

Unterlagen zur Vorlesung

# EINFÜHRUNG IN DIE MEDIZINISCHE INFORMATIK

Verfaßt von: H. Grabner  
Erweitert und ergänzt von: E. Schuster  
Version 2.1 vom 1.7.2002

## TEIL 2

<b>1. GRUNDBEGRIFFE DER STATISTIK.....</b>	<b>2</b>
1.1 SKALEN.....	3
1.2 LOKALISATIONSMABE .....	3
<b>2. GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITSTHEORIE .....</b>	<b>8</b>
2.1 BEDINGTE WAHRSCHEINLICHKEIT.....	11
<b>3. ANWENDUNGEN DER STATISTIK UND WAHRSCHEINLICHKEITSTHEORIE.....</b>	<b>12</b>
3.1 DIE BAYESSCHE FORMEL.....	12
3.2 KORRELATION .....	13
3.3 CLUSTER-ANALYSE (MUSTERERKENNUNG) .....	16
3.3.1 <i>Algorithmen zur Bildung von Klassen</i> .....	18
<b>4. ROC-ANALYSE .....</b>	<b>19</b>
4.1 ROC-KURVEN IN DER MEDIZIN .....	19
4.2 HERLEITUNG EINER ROC-KURVE.....	23
4.2.1 <i>Aus diskreten Daten</i> .....	23
4.2.2 <i>Aus stetigen Daten</i> .....	27
<b>5. ENTSCHEIDUNGSMODELLE UND -STRATEGIEN.....</b>	<b>31</b>
5.1 STATISTISCHE VERFAHREN.....	32
5.2 PRÄDIKATENLOGISCHE VERFAHREN.....	34
5.3 SEQUENTIELLE STRATEGIEN.....	35
5.4 COMPUTERUNTERSTÜTZTE DIAGNOSE.....	37

---

## 1. GRUNDBEGRIFFE DER STATISTIK

---

In vielen Anwendungen der Medizinischen Informatik werden Begriffe verwendet, die der Statistik entnommen sind.

Die Aufgabe der beschreibenden oder **deskriptiven Statistik** besteht darin, empirisch gewonnene Daten

- zu ordnen
- durch Maßzahlen (wie Mittelwert, Streuung, ...) zusammenzufassen und diese
- tabellarisch oder graphisch darzustellen

Die Bedeutung der deskriptiven Statistik für die Medizin ist sehr groß, da das medizinische Wissen in seinen wesentlichen Teilen auf Erfahrungen am kranken Menschen beruht.

Die statistischen Methoden versetzen den Arzt in die Lage, sich auf einzelne Merkmale, wie Symptome oder Messergebnisse eines Patienten zu beschränken. Man ist nicht gezwungen, die Untersuchungen auf die gesamte Menschheit auszudehnen!

Ein Merkmal kann verschiedene **Ausprägungen** annehmen.

Grundsätzlich unterscheidet man:

- qualitative Merkmale
- quantitative Merkmale

**Qualitative** Merkmale sind Merkmale, die sich nicht zahlenmäßig erfassen, sondern nur benennen lassen.

Beispiel:

- Blutgruppe
- Augenfarbe

**Quantitative**<sup>1</sup> Merkmale sind Merkmale, die zahlenmäßig erfaßbar sind.

Beispiel: Ergebnis der Messung des Blutzuckerwertes (Glucose)

Quantitative Merkmale lassen sich weiter unterteilen in

- diskrete Merkmale
- stetige Merkmale

**Diskrete Merkmale** nehmen nur bestimmte Werte an.

Beispiel: Zahl der Kinder in einer Familie.

Alle Zählvorgänge führen zu diskreten Werten.

**Stetige Merkmale** können in einem definierten Bereich jeden reellen Zahlenwert annehmen.

Beispiel:

Die Körpergröße kann beliebig genau gemessen werden. Die Genauigkeit hängt nur von der Meßeinrichtung ab.

---

<sup>1</sup> lat. quantum = Menge

Eine zu große Genauigkeit ist in der Regel sinnlos. Deshalb wird die Körpergröße im allgemeinen in ganzen cm oder in Meter mit 2 Dezimalen angegeben. Auf diese Weise wird das stetige Merkmal klassiert, d.h. in **Klassen** eingeteilt. Die Klasse 186 cm beispielsweise umfaßt alle Größen zwischen 185,500 cm und 186,499 cm. Dabei ist die Klassenbreite nur vom Ziel der jeweiligen statistischen Analyse abhängig. Genauso könnte man auch eine grobe Klassenbreite mit 10 cm wählen. Grundsätzlich gilt, daß bei einer Vergrößerung der Klassenbreite Information verlorengeht.

