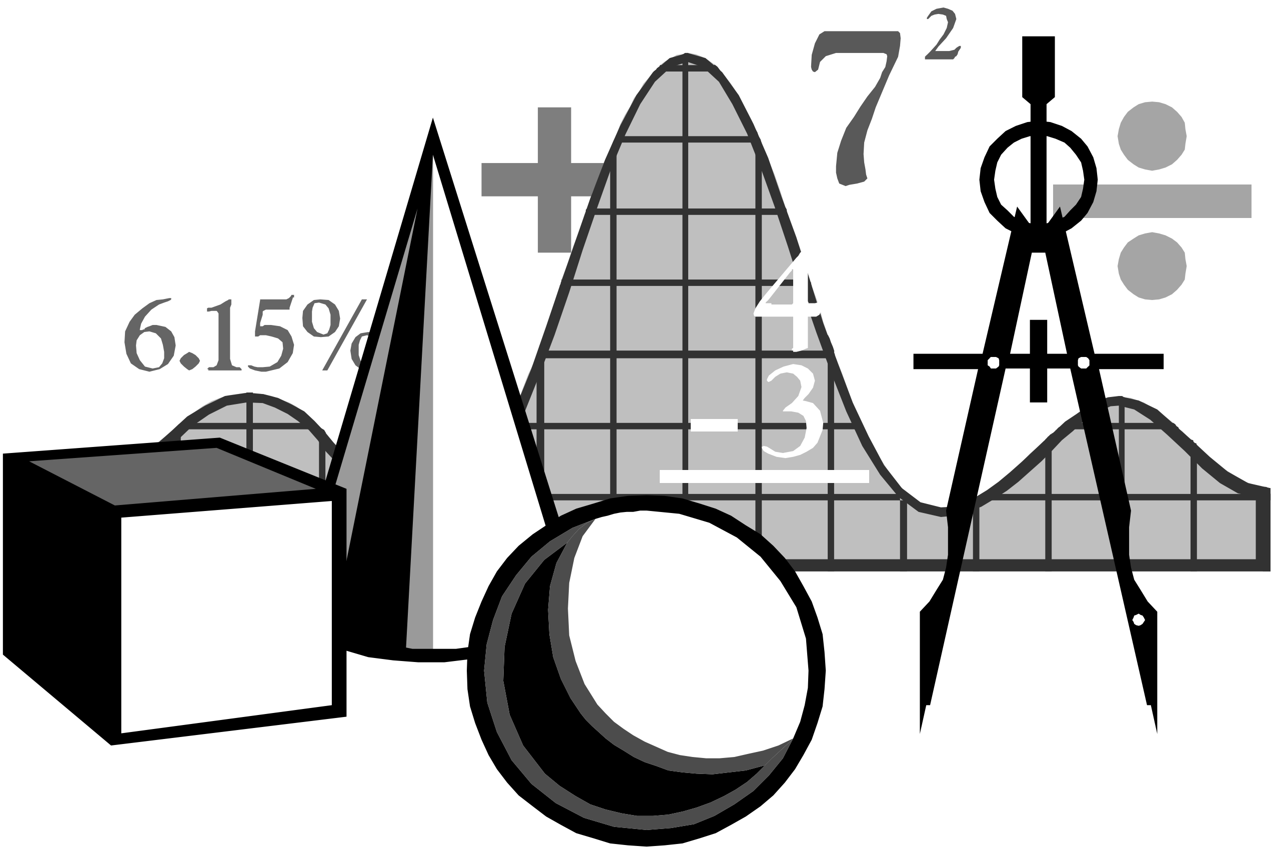
Statistik

für Wirtschaftsinformatiker



Martin Vonwald
Martin@voni.at

Aus den Vorlesungen für Statistik für WInf I+II
an der Technischen Universität Wien
(SS+WS 1995/96)

Inhalt

I Einleitung

1. Grundsätzliches und Historisches

Qualitative Aussagen und quantitative Aussagen
Statistik ist die zahlenmäßige Erfassung und Beschreibung von Massenphänomenen oder nichtdeterministischen Größen.
Beispiele:

- Volkszählungen
- statistische Prognosen

Historische Entwicklung

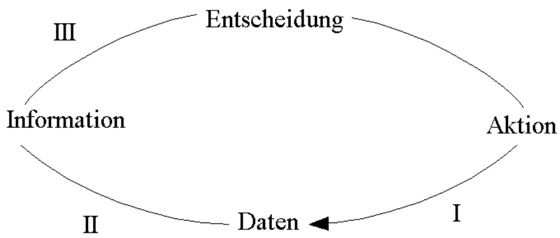
1) Amtliche Erhebung
Altertum: für militärische Zwecke und Geldeintreibung
2) Universitätsstatistik
Im 17 Jhdt. Lehre der Staatsmerkwürdigkeiten
Grundlage für den Vergleich von Staaten
G. Achenwall (1719-1772) Prof. in Göttingen, von dem der Name Statistik stammt.
3) Politische Arithmetik
Suche nach Gesetzmäßigkeiten und sozial-statistischen Daten
4) Stochastik
Im 16. Jhdt.: mathematische Beschreibung von Glücksspielen (J.F.C. Gauß)
Zwei wesentliche Wahrscheinlichkeitsbegriffe:

Klassische objektivistische Wahrscheinlichkeit	Subjektive Wahrscheinlichkeit
A. Kolmogoroff (1903-1987)	B. de Finetti (1906-1985)
statisch	dynamisch

2. Aufgabe und Gegenstand der Statistik

Quantitative Beschreibung gewisser Umweltphänomene: Informationsgewinnung durch Komprimieren von Daten, Erklärungsmodelle für nicht deterministische (=stochastische) kausale Zusammenhänge.
Es gibt zwei Definitionen: Statistik ist

1) eine Methodenlehre, die sich mit der Entstehung und Anwendung von Erhebungen und Versuchen sowie mit einer am stochastischen Modell orientierten Beurteilung der Ergebnisse beschäftigt.
2) das Bestreben, die Dinge so zu sehen, wie sie sind.



Viele Gründe sprechen für Teilerhebungen (billiger,...). Die Vergleichbarkeit von Daten ist oft ein Problem.

Amtliche Statistik
Städte, Gemeinden, Bundesländer, Staat, Regionen, International, Global
Statistische Ämter größerer Städte
Statistische Landesämter
Österreichisches Statistisches Zentralamt: ÖSTAT
Statistisches Amt der EU: EUROSTAT
Statistisches Amt der UNO: UNSO
Statistischer Dienst der UNIDO, OECP, IAEA.
Statistische Dienste: Ministerien, Interessengemeinschaften, Vereine, Gewerkschaften, Umweltbundesamt,...

3. Begleitende Beispiele

B1: Bei Qualitätskontrollen von Warenlieferungen soll aus sogenannten Stichproben (Untersuchung eines Teils der Ware) Information über den Anteil der schlechten Stücke in der gesamten Warenlieferung gewonnen werden.

B2: In der Betriebswirtschaft spielen Lebensdauer von Systemen eine Rolle z.B. für Erneuerung des Fuhrparks. Da die Lebensdauer i.a. nicht detailliert vorhersagbar ist, benötigt man stochastische Modelle.

B3: Abhängigkeit des Bremsweges von der Geschwindigkeit.

B4: Arbeitszeit für ein Werkstück.

II Beschreibende Statistik

4. Merkmalsarten und Datentypen

4.1 Artmäßige Merkmale (Nominaldaten)

keine natürliche Reihenfolge
gleichbedeutendes Nebeneinander
Bsp.: Farbe, Geschlecht, Religion, Nationalität, ...

4.2 Rangmerkmale (Ordinaldaten)

natürliche Rangordnung („größer"-Beziehung)
Bsp.: Güteklassen, Prüfungsnoten, Rangplätze, ...

4.3 Abstandsmerkmale (Intervalldaten)

Neben der Rangordnung können sinnvolle Abstände angegeben werden. Nullpunkt kann willkürlich gewählt werden.
Bsp.: Kalenderrechnung, Zeitmessung, Temperaturmessung, ...

4.4 Merkmale mit Verhältnisskalen (metrische Daten)

Zusätzlich zu den Eigenschaften von Intervalldaten gibt es einen absoluten Nullpunkt. Dadurch wird der Quotient (Verhältnis) zweier Ausprägungen von der gewählten Maßeinheit unabhängig.
Bsp.: Körpergröße, Alter, Einkommen, ...
Bei metrischen Größen unterscheidet man zwischen **diskreten** und **kontinuierlichen** Größen. Manchmal treten auch gemischte Größen auf.
Manchmal werden diskrete Größen durch kontinuierliche Größen beschrieben. Dann spricht man von **quasikontinuierlichen** Größen.

4.5 Eindimensionale und mehrdimensionale Merkmale

Wenn Erhebungseinheiten eine Zahl zugeordnet wird, dann spricht man von einem 1-dimensionalen Merkmal. Wenn der Erhebungseinheit ein Vektor zugeordnet ist, spricht man von einem mehrdimensionalen Merkmal.

5. Aufbereitung von Daten

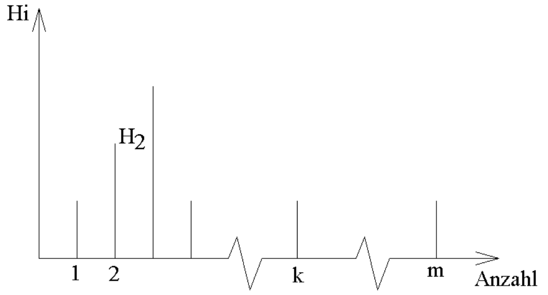
Oft liefern Erhebungen bzw. Versuche eine große Anzahl von Daten, die zu einer überschaubaren Information zusammengefaßt werden sollen. Konzentration der Daten in Kennzahlen, Diagrammen, Tabellen und Verteilungsmodellen.

5.1 Diskrete Größen

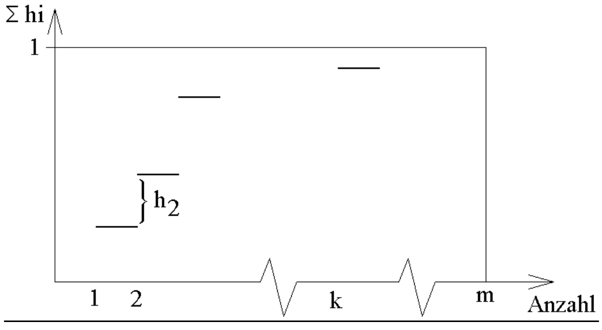
Solche Größen können höchstens abzählbar viele verschiedene Werte annehmen, die sich nicht häufen dürfen. Beobachtete Daten werden in **Strichlisten** und **Stabdiagrammen** zusammengefaßt.

Merkmal i	Anzahl dieser Werte		rel. Häufigkeit	summierte Häufigkeit
	Strichliste	Häufigkeit Hi	$h_i = \frac{H_i}{n}$	$\sum_{j=1}^{i} h_j$
1	||||	4	4/n	h1
2	||	2	2/n	h1+h2
3	|||	3	3/n	
...				
k	|	1	1/n	h1+h2+...+hk
...				
m	||	2	2/n	1
	Summe	n	1	

Graphische Darstellung in einem Stabdiagramm

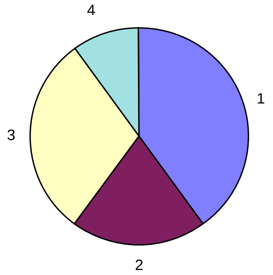


Summenkurve



Kreisdiagramme (Torten-)

Kreisflächen proportional zu den Anteilen (rel. Häufigkeit)



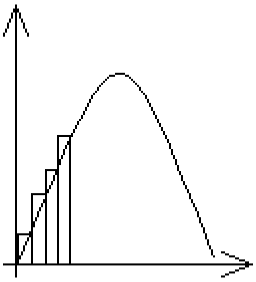
5.2 Kontinuierliche Merkmale

Solche Größen können alle Werte (=Zahlen) aus einem Intervall annehmen (auch unendlich!). Bei größeren Beobachtungszahlen (n>30) ist eine **Klasseneinteilung** zweckmäßig. Zu jeder Klasse stellt man die absolute Häufigkeit H_i und die relative Häufigkeit h_i fest.

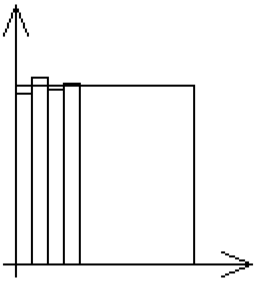
Klasse (a_i,b_i)	Anzahl dieser Werte		rel. Häufigkeit	summierte Häufigkeit
	Strichliste	Häufigkeit Hi	$h_i = \frac{H_i}{n}$	$\sum_{j=1}^{i} h_j$
(3,4]	||||	4	4/n	h1
(4,5]	||	2	2/n	h1+h2
...	...	...	...	...
(13,14]	|||	3	3/n	1
	Summe	n	1	

<u>Typen von Verteilungen</u> (Histogramm)

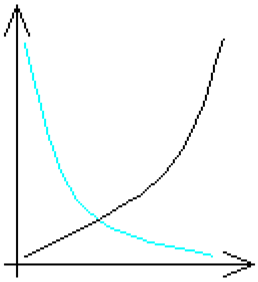
Je nach Gestalt des Histogrammes unterscheidet man:



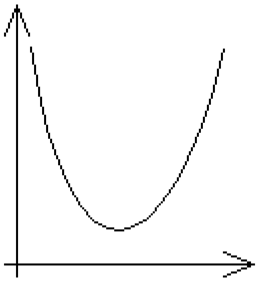
symmetrisch eingipfelig



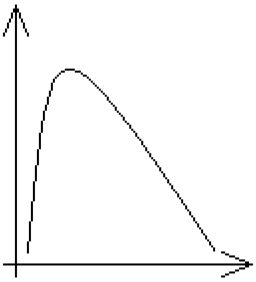
rechteckig



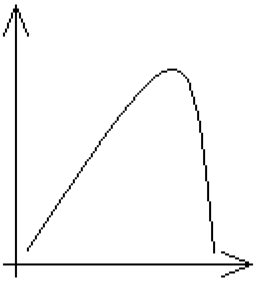
S-förmig



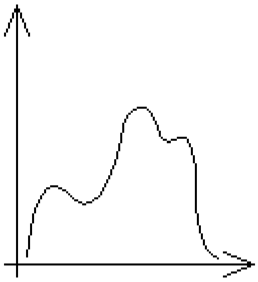
U-förmig



linkssteil, rechtsschief



rechtssteil, linksschief

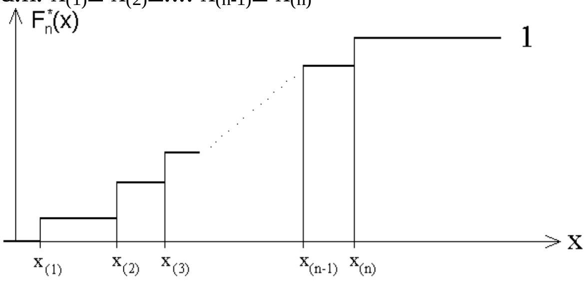


mehrgipfelig

Falls man alle Urdaten eines kontinuierlichen Merkmals in die „Verteilung“ aufnehmen will:

<u>Empirische Verteilungsfunktion</u> $F_n^*(\cdot)$

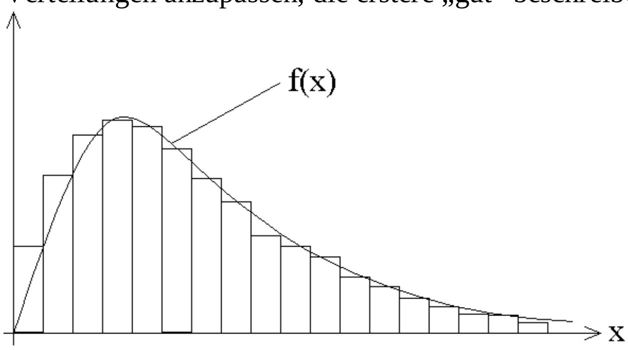
Für die beobachteten (1-dimensionalen) Daten $x_1, \ldots, x_n$ gilt eine sog. **Ordnungsstatistik** $x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(n)}$, d.h. $x_{(1)} \leq x_{(2)} \leq \ldots.\ x_{(n-1)} \leq x_{(n)}$



Treppenfunktion mit den Sprunghöhen $^1/_n$
n...Anzahl der Beobachtungen

Bemerkung: Falls k Beobachtungen identisch sind, so ist dort die Sprunghöhe $^k/_n$.

Ziel der schließenden Statistik ist es, den empirisch gegebenen Verteilungen (Histogramme) „theoretische“ Verteilungen anzupassen, die erstere „gut“ beschreiben.



$f(\cdot)$.....Dichtefunktion (theoretische Verteilung)

6. Statistische Kenngrößen eindimensionaler Häufigkeitsverteilungen

Für n Beobachtungen $x_1, ..., x_n$ einer 1-dimensionalen Merkmals sucht man charakteristische Größen deren Verteilung.

6.1 Lageparameter

a) Mittelwert

Ist das arithmetisches Mittel: $\overline{x}_n = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sind die Daten bereits gruppiert mit Klassenmitten z_j und Häufigkeiten H_j, so gilt:

$$\overline{z} := \frac{\sum z_j \cdot H_j}{\sum H_j} = \sum_{j=1}^{k} z_j \cdot h_j$$

b) Geometrisches Mittel

$$\overline{x}_g := \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

c) Median einer empirisch gegebenen Verteilung

Für Urdaten wird bei ungerader Anzahl, d.h. n=2k+1, der „mittlere" Wert $x_{(k+1)}$ als Median bezeichnet. Bei gerader Anzahl, d.h. n=2k, gilt: $\underset{-}{x} = \frac{x_{(k)} + x_{(k+1)}}{2}$

Für gruppierte Daten ermittelt man das Summenpolygon und definiert den (empirischen) Median als jenen Wert, für den das Summenpolygon den Wert ½ annimmt.

Allgemeiner nennt man für $0<p<1$ jenen Wert x_p, für den das Summenpolygon den Wert p annimmt, das (empirische) **p-Fraktile** (oder 100·p-Perzentile der empirischen Verteilung).

Bemerkung: Der Anteil jener Werte, die kleiner gleich x_p sind ist gleich p.

d) Modus

Falls eine Verteilung einen häufigsten Wert (diskretes Merkmal) hat, so wird dieser als Modus bezeichnet. Im kontinuierlichen Fall ist der Modus als die Mitte jener Histogrammklasse definiert, über der das Histogramm den höchsten Balken hat (falls dies eindeutig). Bei theoretischen Verteilungen verwendet man die Dichtefunktion und bestimmt ihr Maximum.

6.2 Streuungsparameter

Für die Form einer Verteilung sollen ebenfalls charakteristische Werte (Parameter) gefunden werden.

a) Spannweite

Spannweite:=$x_{(n)}-x_{(1)}=x_{(max)}-x_{(min)}$

b) Quartilsabstand $x_{0,75}-x_{0,25}$

Bemerkung: Im Intervall $[x_{0.25};x_{0.75}]$ liegen 50% aller Daten.

c) Mittlere absolute Abweichung MAD

$$\text{MAD} := \frac{1}{n}\sum_{i=1}^{n} \left|x_i - x_{0,5}\right|$$

Bemerkung: Berechnet man $\sum_{i=1}^{n}\left|x_i - x_0\right|$ für irgend einen Wert $x_0 \in R$ so gilt $\sum_{i=1}^{n}\left|x_i - x_{0,5}\right| \le \sum_{i=1}^{n}\left|x_i - x_0\right|$. Bei gruppierten Daten mit k Gruppen, den Gruppenmitten z_j und den absoluten Häufigkeiten H_j gilt:

$$\text{MAD} := \frac{1}{n}\sum_{j=1}^{k} \left|z_j - z_{0,5}\right| \cdot H_j$$

d) Empirische Varianz

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2$$

Bemerkung:
- Bei Stichproben oft Division durch n-1.
- $\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2 \le \sum_{i=1}^{n}\left(x_i - x_0\right)^2 \quad \forall x_0 \in R$
- für gruppierte Daten:

$$s_g^2 = \frac{1}{n}\sum_{j=1}^{k}\left(z_j - \overline{z}_g\right)^2 \cdot H_j$$

$s = \sqrt[+]{s^2}$ heißt empirische Streuung oder empirische Standardabweichung

e) Empirischer Variationskoeffizient

$$VK = \frac{S}{\overline{x}} \qquad \text{Bemerkung: dimensionsloses Streumaß}$$

7. Zweidimensionale Häufigkeitsverteilungen

Für Untersuchungen von Zusammenhängen zwischen eindimensionalen Größen X und Y betrachtet man Beobachtungen (x_i, y_i); i=1(1)n

7.1 Kontingenztafeln

Sind beide Merkmale qualitativer Art, diskret aber nicht der Größe nach zu ordnen.

X.....l Ausprägungen
Y.....m Ausprägungen
n Beobachtungen (x_k, y_k) k=1(1)n

x y	b_1	b_2	...	b_j	...	b_m	$n_{i\bullet} = \sum_{j=1}^{m} n_{ij}$
a_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	$n_{2\bullet}$
...	...						...
a_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	$n_{i\bullet}$
...	...						...
a_l	n_{l1}	n_{l2}	...	n_{lj}	...	n_{lm}	$n_{l\bullet}$
$n_{\bullet j} = \sum_{i=1}^{l} n_{ij}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet m}$	n

n_{ij}=Anzahl der Paare (x_k,y_k) mit $x_k=a_i$ und $y_k=b_j$
Will man die Abhängigkeit bzw. die Unabhängigkeit der beiden Merkmale X und Y untersuchen, so verwendet man die sog. **Unabhängigkeitszahlen** U_{ij}:

$$U_{ij} := \frac{n_i \cdot n_j}{n}$$

Bei statistischer Unabhängigkeit der beiden Merkmale müßte gelten: $n_{ij} \approx n_{i\bullet} n_{\bullet j} / n$

Beispiel: Vierfeldertafel

Falls X und Y nur 2 Warte annehmen können, d.h. m=2, l=2, n Beobachtungen.

X Y	b_1	b_2	
a_1	n_{11}	n_{12}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	$n_{2\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	n

Yule'scher Assoziationskoeffizient $Q = \dfrac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{n_{11} \cdot n_{22} + n_{12} \cdot n_{21}}$

7.2 Empirischer Korrelationskoeffizient

Den Grad eines eventuellen linearen Zusammenhanges zweier metrischer Merkmale X und Y kann man mittels des empirischen Korrelationskoeffizientens r beschreiben.
Der empirische Korrelationskoeffizient für beobachtete Paare ist definiert als:

Beobachtete Paare (x_k, y_k), k=1(1)n

$$r = \frac{\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\left[\sum_{k=1}^{n}(x_k - \bar{x})^2\right] \cdot \left[\sum_{k=1}^{n}(y_k - \bar{y})^2\right]}}$$

Bemerkung:

- Es gilt $-1 \le r \le 1$.
- r ändert sich nicht, wenn man anstelle der Wertepaare (x_k,y_k) Lineartransformationen (u_k,v_k) mit $u_k = \dfrac{x_k - x_0}{c}$ und $v_k = \dfrac{y_k - y_0}{d}$ nimmt.
- Falls $|r|=1$, so liegen alle Paare (x_k,y_k) auf einer Geraden in der x,y-Ebene (r=1: ansteigen; r=-1: fallend).
- r=0 bedeutet nicht keine Abhängigkeit, wohl aber keine lineare Abhängigkeit.

Bei *zweidimensionalen Häufigkeitsverteilungen* gilt:

l Klassen für die x-Werte: $G_1,\ldots,G_l$
m Klassen für die y-Werte: $K_1,\ldots,K_m$
Klassenmitten u_i der x-Werte; i=1(1)l
Klassenmitten z_j der y-Werte; j=1(1)m
Anzahl n_{ij} der Paare (x_k,y_k) mit $x_k \in b_i$ und $y_k \in k_j$

$$r_0 = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}(u_i - \bar{x})(z_j - \bar{y}) \cdot n_{ij}}{\sqrt{\left[\sum_{i=1}^{l}(u_i - \bar{x})^2 \cdot n_{i\bullet}\right]\left[\sum_{j=1}^{m}(z_j - \bar{y})^2 \cdot n_{\bullet j}\right]}}$$

mit der Bedeutung von $n_{i\bullet}$ und $n_{\bullet j}$ aus Abschnitt 7.1.

8. Indexzahlen

Hat man mehrere, sachlich zusammenhängende Reihen von Daten, so möchte man deren Verlauf oft jeweils durch eine Zahl beschreiben. Dies erfolgt durch sog. Indexzahlen.

Der Wertindex gibt eine Aussage über die relative Änderung des Wertes eines Warenbündels (Warenkorb).
1,...,n Warennummern der n Waren des Warenkorbs
$p_{t1},...,p_{tn}$ Preis der Waren (zur Berichtsperiode t)
$q_{t1},...,q_{tn}$ Menge der Waren
Basisperiode t=0

Wert des Warenkorbes in der Periode t: $\sum_{i=1}^{n} p_{ti} q_{ti}$

Wertindex: $I_{0t}^{W} = \frac{\sum_{i=1}^{n} p_{ti} \cdot q_{ti}}{\sum_{i=1}^{n} p_{0i} \cdot q_{0i}}$

Aussagen darüber, wie sich der tatsächliche Aufwand eines 4-Personen-Haushalts ändert, liefern **Preisindizes**.

Preisindex nach Paasche:

$${}_{P}I_{0t}^{P} := \frac{\sum_{i=1}^{n} p_{ti} \cdot q_{ti}}{\sum_{i=1}^{n} p_{0i} \cdot q_{ti}}$$

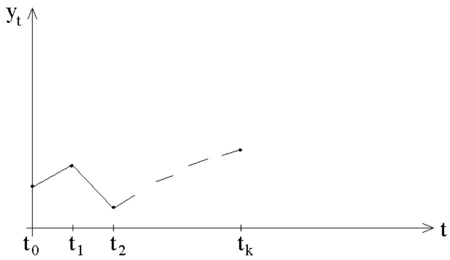
Preisindex nach Laspeyres:

$${}_{L}I_{0t}^{P} := \frac{\sum_{i=1}^{n} p_{ti} \cdot q_{0i}}{\sum_{i=1}^{n} p_{0i} \cdot q_{0i}} \qquad \text{(ÖSTAT)}$$

Bemerkung: Bei ${}_{P}I^{P}_{0t}$ wird von fiktiven Mengen zur Zeit t ausgegangen, bei ${}_{L}I^{P}_{0t}$ zum Zeitpunkt t=0.

9. Zeitreihen

Zeitreihe: y_t, t=1, 2,..., k (glatte Komponente)
Ziel ist die Schätzung einer sog. **Trendfunktion** g_t.
Idee: $y_t=g_t+r_t$
z.B. $g_t=a+bt$
$g_t=a+b\cdot t+c\cdot t^2$
$g_t=a\cdot e^{bt}$



Allgemein: $g_t=\psi(t,a,b,....)$ wobei ψ eine bekannte Funktion mit den Parametern a, b, ... ist.
Die Parameter müssen zur Anpassung der Komponente g_t an die Daten statistisch geschätzt werden.

9.1 Methode der kleinsten Abstandsquadratsumme

Bei dieser Methode wird das Minimum der folgenden Funktion bestimmt: $\sum_{t=1}^{n} (y_t - g_t)^2 \rightarrow \text{Min}$

9.2 Methode der gleitenden Mittelwerte

Es wird unterschieden zwischen:

a) Gleitende Mittelwerte ungerader Ordnung (2k+1)

$$\hat{g}_t := \frac{1}{2k+1} \sum_{j=t-k}^{t+k} y_j$$

b) Gleitende Mittelwerte gerader Ordnung (2k)

$$\hat{g}_t := \frac{1}{2k} \left[\sum_{j=t-k+1}^{t+k-1} y_j + \frac{1}{2}\left(y_{t.k} + y_{t+k}\right) \right]$$

Bemerkung: Die Trendkomponente der ersten k und der letzten k Beobachtungen können mit dieser Methode nicht geschätzt werden.

Beispiel:

$y_1=2$, $y_2=6$, $y_3=1$, $y_4=5$, $y_5=3$, $y_6=7$, $y_7=2$

Gleitender Mittelwert der Ordnung 3:

$2k+1=3 \Rightarrow k=1$

$\hat{g}_2 = \frac{2+6+1}{3} = 3$ $\hat{g}_3 = \frac{6+1+5}{3} = 4$ $\hat{g}_4 = \frac{1+5+3}{3} = 3$ .....

Gleitender Mittelwert der Ordnung 4:

$2k=4 \Rightarrow k=2$ $\hat{g}_3 = \frac{1}{4}\left[y_2 + y_3 + y_4 + \frac{1}{2}\left(y_1 + y_5\right) \right] = \frac{29}{8}$

9.3 Prognosen

Gegeben ist der zeitliche Verlauf einer nichtdeterministischen Größe Y_t (Zeitreihe). Man beobachtet nun die Zeitreihe bis zu einem Zeitpunkt t_n und versucht daraus Aussagen (Prognosen) über $Y_{tn+\Delta t}$ zu einem zukünftigen Zeitpunkt $t_n+\Delta t$ zu gewinnen. Dazu benötigt man eine **stochastische Modellbildung** nach dem Stand des Wissens mit Wahrscheinlichkeiten.

III Wahrscheinlichkeitsräume

10. Beschreibung der Unsicherheiten

Definition: Wenn der Ausgang eines Versuches nicht vorher bestimmt werden kann, spricht man von einem **statistischen Experiment**.

B 1: Lieferung von N elektronischen Bauteilen sind A Stück schlecht (Ausschuß). Entnimmt man eine Stichprobe von N>n Stücken, so ist die Feststellung der Anzahl a von schlechten Stücken in der Stichprobe ein statistisches Experiment.

B 2: Die Lebensdauer von technischen Produkten ist i.a. nicht vorhersehbar. Daher ist die Feststellung der Lebensdauer (z.B.: Glühbirne) ein statistisches Experiment.

B 3: Die Feststellung des Bremsweges bei fester Geschwindigkeit ist ein statistisches Experiment.

Gefragt: Numerische Aussagen (Zahlen), für die Beschreibung der Unsicherheit. Solche Zahlen nennt man **Wahrscheinlichkeiten**.

10.1 Klassische Wahrscheinlichkeitsdefinition

Für B1 wendet man die klassische Wahrscheinlichkeitsdefinition an:

Kann das statistische Experiment auf m endlich viele Arten ausgehen, so bezeichnet $M=\{a_1, a_2,..., a_m\}$ die Menge der möglichen Versuchsausgänge (**Merkmalraum**).

Für eine Teilmenge $E\subseteq M$ mit g Elementen wird die Wahrscheinlichkeit dafür, daß eines dieser Elemente als Versuchsausgang auftritt, definiert als:

$$W(E):=\frac{g}{m} \quad \text{(klassische Wahrscheinlichkeitsdefinition} = \frac{\text{günstig}}{\text{möglich}}$$

10.2 Permutation

Frage: Auf wieviel Arten kann man n Dinge in linearer Ordnung anordnen?

Antwort: Auf n! Arten mit $n!=n\cdot(n-1)\cdot(n-2)\cdot....2\cdot1$.

Beweis:

Für n=1 ist diese Aussage richtig, wie sich durch vollständige Induktion beweisen läßt. Es ist nun zu zeigen, daß die Aussage (die für n richtig ist) auch für n+1 richtig ist.

Angenommen man hat n+1 Dinge. Man erhält dann alle Anordnungen dieser n+1 Dinge, wenn man zu jeder Anordnung von n Stücken alle Einreihungen des (n+1)ten Stückes in diese Anordnung betrachtet.

$$\begin{array}{ccccc} \bullet & \bullet & \bullet .... & \bullet & \bullet \\ a_1 & a_2 & a_3 ...... & a_n & a_{n+1} \end{array}$$

Es gibt n+1 Positionen für a_{n+1}. Daher gibt es zu jeder Anordnung der Elemente $a_1,..., a_n$ n+1 Anordnungen der Elemente $a_1,..., a_n, a_{n+1}$. Daher gibt es insgesamt (n+1)·(Anzahl der n Elemente) mögliche Anordnung.

10.3 Kombinationen ohne Wiederholung

Frage: Auf wieviel Arten kann man aus n Stücken k Stücke auswählen, wenn es nicht auf die Reihenfolge der Auswahl ankommt und jedes Stück nur einmal vorkommt?

Antwort: Auf $\binom{n}{k}$ Arten.

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

Begründung: Aus allen linearen Anordnungen nimmt man die ersten k. Es gibt n! solche Anordnungen, wovon die ersten k! dasselbe Resultat liefern; analog die hinteren (n-k)!.

Anwendung auf B 1:

Die Anzahl der Auswahlen von n Stücken aus N vorhandenen ist $\binom{N}{n} = m$. Die Anzahl der Auswahlen von a Stücken aus A vorhandenen ist $\binom{A}{a}$. Die letzen Auswahlen müssen mit allen Auswahlen von n-a guten Stükken aus N-A insgesamt vorhandenen guten Stücken kombiniert werden. Die Anzahl der Auswahlen von n Stücken, in denen genau a schlechte Stücke und daher n-a gute Stücke sind, ist daher:

$$g = \binom{A}{a}\binom{N-A}{n-a}$$

Die klassische Wahrscheinlichkeitsdefinition liefert daher als Wahrscheinlichkeit des Ereignisses E „genau a schlechte Stücke in der Stichprobe vom Umfang n":

$$W(E) = \frac{\binom{A}{a}\binom{N-A}{n-a}}{\binom{N}{n}} \quad \text{für } a_1 \le a \le a_2 \quad \text{mit} \quad a_1 = \max\{0, n-(N-A)\},\ a_2 = \min\{n, A\}$$

Ist die Menge der möglichen Versuchsausgänge unendlich, so nützt die klassische Wahrscheinlichkeitsdefinition nichts.

10.4 Häufigkeitsinterpretation der Wahrscheinlichkeit

Anwendung auf B 2:

Die Menge der möglichen Versuchsausgänge ist $[0,\infty)$, d.h. der Merkmalraum ist $M=[0,\infty)$. Führt man viele Versuche dazu durch (viele Beobachtungen der Lebensdauer gleichartiger Bauteile), so kann man für ein Intervall $A=[t_1,t_2]\subseteq[0,\infty)$ nach n Beobachtungen von X die sog. **absolute Häufigkeit** $H_n(A)$ (Anzahl der Beobachtungen von X, die in A fallen) angeben. Hält man an A fest und macht sehr viele Beobachtungen ($n\to\infty$), so haben die sogenannten **relativen Häufigkeiten** $h_n(A):=\frac{H_n(A)}{n}$ ein konvergenzartiges Verhalten (empirisches Gesetz der großen Zahlen: siehe Gesetz der großen Zahlen).

Wahrscheinlichkeiten werden als idealisierte, relative Häufigkeiten aufgefaßt. Dazu übernimmt man die Eigenschaften von relativen Häufigkeiten und definiert sogenannte **Wahrscheinlichkeitsverteilungen** auf **Ereignisfeldern ξ**.

Ereignisse sind jene Teilmengen des Merkmalraumes, für deren Wahrscheinlichkeit man sich interessiert. Die Zusammenfassung aller Ereignisse nennt man Ereignisfeld. Solche Ereignisfelder haben sinnvoller Weise folgende Eigenschaften:

1) $0\in\xi$.............unmögliches Ereignis
$M\in\xi$................sicheres Ereignis
2) $A\in\xi \Rightarrow A^C=M\backslash A\in\xi$ (A^C ist dann das zu A komplementäre Ereignis)
3) $A\in\xi\wedge B\in\xi \Rightarrow A\cup B\in\xi$ und $A\cap B\in\xi$

Daraus folgt die analoge Beziehung für k Ereignisse:

$$A_i\in\xi \quad \forall i=1(1)k \Rightarrow \begin{cases} \bigcup_{i=1}^{k} A_i \in\xi \\ \bigcap_{i=1}^{k} A_i \in\xi \end{cases}$$

Aus mathematischen Gründen wird meist verlangt:

$$A_i\in\xi \quad \forall i\in N \Rightarrow \begin{cases} \bigcup_{i=1}^{\infty} A_i \in\xi \\ \bigcap_{i=1}^{\infty} A_i \in\xi \end{cases}$$

Bemerkung: Falls der Merkmalraum mit den reellen Zahlen übereinstimmt, so wird als Ereignisfeld das am wenigsten umfassende System von Teilmengen von R genommen, welches ein Ereignis bildet und alle Intervalle der Form (a,b] enthält. Dieses Ereignissystem wird als System der **Borel-Mengen** bezeichnet.

