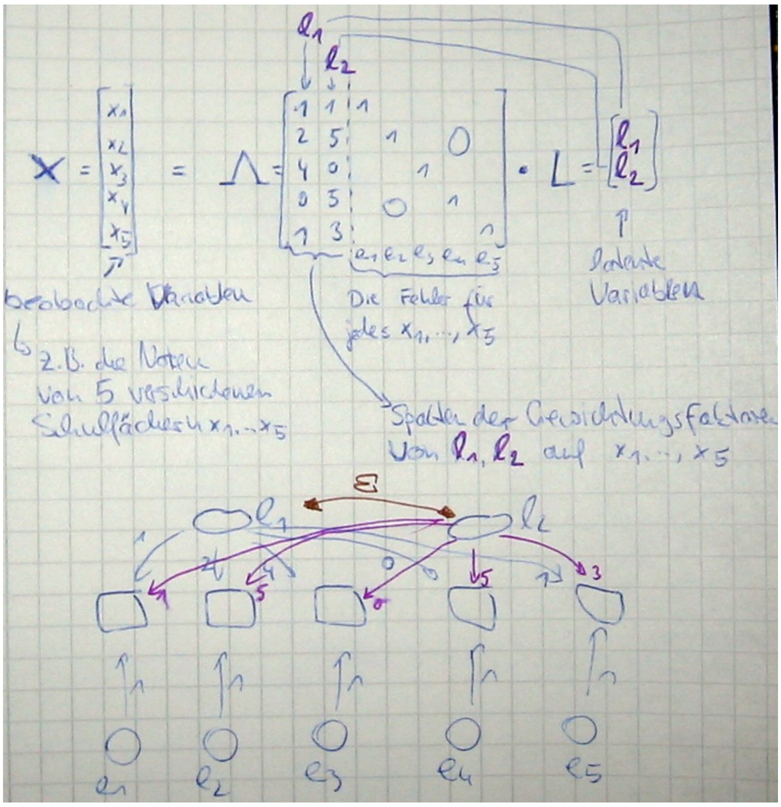
$$\textbf{Strukturmatrix } \boldsymbol{\Lambda} = \textit{Spalte der Varianzen, Spalte der Mittelwerte, je Spalte pro Fehler.}$$
$$\textit{Anzahl Zeilen} = \textit{Anzahl beobachteter Variablen.}$$

$$\textbf{Mittelwertsvektor } \boldsymbol{\mu} = \textit{Pro Zeile die Addition der Mittelwerte von Latenten und}$$
$$\textit{Beobachteten Variablen dieser Zeiteinheit}$$

$$\textbf{Covarianzmatrix } \boldsymbol{\Sigma} = E(XX^T) = CC^T = \Lambda C \Lambda^T$$

Die beobachteten Variablen wurden tatsächlich gemessen, d.h. ein Datensatz hat konkrete Instanzen der beobachteten Variablen.

Für diesen konkreten Datensatz sagen dann die in der Strukturmatrix %LAMBDA angegebenen Faktoren der latenten Variablen aus, wie sehr die latenten Variablen die Werte der beobachteten Variablen beeinflusst haben.


X = x1 x2 x3 x4 x5 = Λ · L = l1 l2
beobachtete Variablen
z.B. die Noten von 5 verschiedenen Schulfächern x1,...,x5
e1 e2 e3 e4 e5
Die Fehler für jedes x1,...,x5
latente Variablen
Spalten der Gewichtungsfaktoren von l1, l2 auf x1,...,x5
l1
l2
e1
e2
e3
e4
e5