Diskrete und qualitative Merkmale können ebenfalls in Klassen zusammengefaßt werden.

## 1.1 Skalen

Eine **Nominalskala** entsteht durch willkürliche Verschlüsselung eines qualitativen Merkmals.

Beispiel (Verschlüsselung des Geschlechts):

1 = männlich	2 = weiblich	3 = unbekannt
M = männlich	W = weiblich	U = unbekannt

Eine **Rang-** oder **Ordinalskala** eines qualitativen Merkmals ist dann gegeben, wenn die Rangfolge definiert ist, nicht jedoch die Differenz zwischen den Merkmalsausprägungen.

Beispiel: Verschlüsselung des Behandlungsergebnisses

1= geheilt 2= gebessert 3= gleichbleibend 4= verschlechtert 5= Tod

Eine **Intervallskala** liegt vor, wenn die Differenzen so definiert sind, daß gleiche Differenzen einem gleichen Merkmalsunterschied entsprechen, wobei aber kein absoluter Nullpunkt existiert.

Beispiel: Grad Celsius

Eine **Verhältnisskala** ist eine Intervallskala mit zusätzlich definiertem absoluten Nullpunkt.

Beispiel: Grad Kelvin, Gewichtsbestimmung

## 1.2 Lokalisationsmaße

Zu den **Lokalisationsmaße** gehören

- die Durchschnittswerte (arithmetischer Mittelwert, Modalwert und Median)
- die Perzentile

Der **arithmetische Mittelwert** ist:

$$\bar{x} = \frac{\text{Summe aller } x\text{-Werte}}{\text{Anzahl aller Werte}}$$

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^n x_i$$

Beispiel (Aufenthaltsdauer von 20 Patienten mit Krankheit X):

In einer Urliste (Tab.1) wird die Aufenthaltsdauer von ausgewählten Patienten festgehalten.

Patient	Aufenthalt (Tage)
1	10
2	8
3	14
4	17
5	21
6	6
7	13
8	10
9	9
10	9
11	9
12	14
13	13
14	27
15	11
16	9
17	10
18	12
19	7
20	11

Tab. 1:

Aus Tab. 1 ergibt sich als arithmetischer Mittelwert  $\bar{x} = \frac{240}{20} = 12$  Tage.

Ordnet man die Urliste nach dem Entlassungstag  $x$  dann ergibt sich Tabelle 2:

Entlassung am Tage $x$	Absolute Häufigkeit $H(x)$	Relative Häufigkeit $h(x)$	Summenhäufigkeit
6	1	0.05	0.05
7	1	0.05	0.10
8	1	0.05	0.15
9	4	0.20	0.35
10	3	0.15	0.50
11	2	0.10	0.60
12	1	0.05	0.65
13	2	0.10	0.75
14	2	0.10	0.85
15			
16			
17	1	0.05	0.90
18			
20			
21	1	0.05	0.95
22			
23			
24			
25			
26			
27	1	0.05	1.00

Tab. 2:

Unter dem **Modalwert** versteht man den am häufigsten auftretenden Wert. Im Beispiel (Tab. 2) ist dies  $\tilde{x} = 9$ . An diesem Tag ist die relative Häufigkeit  $h(\tilde{x}) = 0.2$  am größten.

Der **Median (Zentralwert)** ist jener Wert, der von höchstens der Hälfte aller Werte unterschritten und von höchstens der Hälfte aller Werte überschritten wird.

Im Beispiel (Tab. 2) ist dies  $\bar{x} = 10.5$

(Konvention:

Im Falle, daß  $F(x) = 0,5$  exakt für einen Wert  $x$  - im Beispiel  $x = 10$  - zutrifft, liegt der Median per definitionem in der Mitte von  $x$  und  $(x+1)$ ).

Der Durchschnitt kann also durch 3 Lokalisationsmaße festgelegt werden:

- arithmetischer Mittelwert  $\bar{x} = 12$
- Modalwert  $\tilde{x} = 9$
- Median  $\bar{x} = 10.5$

Eigenschaften:

- Der arithmetische Mittelwert wird sehr stark von sogenannten Ausreißern beeinflusst. Ausreißer können das Ergebnis eines Fehlers sein, z.B. eines Meßfehlers, Übertragungsfehlers.
- Der Modalwert ist gegen Ausreißer absolut unempfindlich
- der Median ist gegen Ausreißer relativ unempfindlich

Am häufigsten wird der arithmetische Mittelwert  $\bar{x}$  verwendet.