Eigenschaften von relativen Häufigkeiten für Ereignisse:

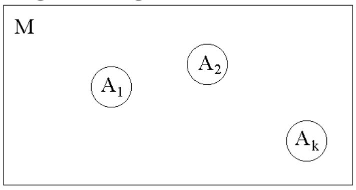
a) Normierung: $h_n(0)=0$
   $h_n(M)=1$

b) $0 \le h_n(A) \le 1 \quad \forall A \in \xi$

c) für einander paarweise ausschließende Ereignisse $A_1,...,A_k$, d.h. $A_i \cap A_j = 0 \quad \forall i \neq j$, gilt:

$$h_n\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} h_n(A_i) \qquad \textbf{Additivität}$$

Begründung: Man betrachte die absoluten Häufigkeiten und das sogenannte Venndiagramm.



Definition: Eine Wahrscheinlichkeitsverteilung W auf einem Ereignisfeld ξ ist eine Zuordnung, die jedem Ereignis $A \in \xi$ eine reelle Zahl W(A) zuordnet, wobei gilt:

1) $0 \le W(A) \le 1 \quad \forall A \in \xi$
2) $W(M)=1$
3) Für jede unendliche Folge A_1, A_2, ... von paarweisen disjunkten Ereignissen gilt:

$$W\left(\bigcup_{i=!}^{\infty} A_i\right) = \sum_{i=1}^{\infty} W(A_i)$$

Mit den voranstehenden Beziehungen heißt (M,ξ,W) ein **Wahrscheinlichkeitsraum**.

Sonderfälle von Wahrscheinlichkeitsräumen:

a) Klassisches Wahrscheinlichkeiten

$M=\{a_1,..., a_m\}$

$\xi=\Re(M)$ (System aller Teilmengen)

Für $\forall A \in \xi$ ist die Wahrscheinlichkeit definiert durch: $W(A) = \dfrac{\text{Anzahl der Elemente von A}}{m}$

b) M ist eine endliche oder abzählbare Menge

$M=\{a_1, a_2,..., a_n\}$, $\xi=\Re(M)$

$$\sum_{i=1}^{N} p_i = 1 \quad \text{mit } i=1(1)N \text{ und } p_i \ge 0$$

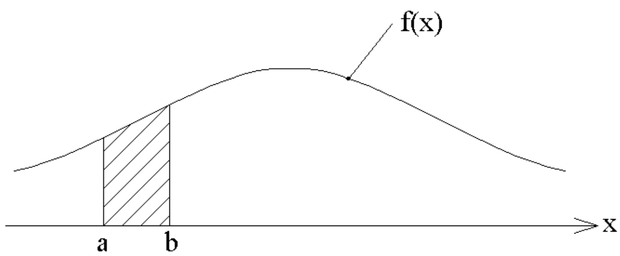
Für ein Ereignis $A \in \xi$ definiert man die Wahrscheinlichkeit von A durch:

$$W(A) = \sum_{a_i \in A} p_i$$

c) M umfaßt alle reellen Zahlen M=R

ξ sei das kleinste Ereignisfeld, das alle Intervalle umfaßt

$$f: R \to [0,\infty) \quad \text{mit} \quad \int_{-\infty}^{\infty} f(x)dx = 1$$



Für ein Ereignis $A \in \xi$ gilt dann:

$$W(A) = \int_A f(x)dx$$

Speziell gilt dann: A=[a,b] $W(A) = \int_a^b f(x)dx$

10.5 Bedingte Wahrscheinlichkeit

Beispiel: Farbenblindheit
A_1 .... Ereignis „Person männlich"
A_2 .... Ereignis „Person farbenblind"

Frage: Wie groß ist die relative Häufigkeit $h_n(A_2|A_1)$ von Farbenblinden unter männlichen Personen?
Gegeben sind: $H_n(A_1 \cap A_2)$=Anzahl der farbenblinden, männlichen Personen
$H_n(A_1)$=Anzahl der männlichen Personen

$$h_n(A_2|A_1) = \frac{H_n(A_1 \cap A_2) \cdot \frac{1}{n}}{H_n(A_1) \cdot \frac{1}{n}} = \frac{h_n(A_1 \cap A_2)}{h_n(A_1)}$$

Daher definiert man in Wahrscheinlichkeitsräumen (M,ξ,W) bedingte Wahrscheinlichkeit folgendermaßen:
Für $A \in \xi$ mit $W(A)>0$ ist dir durch A bedingte Wahrscheinlichkeit von B durch:

$$W(B|A) := \frac{W(A \cap B)}{W(A)}$$

Beispiel:
Zwei Schüler treten zu einer Prüfung an. Der Lehrer schätzt die Wahrscheinlichkeit, daß beide durchfallen mit 0,36 und die Wahrscheinlichkeit, daß beide durchkommen mit 0,16. Wie wahrscheinlich ist es, daß der erste bzw. zweite Schüler die Prüfung besteht?

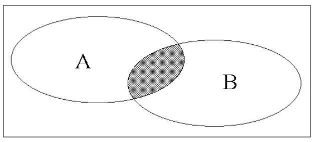
$W(A \cap B)=0{,}16=W(A) \cdot W(B)$
$W(A^C \cap B^C)=0{,}36=1-W(A \cup B)$
$W(A \cup B)=W(A)+W(B)-W(A \cap B)=W(A)+W(B)-W(A) \cdot W(B)=0{,}64 \Rightarrow W(B)=0{,}80-W(A)$
$W(A) \cdot [0{,}80-W(A)]=0{,}16$
$W(A)=0{,}40 \quad W(B)=0{,}40$

11. Eigenschaften von Wahrscheinlichkeitsräumen

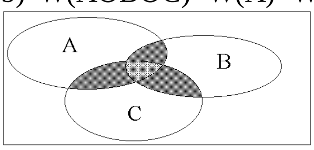
Geg.: Wahrscheinlichkeitsraum (M,ξ,W)
$A, B, H \in \xi$ (sind Ereignisse)

11.1 Additionstheorem

a) $W(A \cup B)=W(A)+W(B)-W(A \cap B)$
Beweis: Venndiagramm



Formal:
$A \cup B = A \cup (A^C \cap B)$
$B = (A \cap B) \cup (A^C \cap B)$
$W(A \cup B)=W(A \cup (A^C \cap B))=W(A)+W(A^C \cap B)$ Additivität, da disjunkte Ereignisse
$W(B)=W((A \cap B) \cup (A^C \cap B))=W(A \cap B)+W(A^C \cap B)$ · — ·
$\Rightarrow W(A \cup B)-W(B)= W(A)+\cancel{W(A^C \cap B)}-(W(A \cap B)+\cancel{W(A^C \cap B)})$
$\Rightarrow W(A \cup B)=W(A)+W(B)-W(A \cap B)$
b) $W(A \cup B \cup C)=W(A)+W(B)+W(C)-W(A \cap B)-W(A \cap C)-W(B \cap C)+W(A \cap B \cap C)$



c) $$W\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{\{i_1,\dots,i_m\} \subseteq \{1,\dots,n\}} (-1)^{m-1} \cdot W\left(A_{i_1} \cap A_{i_2} \cap \dots A_{i_m}\right)$$
Beweis: Vollständige Induktion

11.2 Multiplikationstheorem

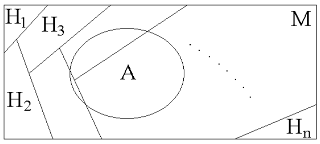
a) $W(A \cap B) = W(A) \cdot W(B|A) = W(B) \cdot W(A|B)$

b) $W(A_1 \cap ... \cap A_n) = W(A_1) \cdot W(A_2|A_1) \cdot W(A_3|A_1 \cap A_2) ... W(A_n|A_1 \cap ... \cap A_{n-1})$ Beweis: bedingte Wahrscheinlichkeit

11.3 Satz von der vollständigen Wahrscheinlichkeit

Sind $H_1, ..., H_n$ einander paarweise ausschließende Ereignisse, von denen jeweils eines eintritt (Zerlegung) , so gilt für jedes Ereignis $A \in \xi$:

$$W(A) = \sum_{k=1}^{n} W(A|H_k) \cdot W(H_k)$$



Beweis:

$$M = \bigcup_{k=1}^{n} H_k$$

$$A = A \cap M = A \cap \left(\bigcup_{k=1}^{n} H_k \right) = \bigcup_{k=1}^{n} (A \cap H_K)$$

$$\Rightarrow W\left(\bigcup_{k=1}^{n} (A \cap H_k) \right) = \sum_{k=1}^{n} W(A \cap H_k)$$

$$\Rightarrow W(A) = \sum_{k=1}^{n} W(A \cap H_k) = W(A|H_k) \cdot W(H_k) = \sum_{k=1}^{n} W(A|H_k) \cdot W(H_k) \quad \text{Multiplikationstheorem}$$

Anwendung:
Eine Produktion erfolgt auf n Maschinen
H_k ....k-te Maschine
A.................Ereignis Ausschußstück
$W(A|H_k)$......Wahrscheinlichkeit, daß die k-te Maschine ein Ausschußstück produziert

1. Wie groß ist die Wahrscheinlichkeit, daß ein zufällig herausgefallenes Stück der Gesamtproduktion Ausschuß ist?

Die Anteile, mit denen die Maschinen produzieren, werden als sogenannte **A-priori-Wahrscheinlichkeiten** $W(H_k)$ aufgefaßt.

$$W(A) = \sum_{k=1}^{n} W(A|H_k) \cdot W(H_k) \qquad \text{Satz von der vollständigen Wahrscheinlichkeit}$$

2. Wie groß ist die Wahrscheinlichkeit, daß ein Ausschußstück von der i-ten Maschine kommt?

Lösung mittels der sog. Bayes'schen Formel.
Anders formuliert lautet die Frage 2: Wie groß ist die Wahrscheinlichkeit einer „Hypothese" H_i nachdem A eingetreten ist?

11.4 Bayes'sche Formel

Unter der Voraussetzung vollständiger Wahrscheinlichkeit gilt:

$$W(H_i|A) = \frac{W(A|H_i) \cdot W(H_i)}{\sum_{k=1}^{n} W(A|H_k) \cdot W(H_k)}$$

Begründung: $W(H_i|A) := \dfrac{W(H_i \cap A)}{W(A)} = \dfrac{W(A|H_i) \cdot W(H_i)}{\sum_{k=1}^{n} W(A|H_k) \cdot W(H_k)}$ Satz der vollst. Wahrscheinlichkeit

Beispiel:

H_k .... k-te Maschine

$W(H_k)$ .... Wahrscheinlichkeit, daß ein zufällig herausgegriffenes Stück von der Maschine k ist.

$W(A|H_k)$ .... Wahrscheinlichkeit, daß die Maschine k Ausschuß produziert.

Bemerkung: $W(H_k)$ heißt A-priori-Wahrscheinlichkeit der H_k

$W(H_k|A)$ heißt **A-posteriori-Wahrscheinlichkeit** der H_k

Beispiel: Herkunft aus n Regionen, Arbeitslosigkeit in den Regionen

12. Stochastische Unabhängigkeit

Definition: Zwei Ereignisse A und B eines Wahrscheinlichkeitsraumes (M,ξ,W) heißen (stochastisch) unabhängig $A\perp B$, wenn der Eintritt eines der beiden Ereignisse die Wahrscheinlichkeit des Eintrittes des anderen Ereignisses nicht verändert, d.h.:

$$A\perp B :\Leftrightarrow \begin{cases} W(A|B) = W(A|B^c) = W(A) \\ W(B|A) = W(B|A^c) = W(B) \end{cases}$$

Es gilt: $A\perp B \Leftrightarrow W(A\cap B) = W(A)\cdot W(B)$

B1: Stichprobe vom Umfang zwei aus einer endlichen Gesamtheit mit N Einheiten.

$M=\{(x,y):x\in\{0,1\},y\in\{0,1\}\}$

0 .... gutes Stück

1 .... schlechtes Stück

$\xi=\Re(M)$

B_1 ..... erstes Stück schlecht

B_2 ..... zweites Stück schlecht

$B_1=\{(1,0),(1,1)\}$

$B_2=\{(0,1),(1,1)\}$

Die Wahrscheinlichkeit ist verschieden, je nachdem ob die Ziehungen *mit Zurücklegen* oder *ohne Zurücklegen* erfolgen.

Angenommen in der gesamten Warensendung sind A Ausschußstücke, so gilt für die Ziehung eines Stückes nach der klassischen Wahrscheinlichkeitsdefinition:

$$W(\{1\}) = \frac{A}{N} \text{ und } W(\{0\}) = \frac{N-A}{N} = 1-\frac{A}{N}$$

Wie ist die Wahrscheinlichkeit für ein Ausschußstück bei der Ziehung eines zweiten Stückes ?

a) Ziehung mit Zurücklegen

$$W(B_2|B_1) = \frac{A}{N} = W(B_2) \qquad W(B_2|B_1^c) = \frac{A}{N} = W(B_2) \quad \Rightarrow \quad B_1\perp B_2$$

b) Ziehung ohne Zurücklegen

$$W(B_2|B_1) = \frac{A-1}{N-1} \neq W(B_2) = \frac{A}{N} \;\Rightarrow\; B_1 \text{ und } B_2 \text{ sind nicht stochastisch unabhängig}$$

Bei Ziehungen mit Zurücklegen erhält man die Wahrscheinlichkeit für einen zusammengesetzten Versuch zu zwei Ziehungen als **Produktwahrscheinlichkeitsraum**.

$$M=M_1\times M_2 \qquad W(B_1\cap B_2)=W_1(B_1)\cdot W_2(B_2)$$

IV Stochastische Größen

13. Stochastische Größen- und Verteilungsfunktionen

Alle nichtdeterministischen Größen werden als stochastische Größen bezeichnet. Eine eindimensionale stochastische Größe nimmt Werte in R an, die nicht exakt vorhersagbar sind.
Nach dem Bestand des Wissens ist die beste Information über solche Größen deren Wahrscheinlichkeitsverteilung.
Stochastische Größen werden meist mit großen lateinischen Buchstaben vom Ende des Alphabets bezeichnet.
Beispiel: Anzahl X der schlechten Stücke in der Stichprobe
Stochastische Größen können als Funktionen von einem Wahrscheinlichkeitsraum nach R aufgefaßt werden.
X muß so sein, daß gilt:

$$X^{-1}(B) \in \xi \quad \text{Meßbarkeit}$$

$$X^{-1}(B) := \{w \in M : X(w) \in B\}$$

Dann kann die Wahrscheinlichkeitsverteilung W_x auf R definiert werden durch:

$$W_X(B) = W\{X \in B\} := W\left(X^{-1}(B)\right) \quad \forall \text{ Intervalle B}$$

Beispiel: Bei unabhängigen Ziehungen erhält man auf dem Produktraum $\{0,1\}^n=\{(x_1,...,x_n):x_1\in\{0,1\}\}$ folgendermaßen eine Wahrscheinlichkeitsverteilung:

Ist $(x_1,...,x_n)$ ein n-Tupel, in der genau k Einser vorkommen (1=schlechtes Stück), so gilt:

$$W((x_1,...,x_n)) = \left(\frac{A}{N}\right)^k \cdot \left(1-\frac{A}{N}\right)^{n-k} \qquad \text{Additivität überprüfen!}$$

Hinweis: Wegen Abschnitt 10.4 braucht man nur zu überprüfen, ob die Summe aller Punktwahrscheinlichkeiten gleich 1 ist (Binomischer Lehrsatz).

$$\binom{n}{k} \quad \text{n - Tupel mit genau k Einsern}$$

$$p = \frac{A}{N} \quad \Rightarrow \quad \sum_{\substack{(x_1,...,x_n): \\ x_i \in \{0,1\}}} W((x_1,...,x_n)) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$$

Interessant ist nur die Anzahl von schlechten Stücken in der Stichprobe: Dies ist eine Funktion auf dem Produktraum $M=\{0,1\}^n$, nämlich:

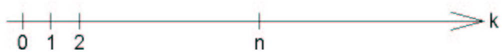
$$(x_1,...,x_n) \to \sum_{i=1}^{n} x_i \qquad X((x_1,...,x_n)) = \sum_{i=1}^{n} x_i \quad \text{stochastische Größe}$$

Die eindimensionale Verteilung von X erhält man folgendermaßen:

$$W\{X=k\} := W\left\{X^{-1}(\{k\})\right\} = W = \qquad \text{alle jene Tupel, in denen genau k Einser sind}$$

$$= \sum_{(x_1,...,x_k):\sum_{i=1}^{n} x_i = k} \underbrace{W((x_1,...,x_n))}_{p^k(1-p)^{n-k}} = \qquad \text{da es genau } \binom{n}{k} \text{ solche n - Tupel gibt}$$

$$= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad \text{für } k = 0(1)n$$



13.1 Verteilungsfunktion einer (eindimensionalen) stochastischen Größe

Mittels der Verteilungsfunktion $F_X(\cdot)$ einer stochastischen Größe X können die Wahrscheinlichkeiten für beliebige Intervalle berechnet werden. Es gilt:

$$F_X(x) := W\{X \le x\} \quad \forall x \in R$$

$$W\{a<X\le b\}=F_X(b)-F_X(a) \quad \text{mit} \quad (a,b]$$

Eigenschaften von Verteilungsfunktionen:

Für die Verteilungsfunktion $F(\cdot)$ gilt:

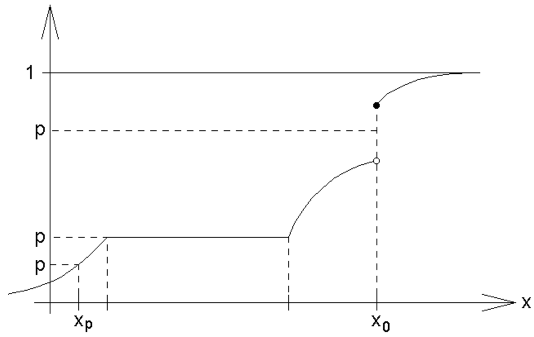
a) $0\le F(x)\le 1 \quad \forall x\in R$.................................... Normierung

b) $x_1<x_2 \quad \Rightarrow F(x_1)\le F(x_2)$................................ Monotonie

c) $\lim\limits_{x\to-\infty} F(x) = 0 \qquad \lim\limits_{x\to+\infty} F(x) = 1$

$\lim\limits_{h\downarrow 0} F(x+h) = F(x) \qquad \forall x \in R$ .... Rechtsstetigkeit

d) $W\{X = x\} = F_X(x) - \lim\limits_{h\downarrow 0} F(x-h)$



Definition: Für $0\le b\le 1$ sind **p-Fraktile** x_p einer eindimensionalen Wahrscheinlichkeitsverteilung mit der Wahrscheinlichkeitsverteilung $F(\cdot)$ folgendermaßen definiert: $p=F(x_p)$ falls $\exists$ (existent)

Bemerkung: Für eine stochastische Größe X ist ein p-Fraktile x_p ein Wert aus R, für den gilt:

$$W\{X\le x_p\}=p \quad \exists x_p$$

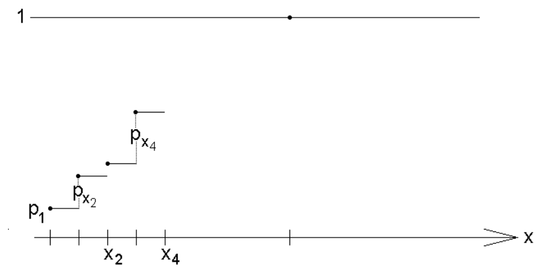
14. Diskrete Verteilungen

Solche Größen nehmen entweder nur endlich verschiedene Werte an oder höchstens abzählbar viele Werte, die sich nicht häufen. So eine Wahrscheinlichkeitsverteilung ist durch ihre sogenannte Punktwahrscheinlichkeit festgelegt.

$p(x)=W\{X=x\}$ mit M_X=Merkmalraum

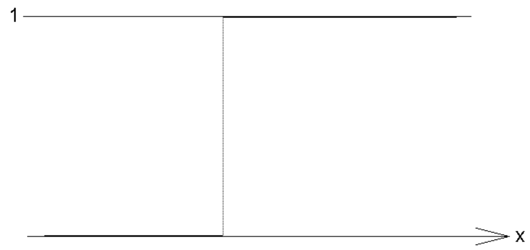
Es muß gelten: $\sum\limits_{x\in M_X} p(x) = 1 \quad \text{mit } 0 \le p(x) \le 1$

Bemerkung: Verteilungsfunktionen diskreter stochastischer Größen sind Treppenfunktionen.



14.1 Dirac-Verteilung δ_μ, $\mu\in$ R

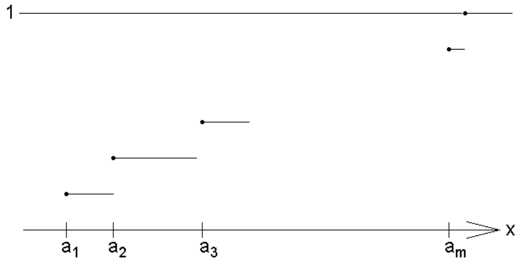
$X\sim\delta_\mu \Rightarrow W\{X=\mu\}=1$ keine Variabilität



14.2 Diskrete Gleichverteilung D_m

$M=\{a_1,...,a_m\}$

$$p(a_i)=\frac{1}{m}$$



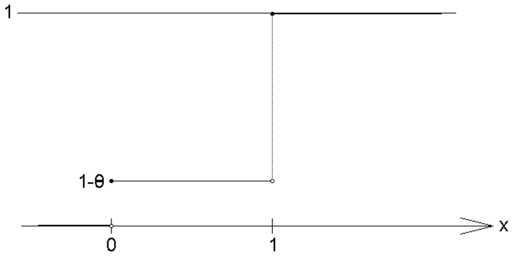
14.3 Alternativverteilung A_θ $(0<\theta<1)$

Sog. Bernoulli-Verteilung

$X\sim A_\theta \Rightarrow M_X=\{0,1\}$

$W\{X=1\}=\theta=p(1)$

$W\{X=0\}=1-\theta=p(0)$



14.4 Binomialverteilung $B_{n,\theta}$

Beispiel: Ziehen mit Zurücklegen

Führt man n unabhängige Alternativversuche durch ($X_i\sim A_\theta$, i=1(1)n) und zählt danach die Summe der X_i ($\Rightarrow$Anzahl X der schlechten Stücke), so gilt $X=\sum X_i$ und X kann Werte aus {0,1,2,..., n} annehmen.

Die Wahrscheinlichkeiten für die stochastische Größe X sind (vgl. Abschnitt 13):

$$W\{X=k\}=\binom{n}{k}\cdot\theta^k\cdot(1-\theta)^{n-k} \qquad \text{für } k = 0(1)n$$

14.5 Hypergeometrische Verteilung $H_{N,A,n}$

Beispiel: Ziehungen ohne Zurücklegen

$$W\{X=a\}=\frac{\binom{A}{a}\cdot\binom{N-A}{n-a}}{\binom{N}{n}} \qquad \text{für } a=a_1(1)a_2$$

A ....besondere Stücke
N ....Gesamtanzahl
n .... Stichprobenumfang

$a_1=\max(0,n-(N-A))$

$a_2=\min(n,A)$

14.6 Poisson-Verteilung P_μ $(\mu>0)$

$X\sim P_\mu \Rightarrow M=N_0$

$$W\{X=k\}=p(k)=\frac{\mu^k\cdot e^{-\mu}}{k!} \qquad \text{für } k=0(1)\infty \qquad \sum_{k=0}^{\infty}p(k)=1$$

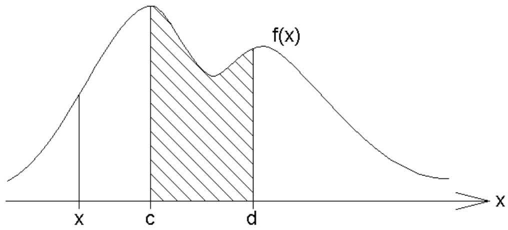
Bemerkung: Approximation

$$H_{N,A,n}\approx \underbrace{B_{n,\frac{A}{N}}}_{n<\frac{N}{10}} \approx \underbrace{P_{n\cdot\frac{A}{N}}}_{n<\frac{N}{10},\ A<\frac{N}{10}}$$

15. Kontinuierliche (stetige) Verteilungen

Eine stochastische Größe X, die alle Werte eines Intervalls annehmen kann und für die es eine Funktion (sog. **Dichtefunktion**) $f:M_X \to [0,\infty)$ gibt, so daß die Wahrscheinlichkeit eines Intervalls (c,d) folgendermaßen gegeben ist:

$$W\{X \in (c,d)\} = \int_c^d f(x)dx \quad \text{mit} \int_{-\infty}^{\infty} f(x)dx = 1$$



heißt kontinuierliche Verteilung mit der Dichtefunktion $f(\cdot)$ (Dichte, Wahrscheinlichkeitsdichte).

Die Verteilungsfunktion zu einer kontinuierlichen Verteilung mit der Dichtefunktion $f(\cdot)$ erhält man durch Integration:

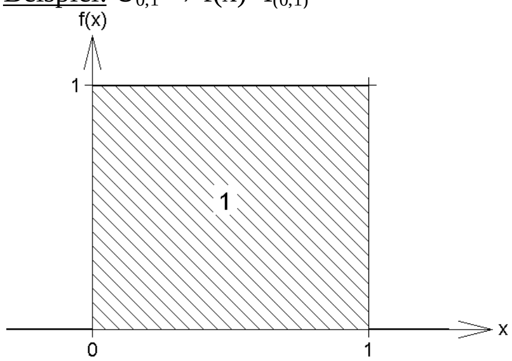
$$F(X) = \int_{-\infty}^{x} f(t)dt$$

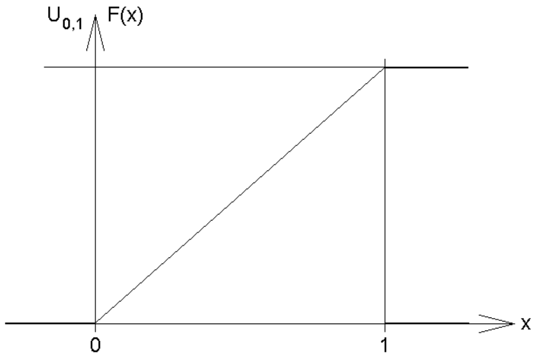
Bemerkung:
1) $f(x)=F'(x)$
2) siehe x in Zeichnung

15.1 Kontinuierliche Gleichverteilung $U_{a,b}$

Auch uniforme oder stetige Gleichverteilung. $X \sim U_{a,b} \Rightarrow M_X=(a,b)$

Beispiel: $U_{0,1} \Rightarrow f(x)=I_{(0,1)}$





Definition: Ist M eine Menge und A eine Teilmenge von M, so ist die **Indikatorfunktion** $I_a(\cdot)$ der Menge A gegeben durch:

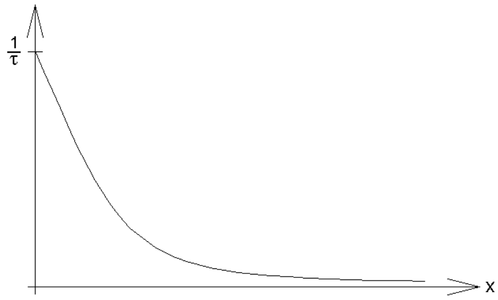
$$I_A(x) = \begin{cases} 1 & \text{für} \quad x \in A \\ 0 & \text{falls} \quad x \notin A \end{cases}$$

Allgemein: $f(x) = \frac{1}{b-a} I_{(a,b)}(x)$

15.2 Exponentialverteilung Ex_τ (τ>0)

$X \sim Ex_\tau \Rightarrow f(x) = \frac{1}{\tau} \cdot e^{-x/\tau} \cdot I_{(0,\infty)}(x)$

$$F(x) = \int_0^x \frac{1}{\tau} \cdot e^{-\xi/\tau} d\xi = 1 - e^{-x/\tau} \qquad \text{für } x \geq 0$$



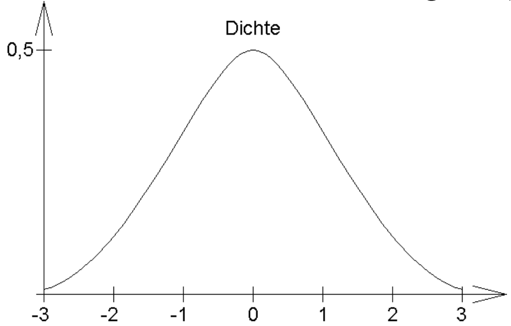
15.3 Standard-Normalverteilung N(0,1)

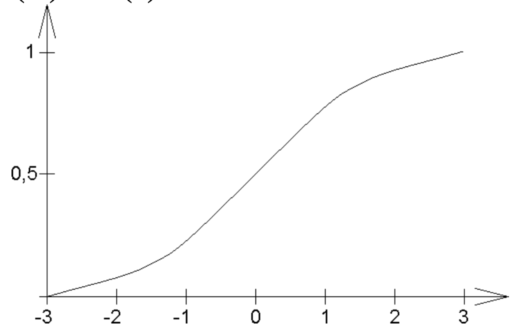
$X \sim N(0,1) \Rightarrow$ Dichtefunktion $\varphi(\cdot)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \forall x \in R$$

Verteilungsfunktion: $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi$

Da die Verteilungsfunktion $\Phi(x)$ nicht elementar integrierbar ist, wird sie tabelliert. Es genügt die Funktionswerte für $x>0$ zu tabellieren, da gilt: $\Phi(0)=½$ und $\Phi(-x)=1-\Phi(x) \quad \forall x>0$





15.4 Allgemeine Normalverteilung N(μ,σ²) (μ∈R, σ²>0)

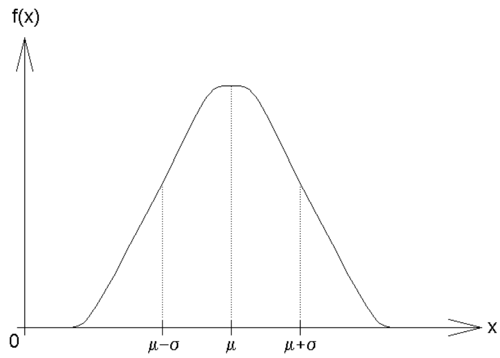
$X \sim N(\mu,\sigma^2)$

Dichtefunktion:

$$f\left(x \middle| \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in R$$

Verteilungsfunktion:

$$F(x) = \int\limits_{-\infty}^{x} f\left(\xi \middle| \mu, \sigma^2\right) d\xi = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{x} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi$$



Die Verteilungsfunktion ist nicht elementar integrierbar, aber auf folgendes Integral rückführbar:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{\xi^2}{2}} d\xi$$

$$\text{Es gilt: } \underset{\mu,\sigma^2}{F}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad \forall x \in R$$

Beweis:

$$\frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz = \qquad \text{Substitution: } \frac{z-\mu}{\sigma} = u^2 \Rightarrow dz = \sigma du^2$$

$$= \frac{1}{\sqrt{2\pi}\cancel{\sigma}} \int\limits_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}u^4} \cancel{\sigma} du^2 = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Es gilt daher: $X \sim N(\mu,\sigma^2)$

$$W\{a \le X \le b\} = \underset{\mu,\sigma^2}{F}(b) - \underset{\mu,\sigma^2}{F}(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Bemerkung:

1) $\Phi(-z)=1-\Phi(z)$
2) Zusammenhang zwischen den Fraktilen z_p der N(0,1) und x_p der $N(\mu,\sigma^2)$:

$$z_p = \frac{x_p - \mu}{\sigma}$$

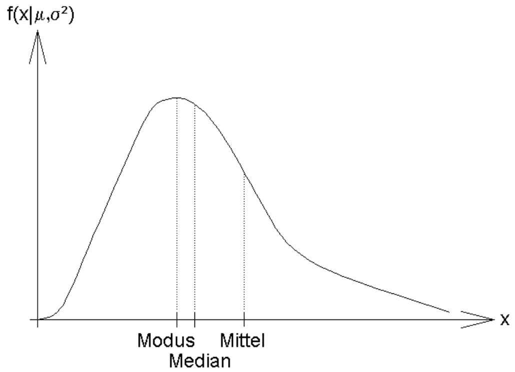
15.5 Logarithmische Normalverteilung LN(μ,σ²) (μ∈R, σ²>0)

$X \sim LN(\mu,\sigma^2) \Leftrightarrow M_X=(0,\infty)$, $\ln X \sim N(\mu,\sigma^2)$
Verteilungsfunktion:

$$F(X) = W\{X \le x\} = W\{\ln X \le \ln x\} = F_{\mu,\sigma^2}(\ln x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

Bemerkung: Die Dichtefunktion erhält man durch Differentiation.

$$f(x) = \frac{d}{dx}\Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

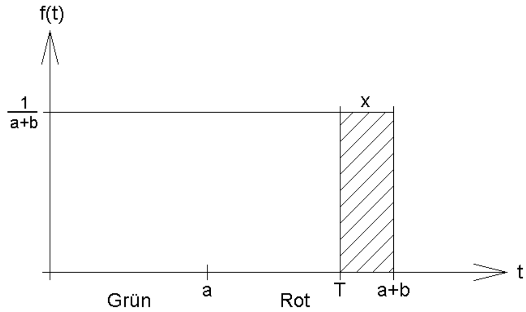


Modus = Maximum
Median = $x_{0,5}$
Mittel = Erwartungswert

16. Gemischte Verteilungen

Prüfung!
Es gibt praktisch wichtige stochastische Größen, deren Wahrscheinlichkeitsverteilung weder diskret noch kontinuierlich ist.
Beispiel: Wartezeit bei einer Ampel
T...Ankunftszeit, a...Grünphase, b...Rotphase
$T \sim U_{0,a+b}$
Wartezeit:

$$X = \begin{cases} 0 \text{ bei Ankunft im Intervall } [0,a) \\ a+b-t \text{ bei Ankunft in } [a,a+b) \end{cases}$$



Wahrscheinlichkeitsverteilung von X:

$$W\{X=0\} = W\{T \in [0,a)\} = \frac{a}{a+b} > 0$$

$$W\{0 < X \le x\} = W\{T \in [a+b-x, a+b)\} = \frac{x}{a+b} \quad \text{für } 0 < x < b$$

17. Erwartungswert von stochastischen Größen bzw. Verteilungen

Zur Beschreibung von Verteilungen sind charakteristische Werte derselben wichtig. Ein solcher Wert ist z.B. der **Erwartungswert**. Der Erwartungswert wird auch Mittel der zugehörigen Verteilung bezeichnet.
Im Beispiel der Lebensdauer wird manchmal ein „zu erwartender“ Wert gewünscht. Man hat endlich viele beobachtete Lebensdauern $x_1,...,x_m$ und den Mittelwert. Angenommen, dies sind Ausprägungen einer stochastischen Größe mit identischen, diskreten Punktwahrscheinlichkeiten, dann gilt:

$$\text{Mittelwert: } \overline{x}_m = \frac{x_1 + x_2 + ...+ x_m}{m} \qquad p(x_i) = \frac{1}{m}$$

$$\frac{1}{m}\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} x_i p(x_i)$$

17.1 Erwartungswert diskreter Verteilungen

Sei $X \sim (M_X, p(x))$ mit $p(x)=W(X=x)$, so wird der Erwartungswert EX von X, auch bezeichnet als **Mittel** der Verteilung von X, folgendermaßen definiert:

$$EX := \sum_{x \in M_X} x \cdot p(x)$$

Beispiel: Alternativverteilung A_θ $(0<\theta<1)$
$M_X=\{0,1\}$
$W\{X=1\}=\theta$ und $W\{X=0\}=1-\theta$
$EX=0\cdot(1-\theta)+1\cdot\theta=\theta$
Beispiel: Binomialverteilung
$X \sim B_{n,\theta}$ $(0<\theta<1)$

$$EX = \sum_{i=0}^{n} i \cdot p(i) = \sum_{i=0}^{n} i \cdot \binom{n}{i} \cdot \theta^i \cdot (1-\theta)^{n-i} = n \cdot \theta$$

Beispiel: Poisson-Verteilung P_μ
$X \sim P_\mu$

$$EX = \sum_{x=0}^{\infty} x \cdot \frac{\mu^x \cdot e^{-\mu}}{x!} = \sum_{x=1}^{\infty} \frac{\mu^x \cdot e^{-\mu}}{(x-1)!} = \mu \cdot e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} = \mu \cdot e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = \mu \cdot e^{-\mu} \cdot e^{\mu} = \mu$$

17.2 Erwartungswert kontinuierlicher Verteilungen

$X \sim (M_X, f(\cdot))$

$$EX := \int_{-\infty}^{\infty} x \cdot f(x)dx$$

Beispiel: Uniforme Verteilung $U_{a,b}$
$X \sim U_{a,b}$ mit $a<b$

$$EX = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{b-a} \cdot I_{(a,b)}(x)dx = \int_a^b x \cdot \frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b xdx = \frac{a+b}{2}$$

Beispiel: Allgemeine Normalverteilung $N(\mu,\sigma^2)$
$X \sim N(\mu,\sigma^2)$

$$EX = \int_{-\infty}^{\infty} x \cdot f\left(x \middle| \mu, \sigma^2\right)dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu$$

Der Erwartungswert ergibt sich aus der Symmetrie dieser Wahrscheinlichkeitsverteilung.
Beispiel: Exponentialverteilung Ex_τ
$X \sim Ex_\tau$

$$EX = \int_0^{\infty} x \cdot \frac{1}{\tau} \cdot e^{-\frac{x}{\tau}} dx = \ldots \text{ partielle Integration } \ldots = \tau$$

17.3 Erwartungswert gemischter Verteilungen

$X \sim (M_X=\{x_1,\ldots,x_m\} \cup \langle a,b \rangle, p(x_i), f(\cdot))$

$$0 < \sum_{i=1}^{m} p(x_i) < 1 \qquad \int_a^b f(x)dx = 1 - \sum_{i=1}^{m} p(x_i)$$

$$EX = \sum_{i=1}^{m} x_i \cdot p(x_i) + \int_a^b x \cdot f(x)dx$$

Beispiel: Erwartungswert der Wartezeit im Beispiel aus Abschnitt 16.

Satz 17.1 Ist X eine stochastische Größe und der Erwartungswert existiert und $c \in R$ so folgt: $E(c \cdot X)=c \cdot EX$

18. Varianz und Streuung einer stochastischen Größe/Verteilung

Sei X eine stochastische Größe. Gesucht ist ein Maß für das Streuverhalten von X (sog. Varianz).

18.1 Varianz von diskreten stochastischen Größen

$X \sim (M_X, p(\cdot))$

$$\mathrm{VarX} := \sum_{x \in M_X} (x - EX)^2 \cdot p(x)$$

Beispiel: Alternativverteilung A_θ

$X \sim A_\theta \Rightarrow EX = \theta$

$\mathrm{VarX} = (0-\theta)^2 \cdot (1-\theta) + (1-\theta)^2 \cdot \theta = \theta \cdot (1-\theta)$

Beispiel: Poisson-Verteilung P_μ

$X \sim P_\mu \Rightarrow EX = \mu$

$$\mathrm{VarX} = \sum_{k=0}^{\infty} (k-\mu)^2 \cdot \frac{\mu^k \cdot e^{-\mu}}{k!} = \sum_{k=0}^{\infty} \left(k^2 - 2\mu k + \mu^2\right) \cdot \frac{\mu^k \cdot e^{-\mu}}{k!} = \sum_{k=0}^{\infty} \left[k(k-1) + k(1-2\mu) + \mu^2\right] \cdot \frac{\mu^k \cdot e^{-\mu}}{k!} =$$

$$= \sum_{k=2}^{\infty} \frac{\mu^k \cdot e^{-\mu}}{(k-2)!} + (1-2\mu) \sum_{k=0}^{\infty} k \cdot \frac{\mu^k \cdot e^{-\mu}}{k!} + \mu^2 \sum_{k=0}^{\infty} k \cdot \frac{\mu^k \cdot e^{-\mu}}{k!} = \mu^2 \sum_{k=2}^{\infty} \frac{\mu^{k-2} \cdot e^{-\mu}}{(k-2)!} + (1-2\mu) \cdot \mu + \mu^2 =$$

$$= \mu^2 + \mu - 2\mu^2 + \mu^2 = \mu$$

Beispiel: Binomialverteilung $B_{n,\theta}$

$X \sim B_{n,\theta} \Rightarrow EX = n \cdot \theta$

$\mathrm{VarX} = n\theta(1-\theta)$

18.2 Varianz kontinuierlicher Verteilungen (stoch. Größen)

$X \sim (M_X, f(\cdot))$

$$\mathrm{VarX} := \int_{M_X} (x - EX)^2 \cdot f(x)\,dx$$

Beispiel: Normalverteilung $N(\mu,\sigma^2)$

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

$\mathrm{VarX} = \sigma^2$

Beispiel: Exponentialverteilung Ex_τ

$X \sim Ex_\tau$

$\mathrm{VarX} = \tau^2$

18.3 Varianz gemischter Verteilungen

$X \sim (M_X = \{x_1, \ldots, x_m\} \cup \langle a,b \rangle, p(x_i), f(\cdot))$

$$\mathrm{VarX} := \sum_{i=1}^{m} (x_i - EX)^2 \cdot p(x_i) + \int_a^b (x - EX)^2 \cdot f(x)\,dx$$

Beispiel: Varianz der Wartezeit im Beispiel aus Abschnitt 16.

Bemerkung: Für die Berechnung von Varianzen gilt ein sog. Verschiebungssatz (vgl. Abschnitt 19).

Satz 18.1 Ist X eine stochastische Größe mit $\exists EX$, $\exists \mathrm{VarX}$ und $c \in R$ so folgt: $\mathrm{Var}(c \cdot X) = c^2 \cdot \mathrm{VarX}$

18.4 Streuung

Ist X eine stochastische Größe (bzw. eine Verteilung auf R) mit $\exists \mathrm{VarX}$, so heißt ${}_{+}\sqrt{\mathrm{VarX}}$ Streuung von X.

Beispiel: Normalverteilung $N(\mu,\sigma^2)$

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

$\mathrm{Streuung}(X) = \sigma$

19. Funktionen einer stochastischen Größe

Sei X ist stochastische Größe mit der Verteilungsfunktion $F_X(\cdot)$. Dann ist die Funktion ψ, definiert als ψ:R⇒R⇒ψ(X) ψ(X):=ψ(Wert(X)), ebenfalls stochastisch.

Beispiel: Maßstabstransformation

Frage:

1) Welche Verteilung hat Y, falls die Verteilung von X bekannt ist?

Für c>0 gilt für die Verteilungsfunktion $F_Y(\cdot)$ von Y=c·X:

$$F_Y(x) = W\{Y \le x\} = W\{c \cdot X \le x\} = W\{X \le \tfrac{x}{c}\} = F_X\left(\tfrac{x}{c}\right)$$

Falls X kontinuierlich verteilt ist mit der Dichtefunktion $f_X(\cdot)$, so kann man die Dichtefunktion $f_Y(\cdot)$ von Y folgendermaßen berechnen:

$$f_Y(x) = \frac{d}{dx} F_Y(x) = \frac{d}{dx}\left[F_X\left(\frac{x}{c}\right)\right] = f_X\left(\frac{x}{c}\right) \cdot \frac{1}{c} \qquad \forall x \in R$$

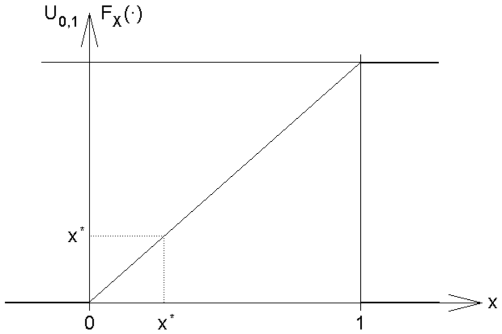
2) Welchen Erwartungswert und welche Varianz hat Y?

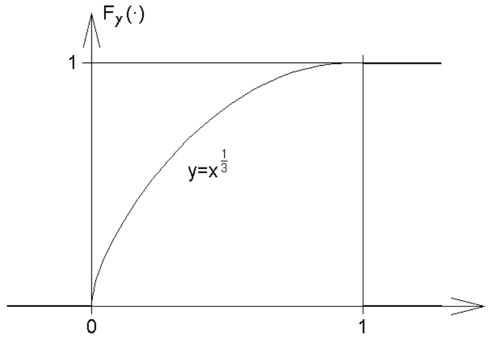
$$E(c \cdot X) = c \cdot EX \quad \text{(vgl. Satz 17.1)}$$
$$Var(c \cdot X) = c^2 \cdot VarX$$

3) Was gilt bei allgemeinen Funktionen $\psi(\cdot)$?

Beispiel: X~$U_{0,1}$ und Y=$X^3F_X(\cdot)$

$$F_Y(x) = W\{Y \le x\} = W\{X^3 \le x\} = W\left\{X \le x^{\frac{1}{3}}\right\}$$





Satz 19.1 Ist X eine 1-dimensionale stochastische Größe mit der Verteilungsfunktion $F_X(\cdot)$ und ψ:R⇒R eine invertierbare und monoton wachsende Funktion, so gilt für die Verteilungsfunktion $F_Y(\cdot)$ der stochastischen Größe Y=ψ(X):

$$F_Y(x) = F_X\left(\psi^{-1}(x)\right) \qquad \forall x \in R$$

Beweis: $F_Y(x) = W\{Y \le x\} = W\{\psi(X) \le x\} = W\{X \le \psi^{-1}(x)\} = F_X\left(\psi^{-1}(x)\right)$

Bemerkung: Im Fall der Existenz einer Dichte $f_X(\cdot)$ für die Verteilung von X und der Differenzierbarkeit der Umkehrfunktion $\psi^{-1}(\cdot)$ von $\psi(\cdot)$ gilt für die Dichte $f_Y(\cdot)$ von Y=ψ(X):

$$f_Y(x) = f_X\left(\psi^{-1}(x)\right)\frac{d}{dx}\psi^{-1}(x)$$

Beispiel: X~N(0,1)

Gesucht ist die Verteilung von $\psi(X)=X^2$.

$$F_{X^2}(x) = W\{X^2 \le x\} = W\{{}_{-}\sqrt{x} \le X \le {}_{+}\sqrt{x}\} = \Phi\left({}_{+}\sqrt{x}\right) - \Phi\left({}_{-}\sqrt{x}\right) = 2 \cdot \Phi\left(\sqrt{x}\right) - 1$$

Durch Differentiation ergibt sich die Dichte folgendermaßen:

$$f_{X^2}(x) = \frac{d}{dx} F_{X^2}(x) = 2 \cdot \Phi'\left(\sqrt{x}\right) \cdot \frac{1}{2} \cdot x^{-\frac{1}{2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x}{2}} \cdot x^{-\frac{1}{2}} \cdot I_{(0,\infty)}(x)$$

Diese Verteilung heißt **Chiquadrat-Verteilung** χ_n^2 mit einem Freiheitsgrad.

Zur Berechnung des Erwartungswertes einer Funktion einer stochastischen Größe gibt es den sog. **Satz vom unbewußten Statistiker**.

Satz 19.2 Ist X eine stochastische Größe und $\psi(\cdot)$ eine reelle Funktion, so gilt für den Erwartungswert der stochastischen Größe $\psi(X)$:

a) Falls X diskret verteilt ist mit den Punktwahrscheinlichkeiten p(x):

$$E[\psi(X)] = \sum_{x:p(x)\geq 0} \psi(x)\cdot p(x)$$

b) Falls X kontinuierlich verteilt ist mit der Dichte $f(\cdot)$:

$$E[\psi(X)] = \int_{-\infty}^{\infty} \psi(x)\cdot f(x)dx$$

Allgemein gilt $E[\psi(X)] \neq \psi(EX)$, es gilt jedoch folgender Satz:

Satz 19.3 Ist X eine stochastische Größe und $a\in R$, $b\in R$, so gilt für die stochastische Größe $a\cdot X+b$:

$$E[a\cdot X+b] = a\cdot EX+b$$

Beweis: Für den kontinuierlichen Fall gilt:

$$E[a\cdot X+b] = \int_{-\infty}^{\infty}(a\cdot x+b)\cdot f(x)dx = a\int_{-\infty}^{\infty} x\cdot f(x)dx + b\int_{-\infty}^{\infty} f(x)dx = a\cdot EX+b$$

Bemerkung: Allgemein gilt für eine lineare Funktion $\sum_{i=1}^{n} c_i\cdot\psi_i(X)$:

$$E\left[\sum_{i=1}^{n} c_i\cdot\psi_i(X)\right] = \sum_{i=1}^{n} c_i\cdot EX$$

Beispiel: $E[2X+3X^2]=2EX+3E(X^2)$

Definition: Ist X eine stochastische Größe mit endlicher Varianz σ^2 und Erwartungswert μ, so heißt die stochastische Größe $\frac{X-\mu}{\sigma}$ die zugehörige **standardisierte** stochastische Größe.

Es gilt:

a) $E\left[\frac{X-\mu}{\sigma}\right]=0$

b) $Var\left[\frac{X-\mu}{\sigma}\right]=1$

Beweis:

a) Aus $\frac{X-\mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$ und Satz 19.3 folgt:

$$E\left[\frac{1}{\sigma}X+\left(-\frac{\mu}{\sigma}\right)\right] = \frac{1}{\sigma}EX + E\left[-\frac{\mu}{\sigma}\right] = \frac{\mu}{\sigma}+\left(-\frac{\mu}{\sigma}\right)=0$$

Beispiel: Berechnung der Verteilungsfunktion der $N(\mu,\sigma^2)$

$X\sim N(\mu,\sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$

Bemerkung: Aufgrund des Satzes vom unbewußten Statistikers gilt bei existierender Varianz $\exists VarX$:
$VarX=E[(X-EX)^2]$

Verschiebungssatz für die Varianzberechnung

Satz 19.4 Für jede stochastische Größe, deren Varianz existiert, gilt:

$$VarX=E[X^2]-(EX)^2$$

$$VarX = E\left[(X-EX)^2\right] = E\left[X^2-2\cdot X\cdot EX+(EX)^2\right] = E\left[X^2\right]-E[2\cdot X\cdot EX]+E\left[E\left[X^2\right]\right] =$$

$$= E\left[X^2\right]-2\cdot EX\cdot EX+(EX)^2 = E\left[X^2\right]-(EX)^2$$

Bemerkung: Manchmal hilft folgende Gleichung zur Varianzberechnung:

$$\mathrm{VarX} = E\left[X(X-1)\right] + EX - (EX)^2$$

Beweis:

$$E\left[X(X-1)\right] + EX - (EX)^2 = E\left[X^2 - X\right] + EX - (EX)^2 = E\left[X^2\right] - EX + EX - (EX)^2 = E\left[X^2\right] - (EX)^2$$

Beispiel: Varianzberechnung der Poisson-Verteilung

$X \sim P_\mu \Rightarrow EX=\mu$

Zuerst berechnet man E[X·(X-1))]

$$E\left[X(X-1)\right] = \sum_{k=0}^{\infty} k\cdot(k-1)\cdot p(k) = \sum_{k=0}^{\infty} k\cdot(k-1)\cdot\frac{\mu^k\cdot e^{-\mu}}{k!} = \sum_{k=2}^{\infty}\frac{\mu^k\cdot e^{-\mu}}{(k-2)!} = e^{-\mu}\cdot\mu^2\cdot\sum_{k=2}^{\infty}\frac{\mu^{k-2}}{(k-2)!} =$$

$$= e^{-\mu}\cdot\mu^2\cdot\sum_{k=0}^{\infty}\frac{\mu^k}{k!} = \mu^2\cdot\sum_{k=0}^{\infty}\frac{\mu^k\cdot e^{-\mu}}{k!} = \mu^2$$

Nach Anwendung der letzten Bemerkung erhält man:

$$\mathrm{VarX} = E\left[X\cdot(X-1)\right] + EX - (EX)^2 = \mu^2 + \mu - \mu^2 = \mu$$

Beispiel: Binomialverteilung

$X \sim B_{n,p} \Rightarrow \mathrm{VarX}=n\cdot p\cdot(1-p)$

Mit Hilfe des Binomischen Lehrsatzes ergibt sich die Varianz folgendermaßen:

$$E\left[X\cdot(X-1)\right] = \sum_{k=0}^{n} k\cdot(k-1)\cdot\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k} = \sum_{k=2}^{n}\frac{n!}{(n-k)!\cdot(k-2)!}\cdot p^k\cdot(1-p)^{n-k} =$$

$$= \sum_{k=0}^{n-2}\frac{n!}{(n-k-2)!\cdot k!}\cdot p^{k+2}\cdot(1-p)^{n-k-2} = n\cdot(n-1)\cdot p^2\cdot\sum_{k=0}^{n-2}\binom{n-2}{k}\cdot p^k\cdot(1-p)^{n-k-2} = n\cdot(n-1)\cdot p^2$$

$$\mathrm{VarX} = E\left[X\cdot(X-1)\right] + EX - (EX)^2 = n\cdot(n-1)\cdot p^2 + n\cdot p - n^2p^2 = n\cdot p\cdot(1-p)$$

Bemerkung: Ist X eine normalverteilte stochastische Größe, d.h. $X \sim N(\mu,\sigma^2)$, und sind c und d reelle Konstanten mit c>0, so folgt:

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

$$F_{cX}(x) = W\{cX \le x\} = W\{X \le \tfrac{x}{c}\} = \Phi\left(\tfrac{x-c\mu}{c\sigma}\right)$$

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

$$F_{X+d}(x) = W\{X + d \le x\} = W\{X \le x - d\} = \Phi\left(\tfrac{x-(\mu+d)}{\sigma}\right)$$

3) $c\cdot X+d \sim N(c\cdot\mu+d, c^2\cdot\sigma^2)$ (siehe 1 und 2)

V Stochastische Vektoren

20. Mehrdimensionale stochastische Größen (stochastische Vektoren)

Der Versuchsausgang wird durch ein m-Tulpel von Zahlen (Vektor) beschrieben.

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \in R^m$$

Beispiel: Messung der Luftqualität an einem bestimmten geographischen Ort.

$$\begin{pmatrix} SO_2-\text{Gehalt} \\ \text{Staubgehalt} \\ CO-\text{Gehalt} \end{pmatrix}$$

Bemerkung:

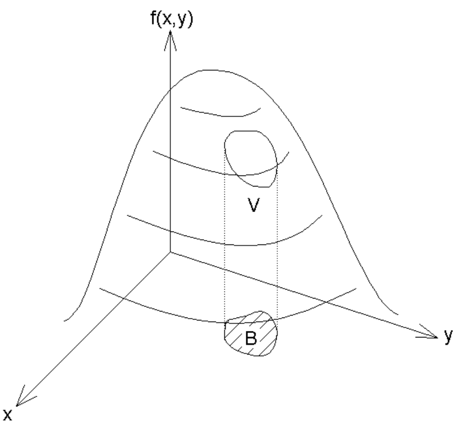
$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$ m-dimensionaler stochastischer Vektor, wobei alle X_i 1-dim. stochastische Größen sind

20.1 Kontinuierliche 2-dimensionale Verteilungen Prüfung!

Die Verteilung $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ wird durch eine 2-dimensionale Dichtefunktion f(.,.) bestimmt.

Für diese Dichtefunktion muß gelten:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \qquad f(x,y) > 0 \quad \forall (x,y) \in R^2$$



Für einen beliebigen Bereich $B \subseteq R^2$ gilt:

$$W(B) := \int_B\int f(x,y)dxdy$$

$$X \sim f(.,.) \Rightarrow W\{X \in B\} = \int_B\int f(x,y)dxdy$$

Speziell für den Fall $[a_1,b_1]\times[a_2,b_2]$ gilt:

$$W(B) = \int_{a_2}^{b_2}\left(\int_{a_1}^{b_1} f(x,y)dx\right)dy$$

Beispiel: 2-dimensionale Normalverteilung

$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\varphi)$ mit $\mu_1 \in R,\ \mu_1 \in R,\ \sigma_1^2 > 0,\ \sigma_2^2 > 0 \quad -1 < \varphi < 1$

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varphi^2}}\exp\left\{-\frac{1}{2(1-\varphi^2)}\cdot\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\varphi\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\} \quad \forall (x,y) \in R^2$$

Bemerkung: Die Parameter μ_1, μ_2, σ_1^2, σ_2^2, φ müssen aus Beobachtungen einer solchen 2-dimensionalen Größe geschätzt werden (siehe Schließende Statistik).

Bemerkung: Es gibt eine Verallgemeinerung für m-dimensionale Größen:

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} \qquad \underline{X} \sim N(\underline{\mu},\Sigma) \quad \text{mit } \underline{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}$$

$\Sigma = (\sigma_{kl})_{\substack{k=1,\dots,m \\ l=1,\dots,m}}$ positiv definite symmetrische quadratische Matrix

Für eine m-dimensionale Normalverteilung mit der Dichtefunktion f gilt:

$$f(x_1,\dots,x_n)=\frac{\sqrt{\mathrm{Det}\Sigma^{-1}}}{(2\pi)^{m/2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{m}\sum_{l=1}^{m}\sigma^{kl}\cdot(x_k-\mu_k)\cdot(x_l-\mu_l)\right\}\quad \forall(x_1,\dots,x_m)\in R^n$$

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m}\\ \vdots & \vdots & \cdots & \vdots\\ \sigma_{m1} & \sigma_{m2} & \vdots & \sigma_{mm}\end{pmatrix}\qquad \text{mit } \sigma_{kl}=\sigma_{lk}$$

$$\Sigma^{-1}=\left(\sigma^{kl}\right)\quad \text{inverse Matrix zu } \Sigma$$

$$\sum_{k=1}^{m}\sum_{l=1}^{m}\sigma^{kl}(x_k-\mu_k)(x_l-\mu_l)\ge 0$$

Für eine 2-dimensionale Normalverteilung gilt:

$$\Sigma=\begin{pmatrix}\sigma_1{}^2 & \varphi\sigma_1\sigma_2\\ \varphi\sigma_1\sigma_2 & \sigma_2{}^2\end{pmatrix}$$

$$\Sigma^{-1}=\frac{1}{\sigma_1{}^2\sigma_2{}^2(1-\varphi^2)}\begin{pmatrix}\sigma_1{}^2 & -\varphi\sigma_1\sigma_2\\ -\varphi\sigma_1\sigma_2 & \sigma_2{}^2\end{pmatrix}$$

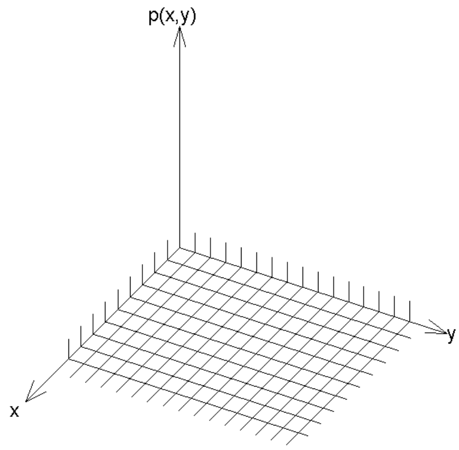
Prüfung: Dichte=?, Volumen=1

20.2 Diskrete 2-dimensionale Verteilung

$\underline{X}=\begin{pmatrix}X\\Y\end{pmatrix}$ ist ein 2-dimensionaler stochastischer Vektor, wobei X und Y jeweils 1-dimensionale stochastische Größen sind. Es gilt: $M_{\underline{X}}\subseteq Z\times Z\subset R^2$

Für die Punktwahrscheinlichkeiten gilt:

$$p(x,y)=W\left\{\begin{pmatrix}X\\Y\end{pmatrix}=\begin{pmatrix}x\\y\end{pmatrix}\right\}\qquad (x,y)\in M_{\underline{X}}\quad \sum p(x,y)=1\quad p(x,y)\ge 0$$



Beispiel: An einem Serienprodukt wird die Anzahl von Fehlern zweier Arten erhoben.
X ....Anzahl der Fehler einer Art (Blankstellen)
Y............Anzahl der Fehler anderer Art (Lötfehler)
Aus der Häufigkeitsverteilung soll ein 2-dimensionales stochastisches Modell konstruiert werden.

Beispiel: Multinomialverteilung $M_{n,\theta_1,\dots,\theta_m}$

$$\begin{pmatrix}X_1\\ \vdots\\ X_m\end{pmatrix}\sim M_{n,\theta_1,\dots,\theta_m}\qquad W\left\{\begin{pmatrix}X_1\\ \vdots\\ X_m\end{pmatrix}=\begin{pmatrix}x_1\\ \vdots\\ x_m\end{pmatrix}\right\}=\frac{n!}{x_1!\cdot\ldots\cdot x_m!}\cdot\theta_1{}^{x_1}\cdot\ldots\cdot\theta_m{}^{x_m}$$

Wobei gilt:

$$\sum_{j=1}^{m}\theta_j=1\qquad 0\le\theta_j\le 1\qquad \sum_{i=1}^{n}x_i=n$$

Diese Verteilung tritt auf bei folgenden Situationen: Versuch wird auf m mögliche Arten durchgeführt mit entsprechenden Wahrscheinlichkeiten $\theta_1,\dots,\theta_m$. Der Versuch wird n-mal durchgeführt (unabhängig).

Frage: Wie groß ist die Wahrscheinlichkeit, daß nach n Beobachtungen die 1. Möglichkeit genau x_1-mal, die 2. Möglichkeit genau x_2-mal, ... und die m-te Möglichkeit genau x_m-mal auftrat?

21. Randverteilungen (Einzelverteilungen)

Von dem stochastischen Vektor $\underline{X}$ sind nur die ersten r Komponenten von Interesse (r<m).

Beispiel: Verkehr ausländischer Fahrzeuge in Österreich

$$\underline{X} = \begin{pmatrix} X_1 & \text{Fahrstrecke in Ö} \\ X_2 & \text{Benzinverbrauch in Ö} \\ X_3 & \text{Aufenthaltsdauer in Ö} \\ X_4 & \text{Gewicht der Fahrzeuge} \end{pmatrix}$$

Für eine andere Analyse ist nur $\underline{X}^* = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ von Interesse.

Gesucht ist die Verteilung von $\underline{X}^*$ $\Rightarrow$ Randverteilung.

21.1 Randverteilung diskreter Verteilungen

Für den Sonderfall m=2 gilt:

$$\underline{X} = \begin{pmatrix} X \\ Y \end{pmatrix} \sim \left(M_{\underline{X}}, p(x,y)\right)$$

Gesucht ist die Randverteilung von X:

$$p_1(x) = W\{X = x\} \qquad \forall x \in M_X$$
$$p_1(x) = \sum_{y \in M_Y} p(x,y)$$

Begründung:
Das Ereignis {X=x} ist eine höchstens abzählbare Vereinigung von Elementarereignissen aus $M_{\underline{X}}$.

$$\{X = x\} = \bigcup_{y \in M_Y} \left\{ \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \right\}$$

$$W(\{X = x\}) = \sum_{y \in M_Y} W\left\{ \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \right\} = \sum_{y \in M_Y} p(x,y)$$

Analog dazu berechnet man die Randverteilung von Y:

$$p_2(y) = W\{Y = y\} = \sum_{x \in M_X} p(x,y)$$

Beispiel:
X ....Alter des Hauses
Y ....Anzahl der Bewohner

$$\underline{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$$

Tabelle zur Darstellung 2-dimensionaler stochastischer Vektoren:

X\Y	0	1	2	·····	n	$p_1(x)$	Zeilensumme
0	p(0,0)	p(0,1)	p(0,2)	·····	p(0,n)	$p_1(0)$	
1	p(1,0)	p(1,1)	p(1,2)	·····	p(1,n)	$p_1(1)$	Randverteilung von X
2	p(2,0)	p(2,1)	p(2,2)	·····	p(2,n)	$p_1(2)$	
:	:	:	:		:	:	
m	p(m,0)	p(m,1)	p(m,2)	·····	p(m,n)	$p_1(m)$	
$p_2(y)$	$p_2(0)$	$p_2(1)$	$p_2(2)$	·····	$p_2(n)$	1	

Spalten-summe Randverteilung von Y

21.2 Randverteilung kontinuierlicher Verteilungen

Beispiel:
X ....Körpergröße
Y.............Körpergewicht

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim f(x,y)$$

Sei $\begin{pmatrix} X \\ Y \end{pmatrix}$ ein 2-dimensionaler stochastischer Vektor (Verteilung).

Gefragt ist die Verteilung der Körpergewicht, d.h. die Randverteilung von Y.
Die Randverteilungen (Einzelverteilungen) von X bzw. Y sind wieder kontinuierliche Verteilungen, deren Dichte $f_1(\cdot)$ bzw. $f_2(\cdot)$ folgendermaßen gegeben sind:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad \text{Dichte von X}$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx \qquad \text{Dichte von Y}$$

Beispiel: Für die 2-dimensionale Normalverteilung gilt:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\varphi\right) \;\Rightarrow\; \begin{cases} X_1 \sim N\left(\mu_1,\sigma_1^2\right) \\ X_2 \sim N\left(\mu_2,\sigma_2^2\right) \end{cases}$$

Für höherdimensionale Normalverteilungen gilt ebenfalls, daß die Randverteilungen Normalverteilungen sind.

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} \sim N\left(\underline{\mu},\Sigma\right) \;\Rightarrow\; X_i \sim N\left(\mu_i,\sigma_{ii}\right) \text{ mit } EX := \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} \text{ und } \Sigma = \left(\sigma_{ij}\right)_{j=1,\dots,m}^{i=1,\dots,m}$$

22. Erwartungswert von Funktionen von stochastischen Größen

Beispiel: Lebensdauer eines Systems ET=?; $T=\psi(X_1,\dots,X_n)$
Sind $X_1,\dots,X_n$ 1-dimensionale stochastische Größen und sei $\psi: R^n \to R$ eine Funktion von n Variablen, so erhält man eine stochastische Größe $T=\psi(X_1,\dots,X_n)$.

Beispiel: Seinen $X_1,\dots,X_n$ 1-dimensionale stochastische Größen, die identisch verteilt sind (Stichprobe).
In diesem Fall gilt:

a) $\overline{X}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ Stichprobenmittel

$\psi(x_1,\dots,x_n) = \frac{x_1+\dots+x_n}{n}$ arithmetisches Mittel

b) $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$ Stichprobenvarianz

Der Erwartungswert der stochastischen Größe $\psi(X_1,\dots,X_n)$ kann folgendermaßen berechnet werden (Satz vom unbewußten Statistiker):

a) Für die diskrete stochastische Größe $\underline{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \sim p(x_1,\dots,x_n)$ gilt:

$$E\psi(x_1,\dots,x_n) = \sum_{(x_1,\dots,x_n)\in M_{\underline{X}}} \psi(x_1,\dots,x_n) \cdot p(x_1,\dots,x_n)$$

b) Für die kontinuierliche Größe $\underline{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \sim f(x_1,\dots,x_n)$ gilt:

$$E\psi(X_1,\dots,X_n) = \int \underset{R^n}{\cdots} \int \psi(x_1,\dots,x_n) \cdot f(x_1,\dots,x_n)dx_1\dots dx_n$$

Bestimmung des Erwartungswertes der Summe von zwei stochastischen Größen: $Z=X+Y=\psi(X,Y)$
Für den kontinuierlichen Fall gilt:

$$EZ = E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)\cdot f(x+y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot f(x+y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\cdot f(x+y)dxdy =$$

$$= \int_{-\infty}^{\infty} x\cdot\left(\int_{-\infty}^{\infty} f(x,y)dy\right)dx + \int_{-\infty}^{\infty} y\cdot\left(\int_{-\infty}^{\infty} f(x,y)dx\right)dy = \int_{-\infty}^{\infty} x\cdot f_1(x)dx + \int_{-\infty}^{\infty} y\cdot f_2(y)dy = EX + EY$$

Beispiel:
X ....Anzahl der verkauften Autos im 1. Monat
Y ....Anzahl der verkauften Autos im 2. Monat
Gefragt ist die zu erwartende Anzahl von verkauften Autos in den 2 Monaten.

$$E(X+Y)=EX+EY$$

Bemerkung: Zur Berechnung des Erwartungswertes einer Summe von stochastischen Größen braucht man die Verteilung der Summe bzw. Summanden nicht (nur die Erwartungswerte).

Allgemein: Sei $X_1,...,X_n$ eine stochastische Größe mit existierendem Erwartungswert ($\exists X_i$) und $a_i \in R$ für $i=1,...,n$, dann gilt:

$$E\left(\sum_{i=1}^{n} a_i \cdot X_i\right) = \sum_{i=1}^{n} a_i \cdot EX_i$$

Die Bildung des Erwartungswertes ist ein lineares Funktional (linearer Operator).

Beispiel: Einfache Berechnung des Erwartungswertes der Binomialverteilung $B_{n,p}$.

$$X \sim B_{n,p} \Rightarrow X = X_1+...+X_n \quad \text{mit } X_i \sim A_p$$

$$EX = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \underbrace{EX_i}_{p} = np$$

23. Kovarianz, Korrelation und Unabhängigkeit stochastischer Größen

23.1 Kovarianz

Sind X und Y zwei 1-dimensionale stochastische Größen, d.h. $\begin{pmatrix} X \\ Y \end{pmatrix}$ ist ein 2-dimensionaler Vektor, so ist die Kovarianz von X und Y definiert als:

$$Cov(X,Y)=E[(X-EX)(Y-EY)]:=\sigma_{xy} \quad \text{falls die Erwartungswerte existieren}$$

Bemerkung:
1) $Cov(X,Y)=E\psi(X,Y)$ im Satz vom unbewußten Statistiker
2) Die Berechnung der Kovarianz erfolgt für diskrete bzw. Kontinuierliche Verteilungen folgendermaßen:
   - Diskrete Größen

$$Cov(X,Y) = \sum_{(x,y)\in M_{\binom{x}{y}}} (x-EX)\cdot(y-EY)\cdot p(x,y)$$

   - Kontinuierliche Größen

$$Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-EX)\cdot(y-EY)\cdot f(x,y)dxdy$$

Bemerkung:

1) Die Kovarianz zweier stochastischer Größen ist ein Maß für deren gegenseitige lineare Kopplung
2) Für mehrere stochastische Größen $X_1,...,X_m$ kann im Falle der Existenz, die Kovarianz für je zwei stochastische Größen X_k und X_l berechnet werden. Diese ergibt $\sigma_{kl}=\mathrm{Cov}(X_k,X_l)$ für k=1(1)m und l=1(1)m mit $k\neq l$, wobei gilt: $\sigma_{lk}=\sigma_{kl}$. Definiert man $\sigma_{kk}=\sigma_k^2=\mathrm{Var}X_k$, so erhält man eine symmetrische, positiv definite quadratische Matrix, genannt **Varianz-Kovarianz-Matrix**,

$$\mathrm{VCov}(X_1,...,X_m)=\Sigma=(\sigma_{kl})_{\substack{l=1,..,m\\k=1,..,m}}$$

der stochastische Größen $X_1,...,X_m$, bzw. des stochastischen Vektors $\underline{X}=\begin{pmatrix}X_1\\ \vdots\\ X_m\end{pmatrix}$.

Beispiel:

1) $\begin{pmatrix}X\\Y\end{pmatrix}\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\varphi)$

$$\mathrm{Cov}(X,Y)=\varphi\cdot\sigma_1\cdot\sigma_2$$

$$\mathrm{VCov}(X,Y)=\begin{pmatrix}\sigma_1^2 & \varphi\sigma_1\sigma_2\\ \varphi\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$$

2) $\underline{X}\sim N(\underline{\mu},\Sigma)\quad\Rightarrow\quad \mathrm{VCov}(\underline{X})=\Sigma$

Satz 23.1 Berechnungen der Varianz von Summen von stochastischen Größen

a) Sind X und Y stochastische Größen mit existierender Varianz, so gilt:
$\mathrm{Var}(X+Y)=\mathrm{Var}X+\mathrm{Var}Y+2\cdot\mathrm{Cov}(X,Y)$
b) Sind $X_1,...,X_n$ stochastische Größen mit existierender Varianz, so gilt:

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{Var}X_i+\sum_{i=1}^n\sum_{\substack{j=1\\i\neq j}}^n \mathrm{Cov}(X_i,X_j)$$

c) Sind $X_1,...,X_n$ stochastische Größen mit existierender Varianz und $c_i\in R\ \forall\ i=1(1)n$, so gilt:

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n c_i^2\cdot\mathrm{Var}X_i+\sum_{i=1}^n\sum_{\substack{j=1\\i\neq j}}^n c_i\cdot c_j\cdot\mathrm{Cov}(X_i,X_j)=$$

$$=\sum_{i=1}^n c_i^2\cdot\mathrm{Var}X_i+2\cdot\sum_{i=1}^n\sum_{\substack{j=1\\i<j}}^n c_i\cdot c_j\cdot\mathrm{Cov}(X_i,X_j)$$ ohne Beweis

23.2 Korrelationskoeffizient

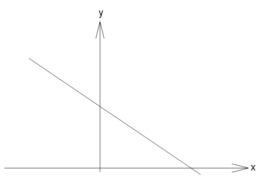
Ein normiertes Maß für die Stärke eines eventuellen näherungsweisen linearen Zusammenhanges zwischen zwei stochastischen Größen X und Y, falls deren Varianzen existieren, ist der Korrelationskoeffizient.

$$\varphi_{x,y}:=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X}\cdot\sqrt{\mathrm{Var}Y}}=\frac{\sigma_{xy}}{\sigma_x\cdot\sigma_y}=\frac{\sigma_{12}}{\sigma_1\cdot\sigma_2}$$

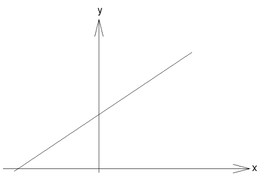
Bemerkung:

1) $-1\le\varphi_{x,y}\le+1$
2) Für $|\varphi_{x,y}|=1$ gilt mit Wahrscheinlichkeit 1, daß ein linearer Zusammenhang zwischen X und Y besteht:

$$W\{Y=k\cdot X+d\}=1$$

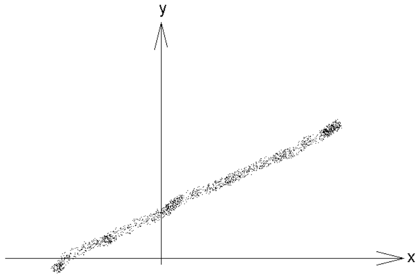
$\varphi_{x,y}=-1$ 

negativ korreliert

$\varphi_{x,y}=+1$ 

positiv korreliert

3) $X \to a \cdot X+b,\ Y \to c \cdot Y+d \Rightarrow \varphi_{aX+b,cY+d}=\varphi_{X,Y}$
4) Für eine 2-dimensionale Gleichverteilung auf (x_i,y_i): siehe Zeichnung



$\varphi_{x,y} \approx 0{,}9$

Beispiel:
$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\varphi\right) \Rightarrow \varphi_{x,y}=\varphi$

Definition: Zwei stochastische Größen X und Y heißen **unkorreliert** falls $\varphi_{x,y}=0$

Satz 23.2 Für die Varianz einer Summe von paarweise unkorrelierten stochastischen Größen $X_1,...,X_n$ gilt:

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}X_i \qquad \text{Beweis: siehe Satz 23.1.b}$$

23.3 Unabhängigkeit stochastischer Größen

Zwei stochastische Größen X und Y heißen (stochastisch) unabhängig, i.Z. $X \perp Y$, wenn ihre gemeinsame Verteilungsfunktion das Produkt der Verteilungsfunktionen der Einzelverteilungen ist, d.h. wenn gilt:

$$F_{(x,y)} = W\{X \le x, Y \le y\} = W\{X \le x\} \cdot W\{Y \le y\} = F_X(x) \cdot F_Y(y) \qquad \forall (x,y) \in R^2$$

Es gilt $X \perp Y$ genau dann, wenn:

a) im diskreten Fall $p(x,y)=p_X(x) \cdot p_Y(y) \quad \forall (x,y) \in M_{\binom{X}{Y}}$

b) im kontinuierlichen Fall $f(x,y)=f_X(x) \cdot f_Y(y) \quad \forall (x,y) \in R^2$

Bemerkung: Die stochastischen Größen $X_1,...,X_n$ sind unabhängig, wenn gilt:

a) im diskreten Fall $p(x_1,...,x_n) = \prod_{i=1}^{n} p_i(x_i)$, wobei $p_i(\cdot)$ die Randwahrscheinlichkeiten zu X_i sind.

b) im kontinuierlichen Fall $g(x_1,...,x_n) = \prod_{i=1}^{n} f_i(x_i)$, wobei $f_i(\cdot)$ die Dichte der Randverteilung von X_i ist.

Es gilt: Sind zwei stochastische Größen X und Y unabhängig, so sind für beliebige (meßbare) Funktionen $\varphi: R \to R$ und $\psi: R \to R$ auch die stochastischen Größen $\varphi(X)$ und $\psi(X)$ unabhängig, i.Z. $X \perp Y \Rightarrow \varphi(X) \perp \psi(Y)$.

Satz 23.3 $X \perp Y \Rightarrow E[\varphi(X) \cdot \psi(Y)] = E[\varphi(X) \cdot E\psi(Y)]$ ohne Beweis

Satz 23.4 $X \perp Y \Rightarrow \varphi_{x,y}=0$

$$\varphi_{x,y} := \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}X} \cdot \sqrt{\operatorname{Var}Y}} = \frac{\operatorname{Cov}(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{E[(X-EX) \cdot (Y-EY)]}{\sigma_x \cdot \sigma_y} = \qquad \text{wegen Satz 23.3}$$

$$= \frac{E[X - EX] \cdot E[Y-EY]}{\sigma_x \cdot \sigma_y} = 0$$

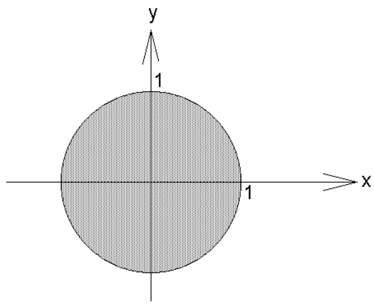
Folgerung: Die Varianz einer Summe von *paarweise unabhängigen* stochastischen Größen ist gleich der Summe der Varianzen der einzelnen stochastischen Größen.

$$X_1,...,X_n \text{ paarweise unabhängig} \to \operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}X_i \qquad \text{wegen Satz 23.1 und 23.4}$$

Bemerkung:

Wenn $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\varphi\right)$, dann gilt: $X \perp Y \Leftrightarrow \varphi=0$ (nur dann!)

Beispiel: 2-dimensionale Dichte



Dichtefunktion: $f(x,y) = \begin{cases} \frac{1}{\pi} & \text{für } x^2 + y^2 \le 1 \\ \text{sonst } 0 \end{cases}$

Randdichten:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \tfrac{1}{\pi}dy = \frac{y}{\pi}\Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} = \frac{2}{\pi}\cdot\sqrt{1-x^2} \quad \text{für } |x| \le 1$$

$$f_Y(y) = \text{ analog zu x } = \frac{2}{\pi}\cdot\sqrt{1-y^2} \quad \text{für } |y| \le 1$$

$f(x,y) \neq f_X(x)\cdot f_Y(y) \Rightarrow$ X und Y sind nicht unabhängig, wohl aber sind sie unkorreliert.

24. Bedingte Verteilungen und bedingte Erwartung

Beispiel:
X ....Preis eines Produktes
Y............Absatzmenge des Produktes
(X,Y) .....diskrete Größe
In Abhängigkeit von x ist die Wahrscheinlichkeitsverteilung von Y von Interesse, sog. bedingte Wahrscheinlichkeit, i.Z. Y|X=x.

24.1 Bedingte diskrete Verteilungen

Man berechnet die bedingte Wahrscheinlichkeit wie folgt:

$$p_{Y|X=x}(y) = W\{Y=y|X=x\} = \frac{W\{X=x,Y=y\}}{W\{X=x\}} = \frac{p(x,y)}{\underbrace{p_1(x)}_{\text{Randw. von X}}} \quad \text{für } p_1(x)>0$$

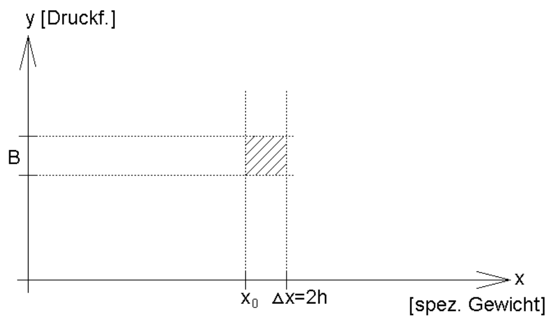
Man schreibt: p(y|x) für die durch X=x bedingte Punktwahrscheinlichkeit von Y.
Analog gilt:

$$p_{X|Y=y}(x) = \frac{p(x,y)}{\underbrace{p_2(y)}_{\text{Randw. von Y}}} = p(x|y)$$

24.2 Bedingte kontinuierliche Verteilungen

Beispiel: Materialprobe $\underline{X} = \begin{pmatrix} X & \text{spez. Gewicht} \\ Y & \text{Bruchfestigkeit} \end{pmatrix}$

In Abhängigkeit des spez. Gewichtes erhält man eine Verteilung der Druckfestigkeit.



f(x,y)......gemeinsame Dichte von $\underline{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$

$W\{X=x_0\}=0$

Für $\Delta x=2h$ gilt:

$$W\{Y\in B|x-h<X<x+h\}=\frac{W(\{Y\in B\}\cap\{x-h<X<x+h\})}{W\{x-h<X<x+h\}}=\frac{\int\limits_B\int\limits_{x-h}^{x+h}f(x,y)dxdy}{\int\limits_{x-h}^{x+h}\int\limits_{-\infty}^{\infty}f(x,y)dydx}\approx$$

$$\approx\frac{\int\limits_B 2h\cdot f(x,y)dy}{\underbrace{2h\cdot\int\limits_{-\infty}^{\infty}f(x,y)dy}_{\text{Randdichte}}}=\int\limits_B\frac{f(x,y)}{f_1(x)}dy$$

Definition: Für die durch X=x bedingte Dichte von Y gilt:

$$f(y|x):=\frac{f(x,y)}{f_1(x)}\quad\forall x: f_1(x)>0\quad\text{Randdichte von X}$$

Analog dazu gilt für die durch Y=y bedingte Dichte von X:

$$f(x|y):=\frac{f(x,y)}{f_2(x)}\quad\forall y: f_2(y)>0\quad\text{Randdichte von Y}$$

Beispiel: (X,Y) sei normalverteilt $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\varphi)$

Es gilt:

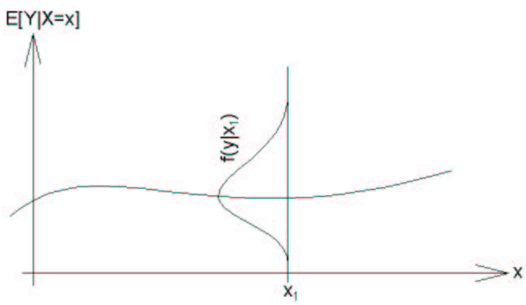
$$Y|X=x\ \sim\ N\left(\mu_2+\varphi\cdot\frac{\sigma_2}{\sigma_1}\cdot(x-\mu_1),(1-\varphi^2)\cdot\sigma_2^2\right)$$

$$X|Y=y\ \sim\ N\left(\mu_1+\varphi\cdot\frac{\sigma_1}{\sigma_2}\cdot(y-\mu_2),(1-\varphi^2)\cdot\sigma_1^2\right)$$

24.3 Bedingte Erwartung

Für ein festes x berechnet man, im Falle der Existenz, den Erwartungswert der bedingten Dichte f(y|x), i.Z. E[Y|X=x]. Die Verbindung dieser bedingten Erwartungswerte heißt bedingte Erwartung oder **Regressionslinie**. Es gilt:

$$E[Y|X=x]=\int\limits_{-\infty}^{\infty}y\cdot f(y|x)dy$$



Beispiel: Für eine 2-dimensionale Normalverteilung $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\varphi)$ gilt:

$$Y|X=x\ \sim\ N\left(\mu_2+\varphi\cdot\frac{\sigma_2}{\sigma_1}\cdot(x-\mu_1),(1-\varphi^2)\cdot\sigma_2^2\right)$$

$$\Rightarrow E[Y|X=x]=\psi(x)=\mu_2+\varphi\cdot\frac{\sigma_2}{\sigma_1}\cdot(x-\mu_1)=\text{Regressionsgerade}$$

Analog dazu berechnet man E[X|Y=y].

24.4 Erwartungswertberechnung mit Hilfe der bedingten Erwartungen

Sind X und Y stochastische Größen, deren entsprechende Erwartungswerte existieren, so können die Erwartungswerte auch mit Hilfe der bedingten Erwartungen berechnet werden.

a) im diskreten Fall: $EY = \sum_{x \in M_x} E\left[Y|X=x\right] \cdot W\left\{X=x\right\}$

b) im kontinuierlichen Fall: $EY = \int_{-\infty}^{\infty} E\left[Y|X=x\right] \cdot f_X(x)dx$

25. Funktionen von stochastischen Vektoren

In Abschnitt 22 wurden Erwartungswerte für Funktionen von stochastischen Vektoren berechnet. Nun will man die Verteilung solcher Funktionen ermitteln.

Gegeben: Verteilung des stochastischen Vektors $\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ und $Z = \psi(X_1, ..., X_n)$.

Gesucht: Verteilung von Z.

Bemerkung: Für n=1 wurde das Problem in Abschnitt 19 behandelt.

25.1 Maximum und Minimum stochastischer Vektoren

Beispiel: Ein Computer besteht aus n Bauteilen, die alle funktionieren müssen, damit der Computer einwandfrei arbeitet. Kann man die Verteilung der Lebensdauer des Computers aus den Verteilungen der Lebensdauern der Bauteile berechnen?

Bei Unabhängigkeit der Lebensdauern der Bauteile geht dies relativ einfach:

Satz 25.1 Sind die 1-dimensionalen stochastischen Größen $X_1,...,X_n$ unabhängig verteilt mit den zugehörigen Verteilungsfunktionen $F_i(\cdot)$, i=1(1)n, so gilt:

a) $W\left\{\max(X_1,...,X_n) \le x\right\} = \prod_{i=1}^{n} F_i(x) \quad \forall x \in R$

b) $W\left\{\min(X_1,...,X_n) \le x\right\} = 1 - \prod_{i=1}^{n}\left[1 - F_i(x)\right] \quad \forall x \in R$

Beweis:

a) Für ein festes x formt man das Ereignis $\left\{\max(X_1,...,X_n) \le x\right\}$ um in $\left\{\max(X_1,..,X_n) \le x\right\} = \bigcap_{i=1}^{n}\left\{X_i \le x\right\}$.

$$\Rightarrow W\left\{\max(X_1,...,X_n) \le x\right\} = W\left\{\bigcap_{i=1}^{n}\left\{X_i \le x\right\}\right\} = \prod_{i=1}^{n} \underbrace{W\left\{X_i \le x\right\}}_{F_i(x)} = \prod_{i=1}^{n} F_i(x)$$

b) Hier behilft man sich folgendermaßen:

$$W\left\{\min(X_1,...,X_n) \le x\right\} = 1 - W\left\{\min(X_1,...,X_n) > x\right\} =$$

$$= 1 - W\left(\bigcap_{i=1}^{n}\left\{X_i > x\right\}\right) = 1 - \prod_{i=1}^{n} W\left\{X_i > x\right\} = 1 - \prod_{i=1}^{n}\left[1 - \underbrace{W\left\{X_i \le x\right\}}_{F_i(x)}\right] \quad \forall x \in R$$

25.2 Faltung diskreter Verteilungen

Gegeben: Anzahl der Kunden in einem Geschäft........................ X, $X \sim p_x(\cdot)$

Anzahl der Kunden in einem anderen Geschäft........... Y, $Y \sim p_y(\cdot)$

Anzahl der Kunden gesamt... X+Y

Gesucht: Verteilung von X+Y

Satz 25.2 Ist X diskret verteilt mit der Punktwahrscheinlichkeit $p_x(\cdot)$ und Y ist diskret verteilt mit der Punktwahrscheinlichkeit $p_y(\cdot)$ und $X \perp Y$, so ist X+Y diskret verteilt mit folgender Punktwahrscheinlichkeit:

$$p_{X+Y}(a) := W\left\{X+Y=a\right\} = \sum_{x \in M_x} p_X(x) \cdot p_X(a-x) \quad \forall a \in M_{X+Y}$$

Beweis:

$$\{X+Y=a\} = \bigcup_{(x,y):x+y=a} \{X=x\} \cap \{Y=y\}$$

$$\Rightarrow W\{X+Y=a\} = W\left[\bigcup_{(x,y):x+y=a} \{X=x\} \cap \{Y=y\}\right] = \sum_{x\in M_X} \underbrace{W\{X=x\}}_{p_X(x)} \cdot \underbrace{W\{Y=a-x\}}_{p_Y(a-x)}$$

Anwendung: **Additionstheorem für Poisson-Verteilungen**

$X\sim P_\mu$, $Y\sim P_\xi$, $X\perp Y \Rightarrow X+Y\sim P_{\mu+\xi}$

Beweis:

$$p_{X+Y}(a) = \sum_{x=0}^{\infty} p_X(x)\cdot p_Y(a-x) = \sum_{x=0}^{a} \frac{\mu^x\cdot e^{-\mu}}{x!}\cdot\frac{\xi^{a-x}\cdot e^{-\xi}}{(a-x)!} = \sum_{x=0}^{a}\frac{1}{a!}\binom{a}{x}\cdot\mu^x\cdot e^{-\mu}\cdot\xi^{a-x}\cdot e^{-\xi} =$$

$$= \frac{e^{-(\mu+\xi)}}{a!}\cdot\underbrace{\sum_{x=0}^{a}\binom{a}{x}\cdot\mu^x\cdot\xi^{a-x}}_{\text{binom. Lehrsatz } (\mu+\xi)^a} = \underbrace{\frac{(\mu-\xi)^a\cdot e^{-(\mu+\xi)}}{a!}}_{\text{Poisson-V. mit Par. } \mu+\xi}$$

25.3 Faltung kontinuierlicher Verteilungen

Sind zwei 1-dimensionale, kontinuierlich verteilte stochastische Größen X und Y gegeben, so ist deren Summe X+Y wieder kontinuierlich verteilt. Die Dichtefunktion der Summe X+Y erhält man mittels folgendem Satz:

Satz 25.3 Ist X kontinuierlich verteilt mit der Dichtefunktion $f(\cdot)$ und Y ebenfalls kontinuierlich verteilt mit der Dichtefunktion $g(\cdot)$ und $X\perp Y$, dann ist X+Y kontinuierlich verteilt mit der Dichtefunktion $h(\cdot)$, wobei gilt:

$$h(z) = \int_{-\infty}^{\infty} f(z-t)\cdot g(t)\,dt$$ **Faltprodukt** der Dichten $f(\cdot)$ und $g(\cdot)$

Beispiel: Seien X und Y uniform verteilt und unabhängig; $X\sim U_{0,1}$, $Y\sim U_{0,1}$, $X\perp Y$

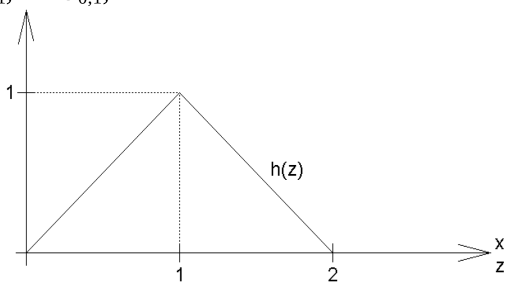
$M_{X+Y}=[0,2]$

$$h(z) = \int_{-\infty}^{+\infty} f(z-t)\cdot g(t)\,dt = \int_{-\infty}^{+\infty} I_{(0,1)}(z-t)\cdot I_{(0,1)}(t)\,dt$$

Für welche Werte t ist der Integrand bei gegebenem x positiv?

Wenn 0<z-t<1 und 0<t<1 bzw. z-1<t<z und 0<t<1.

$$h(z) = \int_{\max(0,z-1)}^{\min(z,1)} 1\cdot dt$$



Für unabhängige stochastische Größen gibt es einige sogenannte Additionstheoreme:

Satz 25.4 Für $X\sim N(\mu_1,\sigma_1^2)$ und $Y\sim N(\mu_2,\sigma_2^2)$ und $X\perp Y$ gilt:

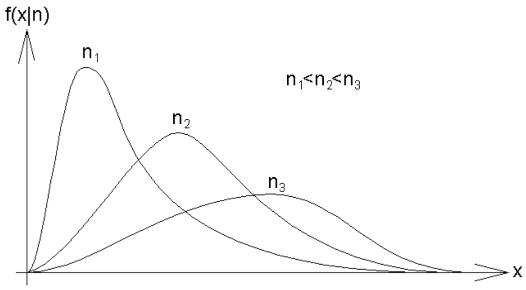
a) $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

b) $X-Y\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$

Beweis: Faltung

Definition: Eine stochastische Größe X hat eine **Chiquadrat-Verteilung** mit n Freiheitsgraden, i.Z. χ_n^2, wenn sie kontinuierlich verteilt ist mit folgender Dichtefunktion:

$$f(x) = \frac{x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}}{\Gamma\left(\frac{n}{2}\right)\cdot 2^{\frac{n}{2}}}\cdot I_{(0,\infty)}(x) \qquad n\in N$$



Dabei ist die Γ-Funktion definiert durch:

$$\Gamma(z) := \int_0^{\infty} t^{z-t}\cdot e^{-t}\,dt \quad \text{für } z>0$$

Diese Funktion ist tabelliert.

Außerdem gilt für sie folgende Rekursionsformel:

$$\Gamma(z) = (z-1)\cdot\Gamma(z-1)$$

$$\Gamma(1) = 1 \quad \text{und} \quad \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi} \Rightarrow \Gamma\left(\tfrac{n}{2}\right) = \left(\tfrac{n}{2}-1\right)\cdot\left(\tfrac{n}{2}-2\right)\cdot\ldots\ldots\cdot\tfrac{3}{2}\cdot\sqrt{\pi}$$

$$\Gamma(n) = (n-1)! \quad \text{für } n \in N$$

<u>Satz 25.5</u> Additionstheorem für χ^2-Verteilungen: Aus $X \sim \chi^2_m$, $Y \sim \chi^2_n$ und $X \perp Y$ folgt $X+Y \sim \chi^2_{m+n}$

Beweis: Faltung

VI Folgen von stochastischen Größen

26. Gesetz der großen Zahlen

Das empirische Gesetz der großen Zahlen zeigt ein konvergenzartiges Verhalten der relativen Häufigkeiten. Diese empirische Konvergenz kann mittels stochastischer Modellbildung genauer beschrieben werden. Zum Beweis benötigt man eine Abschätzung für die Abweichung einer stochastischen Größe von ihrem Erwartungswert. Das ist die sog. **Tschebyscheff´sche Ungleichung**.

Ist X eine eindimensionale stochastische Größe mit existierender Varianz $\exists \operatorname{Var} X=\sigma^2<\infty$ ($\Rightarrow \exists EX=\mu$) so folgt:

$$W\{|X-\mu| \geq \varepsilon\} \leq \frac{\operatorname{Var} X}{\varepsilon^2} \qquad \forall \varepsilon > 0$$

Beweis für kontinuierliche stochastische Größen:

$X \sim f(\cdot)$

$$\frac{\sigma^2}{\varepsilon^2} = \frac{E\left[(X-\mu)^2\right]}{\varepsilon^2} = \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-\mu)^2 \cdot f(x)dx = \frac{1}{\varepsilon^2}\int_{\{x:|x-\mu|\geq\varepsilon\}}(x-\mu)^2 \cdot f(x)dx + \frac{1}{\varepsilon^2}\int_{\{x:|x-\mu|<\varepsilon\}}(x-\mu)^2 \cdot f(x)dx \geq$$

$$\geq \frac{1}{\varepsilon^2}\int_{\{x:|x-\mu|\geq\varepsilon\}}(x-\mu)^2 \cdot f(x)dx \geq \int_{\{x:|x-\mu|\geq\varepsilon\}} 1 \cdot f(x)dx = W\{|X-\mu| \geq \varepsilon\} \qquad \text{da gilt: } |x-\mu| \geq \varepsilon \cdot \left(\frac{x-\mu}{\varepsilon}\right)^2 \geq 1$$

Gesetz der großen Zahlen

Ist X_1, X_2,... ein Folge von unabhängigen stochastischen Größen mit identischen Verteilungen und $\exists EX_i=\mu$, $\exists \operatorname{Var} X_i=\sigma^2$ dann folgt daraus:

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots X_n}{n} \xrightarrow[n\to\infty]{W} EX_i = \mu$$

und weiters

$$\lim_{n\to\infty} W\{|\overline{X}_n - \mu| \geq \varepsilon\} = 0 \quad \forall \varepsilon > 0$$

Bemerkung: Diese Konvergenz heißt **stochastische Konvergenz** oder Konvergenz in der Wahrscheinlichkeit.

Beweis:

$S_n=X_1+.....+X_n$ mit X_i unabhängig $\Rightarrow \operatorname{Var} S_n = \sum_{i=1}^{n} \operatorname{Var} X_i = n \cdot \sigma^2$

$$ES_n = \sum_{i=1}^{n} EX_i = n \cdot \mu \quad \Rightarrow \quad \operatorname{Var} \overline{X}_n = \operatorname{Var}\left(\frac{1}{n} S_n\right) = \frac{1}{n^2}\operatorname{Var} S_n = \frac{1}{n^{\cancel{2}}} \cdot \cancel{n} \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$E\overline{X}_n = E\left(\frac{1}{n} S_n\right) = \frac{1}{n} ES_n = \frac{1}{\cancel{n}} \cdot \cancel{n} \cdot \mu = \mu$$

Anwendung der Tschebyscheff'schen Ungleichung

$$W\{|\overline{X} - \mu| \geq \varepsilon\} \leq \frac{\operatorname{Var} \overline{X}_n}{\varepsilon^2} = \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n} \xrightarrow[n\to\infty]{} 0$$

Beispiel: Sei $B\in\xi$ ein festes Ereignis. Bei wiederholter Durchführung des Experimentes mit (M,ξ,W) gilt:

$$h_n(B)=\frac{H_n(B)}{n}\xrightarrow[n\to\infty]{W}W(B)$$

Zum Beweis spezialisieren wir im Gesetz der großen Zahlen:

$$X_k=I_B=\begin{cases}1\ldots\text{für k-te Beobachtung in B}\\0\ldots\text{für k-te Beobachtung nicht in B}\end{cases}$$

Es gilt: $W\{X_k=1\}=W(B)=EX_k$ da $X_k\sim A_{W(B)}$

Aus dem Gesetz der großen Zahlen folgt:

$$\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow[n\to\infty]{W}EX_k=\frac{H_n(B)}{n}\xrightarrow[n\to\infty]{W}W(B)$$

Beispiel: Münzwerfen W(B)=½

Bemerkung: Die relativen Häufigkeiten eines Ereignisses konvergieren stochastisch gegen seine Wahrscheinlichkeit.

27. Zentraler Grenzverteilungssatz

Dieser zeigt die allgemeine Bedeutung der Normalverteilung.

Satz 27.1 Ist $X_1, X_2,\ldots$ eine Folge von unabhängigen stochastischen Größen mit existierenden Varianzen $\exists\text{Var } X_k\ \forall k\in N$ und genügen die Verteilungsfunktionen $F_k(\cdot)$ der X_k einer bestimmten Bedingung (Lindeberg-Bedingung), so gilt:

Die Folge der Verteilungsfunktionen $G_n(\cdot)$ der standardisierten Summen konvergiert für $n\to\infty$ punktweise (für jedes feste $x\in R$) gegen den Wert $\Phi(x)$ der Verteilungsfunktion der standardisierten Normalverteilung.

$$Z_n=\frac{\sum_{k=1}^{n}X_k-E\left(\sum_{k=1}^{n}X_k\right)}{\sqrt{\text{Var}\left(\sum_{k=1}^{n}X_k\right)}}-G_n(\cdot)\xrightarrow[n\to\infty]{}\Phi(\cdot)$$

Bemerkung:

a) Diese Konvergenz heißt **Verteilungskonvergenz** oder auch „schwache Konvergenz“.

b) $\forall x\in R$ gilt im Fall des Satzes $G_n(x)\xrightarrow[n\to\infty]{}\Phi(x)$

c) Die Lindeberg-Bedingung ist erfüllt wenn:
   I) alle X_k identisch verteilt sind oder
   II) alle $|X_k|\le C\ \forall k\in N$ und $\text{Var}\left(\sum_{k=1}^{n}X_k\right)\xrightarrow[n\to\infty]{}\infty$

d) Die **Lindeberg-Bedingung** lautet:

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{\{x_i:|x-\mu_k|\ge\varepsilon s_n\}}(x-\mu_k)^2\,dF_k(x)\xrightarrow[n\to\infty]{}0\qquad\forall\varepsilon>0$$

$$\text{mit }\ \mu_k=EX_k\qquad s_n^2=\text{Var}\left(\sum_{k=1}^{n}X_k\right)$$

e) Die Verteilung von stochastischen Größen die als Summe eine größere Anzahl unabhängiger stochastischer Größen dargestellt werden können ist näherungsweise eine Normalverteilung, da eine lineare Transformation einer normalverteilten stochastischen Größe wieder normalverteilt ist gilt dies auch approximativ.

$$S = \sum_{k=1}^{n} X_k \quad \text{mit} \quad ES = \mu \quad \text{und} \quad \operatorname{Var} S = \sigma^2$$

und daher gilt approximativ

$$Z = \frac{S-\mu}{\sigma} \overset{\bullet}{\sim} N_{(0,1)} \qquad\qquad S = \sigma Z + \mu \overset{\bullet}{\sim} N_{(\mu,\sigma^2)}$$

f) Für viele Anwendungen ist die Approximation für $n \geq 30$ brauchbar.

Beispiel: Normal-Approximation der $B_{n,p}$ (für $n \cdot p \cdot (1-p) \geq 9$)

$X \sim B_{n,p}$

$Y \sim N_{(n \cdot p, n \cdot p \cdot (1-p))}$

$$W_B\{X = a\} \approx W_N\left\{a - \tfrac{1}{2} < Y < a + \tfrac{1}{2}\right\} =$$

$$= W_N\left\{\frac{a - \frac{1}{2} - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < \underbrace{\frac{y - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}}_{\sim N_{(0,1)}} < \frac{a + \frac{1}{2} - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right\} = \qquad \text{Standardisierung von Y}$$

$$= \Phi\left(\frac{a + \frac{1}{2} - n \cdot p}{\sqrt{n \cdot p(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - n \cdot p}{\sqrt{n \cdot p(1-p)}}\right)$$

Die Parameter müssen geschätzt werden.

28. Erneuerungsprozesse

Betrachtet man die Lebensdauer von Produkten (z.B. Glühlampen) und hat man bei Ausfall gleichartige Ersatzeinheiten, die man unmittelbar „einschaltet“, so kann dies durch folgendes stochastische Modell beschrieben werden:

X_i..........................Lebensdauer der i-ten Einheit

$W\{X_i=0\}<1$ Das Produkt ist zumindest begrenzt einsatzfähig

Definition: Ist X_1, X_2,... eine (unendliche) Folge von unabhängigen stochastischen Größen, die alle dieselbe Verteilungsfunktion $F(\cdot)$ haben und definiert man

$$S_n := \sum_{k=1}^{n} X_k$$

so heißt die Folge (S_n; $n \in N$) der zu (X_i; $i \in N$) gehörende **Erneuerungsprozeß**.

Frage: Wie groß ist die Anzahl N_t der Erneuerungen in dem gegebenen Intervall [0,t], falls unbegrenzt Ersatzeinheiten vorhanden sind?

Definition: Ist (S_n, $n \in N$) der zur Folge (X_n, $n \in N$) gehörende Erneuerungsprozeß, so heißt die Familie (N_t, $t \geq 0$) der zum Erneuerungsprozeß gehörende **Zählprozeß**.

Bemerkung: Die Verteilungsfunktion von S_n erhält man als Faltprodukt $F_n(\cdot)$: $F_n(\cdot)=F*.....*F(\cdot)$ (vgl. Abschnitt 25.3). Für die Dichtefunktion gilt folgendes:

$$f_n(\cdot) = f*.....*f(\cdot)$$

$$f_2(x) = \int_{-\infty}^{+\infty} f(x-t)\cdot f(t)dt$$

Satz 28.1 Mit den Bezeichnungen dieses Abschnittes gilt für die diskreten stochastischen Größen N_t:

a) $W\{N_t=n\}=F_n(t)-F_{n+1}(t) \quad \forall t\geq 0$

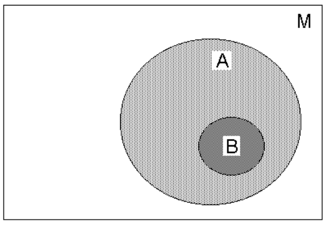
b) Mit $F_0(\cdot)=I_{[0,\infty]}(\cdot)$ und $F_1(\cdot)=F(\cdot)$ gilt:
$W\{N_t\leq n\}=1-F_{n+1}(t)$

Beweis:

a) Ereignis $\{N_t=n\}=\{S_n\leq t\}\backslash\{S_{n+1}\leq t\}$

$N_t=n\}=W\{S_n\leq t\}-W\{S_{n+1}\leq t\}$

es gilt nur, da $\{S_n\leq t\}$ enthalten ist in $\{S_{n+1}\leq t\}$!



ch obiger Bemerkung gilt:

$$W\{S_n\leq t\}= F_n(t)$$
$$W\{N_t=n\}= F_n(t)- F_{n+1}(t)$$

b) $W\{N_t \leq n\} = \sum_{k=0}^{n} W\{N_t = k\} = F_o(t) - F_{n+1}(t)$

Bemerkung: Falls X_i~$Ex\tau$, nennt man den zugehörigen Zählprozeß **Poisson-Zählprozeß**, da folgender Satz gilt:

Satz 28.2 Ist $(N_t,t\geq 0)$ der zum Erneuerungsprozeß $\left(S_n = \sum_{k=1}^{n} X_k, n\in N\right)$ gehörende Zählprozeß, so gilt:

$$\left[N_t \sim P_{t/\tau} \quad \forall t\geq 0\right] \Leftrightarrow \left[X_k \sim Ex_\tau \quad \forall k\in N\right]$$

Beweis mittels Faltung.

29. Stichproben und schließende Statistik

Ist X eine stochastische Größe, so nennt man eine endliche oder unendliche Folge X_1, X_2,....., X_n bzw. X_1, X_2,..... von unabhängigen, ebenso wie X verteilten, stochastischen Größen, eine (theoretische) **Stichprobe** von X. Konkrete Beobachtungen x_1,.....,x_n dieser stochastischen Größen nennt man eine **konkrete Stichprobe**.

Bemerkung: Eine unendliche Folge von stochastischen Größen heißt unabhängig, wenn je endlich viele davon unabhängig sind.

In der schließenden Statistik werden Methoden ermittelt, mit deren Hilfe man die Verteilung von X aufgrund einer Stichprobe (Beobachtung) schätzen bzw. testen kann, ob gewisse Hypothesen über die Verteilung von X haltbar sind. Man kann auch sagen, daß in der schließenden Statistik stochastische Modelle an Daten angepaßt werden.

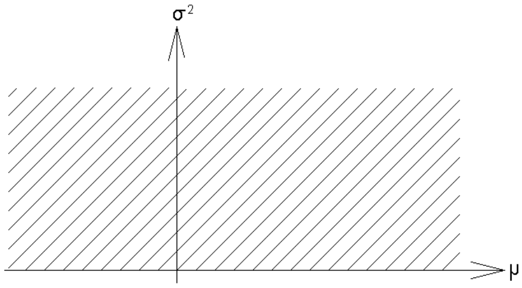
Wenn die Verteilung von X zu einer parametrischen Familie W_θ; $\theta\in\Theta$ gehört, heißt Θ **Paramterraum**. Die statistische Analyse betrifft dann Aussagen über den unbekannten Parameter θ.

Beispiel:

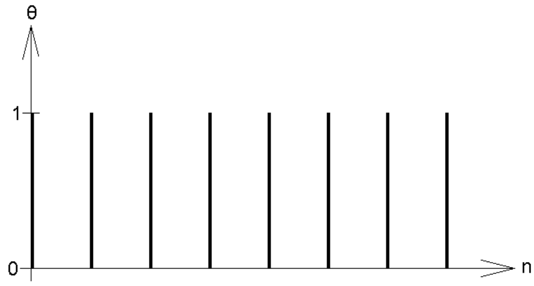
Poisson-Verteilung P_θ — $\Theta=(0,\infty)=N$

Expontential-Verteilung Ex_τ — $\Theta=(0,\infty)=R^+$

Normal-Verteilung $N(\mu,\sigma^2)$ — 2-dimensional



Binomialverteilung $B_{n,\theta}$ — 2-dimensional



Ist M_x der Merkmalraum von X und ist $X_1,.....,X_n$ eine Stichprobe von X so nimmt $\underline{X}=(X_1,.....,X_n)$ Werte im kartesischen Produkt $M_x \cdot M_x \cdot ..... \cdot M_x = M_x^n$ (sog. **Stichprobenraum** zu X) an. Im Sonderfall kontinuierlicher stochastischer Größen $X \sim f(\cdot|\theta)$ ist die gemeinsame Dichtefunktion der Stichprobe $(X_1,.....,X_n)$ folgendermaßen gegeben:

$$g(x_1,\ldots,x_n|\theta)=\prod_{i=1}^{n} f(x_i|\theta) \quad \forall (x_1,\ldots,x_n)\in M_X^n$$

Im Fall diskreter stochastischer Größen $X \sim p(\cdot|\theta)$ gilt für die Punktwahrscheinlichkeiten:

$$p(x_1,\ldots,x_n|\theta)=\prod_{i=1}^{n} p(x_i|\theta) \quad \forall (x_1,\ldots,x_n)\in M_X^n$$

Verfahren der schließenden Statistik

- Punktschätzungen für Parameter
- Bereichsschätzungen für Parameter
- nicht parametrischen Schätzungen
- statistische Tests

Diese Verfahren sind alle Sonderfälle sog. **statistischer Entscheidungen** d (engl. decision). Diese Entscheidungen erhält man mittels sog. Entscheidungsregeln $\delta: M_X^n \to D$. Dabei ist D die Menge der möglichen Entscheidungen.

Für Beobachtungen $x_1,....,x_n$ liefert eine sogenannte **Schätzfunktion** (Entscheidungsregel) $\vartheta: M_X^n \to \Theta$ den **Schätzwert** $\hat{\theta}$ (Entscheidung).

Bemerkung: Die Beziehung zu Folgen stochastischer Größen ergibt sich, wenn man keine feste Anzahl von Beobachtungen einer stochastischen Größe vorgibt, sondern den Stichprobenumfang n anwachsen läßt, d.h. $n \to \infty$.

Frage: Konvergieren für den Stichprobenumfang $n \to \infty$ Schätzungen gegen die zugrunde liegenden Größen? (siehe Abschnitt 30.3).

VII Klassische schließende Statistik

In der klassischen schließenden Statistik geht man davon aus, daß den beobachteten stochastischen Größen X eine „wahre" Wahrscheinlichkeitsverteilung zu Grunde liegt, die man nicht kennt. Diese unbekannte Verteilung will man aufgrund von Beobachtungen von X „schätzen" bzw. will man entscheiden, ob gewisse Aussagen über die Verteilung von X haltbar sind. Wenn die gesuchte Wahrscheinlichkeitsverteilung durch einen endlich-dimensionalen Parameter charakterisiert ist, spricht man von **parametrischer Statistik**, ansonsten von nichtparametrischer Statistik.

30. Klassische Punktschätzungen für Parameter

Zugrunde liegt das stochastische Modell $X\sim f(\cdot|\theta)$ bzw. $p(\cdot|\theta);\theta\in\Theta$ wobei θ_0 der wahre Parameter ist. Zu beachten ist, daß $\vartheta(X_1,....,X_n)$ ebenfalls eine stochastische Größe ist.

Beispiel:
$X\sim Ex_\tau$

Dichte: $f(x|\tau)=\frac{1}{\tau}\cdot e^{-\frac{x}{\tau}}\cdot I_{(0,\infty)}(x)$ und $X_1,.....,X_n$ ist eine Stichprobe von X

Schätzfunktion: $\vartheta(X_1,.....,X_n)=\overline{X}_n=\frac{1}{n}\cdot\sum_{i=1}^{n}X_i$.....Stichprobenmittel

Für eine konkrete Stichprobe $x_1,.....,x_n$ gilt in diesem Fall: $\hat{\theta}=\overline{x}_n=\frac{x_1+.....+x_n}{n}$

Manchmal ist nur ein Teilparameter oder eine Funktion $\tau(\theta)$ des Parameters von Interesse. In diesem Fall spricht man von **Raffung** des Parameters. Die Raffung wird mathematisch durch eine Funktion $\tau:\Theta\to\Theta^*$ beschrieben, wobei Θ^* die Menge aller möglichen gerafften Parameter, sog. geraffter Parameterraum, bezeichnet, d.h.

$$\Theta^*=\{\tau(\theta):\theta\in\Theta\}$$

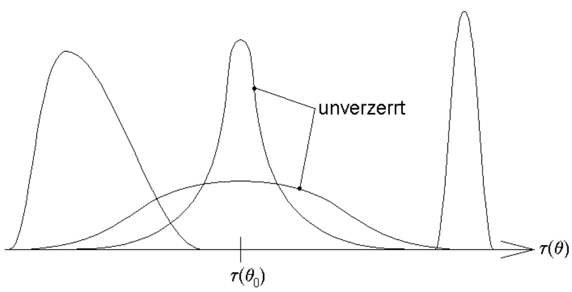
Man muß dabei beachten, daß $\tau(\theta)$ kann auch eine identische Abbildung sein kann ($\tau(\theta)=\theta$).

Beispiel:
$X\sim N(\mu,\sigma^2)$ $\theta=(\mu,\sigma^2)\Rightarrow\tau(\theta)=\mu$ oder $\tau(\theta)=\sigma^2$
Um brauchbare Schätzfunktionen für (geraffte) Parameter zu erstellen, verwendet man verschiedene Güteeigenschaften von Schätzfunktionen.

30.1 Unverzerrtheit

Definition: Die Schätzfunktion $t(X_1,.....,X_n)$ für einen gerafften Parameter $\tau(\theta)$ heißt unverzerrt (erwartungstreu), falls gilt:

$$Et(X_1,.....,X_n)=\tau(\theta)\forall\theta\in\Theta$$



Satz 30.1 Ist $X_1,....,X_n$ eine Stichprobe von X mit existierender Varianz, dann ist

$$\overline{X}_n=t_1(X_1,....,X_n):=\frac{1}{n}\cdot\sum_{i=1}^{n}X_i$$

eine unverzerrte Schätzfunktion für den Erwartungswert EX, sog. **Stichprobenmittel**. Eine unverzerrte Schätzfunktion für die Varianz von X ist die **Stichprobenvarianz**.

$$S_n^2:=\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2=t_2(X_1,....,X_n)\quad \text{da } ES_n^2=.....=\text{VarX}$$

Bemerkung: Für konkrete Beobachtungen $x_1,....,x_n$ erhält man Schätzwerte $\widehat{EX}$ bzw. $\widehat{VarX}$

30.2 Effizienz

Für unverzerrte Schätzfunktionen wäre eine möglichst kleine Varianz wünschenswert.

Definition: Betrachtet man eine Familie T von unverzerrten Schätzfunktionen $t(X_1,....,X_n)$ für einen gerafften Parameter (1-dimensional) $\tau(\theta)\in R$, so heißt eine Schätzfunktion $T^*=t^*(X_1,....,X_n)$ effizient für $\tau(\theta)$, wenn T^* die kleinstmögliche Varianz hat, d.h. wenn gilt:

$$\mathrm{Var}_\theta t^*(X_1,....,X_n)=\min_{t\in T}\mathrm{Var}_\theta t(X_1,....,X_n)\qquad\forall\theta\in\Theta$$

Satz 30.2 Ist X eine stochastische Größe mit endlicher Varianz σ^2 und $X_1,....,X_n$ eine Stichprobe von X so ist $\overline{X}_n$ die effiziente, lineare Schätzfunktion von EX.

Beweis:
Die Menge der linearen Schätzfunktionen ist:

$$T=\left\{t(X_1,....,X_n)=\sum_{i-1}^{n}\alpha_i X_i\quad\text{mit }\alpha_i\in R\right\}$$

Für die Unverzerrtheit muß gelten:

$$Et(X_1,....,X_n)=\mu=EX$$

Es muß gelten $\sum_{i=1}^{n}\alpha_i=1$ da:

$$E\left[\sum_{i=1}^{n}\alpha_i\cdot X_i\right]=\sum_{i=1}^{n}E[\alpha_i\cdot X_i]=\sum_{i=1}^{n}\alpha_i\cdot\underbrace{EX_i}_{=EX=\mu}=\underbrace{\mu\cdot\sum_{i=1}^{n}\alpha_i}_{=\mu\quad\text{da unverzerrt}}$$

Die betrachtete Menge T von Schätzfunktionen ist daher:

$$T=\left\{\sum_{i=1}^{n}\alpha_i\cdot X_i:\sum_{i=1}^{n}\alpha_i=1\right\}$$

Für die Varianz einer Schätzfunktion $t(X_1,....,X_n)$ aus T gilt:

$$\mathrm{Var}\left(\sum_{i=1}^{n}\alpha_i\cdot X_i\right)\overset{\text{nur bei Unabhängigkeit!}}{=}\sum_{i=1}^{n}\mathrm{Var}(\alpha_i\cdot X_i)=\sum_{i=1}^{n}\alpha_i^2\cdot\mathrm{Var}X_i=\sigma^2\cdot\sum_{i=1}^{n}\alpha_i^2\qquad\text{vgl. Satz 18.1 und 23.4}$$

Es gilt, daß $T=t(X_1,....,X_n)=\sum_{i=1}^{n}\alpha_i\cdot X_i$ effizient ist, wenn $\sum_{i=1}^{n}\alpha_i^2$ minimal ist. Um zu sehen, für welche Werte $\alpha_1,....,\alpha_n$ dies gilt, geht man wie folgt vor:

$$\sum_{i=1}^{n}\alpha_i^2=\sum_{i=1}^{n}\left(\alpha_i-\frac{1}{n}+\frac{1}{n}\right)^2=\sum_{i=1}^{n}\left(\alpha_i-\frac{1}{n}\right)^2+\sum_{i=1}^{n}2\cdot\left(\alpha_i-\frac{1}{n}\right)\cdot\frac{1}{n}+\sum_{i=1}^{n}\frac{1}{n^2}=$$

$$=\sum_{i=1}^{n}\left(\alpha_i-\frac{1}{n}\right)^2+\underbrace{\frac{2}{n}\left(\underbrace{\sum_{i=1}^{n}\alpha_i}_{1}-n\cdot\underbrace{\frac{1}{n}}_{1}\right)}_{0}+\frac{1}{n}$$

Dieser Ausdruck wird minimal für $\alpha_i=1/n\ \forall i=1(1)n$, d.h. $\overline{X}_n$ ist die effiziente lineare Schätzfunktion für EX.

Bemerkung: Die Varianz von Schätzfunktionen kann für ein festes n nicht beliebig klein gemacht werden. Unter bestimmten Regularitätsvoraussetzungen gilt die sog. „Ungleichung von **Fréchét-Rao-Cramér**".

$X\sim f(\cdot|\theta);\ \theta\in R;\ \tau(\theta)=$ reele Raffung; $\exists\tau'(\theta)=\frac{d}{d\theta}\tau(\theta)$

$T=t(X_1,....,X_n)$ sei die Schätzfunktion für $\tau(t)$

$$\mathrm{Var}_\theta T\geq\frac{(\tau'(\theta))^2}{n\cdot E\left[\left\{\frac{\partial}{\partial\theta}\ln f(X|\theta)\right\}^2\right]}$$

Falls T unverzerrt bezüglich θ ist, dann gilt:

$$\mathrm{Var}_\theta T\geq\frac{1}{n\cdot E\left[\left(f(x|\theta)\right)^2\right]}$$

30.3 Konsistenz

Konsistenz ist eine Eigenschaft von Folgen $T_n=t_n(X_1,....,X_n)$ von Schätzfunktionen, wobei eine unbeschränkte Stichprobe $X_1,X_2,....$ von X vorliegen muß.

Definition: Die Schätzfolge T_n; $n\in N$ für $\tau(\theta)\in R$ heißt konsistent, falls $T_n \underset{n\to\infty}{\overset{W}{\to}} \tau(\theta) \quad \forall\theta\in\Theta$, d.h.:

$$\lim_{n\to\infty} W_\theta\left\{\tau(\theta)-\varepsilon < T_n < \tau(\theta+\varepsilon)\right\} = 1 \quad \forall\begin{cases}\theta\in\Theta\\ \varepsilon>0\end{cases} \quad \text{sog. stochastische Konvergenz}$$

Satz 30.3 Ist X eine stochastische Größe mit $X\sim W_\theta$ und $\tau(\theta)\in R$, ferner X_1, X_2,.... eine unbegrenzte Stichprobe und $T_n=t_n(X_1,....,X_n)$ eine Schätzfolge mit $\lim\limits_{n\to\infty} E_\theta T_n = \tau_\theta$ und $\lim\limits_{n\to\infty} Var_\theta T_n = 0$, so ist die Folge T_n, $n\in N$ von Schätzfunktionen eine konsistente Schätzfunktion für $\tau(\theta)$.

Beispiel: Unter den Voraussetzungen von Satz 30.3 ist $\overline{X}_n$, $n\in N$ eine konsistente Schätzfolge für EX, d.h. $\overline{X}_n$ konvergiert für $n\to\infty$ stochastisch gegen den Erwartungswert von X.

30.4 Plausibilität

Diese Güteeigenschaft bezieht sich auf konkrete Schätzwerte, nicht auf Schätzfunktionen.

a) Diskretes stochastisches Modell

Ausgehend von einer konkreten Stichprobe $x_1,....,x_n$ betrachtet man die Wahrscheinlichkeiten der beobachteten Werte in Abhängigkeit des Parameters θ. Die Plausibilitätsfunktion $l(\theta;x_1,....,x_n)$ wird durch die gemeinsamen Punktwahrscheinlichkeiten $p(x_1,....,x_n)$ der Stichprobe als Funktion des Parameters θ gebildet.

$$W\left\{\left.\begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix}=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\right|\theta\right\}=\prod_{i=1}^{n} p\left(x_i\middle|\theta\right)=l(\theta;x_1,....,x_n)$$

Der plausible Schätzwert $\hat{\theta}$ für den wahren Parameter θ_0 ist jener Parameterwert für den diese Wahrscheinlichkeit maximal wird, vorausgesetzt dieses Maximum existiert.

b) Kontinuierliches stochastisches Modell

Man faßt die gemeinsame Dichtefunktion $g(x_1,....,x_n|\theta)$ als Funktion von θ auf, d.h.

$$g\left(x_1,....,x_n\middle|\theta\right)=\prod_{i=1}^{n} f\left(x_i\middle|\theta\right)=l(\theta;x_1,....,x_n).$$

Der plausible Schätzwert für θ_0 ist jener Parameterwert $\hat{\theta}$, für den die gemeinsame Dichtefunktion von $(X_1,....,X_n)$ beim Variablenwert $x_1,....,x_n$ maximal wird, vorausgesetzt dieses Maximum existiert.

In beiden Fällen nennt man $l(\theta;x_1,...,x_n)$ die Plausibilitätsfunktion (eng. likelihood). Daher gilt:

$$\hat{\theta}=\arg\max\, l(\theta;x_1,...,x_n) \quad \text{oder} \quad l(\hat{\theta};x_1,...,x_n)=\max_{\theta\in\Theta} l(\theta;x_1,...,x_n).$$

Bemerkung:

1) Plausible Schätzwerte müssen weder existieren noch eindeutig sein.
2) Zur Bestimmung plausibler Schätzwerte kann man häufig die Ableitung der Plausibilitätsfunktion null setzen, d.h.:
$$\frac{\partial\, l(\theta; x_1,....,x_n)}{\partial\theta} = 0 \quad \text{oder} \quad \frac{\partial \ln l(\theta; x_1,....,x_n)}{\partial\theta} = 0$$
3) Im Falle eines mehrdimensionalen Parameters $\theta=(\theta_1,....,\theta_k)$ lautet eine notwendige Bedingung für ein Maximum:
$$\frac{\partial\, l(\theta_1,...,\theta_k; x_1,...,x_n)}{\partial\theta_j} = 0 \quad \forall j = 1(1)k \quad \text{oder} \quad \frac{\partial \ln l(\theta_1,...,\theta_k; x_1,...,x_n)}{\partial\theta_j} = 0 \quad \forall j = 1(1)k$$
4) Plausible Schätzwerte für geraffte Parameter $\tau(\theta)$ erhält man aus dem plausiblen Schätzwert $\hat{\theta}$ mittels $\tau(\hat{\theta})$.

Beispiel:

Die stochastische Größe X_i sein alternativverteilt (X_i~A_θ), wobei θ dem Schlechtanteil in einem Los entspricht (Θ=[0,1]). Es werden n Ziehungen mit Zurücklegen durchgeführt ($x_i \in \{0,1\}$). Die Plausibilitätsfunktion ist:
$$l(\theta; x_1,...,x_n) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}\cdot(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}\cdot(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$
Gesucht ist der plausible Schätzwert $\hat{\theta}$.
$$\ln l(\theta; x_1,...,x_n) = \left(\sum_{i=1}^{n} x_i\right)\cdot\ln\theta + \left(n-\sum_{i=1}^{n} x_i\right)\cdot\ln(1-\theta)$$
$$\frac{\partial \ln\ln l(\theta; x_1,...,x_n)}{\partial\theta} = \left(\sum_{i=1}^{n} x_i\right)\cdot\frac{1}{\theta} + \left(n-\sum_{i=1}^{n} x_i\right)\cdot\frac{1}{1-\theta}\cdot(-1)$$
Nullsetzen der Ableitung der Plausibilitätsfunktion führt auf:
$$(1-\theta)\cdot\sum_{i=1}^{n} x_i = \theta\cdot\left(n-\sum_{i=1}^{n} x_i\right)$$
$$\sum_{i=1}^{n} x_i = \theta\cdot\sum_{i=1}^{n} x_i + n\cdot\theta - \theta\cdot\sum_{i=1}^{n} x_i = n\cdot\theta$$
$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Definition: Eine **plausible Schätzfunktion** ist eine solche, die jeder konkreten Stichprobe den plausiblen Schätzwert zuordnet.

Beispiel: Für die plausible Schätzfunktion einer alternativverteilten Stichprobe (X~A_θ) gilt: $\vartheta(X_1,...,X_n)=\overline{X}_n$

Beispiel: Die stochastische Größe X sei normalverteilt (X~N(μ,σ^2), $\theta=(\mu,\sigma^2)$). Gegeben ist eine konkrete Stichprobe $x_1,...,x_n$. Es gilt:
$$\ln l(\mu,\sigma^2; x_1,...,x_n) = -\frac{n}{2}\cdot\ln(2\pi) - \frac{n}{2}\cdot\ln(\sigma^2) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\cdot\sigma^2}$$
Durch partielle Ableitung mit μ bzw. σ^2 erhält man die Schätzfunktion:
$$\widehat{(\mu,\sigma^2)} = \left(\overline{x}_n, \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\overline{x}_n)^2\right)$$
Bemerkung: Hier ist die plausible Schätzfunktion für die Varianz verzerrt (vgl. Abschnitt 30.1)!

30.5 Anteilsschätzung

Gegeben sind N Einheiten, davon A mit einer bestimmten Eigenschaft. Gefragt ist der Anteil von A in der Gesamtheit, d.h. $\theta=A/N$. Dazu werden n Einheiten untersucht. Die stochastische Größe X gibt die Anzahl der Einheiten, unter den n untersuchten, an, welche die bestimmte Eigenschaft besitzen.

Für den Schätzwert von θ gilt $\hat{\theta}=x/n$ und für die Schätzfunktion gilt $T=X/n$.

a) Ziehungen mit Zurücklegen

Die stochastische Größe X ist binomialverteilt, d.h.:

$$X\sim B_{n,\theta};\ EX=n\cdot\theta;\ ET=1/n\cdot EX=\theta$$

T ist eine unverzerrte Schätzfunktion für θ.

$$\mathrm{VarT}=\mathrm{Var}\left(\frac{X}{n}\right)=\frac{1}{n^2}\mathrm{VarX}=\frac{1}{n^2}\cdot n\cdot\theta\cdot(1-\theta)=\frac{\theta\cdot(1-\theta)}{n}$$

Daher bildet T_n; $n\in N$ eine konsistente Schätzfolge (vgl. Satz 30.3) für θ (plausibler Schätzwert).

b) Ziehungen ohne Zurücklegen

Die stochastische Größe ist hypergeometrisch verteilt, d.h.:

$$X\sim H_{N,A,n};\ EX=n\cdot A/N=n\cdot\theta;\ ET=1/n\cdot EX=\theta$$

T ist eine unverzerrte Schätzfunktion für θ.

$$\mathrm{VarT}=\frac{1}{n^2}\cdot\mathrm{VarX}=\frac{1}{n^2}\cdot n\cdot\frac{A}{N}\cdot\left(1-\frac{A}{N}\right)\cdot\underbrace{\left(1-\frac{n-1}{N-1}\right)}_{\text{Korrekturfaktor für endliche Gesamtheiten}}=\frac{\theta\cdot(1-\theta)}{n}\cdot\left(1-\frac{n-1}{N-1}\right)$$

Im Vergleich zur Ziehung mit Zurücklegen erhält man eine kleinere Varianz. Für den Fall, daß $N\to\infty$ gilt, kann man eine Approximation mit $B_{n,\theta}$ durchführen.

Bemerkung: Für n<200 sind nur zwei Dezimalstellen brauchbar.

c) Eine bessere Analyse ist mittels Bayes-Methoden möglich (vgl. Kapitel VIII).

30.6 Schätzung des Korrelationskoeffizienten bei 2-dim. normalverteilten Daten

Falls die stochastische Größe (X,Y) 2-dimensional normalverteilt ist, gilt:

$$(X,Y)\sim N\left(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\varphi\right)$$

$$\varphi=\frac{\mathrm{Cov}(X,Y)}{\sigma_X\cdot\sigma_Y}$$

$$\mathrm{Cov}(X,Y)=E\left[(X-\mu_X)\cdot(Y-\mu_Y)\right]$$

Satz 30.4 Ist (x_i,y_i); $i=1(1)n$ eine konkrete Stichprobe von der 2-dimensionalen Stichprobe (X,Y), so ist der Stichproben-Korrelationskoeffizient

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x}_n)\cdot(y_i-\bar{y}_n)}{\sqrt{\left[\sum_{i=1}^{n}(x_i-\bar{x}_n)^2\right]\cdot\left[\sum_{i=1}^{n}(y_i-\bar{y}_n)^2\right]}}$$

der plausible Schätzwert für φ.

Der plausible Schätzwert für den 5-dimensionalen Parameter θ ist:

$$\theta=\begin{pmatrix}\mu_X\\ \mu_Y\\ \sigma_X^2\\ \sigma_Y^2\\ \varphi\end{pmatrix}\Rightarrow\hat{\theta}=\begin{pmatrix}\bar{x}_n\\ \bar{y}_n\\ \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2\\ \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2\\ r\end{pmatrix}$$

30.7 Bemerkungen zu den Güteeigenschaften von Schätzfunktionen

Verschiedene Güteeigenschaften können verschiedene Schätzwerte bzw. Schätzfunktionen ergeben.

a) Unter bestimmten, allgemeinen Voraussetzungen stimmen plausible und effiziente Schätzfunktion überein.
b) Unter relativ allgemeinen Voraussetzungen bildet eine Folge von plausiblen Schätzfunktionen eine konsistente Schätzfolge.
c) Unter bestimmten Voraussetzungen sind plausible Schätzfunktionen asymptotisch normalverteilt (Approximation Verteilung von Schätzfunktionen).

31. Nichtparametrische Schätzungen der Verteilungsfunktion

Ist die Annahme einer parametrischen Verteilungsfamilie für eine stochastische Größe nicht gerechtfertigt, so kann man im 1-dimensionalen Fall die Verteilungsfunktion schätzen. Für n reelle Zahlen $x_1,....,x_n$ betrachtet man die diskreten Gleichverteilung mit den Merkmalswerten $x_1,....,x_n$ (nicht äquidistant). Ordnet man die Zahlen der Größe nach, so erhält man die **geordnete Stichprobe** $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$. Die Verteilungsfunktion zu dieser diskreten Gleichverteilung ist die empirische Verteilungsfunktion $F_n^*(\cdot)$ (vgl. Abschnitt 4.2).

$$F_n^*(x) = \begin{cases} 0 \text{ für } x < x_{(1)} \\ \frac{i}{n} \text{ für } x \in \left[x_{(i)}, x_{(i+1)}\right) \; i = 1(1)n-1 \\ 1 \text{ für } x \ge x_{(n)} \end{cases}$$

Vor Beobachtung von $X \sim F(\cdot)$ kann man eine stochastische Größe als Schätzfunktion angeben, die für konkrete Beobachtungen $x_1,....,x_n$ genau die empirische Verteilungsfunktion $F_n^*(\cdot)$ ergibt.
Ist $X_1, X_2,...., X_n$ eine Stichprobe von X, so definiert man für ein festes $x \in R$:

$$F_n^*(x|X_1,..., X_n) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$

Die beiden folgenden Sätze zeigen, daß die empirische Verteilungsfunktion eine gute Schätzung für die zugrundeliegende Verteilungsfunktion ist.

Satz 31.1 X sei eine 1-dimensionale stochastische Größe mit der Verteilungsfunktion $F(\cdot)$. Wenn $X_1,...,X_n$ eine Stichprobe von X ist, so gilt für jedes feste $x \in R$:

a) Diese Schätzfolge $F_n^*(x|X_1,...,X_n)$ ist eine unverzerrte Schätzfunktion für F(X).
b) $F_n^*(x|X_1,...,X_n)$ ist eine konsistente Schätzfolge.

Beweis:

a) $$EF_n^*(x|X_1,..., X_n) = E\frac{1}{n}\cdot\sum_{i=1}^{n} I_{(-\infty,x]}(X_i) = \frac{1}{n}\cdot\sum_{i=1}^{n} \underbrace{E\underbrace{I_{(-\infty,x]}(X_i)}_{\sim A_{F(x)}}}_{F(x)} = \frac{1}{n}\cdot n\cdot F(x) = F(x)$$

b) Z_i sei unabhängig und identisch verteilt wie X_i.

$$F_n^*(x|X_1,..., X_n) = \frac{\overbrace{I_{(-\infty,x]}(X_1)}^{Z_1}+....+\overbrace{I_{(-\infty,x]}(X_n)}^{Z_n}}{n} \xrightarrow{W} EZ_i = F(x) \quad \text{Gesetz der großen Zahlen}$$

Die Folge $F_n^*(x)$ der Werte der empirischen Verteilungsfunktion zu festem x konvergiert für $n \to \infty$ stochastisch gegen den Wert F(x) der zugrundeliegenden Verteilungsfunktion an dieser Stelle.

Fundamentalsatz der Statistik

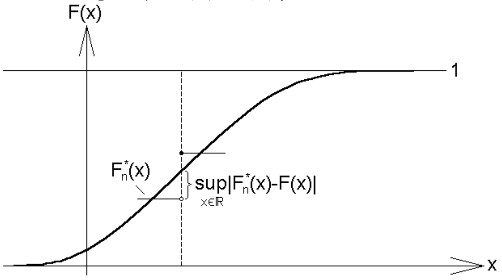
Satz 31.2

Unter den Voraussetzungen von Satz 31.1 gilt:

$$W\left\{\lim_{n\to\infty}\sup_{x\in R}\left|F_n^*(x)-F(x)\right|=0\right\}=1 \qquad F_n^*(x) \text{ wird verwendet für } F_n^*(x|X_1,...,X_n)$$

Bemerkungen:

1) Mit Wahrscheinlichkeit 1 konvergiert die Folge der empirischen Verteilungsfunktionen gleichmäßig in x gegen die zugrundeliegende Verteilungsfunktion.
2) Da der extreme Abstand zwischen $F_n^*(\cdot)$ und $F(\cdot)$ nicht angenommen werden muß, ist es erforderlich statt max|...| die Funktion sup|...| zu verwenden. Damit betrachtet man die kleinste obere Schranke der Abweichungen $|F_n^*(x)-F(x)|$.



32. Klassische Punktschätzung für Parameter

Dient zur quantitative Analyse über die Güte von Parameterschätzungen. Dazu sucht man Teilbereiche des Parameterraumes, in denen der wahre Parameter θ_0 mit hoher Wahrscheinlichkeit liegt.

Definition: Ist $X\sim W_\theta$, $\theta\in\Theta$ ein stochastisches Modell, so heißt eine Funktion $\kappa(X_1,...,X_n)$, die jeder Stichprobe $X_1,....,X_n$ von X eine Teilmenge $\Theta^*\subseteq\Theta$ derart zuordnet, daß gilt:

$$W_\theta\left\{\theta\in\Theta^*=\kappa(X_1,...,X_n)\right\}=1-\alpha \qquad \forall\theta\in\Theta$$

eine **Konfidenzfunktion** für den Parameter θ mit Überdeckungswahrscheinlichkeit $1-\alpha$. Die Konstante α ist eine Zahl nahe 0.

Bemerkung: Vor Erhebung einer konkreten Stichprobe $x_1,...,x_n$ ist $\kappa(X_1,....,X_n)$ ein stochastischer Bereich. Nach Erhebung der Daten ist $\kappa(x_1,...,x_n)$ eine Teilmenge von Θ. Praktisch besonders wichtig sind Konfidenzbereiche für 1-dimensionale (geraffte) Parameter, sog. **Konfidenzintervalle**.

Beispiel: Die stochastische Größe X sei normalverteilt $X\sim N(\mu,\sigma^2)$. Für die Stichprobe $X_1,....,X_n$ von X gilt:

$$\overline{X}_n \sim N\left(\mu,\frac{\sigma_0^2}{n}\right) \quad\Rightarrow\quad \frac{\overline{X}_n-\mu}{\sigma_0/\sqrt{n}}\sim N(0,1)$$

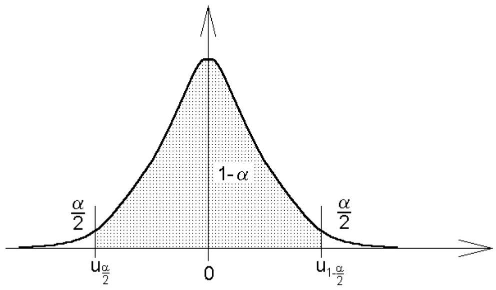
Bemerkung: Die stochastische Größe $\frac{\overline{X}_n-\mu}{\sigma_0/\sqrt{n}}=Q$ ist eine Funktion der Stichprobe und vom unbekannten Parameter abhängig. Ihre Verteilung hängt nicht vom unbekannten Parameter (μ,σ_0^2) ab. Solche Größen sind der „Angelpunkt“ für die Konstruktion von Konfidenzintervallen und man nennt sie **Pivot-Größen**. Prüfung!

32.1 Pivot-Größen-Methode

Es wird nun versucht, ein Konfidenzintervall für vorheriges Beispiel zu konstruieren. Dabei gilt, daß $Q=q(X_1,...,X_n;\mu)$ eine Pivot-Größe mit kontinuierlicher Verteilung ist. Es gilt:

$$W\left\{u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n-\mu}{\sigma_0/\sqrt{n}} \leq u_{1-\frac{\alpha}{2}}\right\}=1-\alpha \quad \forall\left(\mu,\sigma_0^2\right)$$

Dabei ist u_p das entsprechende p-Fraktile der Standard-Normalverteilung.



Eine äquivalente Umformung der Doppelungleichung

$$u_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n-\mu}{\sigma_0/\sqrt{n}} \leq u_{1-\frac{\alpha}{2}}$$

führt auf:

$$u_{\frac{\alpha}{2}}\cdot\frac{\sigma_0}{\sqrt{n}} \leq \overline{X}_n-\mu \leq u_{1-\frac{\alpha}{2}}\cdot\frac{\sigma_0}{\sqrt{n}}$$

$$-\overline{X}_n+u_{\frac{\alpha}{2}}\cdot\frac{\sigma_0}{\sqrt{n}} \leq -\mu \leq -\overline{X}_n+u_{1-\frac{\alpha}{2}}\cdot\frac{\sigma_0}{\sqrt{n}}$$

$$\overline{X}_n-u_{1-\frac{\alpha}{2}}\cdot\frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \overline{X}_n-u_{\frac{\alpha}{2}}\cdot\frac{\sigma_0}{\sqrt{n}}$$

Aufgrund der Symmetrie der Normalverteilung $\left(u_{\frac{\alpha}{2}}=-u_{1-\frac{\alpha}{2}}\right)$ gilt:

$$\overline{X}_n-\frac{\sigma_0}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X}_n+\frac{\sigma_0}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}}$$

$$\Rightarrow W\left\{\overline{X}_n-\frac{\sigma_0}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X}_n+\frac{\sigma_0}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}}\right\}=1-\alpha \quad \forall\left(\mu,\sigma_0^2\right)$$

Daher ist ein gültiges Konfidenzintervall für μ mit der Überdeckungswahrscheinlichkeit $1-\alpha$ gleich:

$$\left[\overline{X}_n-\frac{\sigma_0}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}},\overline{X}_n+\frac{\sigma_0}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}}\right]$$

Allgemein:

Es sei $Q=q(X_1,...,X_n;\theta)$ eine Pivot-Größe mit kontinuierlicher Verteilung, weiters $1-\alpha$ eine Zahl nahe bei 1 und q_1 und q_2 zwei p-Fraktile und es gilt:

$$W_\theta\{q_1\leq q(X_1,...,X_n;\theta)\leq q_2\}=1-\alpha \quad \forall\theta\in\Theta$$

Wenn man die Doppelungleichung identisch umformen kann, so daß θ in der Mitte steht, so erhält man

$$W\{t_1(X_1,...,X_n)\leq\theta\leq t_2(X_1,...,X_n)\}=1-\alpha \quad \forall\theta\in\Theta$$

und das stochastische Intervall

$$[T_1=t_1(X_1,...,X_n),T_2=t_2(X_1,...,X_n)]$$

ist ein Konfidenzintervall für θ mit der Überdeckungswahrscheinlichkeit $1-\alpha$.

Falls θ 2-dimensional ist, so konstruiert man Konfidenzintervalle für geraffte Parameter. Die obigen Umformungen erhalten die Form

$$W_\theta\{q_1\leq q(X_1,...,X_n;\theta)\leq q_2\}=1-\alpha \quad \forall\theta\in\Theta$$

und nach identischer Umformung (falls möglich) erhält man:

$$W\{T_1=t1(X_1,...,X_n)\leq\tau(\theta)\leq T_2=t_2(X_1,...,X_n)\}=1-\alpha \quad \forall\theta\in\Theta$$

32.2 Konfidenzintervalle für den Parameter der Exponentialverteilung

Satz 32.1 Sind die stochastischen Größen $X_1,...,X_n$ unabhängig verteilt nach Ex_τ, so gilt:

$$\frac{2 \cdot \sum_{i=1}^{n} X_i}{\tau} \sim \chi^2_{2n} \qquad \text{Bemerkung: } \frac{2 \cdot \sum_{i=1}^{n} X_i}{\tau} \text{ ist eine Pivot-Größe.}$$

Satz 32.2 Für eine stochastische Größe $X \sim Ex_\tau$ und der Stichprobe $X_1,...,X_n$ von X erhält man ein Konfidenzintervall für den Parameter θ mit der Überdeckungswahrscheinlichkeit $1-\alpha$ folgendermaßen:

$$\left[\frac{2 \cdot \sum_{i=1}^{n} X_i}{\chi^2_{2n;1-\frac{\alpha}{2}}} ; \frac{2 \cdot \sum_{i=1}^{n} X_i}{\chi^2_{2n;\frac{\alpha}{2}}}\right]$$

Beweis: Ausgehend von

$$W\left\{\chi^2_{2n;\frac{\alpha}{2}} \leq \frac{2 \cdot \sum_{i=1}^{n} X_i}{\tau} \leq \chi^2_{2n;1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

erhält man durch identische Umformung:

$$W\left\{\frac{2 \cdot \sum_{i=1}^{n} X_i}{\chi^2_{2n;1-\frac{\alpha}{2}}} \leq \tau \leq \frac{2 \cdot \sum_{i=1}^{n} X_i}{\chi^2_{2n;\frac{\alpha}{2}}}\right\} = 1-\alpha$$

32.3 t-Verteilung und Gamma-Verteilung

Für die weiteren Überlegungen werden noch zwei Typen von Verteilungen benötigt.
t-Verteilung t_n (Student-Verteilung mit n Freiheitsgraden) $n \in N$
Die Dichtefunktion ist gegeben durch:

$$f(x|n) = k \cdot \frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}} \qquad \forall x \in R$$

Dabei ist k folgendermaßen definiert:

$$k = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n \cdot \pi} \cdot \Gamma\left(\frac{n}{2}\right)}$$

Die Funktion $\Gamma(\cdot)$ ist die **Gammafunktion** und folgendermaßen definiert:

$$\Gamma(z) := \int_{o}^{\infty} t^{z-1} \cdot e^{-t} dt$$

Die Gammafunktion ist tabelliert. Außerdem gilt folgendes:

$$\Gamma(n) = (n-1)! \qquad \forall n \in N$$

$$\Gamma(z) = (z-1) \cdot \Gamma(z-1)$$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

Bemerkung: Für $n \to \infty$ geht die Folge der Dichtefunktion $f(x|n)$ gegen die Standard-Normalverteilungsdichte:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \qquad \forall x \in R$$

Gamma-Verteilung $\gamma(\alpha,\beta)$ $\alpha>0$, $\beta>0$
Die Dichtefunktion ist gegeben durch:

$$f(x|\alpha,\beta)=\frac{x^{\alpha-1}\cdot e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\cdot\beta^{\alpha}}\cdot I_{(0,\infty)}(x) \qquad \text{(vgl. Abschnitt 25.3)}$$

Die Gammafunktion enthält u.a. die Chiquadrat-Verteilung:

$$\gamma\left(\tfrac{n}{2},2\right)=\chi^2_n$$

Die Verteilungsfunktion, die sog. **Unvollständige Gammaverteilung** ist tabelliert.

32.4 Konfidenzintervalle für die Parameter der Normalverteilung

Satz 32.3 Sind die stochastischen Größen $X_1,....,X_n$ unabhängig und identisch nach $N(\mu,\sigma^2)$ verteilt (d.h. eine Stichprobe von $X\sim N(\mu,\sigma^2)$), so gilt:

a) $\sqrt{n}\cdot\frac{\overline{X}_n-\mu}{S_n}\sim t_{n-1}$

b) $\frac{(n-1)\cdot S_n^2}{\sigma^2}\sim\chi^2_{n-1}$

Bemerkung: Die stochastischen Größen in a) und b) sind Funktionen von der Stichprobe und vom unbekannten Parameter. Ihre Verteilung hängt nicht vom unbekannten Parameter ab. Diese beiden stochastischen Größen sind daher Pivot-Größen.

Satz 32.4 Für eine normalverteilte stochastische Größe $X\sim N(\mu,\sigma^2)$ und einer Stichprobe $X_1,....,X_n$ von X erhält man ein Konfidenzintervall für den gerafften Parameter μ bzw. σ^2 folgendermaßen:

a) Konfidenzintervall für μ mit Überdeckungswahrscheinlichkeit $1-\alpha$

$$\left[\overline{X}_n-\frac{S_n}{\sqrt{n}}\cdot t_{n-1;1-\frac{\alpha}{2}},\overline{X}_n+\frac{S_n}{\sqrt{n}}\cdot t_{n-1;1-\frac{\alpha}{2}}\right]$$

b) Konfidenzintervall für σ^2 mit Überdeckungswahrscheinlichkeit $1-\alpha$

$$\left[\frac{(n-1)\cdot S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}},\frac{(n-1)\cdot S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Beweis:

a) Nach Satz 32.3a gilt:

$$\sqrt{n}\cdot\frac{\overline{X}_n-\mu}{S_n}\sim t_{n-1}$$

Das ist eine Pivot-Größe mit dem Parameter $\theta=(\mu,\sigma^2)$. Daher gilt weiters:

$$W_\theta\left\{t_{n-1;\frac{\alpha}{2}}\le\sqrt{n}\cdot\frac{\overline{X}_n-\mu}{S_n}\le t_{n-1;1-\frac{\alpha}{2}}\right\}=1-\alpha \qquad \forall\theta\in\Theta$$

Eine Umformung der Doppelungleichung führt auf:

$$W_\theta\left\{\overline{X}_n-\frac{S_n}{\sqrt{n}}\cdot t_{n-1;1-\frac{\alpha}{2}}\le\mu\le\overline{X}_n-\frac{S_n}{\sqrt{n}}\cdot t_{n-1;\frac{\alpha}{2}}\right\}=1-\alpha$$

Unter Berücksichtigung von $t_{n-1;\frac{\alpha}{2}}=-t_{n-1;1-\frac{\alpha}{2}}$ erhält man:

$$W_\theta\left\{\overline{X}_n-\frac{S_n}{\sqrt{n}}\cdot t_{n-1;1-\frac{\alpha}{2}}\le\mu\le\overline{X}_n+\frac{S_n}{\sqrt{n}}\cdot t_{n-1;1-\frac{\alpha}{2}}\right\}=1-\alpha$$

Und somit ist das Konfidenzintervall für μ mit einer Überdeckungswahrscheinlichkeit 1-α gleich:

$$\left[\overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}, \overline{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right]$$

b) Nach Satz 32.3b gilt:

$$\frac{(n-1) \cdot S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Dies ist ebenfalls eine Pivot-Größe. Daher gilt analog:

$$W_\theta\left\{\chi^2_{n-1;\frac{\alpha}{2}} \le \frac{(n-1) \cdot S_n^2}{\sigma^2} \le \chi^2_{n-1;1-\frac{\alpha}{2}}\right\} = 1-\alpha \quad \forall \theta \in \Theta$$

Umformung der Doppelgleichung führt auf das Konfidenzintervall für σ^2:

$$\left[\frac{(n-1) \cdot S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1) \cdot S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Bemerkung: Für eine konkrete Stichprobe $x_1,....,x_n$ erhält man durch Einsetzen in das Konfidenzintervall ein **konkretes Intervall**.

32.5 Approximative Konfidenzintervalle

Für große Stichprobenumfänge können unter bestimmten Voraussetzungen approximative Konfidenzintervalle berechnet werden.

Satz 32.5 Ist W_θ, $\theta \in \Theta$ ein stochastisches Modell mit $\Theta=(a,b) \subseteq R$ und die Folge (T_n), $n \in N$ eine Schätzfolge für θ_0 mit

$$\frac{T_n - ET_n}{\sqrt{VarT_n}} \underset{n\to\infty}{\overset{Vtlg.}{\to}} N(0,1)$$

und lassen sich die Doppelungleichungen

$$u_{\frac{\alpha}{2}} \le \frac{T_n - E_\theta T_n}{\sqrt{Var_\theta T_n}} \le u_{1-\frac{\alpha}{2}} \qquad u_p\text{: p-Fraktile der Standard-Normalverteilung N(0,1)}$$

äquivalent umformen in die Gestalt $\vartheta_1(T_n,\alpha) \le \theta \le \vartheta_2(T_n, \alpha)$, so strebt die Folge $[\vartheta_1(T_n,\alpha), \vartheta_2(T_n, \alpha)]$ für $n\to\infty$ gegen ein Konfidenzintervall für θ mit Überdeckungswahrscheinlichkeit 1-α.

Begründung:
Aus der Verteilungskonvergenz folgt:

$$W_\theta\left\{u_{\frac{\alpha}{2}} \le \frac{T_n - E_\theta T_n}{\sqrt{Var_\theta T_n}} \le u_{1-\frac{\alpha}{2}}\right\} \xrightarrow[n\to\infty]{} \Phi\left(u_{1-\frac{\alpha}{2}}\right) - \Phi\left(u_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1-\alpha$$

Das heißt, die Überdeckungswahrscheinlichkeit strebt für $n\to\infty$ gegen 1-α.

Beispiel: Anteilsschätzung bei großem Stichprobenumfang

Die stochastische Größe X sei alternativverteilt $X \sim A_\theta$ und $X_1,....,X_n$ sei eine Stichprobe von X und das Mittel $\overline{X}_n$ sei die Schätzfunktion für θ. Dann gilt:

$$\frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} \underset{n\to\infty}{\overset{Vt\,lg.}{\longrightarrow}} N(0,1) \qquad \text{(siehe Zentraler Grenzverteilungssatz)}$$

$$\frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{\left(\frac{1}{n}\right)^2 \cdot Var\left(\sum_{i=1}^{n} X_i\right)}} = \frac{\overline{X}_n - E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\sqrt{Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}} = \frac{\overline{X}_n - E\overline{X}_n}{\sqrt{Var\overline{X}_n}} \underset{n\to\infty}{\overset{Vt\,lg.}{\longrightarrow}} N(0,1)$$

Unter Berücksichtigung von $Var\overline{X}_n = \frac{\theta \cdot (1-\theta)}{n}$ gilt:

$$W_\theta\left\{u_{\frac{\alpha}{2}} \le \frac{\overline{X}_n - \theta}{\sqrt{\frac{\theta \cdot (1-\theta)}{n}}} \le u_{1-\frac{\alpha}{2}}\right\} \approx 1-\alpha$$

Auflösen der Doppelungleichung unter Berücksichtigung von $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$

$$\frac{\left(\overline{X}_n - \theta\right)^2}{\frac{\theta \cdot (1-\theta)}{n}} \le u^2_{1-\frac{\alpha}{2}} \quad \Rightarrow \quad n \cdot \left(\overline{X}_n - \theta\right)^2 \le u^2_{1-\frac{\alpha}{2}} \cdot \theta \cdot (1-\theta)$$

Auflösen der quadratischen Gleichung führt auf:

$$\vartheta_1\left(\overline{X}_n, \alpha\right) = \frac{2 \cdot n \cdot \overline{X}_n + d^2 - d \cdot \sqrt{4 \cdot n \cdot \overline{X}_n + d^2 - 4 \cdot n \cdot \overline{X}_n^2}}{2 \cdot \left(n + d^2\right)} \qquad d = u_{1-\frac{\alpha}{2}}$$

$$\vartheta_1\left(\overline{X}_n, \alpha\right) = \frac{2 \cdot n \cdot \overline{X}_n + d^2 + d \cdot \sqrt{4 \cdot n \cdot \overline{X}_n + d^2 - 4 \cdot n \cdot \overline{X}_n^2}}{2 \cdot \left(n + d^2\right)}$$

Man erhält das Konfidenzintervall $\left[\vartheta_1\left(\overline{X}_n, \alpha\right), \vartheta_2\left(\overline{X}_n, \alpha\right)\right]$. Nun kürzt man durch 2·n und vernachlässigt die Glieder konst/n und erhält dann als Näherung für das approximative Konfidenzintervall mit Überdeckungswahrscheinlichkeit 1-α:

$$\left[\overline{X}_n - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\overline{X}_n \cdot \left(1 - \overline{X}_n\right)}{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\overline{X}_n \cdot \left(1 - \overline{X}_n\right)}{n}}\right]$$

33. Statistische Tests

Für gegebene Beobachtungen $x_1,...,x_n$ einer stochastischen Größe X möchte man möglichst objektiv entscheiden, ob gewisse Annahmen über die Wahrscheinlichkeitsverteilung W_X von X (sog. statische Hypothese) haltbar sind. Formale Entscheidungsverfahren dieses Art nennt man **statistische Tests**.
Ist W={W:W mögliche Verteilungen von X} die Menge aller in Frage kommenden Wahrscheinlichkeitsverteilungen, so heißt jede Teilmenge $\mathsf{H}\subseteq\mathsf{W}$ eine **statistische Hypothese**.

Beispiel: Sei X eine 1-dimensionale stochastische Größe mit kontinuierliches Verteilung, d.h. W={W:W kontinuierliche Verteilung auf (R,B)}, so ist $\mathsf{H}=\{N(\mu,\sigma^2):\mu, \sigma^2$ beliebig} eine statistische Hypothese.

Eine Hypothese heißt **einfach**, falls sie nur aus einer Wahrscheinlichkeitsverteilung besteht, d.h. $\mathsf{H}:X\sim W_\theta$ mit W_θ bekannt. Ansonsten heißen Hypothesen **zusammengesetzt**. Falls parametrische Familien von Wahrscheinlichkeitsverteilungen betrachtet werden, spricht man von **Parameter-Hypothesen**.

Beispiel:
Kann von einer stochastischen Größe X angenommen werden, daß sie eine Poisson-Verteilung (Poisson-Prozeß) hat, d.h. $X\sim P_\theta$, $\theta>0$, $\Theta=(0,\infty)$, so gilt: $\mathsf{W}=\{P_\theta:\theta>0\}$. Eine einfache Hypothese wäre z.B.:

$$\mathsf{H}=\{P_{3.2}\} \quad \text{oder} \quad \mathsf{H}:X\sim P_{3.2} \quad \text{oder} \quad \mathsf{H}:\theta=3.2$$

Eine zusammengesetzte Hypothese wäre z.B.:

$$\mathsf{H}=\{P_\theta:\theta\geq 3\} \quad \text{oder} \quad \mathsf{H}:\theta\geq 3$$

Manchmal werden auch Alternativen (sog. **Gegenhypothesen**) H_1 formuliert. Für die ursprüngliche Hypothese schreibt man denn H_0 (sog. **Nullhypothese**).

Beispiel: Die Lebensdauer X habe eine Exponentialverteilung Ex_τ.
Einfache Nullhypothese: $\mathsf{H}_0:\tau=2$
Zusammengesetzte Gegenhypothese: $\mathsf{H}_1: \tau> 2$

33.1 Statistische Tests und Verwerfungsräume

Die Entscheidung, ob eine Hypothese H bezüglich einer stochastischen Größe X angenommen werden kann oder abzulehnen ist, hängt von der Stichprobe $x_1,....,x_n$ von X ab. Daher wird eine Teilmenge V des Stichprobenraumes M_X^n gesucht, so daß Stichproben aus V gegen die Hypothese sprechen. Die Teilmenge V wird **Verwerfungsraum** genannt. Für $(x_1,...,x_n)\in V$ wird H verworfen, für $(x_1,...,x_n)\notin V$, d.h. $(x_1,...,x_n)\in V^c=M_X^n\setminus V$ wird die Hypothese angenommen.
Oft läßt sich die Entscheidung über die Annahme oder Ablehnung einer Hypothese mit Hilfe einer Statistik $T=t(X_1,...,X_n)$ treffen. Dann nennt man T eine **Teststatistik**.

33.2 Fehler erster und zweiter Art

Für ein Testproblem gibt es zwei Fehlerarten bei der Entscheidung. Entweder die Hypothese H_0 ist richtig und wird verworfen (**Fehler erster Art**) oder H_0 ist falsch und wird angenommen (**Fehler zweiter Art**). Um gute Testentscheidungen zu konstruieren, versucht man die Wahrscheinlichkeiten von Fehlern klein zu halten. Symbolisch schreibt man:

$\alpha=W\{\mathsf{H}_0 \text{ verworfen} | \mathsf{H}_0 \text{ richtig}\}$ Wahrscheinlichkeit für einen Fehler erster Art
$\beta=W\{\mathsf{H}_0 \text{ angenommen} | \mathsf{H}_0 \text{ falsch}\}$ Wahrscheinlichkeit für einen Fehler zweiter Art

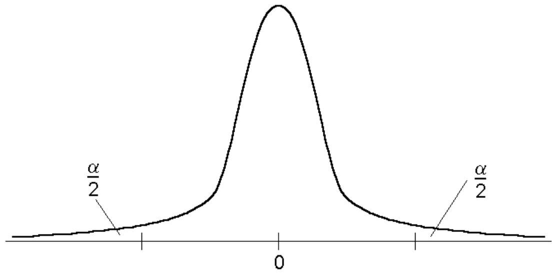
Wünschenswert wären Tests, die beide Fehlerwahrscheinlichkeiten möglichst klein machen, sog. **schärfste/beste Tests**. Leider existieren solche Tests nur selten.
Für einfache Nullhypothesen $\mathsf{H}_0:X\sim W_0$ (W_0 ist eine feste, bekannte Wahrscheinlichkeitsverteilung) gilt:

$$\alpha=W\{(X_1,...,X_m)\in V\}$$

Beispiel: Die stochastische Größe X sei normalverteilt $X \sim N(\mu,\sigma_0^2)$, wobei der Parameter σ_0^2 bekannt ist. Gegeben ist eine Stichprobe $X_1,\dots,X_n$ von X und die Hypothese $H_0{:}\mu=\mu_0$, wobei μ_0 gegeben ist. Gesucht ist nun ein Test für einen Fehler erster Art.

Bei richtiger Hypothese H_0 gilt:

$$\overline{X}_n \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right) \Rightarrow T = \frac{\overline{X}_n - \mu_0}{\sigma_0 / \sqrt{n}} \sim N(0,1)$$



Um einen Test für H_0 zu konstruieren, verwendet man die Teststatistik T, denn für richtige H_0 gilt:

$$W_{\mu_0}\left\{u_{\frac{\alpha}{2}} \le T \le u_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

Ein Test für $H_0{:}\mu=\mu_0$ mit Wahrscheinlichkeit α eines Fehlers erster Art ist: Verwerfe die Hypothese H_0 für $T \notin [u_{\alpha/2}, u_{1-\alpha/2}]$. Der Verwerfungsraum (unter Berücksichtigung der Symmetrie der Normalverteilung) dieses Tests ergibt sich daher durch:

$$V = \left\{(x_1,\dots,x_n): \left|\frac{\overline{x}_n - \mu_0}{\sigma_0/\sqrt{n}}\right| \ge u_{1-\frac{\alpha}{2}}\right\}.$$

Begründung: Wegen obiger Gleichung für $W_{\mu 0}$ fällt die Teststatistik T bei richtiger Hypothese H_0 mit der kleinen Wahrscheinlichkeit α nicht in das Intervall $\left[u_{\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}\right]$ (dies ist die Wahrscheinlichkeit der Verwerfung).

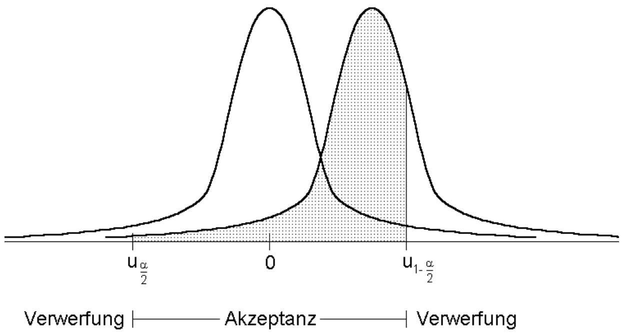
Bemerkung: Die Wahrscheinlichkeit eines Fehlers zweiter Art kann hier sehr groß werden, wenn das tatsächliche μ nahe bei μ_0 liegt aber $\mu_0 \ne \mu$ ist.

Im letzten Beispiel gilt, im Falle, daß der wahre Parameter μ_1 ist und $\mu_1 > \mu_0$, für die Wahrscheinlichkeit eines Fehlers zweiter Art:

W{ $H_0{:}\mu=\mu_0$ angenommen| H_0 falsch, da $\mu=\mu_1$}

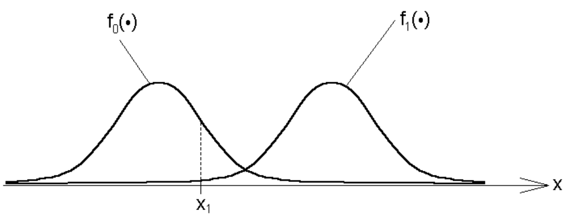
$$T = \frac{\overline{X}_n - \mu_0}{\sigma_0/\sqrt{n}} \sim N\left(\frac{\mu_1-\mu_0}{\sigma_0/\sqrt{n}}, 1\right)$$

Die schraffierte Fläche in nebenstehender Abbildung zeigt die Wahrscheinlichkeit für die Annahme der falschen Hypothese.



33.3 Plausibilitätsquotienten-Test

Gegeben sind zwei Hypothesen für die stochastische Größe X. Die erste Hypothese besagt $H_0{:}X \sim f_0(\cdot)$ und die zweite lautet $H_1{:}X \sim f_1(\cdot)$. Der Plausibilitätsgedanke bei einer Beobachtung ist nun, daß man diejenige Hypothese wählt, deren Dichtefunktion den größeren Wert aufweist, d.h.:



- Annahme von H_0 falls $f_1(x_1) < f_0(x_1)$
- Annahme von H_1 falls $f_1(x_1) > f_0(x_1)$

Im Fall von n Beobachtungen zieht man die gemeinsame Dichte der Stichprobe $X_1,\dots,X_n$ heran:

$$g_j(x_1,\dots,x_n) = \prod_{i=1}^{n} f_j(x_i) = l_j(x_1,\dots,x_n) \qquad \text{für } j = 0 \text{ oder } j = 1$$

Bemerkung: Für parametrische Modelle ist $f_j(\cdot) = f(\cdot|\theta_j)$ (Plausibilitätsfunktion).

Definition: Ein **Plausibilitätsquotienten-Test** für die einfache Hypothese H_0:X~$f_0(\cdot)$ gegen die einfache H_1:X~$f_1(\cdot)$ ist ein Test, dessen Verwerfungsraum folgendermaßen gegeben ist:

$$V = \left\{ (x_1, \ldots, x_n) : \frac{l_1(x_1, \ldots, x_n)}{l_0(x_1, \ldots, x_n)} \geq k_\alpha \right\}$$

Dabei ist die Konstante k_α durch die Wahrscheinlichkeit α eines Fehlers erster Art bestimmt wird, d.h.:

$$W_0\left\{ \frac{l_1(X_1, \ldots, X_n)}{l_0(X_1, \ldots, X_n)} \geq k_\alpha \right\} = \alpha$$

Der Quotient $\lambda(x_1, \ldots, x_n) = \dfrac{l_1(x_1, \ldots, x_n)}{l_0(x_1, \ldots, x_n)}$ heißt **Plausibilitätsquotient**.

Beispiel: Angenommen für die Verteilung einer stochastischen Größe kommt eine Normalverteilung X~N(μ,σ^2) in Frage, entweder mit dem Parameter μ_0 oder dem Parameter $\mu_1>\mu_0$, wobei σ^2 bekannt ist. In diesem Fall gilt:

$$f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\cdot\frac{(x-\mu_0)^2}{\sigma^2}} \qquad f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\cdot\frac{(x-\mu_1)^2}{\sigma^2}}$$

$$l_0(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \cdot \exp\left\{ -\frac{1}{2\cdot\sigma^2} \cdot \sum_{i=1}^n (x_i - \mu_0)^2 \right\}$$

$$l_1(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \cdot \exp\left\{ -\frac{1}{2\cdot\sigma^2} \cdot \sum_{i=1}^n (x_i - \mu_1)^2 \right\}$$

$$\lambda(x_1, \ldots, x_n) = \frac{l_1(x_1, \ldots, x_n)}{l_0(x_1, \ldots, x_n)} = \exp\left\{ \frac{1}{2\cdot\sigma^2} \cdot \left[\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \mu_1)^2 \right] \right\}$$

$$\lambda(x_1, \ldots, x_n) \begin{Bmatrix} \geq \\ < \end{Bmatrix} k_\alpha \Leftrightarrow \ln\lambda(x_1, \ldots, x_n) \begin{Bmatrix} \geq \\ < \end{Bmatrix} \underbrace{\ln k_\alpha}_{k'_\alpha} \quad \Rightarrow$$

$$\sum_{i=1}^n \left\{ (x_i - \mu_0)^2 - (x_i - \mu_1)^2 \right\} \begin{Bmatrix} \geq \\ < \end{Bmatrix} \underbrace{2\cdot\sigma^2\cdot k'_\alpha}_{k''_\alpha} \quad \Rightarrow$$

$$n\cdot(\mu_0^2 - \mu_1^2) + 2\cdot(\mu_1 - \mu_0)\cdot\sum_{i=1}^n x_i \begin{Bmatrix} \geq \\ < \end{Bmatrix} k''_\alpha \quad \Rightarrow$$

$$2\cdot(\mu_1 - \mu_0)\cdot\sum_{i=1}^n x_i \begin{Bmatrix} \geq \\ < \end{Bmatrix} k'''_\alpha \quad \Rightarrow$$

$$\sum_{i=1}^n x_i \begin{Bmatrix} \geq \\ < \end{Bmatrix} k^*$$

Die Entscheidung ist hier abhängig von der Teststatistik $T = \sum_{i=1}^n X_i$. Die Verteilung von T kann man angeben:

$$\sum_{i=1}^n X_i \sim N(n\cdot\mu_0, n\cdot\sigma^2)$$

Zur Bestimmung der Konstanten k_α* zieht man die Wahrscheinlichkeit α eines Fehlers erster Art heran (d.h. die Wahrscheinlichkeit der Verwerfung der richtigen Hypothese). Es gilt:

$$\alpha = W_0\left\{\sum_{i=1}^{n} X_i \geq k_\alpha^*\right\} \quad \text{wobei } k_\alpha^* \text{ das } (1-\alpha)\text{-Fraktile der } N(n\cdot\mu_0, n\cdot\sigma^2).$$

Wegen des Zusammenhanges zwischen p-Fraktilen x_p einer $N(\mu,\sigma^2)$-Verteilung und dem p-Fraktile u_p der $N(0,1)$-Verteilung gilt in unserem Fall:

$$u_p = \frac{x_p - \mu}{\sigma} \;\Rightarrow\; u_p = \frac{k_\alpha^* - n\cdot\mu_0}{\sigma\cdot\sqrt{n}} \;\Rightarrow\; k_\alpha^* = \sigma\cdot\sqrt{n}\cdot u_{1-\alpha} + n\cdot\mu_0$$

Der Verwerfungsraum ist daher gegeben durch:

$$V = \left\{(x_1,....,x_n): \sum_{i=1}^{n} x_i \geq \sigma\cdot\sqrt{n}\cdot u_{1-\alpha} + n\cdot\mu_0\right\}$$

Für den Stichprobenraum gilt $M_X = V \cup V^c$. Die Hypothese wird verworfen, falls die konkrete Stichprobe in den Verwerfungsraum fällt, ansonsten wird sie angenommen. Oft wird die Entscheidung ob verwerfen oder annehmen vom Wert abhängig gemacht, den eine sog. **Teststatistik** annimmt.

Bemerkung: Für den letzten Test ist bei gegebener Wahrscheinlichkeit α eines Fehlers 1. Art die Wahrscheinlichkeit eines Fehlers 2. Art minimal. Das folgt aus einem grundlegendem Satz von Neyman und E. Pearson.

34. Tests für Normalverteilungen

In diesem Abschnitt wird ausnahmslos vorausgesetzt, daß die betrachteten stochastischen Größen eine Normalverteilung haben. Da im allgemeinem die Parameter unbekannt sind, benötigt man Tests, die dies berücksichtigen.

34.1 t-Test für das Mittel einer Normalverteilung

Prüfung!

Satz 34.1 Ist X normalverteilt $X \sim N(\mu,\sigma^2)$ mit unbekanntem μ und σ^2 und $x_1,...,x_n$ ist eine konkrete Stichprobe von X, so ist ein Test für die Hypothese $H_0: \mu = \mu_0$ mit Wahrscheinlichkeit α eines Fehlers 1.Art durch folgenden Verwerfungsraum gegeben:

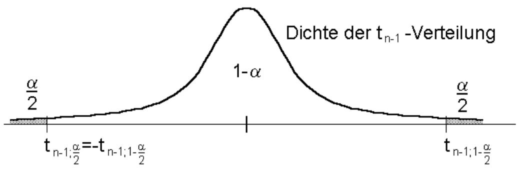
$$V = \left\{(x_1,...,x_n): \frac{\sqrt{n}\cdot\left|\overline{x}_n - \mu_0\right|}{s_n} \geq t_{n-1;1-\frac{\alpha}{2}}\right\}$$

Die Teststatistik T ist in diesem Fall gegeben durch:

$$T = \frac{\sqrt{n}\cdot\left|\overline{X}_n - \mu_0\right|}{S_n}$$

Beweis: Wenn $X \sim N(\mu_0,\sigma^2)$ ist, gilt wegen Satz 32.3:

$$\frac{\sqrt{n}\cdot\left(\overline{X}_n - \mu_0\right)}{S_n} \sim t_{n-1} \quad \text{(bei richtiger Nullhypothese)}$$



Wegen der Symmetrie der Dichtefunktion der t_{n-1}-Verteilung gilt für die Wahrscheinlichkeit α eines Fehlers 1.Art:

$$W\{\text{Verwerfung der richtigen } H_0\} = W\{\text{Fehler 1.Art}\} = W_{\mu_0,\sigma^2}\left\{\frac{\sqrt{n}\cdot\left|\overline{X}_n - \mu_0\right|}{S_n} \geq t_{n-1;1-\frac{\alpha}{2}}\right\} = \alpha$$

Der **kritische Bereich** für die Teststatistik T ist $(-\infty, -t_{n-1;1-\alpha/2}] \cup [t_{n-1;1-\alpha/2}, \infty)$. Dieser ist eine Teilmenge des Merkmalraums der Teststatistik (meist eine Teilmenge von R). Wenn die konkrete Teststatistik im kritischen Bereich liegt, so führt dies zu Ablehnung der Nullhypothese. Der Verwerfungsraum ist dagegen eine Teilmenge des Stichprobenraums. Die Ablehnung erfolgt, wenn die konkrete Stichprobe im Verwerfungsraum liegt.

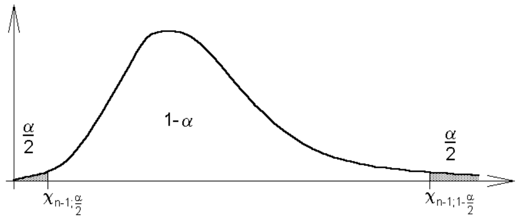
34.2 Test für die Varianz einer Normalverteilung

Satz 34.2 Ist X normalverteilt $X \sim N(\mu,\sigma^2)$ mit unbekanntem μ und σ^2 und ist $x_1,....,x_n$ eine konkrete Stichprobe von , so ist ein Test für die Hypothese $H_0:\sigma^2=\sigma_0^2$ mit Wahrscheinlichkeit α eines Fehlers 1. Art gegeben durch den Verwerfungsraum:

$$V = \left\{ (x_1,....,x_n) : \frac{(n-1)\cdot s_n^2}{\sigma_0^2} \notin \left[\chi^2_{n-1;\frac{\alpha}{2}}, \chi^2_{n-1;1-\frac{\alpha}{2}} \right] \right\}$$

Beweis: Wenn $X \sim N(\mu,\sigma_0^2)$ ist, dann folgt aus Satz 32.3:

$$\frac{(n-1)\cdot S_n^2}{\sigma_0^2} \sim \chi^2_{n-1}$$



Für die Wahrscheinlichkeit eines Fehlers 1.Art gilt:

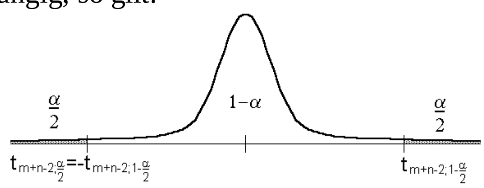
$$W\left\{ \frac{(n-1)\cdot S_n^2}{\sigma_0^2} \notin \left[\chi^2_{n-1;\frac{\alpha}{2}}, \chi^2_{n-1;1-\frac{\alpha}{2}} \right] \right\} = \alpha$$

Bemerkung: Die beiden vorangegangenen Tests behandeln das Problem der Verteilung einer Meßreihe. Hat man Daten aus zwei unabhängigen Versuchsreihen, so soll entschieden werden, ob diese zwei Meßreihen $x_1,...,x_m$ bzw. $y_1,...,y_n$ derselben Verteilung entstammen. Diese Fragestellung heißt **Zwei-Stichproben-Problem**.

34.3 t-Test für die Gleichheit zweier Normalverteilungen mit identischer Varianz

Satz 34.3 Ist $X_1,...,X_m$ eine Stichprobe einer $N(\mu_x,\sigma^2)$-Verteilung und ist $Y_1,...,Y_n$ eine Stichprobe einer $N(\mu_y,\sigma^2)$-Verteilung und sind beide Stichproben voneinander unabhängig, so gilt:

$$Z = \frac{\overline{X}_m - \overline{Y}_n - \mu_x - \mu_y}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\cdot\frac{(m-1)\cdot S_x^2 + (n-1)\cdot S_y^2}{m+n-2}}} \sim t_{m+n-2}$$



Ein Test für die Hypothese $H_0:\mu_x=\mu_y$ mit Wahrscheinlichkeit α eines Fehlers 1. Art ist durch die Teststatistik Z gegeben.

Die $H_0:\mu_x=\mu_y$ wird verworfen, falls für die zwei Meßreihen $x_1,...,x_m$ und $y_1,...,y_n$ gilt:

$$|z| = \left| \frac{\overline{x}_m - \overline{y}_n}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\cdot\frac{(m-1)\cdot s_x^2 + (n-1)\cdot s_y^2}{m+n-2}}} \right| \geq t_{m+n-2;1-\frac{\alpha}{2}}$$

Beweis: Bei richtiger Nullhypothese H_0 gilt:

$$W\left\{ -t_{m+n-2;1-\frac{\alpha}{2}} \leq Z \leq t_{m+n-2;1-\frac{\alpha}{2}} \right\} = 1-\alpha$$

und für die Verwerfungsraumwahrscheinlichkeit gilt daher:

$$W\left\{ Z \notin \left(-t_{m+n-2;1-\frac{\alpha}{2}}, t_{m+n-2;1-\frac{\alpha}{2}} \right) \right\} = \alpha$$

34.4 F-Test für die Gleichheit der Varianzen zweier Normalverteilungen

Satz 34.4 Ist $X_1,...,X_m$ eine Stichprobe einer $N(\mu_x,\sigma_x^2)$ und $Y_1,...,Y_n$ eine Stichprobe einer $N(\mu_y,\sigma_y^2)$ und sind die beiden Stichproben voneinander unabhängig, so gilt:

$$\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F_{m-1,n-1}$$

Dabei bedeutet $F_{m,n}$ die **F-Verteilung** mit m und n Freiheitsgraden. Diese Wahrscheinlichkeitsverteilung auf R^+ hat die Dichtefunktion:

$$f(x|m,n) = k \cdot \frac{x^{\frac{m}{2}-1}}{(m \cdot x + n)^{\frac{m+n}{2}}} \cdot I_{(0,\infty)}(x) \qquad \text{mit } k = \frac{m^{\frac{m}{2}} \cdot n^{\frac{n}{2}} \cdot \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)}$$

Diese Verteilung ist ausführlich tabelliert.

Ein Test für die Hypothese $H_0:\sigma_x^2=\sigma_y^2$ mit Wahrscheinlichkeit α eines Fehlers 1. Art ist mit Hilfe der Teststatistik Z

$$Z = \frac{S_x^2}{S_y^2}$$

folgendermaßen gegeben. Die Nullhypothese H_0 wird verworfen, falls für die konkreten Meßreihen $x_1,...,x_m$ und $y_1,...,y_n$ gilt:

$$z \notin \left[F_{m-1,n-1,\frac{\alpha}{2}}, F_{m-1,n-1,1-\frac{\alpha}{2}}\right] \qquad \text{wobei } F_{m,n;p} \text{ das p-Fraktile der } F_{m,n} \text{ ist.}$$

Bemerkung: Um die Gleichheit der Varianzen mehrerer Meßreihen zu testen, gibt es Verfahren der sog. „**Varianz-Analyse**“.

34.5 Unabhängigkeit für 2-dimensionale Normalverteilungen

Prüfung: Zusammenhang zwischen Unabhängigkeit und Unkorreliertheit

Der stochastische Vektor (X,Y) habe die Verteilung $(X,Y)\sim N(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\varphi)$. Die stochastischen Größen X und Y seien unabhängig und $\varphi=0$. Gesucht ist ein Test, für die Unabhängigkeit von X und Y.

Der Stichproben-Korrelationskoeffizient R (vgl. 30.6) ist der plausible Schätzwert für φ. R ist als Schätzfunktion eine stochastische Größe. Zur Konstruktion eines Tests für $\varphi=0$, was im Fall von Normalverteilungen auch Unabhängigkeit bedeutet, ist folgender Satz wichtig:

Satz 34.5 Unter den Voraussetzungen dieses Abschnittes gilt unter der Bedingung $\varphi=0$ für n Beobachtungen:

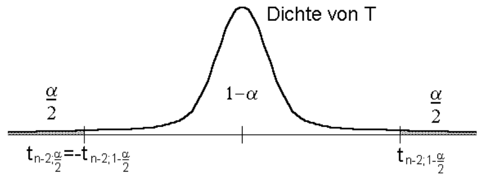
$$\frac{\sqrt{n-2} \cdot R}{\sqrt{1-R^2}} \sim t_{n-2}$$

Ein Test für die Unabhängigkeit der Komponenten X und Y eines 2-dimensionalen normalverteilten stochastischen Vektors (X,Y) mit Irrtumswahrscheinlichkeit α ist folgendermaßen gegeben: Die Hypothese $H_0:\varphi=0$ wird verworfen, falls für die konkrete Stichprobe $(x_1,y_1),...,(x_n,y_n)$ gilt:

$$\left|\frac{\sqrt{n-2} \cdot r}{\sqrt{1-r^2}}\right| \geq t_{n-2;1-\frac{\alpha}{2}}$$

Beweis: Bei Normalverteilungen bedeutet Unabhängigkeit auch Unkorreliertheit.

Nach Satz 34.5 gilt bei richtiger Nullhypothese H_0:



$$\frac{\sqrt{n-2}\cdot R}{\sqrt{1-R^2}} \sim t_{n-2}$$

Daher gilt bei richtige Nullhypothese H_0:

$$W\left\{-t_{n-2;1-\frac{\alpha}{2}} \le \frac{\sqrt{n-2}\cdot R}{\sqrt{1-R^2}} \le t_{n-2;1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

Die Verwerfungswahrscheinlichkeit der Nullhypothese H_0 bei richtiger H_0 ist daher α.

35. Chiquadrat-Anpassungs-Test

Hat man eine größere Stichprobe einer stochastischen Größe X und möchte testen, ob die Wahrscheinlichkeitsverteilung von X eine ganz bestimmte ist, so ist der χ^2-Test für einfache Hypothese geeignet.

35.1 Chiquadrattest für einfache Hypothesen

Satz 35.1 Ist $X_1,\ldots,X_n$ eine Stichprobe einer stochastischen Größe X mit der Wahrscheinlichkeitsverteilung W auf dem Merkmalraum (M,ξ) und $E_1,\ldots,E_r$ eine Zerlegung des Merkmalraumes in Ereignisse ($E_i \cap E_j = 0$, $i \neq j$), so bezeichnet w_j die Wahrscheinlichkeit, daß X in E_j fällt und Y_j die absoluten Häufigkeiten des Ereignisses E_j nach n Beobachtungen von X. Mit diesen Bezeichnungen gilt:

$$\lim_{n\to\infty} W\left\{ Z_n = \sum_{j=1}^{r} \frac{\left(Y_j - n\cdot w_j\right)^2}{n\cdot w_j} \le \chi^2_{r-1;p} \right\} = p \quad \forall p \in (0,1)$$

Das heißt, daß Z_n asymptotisch nach χ^2_{r-1} verteilt ist.

Mit den Bezeichnung aus Satz 35.1 ist ein Test für die Nullhypothese H_0:X~W durch folgenden Verwerfungsraum gegeben:

$$V = \left\{(x_1,\ldots,x_n): z_n \ge \chi^2_{r-1;1-\alpha}\right\}$$

Die Wahrscheinlichkeit für einen Fehler erster Art konvergiert für $n\to\infty$ gegen α (sog. **approximativer Test**).

Bemerkung: Damit die Approximation hinreichend genau ist, soll für die Zerlegung des Merkmalraumes in die Ereignisse $E_1,\ldots,E_r$ gelten: $n\cdot w_j \ge 5 \quad \forall j=1(1)r$

35.2 Chiquadrattest für zusammengesetzte Hypothesen

Soll für eine Stichprobe $X_1,\ldots,X_n$ von X getestet werden, ob die Verteilung von X aus einer parametrischen Familie $W_\theta = F(x|\theta)$ mit dem wahren Parameter θ_0 ist, i.Z. H_0:X~W_θ, wobei θ ein s-dimensionaler Parameter ist, so gilt folgender Satz:

Satz 35.2 Es gelten die Bezeichnungen von Satz 35.1, aber X~W_θ mit $\theta \in \Theta \subseteq R^s$, wobei dieses Θ ein offenes, nicht ausgeartetes Intervall des R^s ist. Weiters gelte die Bezeichnung $w_j(\theta) = W_\theta(E_j)$, j=1(1)r. Dann gilt mit dem plausiblen Schätzwert für den Parameter θ_0:

$$Z_n = \sum_{j=1}^{r} \frac{\left[Y_j - n\cdot w_j\left(\hat{\theta}_n\right)\right]^2}{n\cdot w_j\left(\hat{\theta}_n\right)} \quad \text{ist asymptotisch verteilt nach } \chi^2_{r\text{-}s\text{-}1}$$

Dabei bezeichnet $\hat{\theta}_n$ den plausiblen Schätzwert für den Parameter θ_0.

Mit obigen Bezeichnungen ist ein Test für die Hypothese H_0:X~W_θ mit $\theta \in \Theta$, dessen Wahrscheinlichkeit für einen Fehler erster Art mit $n\to\infty$ gegen α konvergiert, durch folgenden Verwerfungsraum gegeben.

$$V = \left\{(x_1,\ldots,x_n): z_n \ge \chi^2_{r-s-1;1-\alpha}\right\}$$

Beispiel: Es ist zu testen, ob die täglichen Anzahl von Ausfällen je Tablettenpresse eines pharmazeutischen Betriebes Poisson-verteilt ist. Dazu wurden zwanzig Pressen fünfzig Tage lang beobachtet.

Anzahl der Ausfälle	Absolute Häufigkeit	Absolute Häufigkeiten der Klassen
0	305	E1: 305
1	365	E2: 365
2	210	E3: 210
3	80	E4: 80
4	28	E5: 28
5	10	
6	1	E6: 12
≥7	1	

Der Wert für z_n ergibt sich zu 3,31 (abhängig von der Klasseneinteilung!). Der Wert des p-Fraktiles der $\chi^2_{4;0,95}$-Verteilung ist gleich 9,49. Daher wird die Hypothese angenommen.

36. Kolmogoroff-Smirnov-Test

Falls die Annahme einer parametrischen Verteilungsfamilie für eine stochastische Größe nicht gerechtfertigt ist, hilft bei *kontinuierlichen* Verteilungen folgender Satz, um eine Entscheidung über die Gleichheit der Verteilungen zweier Stichproben zu treffen.

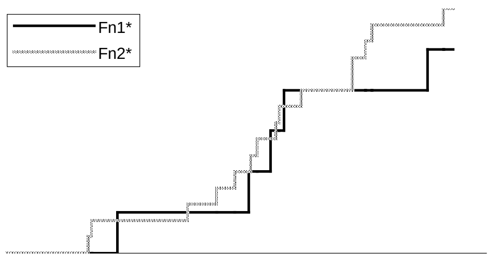
Satz 36.1 Sind $X_{11},...,X_{1n1}$ und $X_{21},...,X_{2n2}$ zwei voneinander unabhängige Stichproben einer kontinuierlichen Verteilung auf R und sind F^*_{n1} und F^*_{n2} die entsprechenden empirischen Verteilungsfunktionen, so gilt:

$$n := \frac{n_1 \cdot n_2}{n_1 + n_2}$$

$$\lim_{\substack{n_1 \to \infty \\ n_2 \to \infty}} W\left\{\sqrt{n} \cdot \max_{x \in R}\left|F^*_{n_1}(x) - F^*_{n_2}(x)\right| \leq z\right\} = \sum_{k=\infty}^{\infty} (-1)^k \cdot e^{-2 \cdot k^2 \cdot z^2} = H(z) \quad \forall z > 0$$

Die Funktion H und ihre Fraktilen sind tabelliert.
Vor Erhebung der Daten ist F^*_n eine stochastische Größe.

$$F^*_n(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$



Mit den Bezeichnungen von Satz 36.1 kann ein Test für die Gleichheit der Verteilungen von zwei unabhängigen Stichproben mittels der Teststatistik

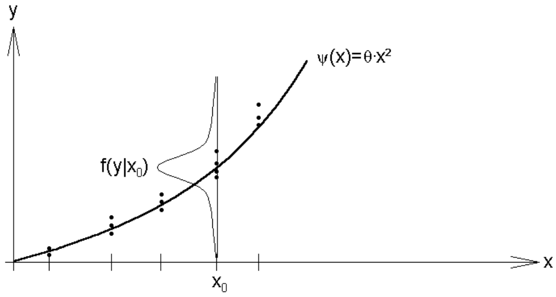
$$D_n = \sqrt{n} \cdot \max_{x \in R}\left|F^*_{n_1}(x) - F^*_{n_2}(x)\right|$$

konstruiert werden. Verwirft man die Hypothese der Gleichheit der Verteilungen für $d_n \geq k_{1-\alpha}$, wobei $k_{1-\alpha}$ das 1-α-Fraktile der H-Verteilung ist, so ist dies ein Test für die Gleichheit der beiden Verteilungen, dessen Irrtumswahrscheinlichkeit für $n_1 \to \infty$ und $n_2 \to \infty$ gegen α konvergiert. Für kleinere n gibt es adaptierte Tabellen.

37. Klassische Regressionsrechnung

Zu Beschreibung kausaler, aber nicht deterministischer Zusammenhänge (also stochastischer Zusammenhänge), verwendet man stochastische Modelle.

Beispiel: Bremsweg in Abhängigkeit der Geschwindigkeit



Stochastisches Modell: Jedem x wird eine stochastische Größe Y_x zugeordnet, deren Wahrscheinlichkeitsverteilung außer von x auch von einem (ein- oder mehrdimensionalen) Parameter θ abhängen kann. Die Zusammenhangskurve

$$\psi(x) = E_\theta Y_x$$

heißt **Regressionsfunktion** von Y bezüglich x. Die stochastische Größe Y_x ist zusammengesetzt:

$$Y_x = y(x) + U_x$$

wobei $\psi(x)$ eine (deterministische) Funktion ist und U_x eine stochastische Größe mit $Eu_x=0$.

Beispiele für Regressionsfunktionen

1) Regressionsgeraden

$Y_x=\alpha+\beta\cdot x+U_x \quad$ mit $x\in[a,b]$

2) Polynomische Regression (Sonderfall der linearen Regression)

$$Y_x = \theta_0 + \theta_1 \cdot x + \theta_2 \cdot x^2 + \ldots + \theta_k \cdot x^k \quad \text{mit } x \in [a,b]$$

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

3) Multiple Regression

Falls die unabhängige Variable $\underline{x}$ mehrdimensional ist, spricht man von multipler Regression.

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \quad \text{und} \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

$$E_\theta Y_{\underline{x}} = \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \ldots + \theta_k \cdot x_k$$

Vektorschreibweise: $Y_{\underline{x}} = \underline{x} \cdot \theta + U_{\underline{x}}$

Ist die abhängige Größe $Y_{\underline{x}}$ auch mehrdimensional, so spricht man von **multivarianter Regression**. Der Parameter θ muß aus den Daten $(x_{i1},\ldots,x_{ik};y_i)$; i=1(1)n geschätzt werden.

Definition: Eine Regressionsfunktion $\psi(x)$ heißt **linear**, falls sie in den Parametern linear ist. Die Regressionsfunktionen in den vorangehenden Beispielen waren alle linear.

37.1 Regressionsgeraden

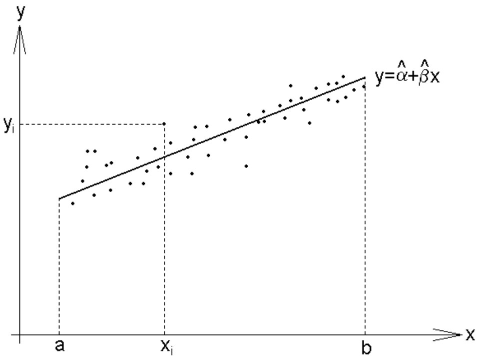
Für diesen Fall gilt:

$$Y_x = \alpha + \beta \cdot x + U_x \qquad x \in R \qquad \theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \alpha, \beta \in R$$

Für konkrete Daten $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$ muß man α und β schätzen. Ohne stochastische Modell könnte man die Methode der kleinsten Abstandsquadratsumme heranziehen und so eine **Ausgleichsgerade** $y = \hat{\alpha} + \hat{\beta} \cdot x$ ermitteln. Die Lösung des Ausgleichproblems

$$\sum_{i=1}^{n} (y_i - \alpha - \beta \cdot x) \rightarrow \min$$

liefert eine „Schätzung“ der Parameters θ, die auch statistisch optimal ist (vgl. Satz 37.1).



Definition: Eine Schätzfunktion T_j für den Parameter θ_j in einem Regressionsmodell heißt **linear**, wenn

$$T_j = \sum_{i=1}^{n} c_{ji} \cdot Y_i$$

wobei die c_{ji} nur von den bekannten x_i abhängen. Im Spezialfall der Regressionsgeraden bedeutet dies für die Schätzfunktion A und B für α und β:

$$A = \sum_{i=1}^{n} c_i \cdot Y_i \qquad B = \sum_{i=1}^{n} d_i \cdot Y_i$$

Satz von Gauß-Markoff

Satz 37.1 Sind die stochastischen Größen $Y_1,...,Y_n$ unkorreliert und gilt:

$$EY_i = \alpha + \beta \cdot x_i \quad \text{mit} \quad VarY_i = \sigma^2$$

wobei die Werte $x_1,..,x_n$ bekannt und nicht alle gleich sind und die Parameter α, β und σ^2 unbekannt sind, so gilt:

$$A = \frac{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i \cdot Y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad \text{und} \qquad B = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot Y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} Y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Dies sind sowohl die effizienten, linearen Schätzfunktionen für die Parameter α und β als auch für beobachtete Werte $y_1,....,y_n$ die Lösung des Ausgleichsproblems bezüglich minimaler Abstandsquadratsumme. Für diese Schätzfunktionen gilt weiters:

$$VarA = \frac{\sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} \cdot \sigma^2 \qquad \text{und} \qquad VarB = \frac{n}{n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} \cdot \sigma^2$$

$$Cov(A,B) = -\frac{\sum_{i=1}^{n} x_i}{n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} \cdot \sigma^2$$

Eine unverzerrte Schätzfolge für σ^2 ist gegeben durch:

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - A - B \cdot x_i)^2}{n-2}$$

Bemerkung: Mit den Bezeichnungen:

$$\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \qquad \underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad \kappa = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

gilt in Matrixschreibweise für die Schätzfunktionen folgendes:

$$\hat{\theta} = \begin{pmatrix} A \\ B \end{pmatrix} = \left(\kappa^T \cdot \kappa\right)^{-1} \cdot \kappa^T \cdot \underline{Y}$$

Der Wert EY_x der Regressionsfunktion an der Stelle x wird durch

$$\widehat{E_\theta Y_x} = A + B \cdot x$$

geschätzt. Es gilt für diese Schätzung:

$$A + B \cdot x = \overline{Y} + B \cdot \left(x - \overline{x}\right)$$

Sind die abhängigen Größen im Regressionsmodell normalverteilt mit $Y_i \sim N(\alpha+\beta \cdot x, \sigma^2)$, so gilt folgender Satz:

Satz 37.2 Unter den Voraussetzungen des Satzes 37.1 und normalverteilten Y_i gilt:

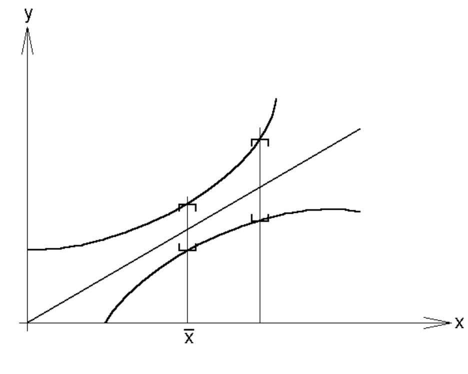
- A und B sind auch die plausiblen Schätzfunktionen für α und β.
- $A \sim N(\alpha, VarA)$ und $B \sim N(\beta, VarB)$
- $\dfrac{\sum_{i=1}^{n}\left(Y_i - A - B \cdot x_i\right)^2}{\sigma^2} = \dfrac{(n-2) \cdot S^2}{\sigma^2} \sim \chi^2_{n-2}$ (Pivot-Größe)

Konfidenzintervalle für die Regressionsparameter und den Wert der Regressionsgeraden an einer Stelle x sind im Fall von Normalverteilungen durch folgenden Satz gegeben.

Satz 37.3 Unter den Voraussetzungen des Satzes 37.2 erhält man folgende Konfidenzintervalle für β, σ^2 und EY_x mit Überdeckungswahrscheinlichkeit $1-\delta$:

- Konfidenzintervall für β: $\left[B - \dfrac{S \cdot t_{n-2;1-\frac{\delta}{2}}}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}}, B + \dfrac{S \cdot t_{n-2;1-\frac{\delta}{2}}}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}} \right]$
- Konfidenzintervall für σ^2: $\left[\dfrac{(n-2) \cdot S^2}{\chi^2_{n-2;1-\frac{\delta}{2}}}, \dfrac{(n-2) \cdot S^2}{\chi^2_{n-2;\frac{\delta}{2}}} \right]$
- Konfidenzintervall für EY_x:

$$\left[\overline{Y} + B \cdot \left(x - \overline{x}\right) - S \cdot \sqrt{\frac{1}{n} + \frac{\left(x - \overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}} \cdot t_{n-2;1-\frac{\delta}{2}}, \overline{Y} + B \cdot \left(x - \overline{x}\right) + S \cdot \sqrt{\frac{1}{n} + \frac{\left(x - \overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}} \cdot t_{n-2;1-\frac{\delta}{2}} \right]$$



37.2 Multiple lineare Regression

Die Einstellgröße und der Parametervektor sind k-dimensional:

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \qquad \text{und} \qquad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

Wenn für die Regressionsfunktion gilt

$$\psi(\underline{x}) = \sum_{j=1}^{k} \theta_j \cdot x_j = \underline{x}' \cdot \theta = EY_{\underline{x}}$$

so ist das Regressionsmodell gegeben durch:

$$\begin{aligned} &Y_{\underline{x}} = \underline{x}' \cdot \theta + U_{\underline{x}} \qquad \text{mit} \qquad EU_{\underline{x}} = 0 \\ &\mathrm{Var}Y_{\underline{x}} \equiv \sigma^2 \qquad \text{(Homoskedastizität)} \end{aligned} \qquad \text{(R)}$$

Für Beobachtungen $(x_{i1}, x_{i2}, ..., x_{ik}; y_k)$; i=1(1)n erhält man mit Hilfe der Methode der kleinsten Abstandsquadratsumme die Schätzungen $\hat{\theta}_1, \ldots, \hat{\theta}_k$ (Anpassung einer Hyperebene)

$$\hat{\psi}(\underline{x}) = \sum_{j=1}^{k} \hat{\theta}_j \cdot x_j$$

an die Daten durch:

$$\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} \hat{\theta}_j \cdot x_{ij} \right)^2 \to \min$$

In Matrixschreibweise erhält man:

$$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad \text{und} \qquad \kappa = \left(x_{ij} \right)_{\substack{i=1,\ldots,n \\ j=1,\ldots,k}}$$

und es gilt:

$$\left(\underline{Y} - \kappa \cdot \theta \right)^T \cdot \left(\underline{y} - \kappa \cdot \theta \right) \to \min$$

Für den Rang κ=k erhält man als Lösung der Minimierungsaufgabe den Parametervektor

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{pmatrix} \qquad \text{mit} \qquad \hat{\theta} = \left(\kappa' \cdot \kappa \right)^{-1} \cdot \kappa^T \cdot \underline{y}$$

Der Beweis ergibt sich durch die Lösung der Gauss´schen Normalgleichungen.
Die Lösung $\hat{\theta}_1, \ldots, \hat{\theta}_k$ sind unter gewissen Voraussetzungen auch statistisch optimal.

Gauß-Markoff-Theorem

Satz 37.4 Mit den Bezeichnungen dieses Abschnittes und einer Stichprobe $(x_{i1},...,x_{ik};y_i)$, i=1(1)n des Regressionsmodells (R) und bei Gültigkeit der folgenden Voraussetzungen

1) Rang κ=k
2) $Y_1,...,Y_n$ sind unkorreliert
3) $VarY_i \equiv \sigma^2$ (alle Varianzen sind gleich)
4) $EY_i = \sum_{j=1}^{k} \theta_j \cdot x_{ij} \quad \forall i = 1(1)n$
5) der Raum der möglichen Parameterwerte ist keine Hyperebene des R^k

sind die eindeutig bestimmten effizienten, linearen Schätzfunktionen T_j für die θ_j, j=1(1)k gegeben durch:

$$\hat{\theta} = \begin{pmatrix} T_1 \\ \vdots \\ T_k \end{pmatrix} = \left(\kappa^T \cdot \kappa\right)^{-1} \cdot \kappa^T \cdot \underline{Y} \qquad \text{mit} \qquad \underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

Eine unverzerrte Schätzfunktion für σ^2 ergibt sich zu:

$$S^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{k} x_{ij} \cdot \hat{\theta}_j\right)^2}{n-k}$$

Beweis: Extrema unter Nebenbedingung

Bemerkung: Aus konkreten Daten erhält man die Schätzwerte wie aus dem Ausgleichsproblem. Der Schätzwert für die Varianz ist:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{ij} \cdot \hat{\theta}_j\right)^2}{n-k}$$

37.3 Prognose im Regressionsmodell

Es gilt:

$$\hat{Y}_{\underline{x}} := \underline{x}^T \cdot \hat{\theta} = \sum_{j=1}^{k} \hat{\theta}_j \cdot x_j$$

Erwartungswert und Varianz sind dann gegeben durch:

$$E\hat{Y}_{\underline{x}} = EY_{\underline{x}}$$

$$Var\hat{Y}_{\underline{x}} = \underline{x}^T \cdot \left(\kappa^T \cdot \kappa\right)^{-1} \cdot \underline{x} \cdot \sigma^2$$

Beweis:

a) $$E\hat{Y}_{\underline{x}} = E\left(\sum_{j=1}^{k} \hat{\theta}_j \cdot x_j\right) = \sum_{j=1}^{k} E\left(x_j \cdot \hat{\theta}_j\right) = \sum_{j=1}^{k} x_j \cdot \underbrace{E\hat{\theta}_j}_{\theta_j \text{ da unverzerrt}} = \sum_{j=1}^{k} x_j \cdot \theta_j = EY_{\underline{x}}$$

b) $$Var\hat{Y}_{\underline{x}} = Var\left(\underline{x}^T \cdot \hat{\theta}\right) = Var\left(\underbrace{\underline{x}^T \cdot \left(\kappa^T \cdot \kappa\right)^{-1} \cdot \kappa^T}_{\underline{z}^T} \cdot \underline{Y}\right) = Var\left(\underline{z}^T \cdot \underline{Y}\right) = Var\left(\sum_{i=1}^{n} z_i \cdot Y_i\right) =$$

da die Varianz-Kovarianz-Matrix gleich $VCov(\underline{Y}) = \sigma^2 \cdot I_n$ ist, gilt weiters:

$$= \sum_{i=1}^{n} Var\left(z_i \cdot Y_i\right) = \sum_{i=1}^{n} z_i^2 \cdot \underbrace{VarY_i}_{\sigma^2} = \underline{z}^T \cdot z \cdot \sigma^2 = \sigma^2 \cdot \left(\underline{x}^T \cdot \left(\kappa^T \cdot \kappa\right)^{-1} \cdot \kappa^T \cdot \underline{x} \cdot \left(\left(\kappa^T \cdot \kappa\right)^{-1}\right)^T \cdot \kappa\right) =$$

$$= \sigma^2 \cdot \left(\underline{x}^T \cdot \left(\kappa^T \cdot \kappa\right)^{-1} \cdot \left(\kappa^T \cdot \kappa\right) \cdot \underline{x} \cdot \left(\left(\kappa^T \cdot \kappa\right)^{-1}\right)^T\right) = \underline{x}^T \cdot \left(\kappa^T \cdot \kappa\right)^{-1} \cdot \underline{x} \cdot \sigma^2$$

VIII Elemente der Bayes-Statistik

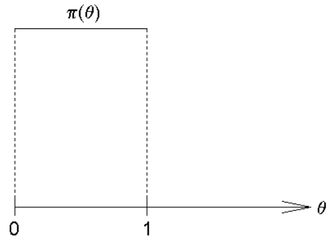
38. Allgemeine und A-priori-Verteilungen

In der Bayes'schen Statistik werden alle unbekannten Größen durch stochastische Größen mit zugehöriger Wahrscheinlichkeitsverteilung beschrieben. Für ein stochastisches Modell $X \sim f(\cdot|\theta)$ mit Parameterraum Θ wird die Unsicherheit bezüglich des Parameters θ, der durch eine stochastische Größe $\tilde{\theta}$ beschrieben wird, mittels einer sog. **A-priori-Verteilung** $\pi(\theta)$ von $\tilde{\theta}$ ausgedrückt.

Beispiel: Anteilsschätzung

Falls keine spezielle Vorinformation über den Parameter θ vorliegt, kann man eine uniforme Verteilung über den Parameterraum $\Theta=[0,1]$ verwenden. Das bedeutet:

$$\pi(\theta) \mathrel{\hat{=}} U_{0,1}$$



In vielen Fällen weiß man jedoch a-priori mehr als nur $0 \leq \theta \leq 1$, z.B. $0 < \theta \ll 1$. Dies kann man durch eine A-priori-Dichte in Form einer **Betaverteilung** ausdrücken.

Betaverteilung Be(α,β) $\alpha>0$, $\beta>0$

Die Dichtefunktion der Betaverteilung ist gegeben durch:

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} \cdot I_{(0,1)}(x)$$

Dabei ist $B(\cdot,\cdot)$ die Betafunktion:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Beispiel: Lebensdauer $X \sim Ex_\tau$

Vorinformation über die zu erwartende Lebensdauer τ kann durch eine A-priori-Verteilung (Gammaverteilung $\gamma(\alpha,\lambda)$) ausgedrückt werden.

$$\pi(\tau) = \gamma(\alpha,\lambda)$$

Dabei sind die beiden Parameter α und λ die sog. **Hyperparameter**.

Wie modifiziert sich nun die A-priori-Verteilung, wenn man eine Stichprobe einer stochastischen Größe hat?

39. Bayes'sches Theorem und A-posteriori-Verteilung

Zu einem stochastischen Modell $X \sim f(\cdot|\theta)$, $\theta \in \Theta$ und einer A-priori-Verteilung $\pi(\cdot)$ von $\tilde{\theta}$ liege eine konkrete Stichprobe $x_1,...,x_n$ mit $x_i \in M_X$, $i=1,...,n$ vor. Nach dem Stichprobenergebnis (Daten D) wird nun eine neue Einschätzung von $\tilde{\theta}$ gesucht, die sog. **A-posteriori-Verteilung** $\pi(\theta|D)$. Die A-posteriori-Dichte beschreibt das Wissen über θ nach Beobachtung der Daten. Die Transformation der A-priori-Verteilung $\pi(\theta)$ in die A-posteriori-Verteilung unter Verwendung der Information aus der Stichprobe D geschieht mit Hilfe des **Bayes'schen Theorems**:

39.1 Bayes'sches Theorem für den diskreten Fall

Falls für θ nur endlich viele Werte möglich sind, ist die A-priori-Verteilung $\pi(\theta)$ folgendermaßen gegeben:

$$\begin{pmatrix} \theta_1 & \cdots & \theta_k \\ p_1 & \cdots & p_k \end{pmatrix} \qquad \text{mit } \sum_{j=1}^{k} p_j = 1 \quad \text{und} \quad \pi(\theta_j) = p_j$$

Für Beobachtungen von $X \sim W(\cdot|\theta)$, d.h. einer konkreten Stichprobe, beschreibt man die durch die Daten bedingten Wahrscheinlichkeiten der θ_j mittels der Bayes'schen Formel

$$\pi(\theta_j|D) = \frac{W(D|\theta_j) \cdot \pi(\theta_j)}{\sum_{j=1}^{k} W(D|\theta_j) \cdot \pi(\theta_j)}$$

wobei $W(D|\theta_j)$ die Wahrscheinlichkeit der Stichprobe $x_1,...,x_n$ bei dem Parameterwert θ_j darstellt. Im Falle einer einfachen Stichprobe einer stochastischen Größe $X \sim p(\cdot|\theta)$ gilt:

$$W(D|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$$

Die durch die Punktwahrscheinlichkeiten $\pi(\theta_j|D)$ gegebene Wahrscheinlichkeitsverteilung ist die A-posteriori-Verteilung.

Bemerkung: Im Beispiel der Anteilsschätzung kann der Parameter θ nur endlich viele Werte annehmen. Die Berechnung ist allerdings umständlich und eine kontinuierliche Beschreibung ist hinreichend genau und einfacher.

39.2 Bayes'sches Theorem für den kontinuierlichen Fall

Bei einem kontinuierlichen stochastischen Modell $X \sim f(x|\theta)$, $\theta \in \Theta$ wird die A-priori-Verteilung von $\tilde{\theta}$ durch eine A-priori-Dichte $\pi(\theta)$ beschrieben.

Hat man nur eine Beobachtung x von X so ist die 2-dimensionale stochastische Größe $(X, \tilde{\theta})$ kontinuierlich verteilt mit der gemeinsamen Dichtefunktion $g(x,\theta)$. Die Dichtefunktion erhält man nach der Definition von bedingten Dichten als

$$g(x,\theta) = f(x|\theta) \cdot \pi(\theta)$$

Hat man eine Beobachtung x von X so läßt sich damit die durch X=x bedingte Dichte $\pi(\theta|x)$ von $\tilde{\theta}$ berechnen. Die bedingte Dichte $\pi(\theta|x)$ ist definiert durch

$$\pi(\theta|x) = \frac{g(x,\theta)}{f_1(x)} \qquad \text{für } f_1(x) > 0$$

wobei $f_1(\cdot)$ die Randdichte von X ist. Weiters gilt

$$f_1(x) = \int_{\Theta} g(x,\theta)d\theta$$

und daraus folgt:

$$\pi(\theta|x) = \frac{f(x|\theta) \cdot \pi(\theta)}{\int_{\Theta} f(x|\theta) \cdot \pi(\theta)d\theta}$$

Hat man eine Stichprobe vom Umfang n (Daten $D=(x_1,...,x_n)$), so benötigt man die gemeinsame Dichtefunktion $h(x_1,...,x_n|\theta)$, die im Fall einer einfachen Stichprobe von folgender Form ist:

$$h(x_1,...,x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = l(\theta; x_1,...,x_n)$$

$$g(x_1,...,x_n;\theta) = l(\theta; x_1,...,x_n) \cdot \pi(\theta)$$

Satz 39.1 Mit den Bezeichnungen dieses Abschnittes gilt für die A-posteriori-Verteilung

$$\pi(\theta|D) = \frac{\pi(\theta) \cdot l(\theta; D)}{\int\limits_{\Theta} \pi(\theta) \cdot l(\theta; D) d\theta}$$

Da der Nenner dieses Bruchs eine Konstante ist, schreibt man das Bayes'sche Theorem auch folgendermaßen:

$$\pi(\theta|D) \propto \pi(\theta) \cdot l(\theta; D)$$

40. Konjugierte Verteilungsfamilien

Eine Familie $\mathcal{F}$ von Wahrscheinlichkeitsverteilungen für den Parameter θ eines stochastischen Modells $X \sim f(x|\theta)$ heißt zugehörige **konjugierte A-priori-Familie**, falls für jedes $\pi(\theta) \in \mathcal{F}$ und beliebige Daten auch die A-posteriori-Dichte $\pi(\theta|D) \in \mathcal{F}$.

Beispiel: Anteilsschätzung

Die Betaverteilungen bilden zur Alternativverteilung eine konjugierte A-priori-Familie. Bei einer Alternativverteilung können die Punktwahrscheinlichkeiten in folgender Form geschrieben werden:

$$p(x|\theta) = \theta^{x} \cdot (1-\theta)^{1-x} \qquad \forall x \in \{0,1\}$$

Für die A-priori-Verteilung $\pi(\theta) \mathrel{\hat{=}} Be(\alpha, \beta)$ gilt

$$\pi(\theta) \propto \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1}$$

und für die Plausibilitätsfunktion:

$$l(\theta; D) = \prod_{i=1}^{n} \theta^{x_i} \cdot (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i} \cdot (1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Die Anwendung des Bayes'schen Theorems führt auf die A-posteriori-Verteilung:

$$\pi(\theta|D) \propto \theta^{\alpha+\sum_{i=1}^{n} x_i - 1} \cdot (1-\theta)^{b+n-\sum_{i=1}^{n} x_i - 1} \mathrel{\hat{=}} Be\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i\right)$$

Beispiel: Lebensdauer

Zur Exponentialverteilung bilden die Gammaverteilungen eine konjugierte A-priori-Familie. Unter folgenden Voraussetzungen

$$X \sim f(x|\lambda) = \lambda \cdot e^{-\lambda \cdot x} \cdot I_{(0,\infty)}(x)$$

$$\pi(\lambda) \propto \lambda^{\alpha-1} \cdot e^{-\beta\cdot\lambda}$$

gilt für die A-posteriori-Verteilung:

$$\pi(\lambda|D) \propto \lambda^{\alpha+n-1} \cdot e^{-\lambda \cdot \left(b + \sum_{i=1}^{n} x_i\right)} \mathrel{\hat{=}} \gamma\left(\alpha + n, \beta + \sum_{i=1}^{n} x_i\right)$$

Bemerkung: Die Statistik $\sum_{i=1}^{n} x_i$ ist in beiden obigen Beispielen hinreichend zur Berechnung der A-posteriori-Dichte. Solche Statistiken nennt man **suffiziente Statistiken**. Prüfung!

41. Verwendung der A-posteriori-Verteilung

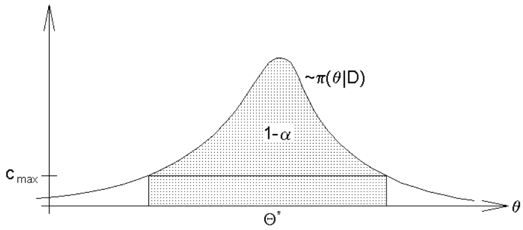
Alle wesentlichen Information über den Parameter θ eines stochastischen Modells $X \sim f(\cdot|\theta), \theta \in \Theta$, die aus erhobenen Daten gewonnen wird, ist in der A-posteriori-Verteilung $\pi(\theta|D)$ enthalten. Daher kann diese für Prognosen, Punktschätzungen, Bereichsschätzungen und Tests (allg. Entscheidungen) herangezogen werden.

41.1 HPD-Bereiche

Sogenannte höchste A-posteriori-Dichte-Bereiche für den Parameter θ eines stochastischen Modells $f(x|\theta)$ entsprechen den Konfidenzbereichen, sind aber anders definiert. Für die A-posteriori-Dichte $\pi(\theta|D)$ ist ein HPD-Bereich $\Theta^* \subset \Theta$ mit Sicherheit $1-\alpha$, $0<\alpha<1$, definiert durch:

1) $\int_{\Theta^*} \pi(\theta|D)\,d\theta = 1-\alpha$

2) $\pi(\theta|D) \geq c_{max} \qquad \forall \theta \in \Theta^*$, wobei c_{max} die größtmögliche reelle Zahl ist, für die 1) erfüllt ist.



Bemerkung: $W\{\tilde{\theta} \in \Theta^* | D\} = 1-\alpha$

41.2 A-posteriori-Bayes-Schätzer

Prüfung!

Falls *ein* Parameterwert als Schätzwert für θ gewünscht wird, kann man einen charakteristischen Wert der A-posteriori-Verteilung $\pi(\theta|D)$ heranziehen.

Definition: Der A-posteriori-Bayes-Schätzer $\hat{\theta}$ für einen 1-dimensionalen Parameter $\theta \in R$ ist der Erwartungswert der A-posteriori-Verteilung von $\tilde{\theta}$, d.h.:

$$\hat{\theta} = E_{\pi(\theta|D)}\tilde{\theta} = \int_{\Theta} \theta \cdot \pi(\theta|D)\,d\theta$$

Falls ein geraffter Parameter $\tau(\theta) \in R$ geschätzt werden soll, ist der diesbezügliche A-posteriori-Bayes-Schätzer durch den Erwartungswert von $\tau(\tilde{\theta})$ definiert:

$$\widehat{\tau(\theta)} = \int_{\Theta} \tau(\theta) \cdot \pi(\theta|D)\,d\theta$$

Beispiel: Anteilsschätzung

Für die A-priori-Dichte wird eine uniforme Gleichverteilung angenommen, d.h. $\pi(\theta) \hat{=} U_{0,1}$. Es gilt:

$\pi(\theta|D) \hat{=} Be(\cdot|\cdot)$

$$\hat{\theta} = \frac{\int_0^1 \theta \cdot \theta^{\sum_{i=1}^n x_i} \cdot (1-\theta)^{n-\sum_{i=1}^n x_i}\,d\theta}{\int_0^1 \theta^{\sum_{i=1}^n x_i} \cdot (1-\theta)^{n-\sum_{i=1}^n x_i}\,d\theta} = \frac{B\left(\sum_{i=1}^n x_i + 2, n - \sum_{i=1}^n x_i + 1\right)}{B\left(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1\right)} = \frac{\sum_{i=1}^n x_i + 1}{n+2} \qquad \text{(verzerrt!)}$$

41.3 A-posteriori-Wahrscheinlichkeiten von statistischen Hypothesen

Man geht von einer Parameterhypothese aus:

$$\Theta_0 \subset \Theta \qquad \Theta_1 \subset \Theta_0^c$$
$$H_0 : \theta \in \Theta_0 \qquad H_1 : \theta \in \Theta_1$$

Um zu entscheiden, welche der Hypothesen angenommen werden kann, berechnet man die A-posteriori-Wahrscheinlichkeiten der Hypothesen.

$$\alpha_0 = W\left\{\tilde{\theta} \in \Theta_0 \middle| D\right\}$$
$$\alpha_1 = W\left\{\tilde{\theta} \in \Theta_1 \middle| D\right\}$$

Die **relative Plausibilität** $\frac{\alpha_1}{\alpha_0}$ wird zur Entscheidung herangezogen. Dies kann mit Hilfe der Plausibilitätsfunktion $l(\theta;D)$ und der A-priori-Dichte $\pi(\theta)$ folgendermaßen berechnet werden:

$$\frac{\alpha_1}{\alpha_0} = \frac{\int_{\Theta_1} \pi(\theta|D)\,d\theta}{\int_{\Theta_0} \pi(\theta|D)\,d\theta} = \frac{\int_{\Theta_1} \pi(\theta) \cdot l(\theta;D)\,d\theta}{\int_{\Theta_0} \pi(\theta) \cdot l(\theta;D)\,d\theta}$$

41.4 Prädiktivverteilungen

In der Prädiktivdichte ist alle Information aus dem stochastischen Modell, der A-priori-Verteilung und den beobachteten Daten verarbeitet. Die Prädiktivdichte entspricht der Randdichte von X der gemeinsamen Dichte von $(X, \tilde{\theta})$:

$$f(x|D) = \int_{\Theta} g(x,\theta)\,d\theta = \int_{\Theta} f(x|\theta) \cdot \pi(\theta|D)\,d\theta$$

<u>Beispiel:</u> Prognose des Energiebedarfs

Die stochastische Größe X sei normalverteilt mit unbekanntem Mittel, d.h.: $X \sim N(\theta,\sigma_0^2)$ mit σ_0^2 bekannt. Die A-priori-Verteilung entspricht folgender Normalverteilung:

$$\pi(\theta) \mathrel{\hat{=}} N(\mu,\tau^2)$$

Daraus ergibt sich die A-posteriori-Verteilung

$$\pi(\theta|D) \mathrel{\hat{=}} N\left(\frac{\sigma_0^2}{n \cdot \tau^2 + \sigma_0^2} \cdot \mu + \frac{\tau^2 \cdot \sum_{i=1}^{n} x_i}{n \cdot \tau^2 + \sigma_0^2}, \frac{\tau^2 \cdot \sigma_0^2}{n \cdot \tau^2 + \sigma_0^2}\right)$$

und daraus die gemeinsame Dichte und weiters die Prädiktivdichte.

42. Bayes'sche Entscheidungsregeln

Entscheidungen sind oft mit Verlusten verbunden.
Beispiel: Warenlieferung von N Stücken
Die möglichen Entscheidungen sind entweder Annahme d_0 oder Ablehnung d_1. Der Anteil der schlechten Stücke in der Warenlieferung ist gleich:

$$\theta = \frac{A}{N}$$

Weiters liegt eine A-priori-Dichte $\pi(\theta)$ von $\tilde{\theta}$ vor und eine Stichprobe vom Umfang n. Daraus ergibt sich die A-posteriori-Dichte $\pi(\theta|D)$. Für jedes gute Stück erzielt man den Gewinn G und für jedes schlechte Stück den Verlust K. Gesucht ist nun die *optimale* Entscheidung. Dazu berechnet man den Verlust $L(\theta,d_j)$ (Verlust bei Anteil θ und Entscheidung d_j, j=0,1).

$$L(\theta, d_0) = \theta \cdot N \cdot K - (1-\theta) \cdot N \cdot G$$
$$L(\theta, d_1) = (1-\theta) \cdot N \cdot G$$

Zur Findung der optimalen Entscheidung berechnet man den A-posteriori-Erwartungswert der Verlustfunktion $L(\theta,d_j)$:

$$E_{\pi(\theta|D)} L(\tilde{\theta}, d_j) = \int_0^1 L(\theta, d_j) \cdot \pi(\theta|D) d\theta \qquad j = 0,1$$

Die Bayes'sche Entscheidung ist jene, welche den kleineren zu erwartenden Verlust liefert (j=0 oder j=1).

Bemerkung: Im obigen Beispiel wird man am Gewinn (Nutzen) interessiert sein. Dazu berechnet man die Nutzenfunktion $U(\theta,d)$:

$$U(\theta, d_0) = (1-\theta) \cdot N \cdot G - \theta \cdot N \cdot K$$
$$U(\theta, d_1) = 0$$

Im Fall dieser Nutzenfunktion wird der a-posteriori zu erwartende Nutzen maximiert. Im konkreten Beispiel kommt es daher zur Entscheidung d_0 falls gilt:

$$E_{\pi(\theta|D)} U(\tilde{\theta}, d_0) > 0$$

42.1 Entscheidungen und Entscheidungsregeln

Das Konzept der Entscheidungstheorie ist sehr allgemein. Das Modell dazu sind Entscheidungsregeln und Entscheidungen. Entscheidungsregeln sind Vorschriften, nach denen konkrete Entscheidungen getroffen werden.

$$d = \delta(x_1, ..., x_n)$$

Beispiele für Entscheidungsregeln:
- Schätzfunktionen
- Konfidenzfunktionen
- Testvorschriften

$\tau(X_1, ..., X_n) = I_V(X_1, ..., X_n)$ ($I_V(...)$: Indikatorfunktion des Verwerfungsraumes $\Rightarrow$ 0=Annahme)

Beispiele für Entscheidungen:
- konkrete Schätzwerte
- konkrete Konfidenzbereiche
- Ablehnung oder Annahme einer statistischen Hypothese

Zur Ermittlung guter Entscheidungsregeln zieht man den zu erwartenden Verlust heran:

$$EL\big(\theta, \delta(X_1, \ldots, X_n)\big) = R(\theta, \delta)$$

Dabei stellt $R(\theta,\delta)$ die sog. **Risikofunktion** der Entscheidungsregel δ dar.

Bemerkung: Vor Erhebung der Daten ist $L(\theta,\delta(X_1,\ldots,X_n))$ eine stochastische Größe.

Der Vergleich von Entscheidungsregeln erfolgt mit Hilfe der Risikofunktion. Dabei wird die Verteilung $\pi(\theta)$ des Parameters $\tilde{\theta}$ herangezogen. Das sog. **Bayes'sche Risiko** $r(\pi,\delta)$ zu einer Entscheidungsregel δ bei gegebener A-priori-Dichte ist:

$$r(\pi, \delta) = E_{\pi(\theta)} R(\tilde{\theta}, \delta) = \int_{\Theta} R(\theta, \delta) \cdot \pi(\theta) d\theta$$

Definition: Eine Bayes'sche Entscheidungsregel δ^* in einer Klasse Δ von betrachteten Entscheidungsregeln ist eine solche, die minimales Bayes'sches Risiko hat, d.h.:

$$r(\pi, \delta^*) = \min_{\delta \in \Delta} r(\pi, \delta)$$

Bemerkung: Im begleitenden Beispiel 1 (Anteilsschätzung) am Anfang dieses Abschnittes wurde die A-posteriori-Dichte von $\tilde{\theta}$ verwendet. Dazu gilt folgender Satz:

Satz 42.1 Bayes'sche Entscheidungsregeln δ entstehen, wenn man zu jeder konkreten Stichprobe jene Entscheidung d nimmt, welche den a-posteriori zu erwartenden Verlust

$$E_{\pi(\theta|D)} L(\tilde{\theta}, \delta) = \int_{\Theta} L\big(\theta, \delta(x_1, \ldots, x_n)\big) \cdot \pi(\theta | x_1, \ldots, x_n) d\theta$$

minimiert.

42.2 Bayes-Schätzer (unter Einbeziehung von Verlustbetrachtungen)

Bei gegebener A-priori-Information und Verlustfunktion ist die optimale Punktschätzung für den Parameter θ (bzw. den gerafften Parameter $\tau(\theta)$) gesucht. Dabei gilt mit den konkreten Daten $x_1,\ldots,x_n$:

Der Bayes-Schätzer für θ bezüglich der **Verlustfunktion** $L(\cdot,\cdot)$ ist jener, der den a-posteriori zu erwartenden Verlust

$$E_{\pi(\theta|D)} L\big(\tilde{\theta}, \vartheta(x_1, \ldots, x_n)\big)$$

minimiert. Für eine quadratische Verlustfunktion kann man den optimalen Schätzwert allgemein angeben:

Satz 42.2 Mit den Bezeichnungen dieses Abschnittes ist die Bayes'sche Schätzung für einen 1-dimensionalen Parameter θ bei quadratischer Verlustfunktion

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$$

gleich dem Erwartungswert der A-posteriori-Verteilung $\pi(\theta|D)$ von $\tilde{\theta}$:

$$\hat{\theta} = E_{\pi(\theta|D)} \tilde{\theta} = \int_{\Theta} \theta \cdot \pi(\theta | D) d\theta$$

Für geraffte Parameter $\tau(\theta)$ gilt analog:

$$\widehat{\tau(\theta)} = E_{\pi(\theta|D)} \tau(\tilde{\theta}) = \int_{\Theta} \tau(\theta) \cdot \pi(\theta | D) d\theta$$

Beweis: Der Bayes-Schätzer bezüglich quadratischer Verlustfunktion muß jeder konkreten Stichprobe jenen Schätzwert d (Entscheidung) zuordnen, der den a-posteriori zu erwartenden Verlust $E_{\pi(\theta|D)}\left(\tilde{\theta}-d\right)^2$ minimal macht.

Mit der Kurzbezeichnung E_D für $E_{\pi(\theta|D)}$ gilt:

$$E_D\left(\tilde{\theta}-d\right)^2 = E_D\left(\left[\tilde{\theta}-E_D\tilde{\theta}\right]+\left[E_D\tilde{\theta}-d\right]\right)^2 =$$

$$= E_D\left[\left(\tilde{\theta}-E_D\tilde{\theta}\right)^2+2\left(\tilde{\theta}-E_D\tilde{\theta}\right)\left(E_D\tilde{\theta}-d\right)+\left(E_D\tilde{\theta}-d\right)^2\right]=$$

$$= E_D\left[\left(\tilde{\theta}-E_D\tilde{\theta}\right)^2\right]+2\cdot E_D\left[\left(\tilde{\theta}-E_D\tilde{\theta}\right)\left(E_D\tilde{\theta}-d\right)\right]+E_D\left[\left(E_D\tilde{\theta}-d\right)^2\right]=$$

$$= \mathrm{Var}_D\hat{\theta}+2\cdot E_D\left(\tilde{\theta}-E_D\tilde{\theta}\right)\left(E_D\tilde{\theta}-d\right)+\left(E_D\tilde{\theta}-d\right)^2 =$$

$$= \mathrm{Var}_D\hat{\theta}+2\cdot\underbrace{\left(E_D\tilde{\theta}-E_D\tilde{\theta}\right)}_{=0}\left(E_D\tilde{\theta}-d\right)+\left(E_D\tilde{\theta}-d\right)^2 =$$

$$= \mathrm{Var}_D\hat{\theta}+\underbrace{\left(E_D\tilde{\theta}-d\right)^2}_{\geq 0} \qquad \text{wird minimal für} \quad E_D\tilde{\theta}=d$$

Bemerkung: Für die Verlustfunktion $L(\theta,d)=(\theta-d)$ ist der Median der A-posteriori-Verteilung die Bayes'sche Entscheidung für die Schätzung von $\theta\in R$.

42.3 Angewandte Bayes'sche Entscheidungen

Beispiel: Ein Autohändler muß entscheiden, wieviele Autos er bestellen soll. Er kann pro verkauftem Auto mit 5 GE Gewinn rechnen, wenn er das bestellte Auto vor Einführung eines neuen Modells verkauft. Wenn er nach Einführung noch Autos auf Lager hat, kann er diese nur mit einem Verlust von 3 GE pro Stück verkaufen. Wie viele Autos soll er bestellen ?

Wir verwenden folgende Bezeichnungen:

θ ....Anzahl der verkauften Autos (vor dem neuen Modell)

d ....Anzahl der bestellten Autos. Dabei stellen $d_1,\dots,d_m$ die möglichen Entscheidungen dar.

$\tilde{\theta}$ ....stochastische Größe mit einer A-priori-Verteilung $\pi(\theta)$ auf N_0, z.B. $M_{\tilde{\theta}}=\{0,1,2,\dots,k\}$.

$U(\theta,d_j)$ ...Nutzen (Gewinn) der Entscheidung d_j, falls $\tilde{\theta}=\theta$.

Zur Entscheidungsfindung zieht man den zu erwartenden Nutzen der einzelnen Entscheidungen d_j heran:

$$\overline{U}\left(d_j\right) := E_{\pi(\cdot)}U\left(\tilde{\theta},d_j\right)=\sum_{i=0}^{k}U\left(i,d_j\right)\cdot\pi(i)$$

Die Bayes'sche Entscheidung ist nun jenes d_j, für das $\overline{U}\left(d_j\right)$ maximal ist. Prüfung

Die Entscheidung hängt wesentlich von der (subjektiven) Wahrscheinlichkeitverteilung $\pi(\cdot)$ auf $M_{\tilde{\theta}}$ ab. Daher ist Expertenwissen und Erfahrung von besonderer Bedeutung.

Im Beispiel ist der Nutzen gleich:

$$U(\theta, d_j) = \begin{cases} \theta \cdot 5 - (d_j - \theta) \cdot 3 & \text{für } \theta \leq d_j \\ d_j \cdot 5 & \text{für } \theta > d_j \end{cases}$$

In einer Tabelle dargestellt sieht dies folgendermaßen aus:

$d_j \backslash \theta$	0	1	2	3	.....	k
0	0	0	0	0	.....	0
1	-3	5	5	5	.....	5
2	-6	2	10	10	.....	10
3	-9	.....	.....	.....	.....	:
:	:	.....	.....	.....	.....	:
m	$-3 \cdot m$	.....	.....	.....	.....	$5 \cdot m$
	$\pi(0)$	$\pi(1)$	.....	.....	.....	$\pi(k)$

<u>Bemerkung:</u> Man kann auch den Verlust einbeziehen, wenn man zu wenige Autos kauft. Die Nutzenfunktion hat dann folgender Form:

$$U(\theta, d_j) = \begin{cases} \theta \cdot 5 - (d_j - \theta) \cdot 3 & \text{für } \theta \leq d_j \\ d_j \cdot 5 - (\theta - d_j) \cdot 5 & \text{für } \theta > d_j \end{cases}$$

<u>Allgemeiner Fall diskreter Entscheidungsmöglichkeiten</u>

Wir verwenden folgende Bezeichnungen:

θ ....Zustand eines Systems (diskret oder kontinuierlich)
Θ ....Raum der möglichen Zustände
$\pi(\theta)$ ....A-priori-Verteilung (A-priori-Information) über θ
$d_1,\dots,d_m$ ....mögliche Entscheidungen
$U(\theta,d_j)$ ....Nutzen der Entscheidung d_j, falls das System im Zustand θ ist

Die Entscheidungsgrundlage bildet der Erwartungswert des Nutzens der Entscheidung d_j.

$$E_{\pi(\cdot)} U(\tilde{\theta}, d_j) = \overline{U}(d_j)$$

Die optimale Entscheidung im Bayes'schen Sinne ist jenes D_j mit größtem $\overline{U}(d_j)$. Im Fall diskreter Zustände $\theta_1,\dots,\theta_k$ gilt:

$$\overline{U}(d_j) = \sum_{i=1}^{k} U(\theta_i, d_j) \cdot \pi(\theta_i) \qquad \text{mit } \pi(\theta_i)=W\{\tilde{\theta}=\theta_i\}$$

und im kontinuierlichen Fall (d.h. Θ kontinuierlich):

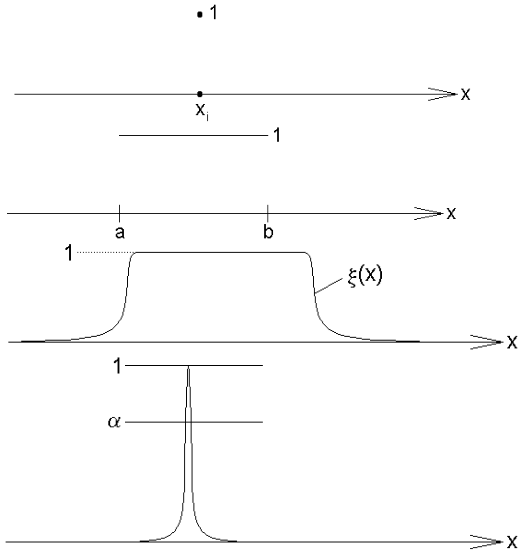
$$\overline{U}(d_j) = \int_{\Theta} U(\theta, d_j) \cdot \pi(\theta) d\theta \qquad \text{mit } \pi(\theta) \text{ gleich der Dichte von } \tilde{\theta}$$

<u>Bemerkung:</u> Man wird für $\pi(\cdot)$ jeweils die aktuellsten Informationen heranziehen ($\pi(\theta|D)$).

IX Unscharfe Daten

43. Charakterisierende Funktionen

Viele Messungen führen nicht zu exakten Zahlen, sondern sind mehr oder weniger unscharf (fuzzy). Zur Beschreibung ist das Konzept **unscharfer Zahlen** nützlich.
Exakte reelle Zahlen und Intervalle sind in eindeutiger Weise charakterisiert durch deren Indikatorfunktion:



$$x_i \in R \;\rightarrow\; I_{\{x_i\}}(\cdot)$$

$$[a,b] \subseteq R \;\rightarrow\; I_{[a,b]}(\cdot)$$

Da die Ränder oft unscharf sind, läßt man auch allgemeine Funktionen als Indikatorfunktionen zu. Dabei wird $\xi(\cdot)$ als **charakterisierende Funktion** bezeichnet.

Weiters gilt, daß für die unscharfe Zahl x^* bei gegebener charakterisierender Funktion $\xi(\cdot)$ die sog. **α-Schnitte** gegeben sind durch:

$$B_\alpha\left(x^*\right) := \left\{x : \xi(x) \geq \alpha\right\}$$

Dabei ist zu beachten, daß alle α-Schnitte abgeschlossene Intervalle sein müssen.

44. Kombination unscharfer Beobachtungen

Im Fall unscharfer Daten $x_1^*,\ldots,x_n^*$ gilt für das unscharfe $\underline{x}^* \in M_X{}^n$:

$$\xi_{\underline{x}^*}\left(x_1,\ldots,x_n\right) = K_n\left(\xi_1\left(x_1\right),\ldots,\xi_n\left(x_n\right)\right)$$

$$\xi\left(x_1,\ldots,x_n\right) = \min_{i=1(1)n} \xi_i\left(x_i\right)$$

Damit kann man Statistik auch für unscharfe Daten machen.

Anhang

1. Standard-Normalverteilung

1.1 Verteilungsfunktion Φ(x) der Standard-Normalverteilung N(0,1)

Es gilt:

- $\Phi(-x)=1-\Phi(x)$
- $F\left(x \middle| \mu,\sigma^2\right)=\Phi\left(\frac{x-\mu}{\sigma}\right)$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

1.2 Quantile u_p

$(X \sim N(0,1) \Rightarrow W(X \leq u_p)=p)$

p	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	0,975	0,99	0,995
u_p	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0364	1,2816	1,6449	1,9600	2,3263	2,5758

1.3 A-posteriori-Verteilung für das unbekannte Mittel

Wenn $X_1,...,X_n$ eine Stichprobe einer Normalverteilung mit bekannter Varianz σ_0^2 und unbekannten Mittel θ ist und θ a-priori normalverteilt ist mit Mittel μ und Varianz τ^2, so gilt:

$$\pi\left(\theta \middle| D\right) \equiv N\left(\frac{\sigma_0^2 \cdot \mu + n \cdot \tau^2 \cdot \overline{x}_n}{\sigma_0^2 + n \cdot \tau^2}, \frac{\sigma_0^2 \cdot \tau^2}{\sigma_0^2 + n \cdot \tau^2}\right)$$

1.4 Rechenregeln

Wenn $X \sim N(\mu,\sigma^2)$ und $Y=\alpha \cdot X+\beta$ $(\alpha>0, \beta \in R)$ dann gilt: $Y \sim N(\alpha \cdot \mu+\beta,\alpha^2 \cdot \sigma^2)$

Wenn $X \sim N(\mu_X,\sigma_X^2)$ und $Y \sim N(\mu_Y,\sigma_Y^2)$ dann gilt: $X+Y \sim N(\mu_X+\mu_Y,\sigma_X^2+\sigma_Y^2)$ (Additionstheorem)

2. Chiquadrat-Verteilung

2.1 Quantile

$(X \sim \chi_n^2 \Rightarrow W(X \le x) = p)$

n\p	0,005	0,01	0,025	0,05	0,10	0,20	0,25	0,30	0,40
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,0642	0,1015	0,1485	0,2750
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,4463	0,5754	0,7133	1,0217
3	0,0717	0,1148	0,2158	0,3518	0,5844	1,0052	1,2125	1,4237	1,8692
4	0,2070	0,2971	0,4844	0,7107	1,0636	1,6488	1,9226	2,1947	2,7528
5	0,4118	0,5543	0,8312	1,1455	1,6103	2,3425	2,6746	2,9999	3,6555
6	0,6757	0,8721	1,2373	1,6354	2,2041	3,0701	3,4546	3,8276	4,5702
7	0,9893	1,2390	1,6899	2,1673	2,8331	3,8223	4,2549	4,6713	5,4932
8	1,3444	1,6465	2,1797	2,7326	3,4895	4,5936	5,0706	5,5274	6,4226
9	1,7349	2,0879	2,7004	3,3251	4,1682	5,3801	5,8988	6,3933	7,3570
10	2,1558	2,5582	3,2470	3,9403	4,8652	6,1791	6,7372	7,2672	8,2955
11	2,6032	3,0535	3,8157	4,5748	5,5778	6,9887	7,5841	8,1479	9,2373
12	3,0738	3,5706	4,4038	5,2260	6,3038	7,8073	8,4384	9,0343	10,1820
13	3,5650	4,1069	5,0087	5,8919	7,0415	8,6339	9,2991	9,9257	11,1291
14	4,0747	4,6604	5,6287	6,5706	7,7895	9,4673	10,1653	10,8215	12,0785
15	4,6009	5,2294	6,2621	7,2609	8,5468	10,3070	11,0365	11,7212	13,0298
16	5,1422	5,8122	6,9077	7,9616	9,3122	11,1521	11,9122	12,6243	13,9827
17	5,6973	6,4077	7,5642	8,6718	10,0852	12,0023	12,7919	13,5307	14,9373
18	6,2648	7,0149	8,2307	9,3904	10,8649	12,8570	13,6753	14,4399	15,8932
19	6,8439	7,6327	8,9065	10,1170	11,6509	13,7158	14,5620	15,3517	16,8504
20	7,4338	8,2604	9,5908	10,8508	12,4426	14,5784	15,4518	16,2659	17,8088
21	8,0336	8,8972	10,2829	11,5913	13,2396	15,4446	16,3444	17,1823	18,7683
22	8,6427	9,5425	10,9823	12,3380	14,0415	16,3140	17,2396	18,1007	19,7288
23	9,2604	10,1957	11,6885	13,0905	14,8480	17,1865	18,1373	19,0211	20,6902
24	9,8862	10,8563	12,4011	13,8484	15,6587	18,0618	19,0373	19,9432	21,6525
25	10,5196	11,5240	13,1197	14,6114	16,4734	18,9397	19,9393	20,8670	22,6156
30	13,7867	14,9535	16,7908	18,4927	20,5992	23,3641	24,4776	25,5078	27,4416
40	20,7066	22,1642	24,4331	26,5093	29,0505	32,3449	33,6603	34,8719	37,1340
50	27,9908	29,7067	32,3574	34,7642	37,6886	41,4492	42,9421	44,3133	46,8638
60	35,5344	37,4848	40,4817	43,1880	46,4589	50,6406	52,2938	53,8091	56,6200
70	43,2753	45,4417	48,7575	51,7393	55,3289	59,8978	61,6983	63,3460	66,3961
80	51,1719	53,5400	57,1532	60,3915	64,2778	69,2070	71,1445	72,9153	76,1879
90	59,1963	61,7540	65,6466	69,1260	73,2911	78,5584	80,6247	82,5111	85,9925
100	67,3275	70,0650	74,2219	77,9294	82,3581	87,9453	90,1332	92,1290	95,8078

n\p	0,50	0,60	0,70	0,75	0,80	0,90	0,95	0,975	0,99	0,995
1	0,4549	0,7083	1,0742	1,3233	1,6424	2,7055	3,8415	5,0239	6,6349	7,8794
2	1,3863	1,8326	2,4079	2,7726	3,2189	4,6052	5,9915	7,3778	9,2104	10,5965
3	2,3660	2,9462	3,6649	4,1083	4,6416	6,2514	7,8147	9,3484	11,3449	12,8381
4	3,3567	4,0446	4,8784	5,3853	5,9886	7,7794	9,4877	11,1433	13,2767	14,8602
5	4,3515	5,1319	6,0644	6,6257	7,2893	9,2363	11,0705	12,8325	15,0863	16,7496
6	5,3481	6,2108	7,2311	7,8408	8,5581	10,6446	12,5916	14,4494	16,8119	18,5475
7	6,3458	7,2832	8,3834	9,0371	9,8032	12,0170	14,0671	16,0128	18,4753	20,2777
8	7,3441	8,3505	9,5245	10,2189	11,0301	13,3616	15,5073	17,5345	20,0902	21,9549
9	8,3428	9,4136	10,6564	11,3887	12,2421	14,6837	16,9190	19,0228	21,6660	23,5893
10	9,3418	10,4732	11,7807	12,5489	13,4420	15,9872	18,3070	20,4832	23,2093	25,1881
11	10,3410	11,5298	12,8987	13,7007	14,6314	17,2750	19,6752	21,9200	24,7250	26,7569
12	11,3403	12,5838	14,0111	14,8454	15,8120	18,5493	21,0261	23,3367	26,2170	28,2997
13	12,3398	13,6356	15,1187	15,9839	16,9848	19,8119	22,3620	24,7356	27,6882	29,8193
14	13,3393	14,6853	16,2221	17,1169	18,1508	21,0641	23,6848	26,1189	29,1412	31,3194
15	14,3389	15,7332	17,3217	18,2451	19,3107	22,3071	24,9958	27,4884	30,5780	32,8015
16	15,3385	16,7795	18,4179	19,3689	20,4651	23,5418	26,2962	28,8453	31,9999	34,2671
17	16,3382	17,8244	19,5110	20,4887	21,6146	24,7690	27,5871	30,1910	33,4087	35,7184
18	17,3379	18,8679	20,6014	21,6049	22,7595	25,9894	28,8693	31,5264	34,8052	37,1564
19	18,3376	19,9102	21,6891	22,7178	23,9004	27,2036	30,1435	32,8523	36,1908	38,5821
20	19,3374	20,9514	22,7745	23,8277	25,0375	28,4120	31,4104	34,1696	37,5663	39,9969
21	20,3372	21,9915	23,8578	24,9348	26,1711	29,6151	32,6706	35,4789	38,9322	41,4009
22	21,3370	23,0307	24,9390	26,0393	27,3015	30,8133	33,9245	36,7807	40,2894	42,7957
23	22,3369	24,0689	26,0184	27,1413	28,4288	32,0069	35,1725	38,0756	41,6383	44,1814
24	23,3367	25,1064	27,0960	28,2412	29,5533	33,1962	36,4150	39,3641	42,9798	45,5584
25	24,3366	26,1430	28,1719	29,3388	30,6752	34,3816	37,6525	40,6465	44,3140	46,9280
30	29,3360	31,3159	33,5302	34,7997	36,2502	40,2560	43,7730	46,9792	50,8922	53,6719
40	39,3353	41,6222	44,1649	45,6160	47,2685	51,8050	55,7585	59,3417	63,6908	66,7660
50	49,3349	51,8916	54,7228	56,3336	58,1638	63,1671	67,5048	71,4202	76,1538	79,4898
60	59,3347	62,1348	65,2265	66,9815	68,9721	74,3970	79,0820	83,2977	88,3794	91,9518
70	69,3345	72,3583	75,6893	77,5766	79,7147	85,5270	90,5313	95,0231	100,4251	104,2148
80	79,3343	82,5663	86,1197	88,1303	90,4053	96,5782	101,8795	106,6285	112,3288	116,3209
90	89,3342	92,7614	96,5238	98,6499	101,0537	107,5650	113,1452	118,1359	124,1162	128,2987
100	99,3341	102,9459	106,9058	109,1412	111,6667	118,4980	124,3421	129,5613	135,8069	140,1697

2.2 Rechenregeln

Es gilt: $\gamma\left(\frac{n}{2}, 2\right) \hat{=} \chi_n^2$

3. t-Verteilung

3.1 Quantile

($X \sim t_n \Rightarrow W(X \leq x) = p$)

n\p	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	0,975	0,99	0,995
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617

4. F-Verteilung

4.1 0,975-Quantile

($X \sim F_{m,n} \Rightarrow W(X \leq x) = p$) $F_{m;n;\alpha} = \frac{1}{F_{n;m;1-\alpha}}$

n\m	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	120	∞
1	647,79	799,48	864,15	899,6	921,83	937,11	948,2	956,64	963,28	968,63	984,87	993,08	1001,4	1005,6	1009,8	1014	1018,3
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,43	39,45	39,46	39,47	39,48	39,49	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,25	14,17	14,08	14,04	13,99	13,95	13,90
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,66	8,56	8,46	8,41	8,36	8,31	8,26
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,43	6,33	6,23	6,18	6,12	6,07	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,27	5,17	5,07	5,01	4,96	4,90	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,57	4,47	4,36	4,31	4,25	4,20	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,10	4,00	3,89	3,84	3,78	3,73	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,77	3,67	3,56	3,51	3,45	3,39	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,52	3,42	3,31	3,26	3,20	3,14	3,08
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,86	2,76	2,64	2,59	2,52	2,46	2,40
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,57	2,46	2,35	2,29	2,22	2,16	2,09
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,31	2,20	2,07	2,01	1,94	1,87	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,18	2,07	1,94	1,88	1,80	1,72	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,06	1,94	1,82	1,74	1,67	1,58	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	1,94	1,82	1,69	1,61	1,53	1,43	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,83	1,71	1,57	1,48	1,39	1,27	1,00

5. Gammaverteilung

Wenn $X_1 \sim \gamma(\alpha_1,\beta)$ und $X_2 \sim \gamma(\alpha_2,\beta)$ dann gilt: $X_1+X_2 \sim \gamma(\alpha_1+\alpha_2,\beta)$

Wenn $X \sim \gamma(\alpha,\beta)$ und $c \in R$ dann gilt: $c \cdot X \sim \gamma(\alpha, c \cdot \beta)$

Es gilt: $\gamma\left(\frac{n}{2},2\right) \hat{=} \chi^2_n$

Symbolverzeichnis

(M,ξ,W) ... Wahrscheinlichkeitsraum
$\sim$ ... verteilt nach
$\perp$ ... stochastisch unabhängig
$\exists$ ... existiert
$\forall$ ... für alle

α ... Wahrscheinlichkeit eines Fehlers erster Art
β ... Wahrscheinlichkeit eines Fehlers zweiter Art
$\mathcal{B}$ ... System der Borel-Mengen in R
φ ... Dichte der Normalverteilung
$\varphi_{x,y}$ ... Korrelationskoeffizient
χ_n^2 ... Chiquadrat-Verteilung
δ_μ ... Dirac-Verteilung
$\in$ ... Element von
$\notin$ ... kein Element von
$\mathcal{F}$ ... Familie von Wahrscheinlichkeitsverteilungen
$\gamma(\cdot,\cdot)$ ... Gamma-Verteilung
$\mathcal{H}$ ... Hypothese
$\mathcal{H}_0$ ... Nullhypothese
$\mathcal{H}_1$ ... Gegenhypothese
λ ... Plausibilitätsquotient
μ ... Mittel
$\Phi(\cdot)$ ... Verteilungsfunktion der Normalverteilung
$\Gamma(\cdot)$ ... Gammafunktion
$\pi(\theta)$ ... A-priori-Verteilung
$\pi(\theta|D)$ ... A-posteriori-Verteilung
θ ... allg. Parameter
Θ ... Parameterraum
$\Re(M)$ ... System aller Teilmengen
σ ... Streuung, Standardabweichung
σ^2 ... Varianz
σ_{xy} ... Kovarianz
$\tau(\theta)$ ... geraffter Parameter
ξ ... Ereignisfeld

A_θ ... Alternativverteilung
$B_{n,\theta}$ ... Binomialverteilung
Cov ... Kovarianz
D ... Daten
D_m ... Diskrete Gleichverteilung
Ex_τ ... Exponentialverteilung
EX ... Erwartungswert von X
$F(\cdot)$ ... Verteilungsfunktion
$F_{m,n}$ ... F-Verteilung
$F_{m,n,p}$ ... p-Fraktile der F-Verteilung
g_t ... Trendfunktion
$\hat{g}_t$ ... gleitender Mittelwert
$h_n(\cdot)$ ... relative Häufigkeit
$H_n(\cdot)$ ... absolute Häufigkeit
$H_{N,A,n}$ ... Hypergeometrische Verteilung
${}_L I_{0t}^P$ ... Preisindex nach Laspeyres
${}_P I_{0t}^P$ ... Preisindex nach Paasche
I_{0t}^W ... Wertindex
$l(\theta;x_1,\ldots,x_n)$ ... Plausibilitätsfunktion
$LN(\cdot,\cdot)$ ... logarithmische Normalverteilung
M ... Merkmalraum
$M_{n,\theta_1,\ldots,\theta_m}$ ... Multinomialverteilung
MAD ... mittlere absolute Abweichung
$N(\cdot,\cdot)$ ... Normalverteilung
$p(\cdot)$ ... Punktwahrscheinlichkeit
P_μ ... Poisson-Verteilung
Q ... Yule'scher Assoziationskoeffizient
r ... empirischer Korrelationskoeffizient
R ... Stichproben-Korrelationskoeffizient
s ... empirische Streuung/Standardabweichung
s^2 ... empirische Varianz
s_g^2 ... empirische Varianz für gruppierte Daten
S_n^2 ... Stichprobenvarianz
T ... Schätzfunktion
Teststatistik
$\mathcal{T}$ ... Familie von unverzerrten Schätzfunktionen
t_n ... t-Verteilung (Student-)
$U_{\cdot,\cdot}$ ... kontinuierliche (uniforme) Gleichverteilung
U_{ij} ... Unabhängigkeitszahl
u_p ... p-Fraktile der Standard-Normalverteilung
V ... Verwerfungsraum
VarX ... Varianz von X
VCov ... Varianz-Kovarianz-Matrix
VK ... empirischer Variationskoeffizient
W ... Wahrscheinlichkeitsverteilung
W(E) ... Wahrscheinlichkeit des Ereignisses E
$\underset{\sim}{x}$ ... Median
$\bar{x}_g$ ... geometrisches Mittel
$\bar{x}_n$ ... arithmetisches Mittel
x_p ... p-Fraktile (i.a. der Normalverteilung)
$\bar{X}_n$ ... Stichprobenmittel

Index

#

A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

V

W

Y