Mit Hilfe der **Dispersions-** oder **Streuungsmaße** wird die Streuung der Meßwerte angegeben. Das einfachste Streuungsmaß ist die Spannweite (Variationsbreite, Range). Diese ist definiert als die Differenz zwischen dem höchsten und dem niedrigsten Wert.

Beispiel:

Aus der Urliste (Tab. 1) ergibt sich die Spannweite zu 27 Tage - 6 Tage = 21 Tage.

Der große Nachteil der Spannweite besteht darin, daß sie von Ausreißern sehr stark beeinflusst wird.

Als häufigstes Dispersionsmaß wird die empirische Varianz (abgekürzt Varianz) verwendet. Sie ist definiert als durchschnittliches Abweichungsquadrat:

$$s^2 = \frac{1}{n-1} \times \sum_{i=1}^n (x_i - \bar{x})^2$$

$x_i$  ..... i-ter Meßwert  
 $\bar{x}$  ..... arithmetischer Mittelwert  
 $n$  ..... Zahl der Meßwerte

Im Nenner wird oft statt  $(n-1)$  nur  $n$  angegeben.

Es gilt:

- Varianz innerhalb der Stichprobe: Nenner  $n$
- Schätzwert für die Varianz innerhalb der Grundgesamtheit: Nenner  $n-1$

Berechnung der Varianz aus Tab. 2:

$x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$H(x_i)$	$H(x_i) \times (x_i - \bar{x})^2$
6	-6	36	1	36
7	-5	25	1	25
8	-4	16	1	16
9	-3	9	4	36
10	-2	4	3	12
11	-1	1	2	2
12	0	0	1	0
13	1	1	2	2
14	2	4	2	8
17	5	25	1	25
21	9	81	1	81
27	15	225	1	225
			20	468

$$\text{Varianz } s^2 = \frac{1}{n-1} \times \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{468}{19} = 24,63$$

Vielfach wird statt der Varianz  $s^2$  die Quadratwurzel derselben verwendet. Diese wird **Standardabweichung** genannt.

Im Beispiel ergibt sich  $s = \sqrt{24,63} = 4,96$

Sie hat große Bedeutung, weil im Falle der **Gauß-** oder **Normalverteilung**<sup>2</sup>

- $\approx 68\%$  der Werte im Intervall  $\bar{x} \pm s$   
also in einer Entfernung von höchstens einer Standardabweichung  $s$  vom Mittelwert  $\bar{x}$
- $\approx 95\%$  der Werte im Intervall  $\bar{x} \pm 2 \times s$
- $\approx 99,7\%$  der Werte im Intervall  $\bar{x} \pm 3 \times s$  liegen.

Der Bereich von einer Standardabweichung vom Mittelwert ist somit  
 $12 - 4,96 = 7,04$  und  $12 + 4,96 = 16,96$   
 also  $[7,04, 16,96]$ . In diesen Bereich fallen 15 Patienten.

Der **Variationskoeffizient** gibt an, wieviel Prozent vom arithmetischen Mittelwert die Standardabweichung ausmacht:

$$\text{Variationskoeffizient} = \frac{s}{\bar{x}} \times 100\%$$

Im Beispiel ergibt er sich zu  $(4,96/12) \cdot 100\% = 41,33\%$

Vorteil (des Variationskoeffizienten):

- Er ist von der gewählten Einheit unabhängig.  
Aus diesem Grund wird er sehr häufig in biochemischen Labors verwendet, da bei diesen Messungen die verschiedensten Einheiten auftreten: mg, mval, mmol, I.E. etc.

<sup>2</sup> die am häufigsten auftretende Verteilung

In vielen Fällen erweist es sich als günstig, die Daten in einem Diagramm graphisch (z. B. Stabdiagramm, Kreis-, Tortendiagramm,...) darzustellen.

Daneben gibt es Darstellungen in Form von **Histogrammen**.

Die **Verteilungsfunktion** (kumulative Häufigkeitsfunktion, Summenhäufigkeitsfunktion)  $F(x)$  gibt an, welcher Anteil aller Werte kleiner oder gleich dem Wert  $x$  ist. Man bezeichnet die Verteilungsfunktion auch als.

In Tab. 2 ist die Verteilungsfunktion  $F(x)$  in der letzten Spalte angeführt. Diese läßt sich ebenfalls sehr anschaulich graphisch darstellen. Bei diskreten Daten erhält man stets eine Treppenkurve, wobei die Höhe der Stufen im Beispiel der Zahl der am jeweiligen Tag entlassenen Patienten entspricht.

Oft will man wissen, welcher  $x$ -Wert von z.B. 25% aller Werte unterschritten oder erreicht wird, d.h. man sucht bei bekanntem  $F(x)$ -Wert den zugehörigen  $x$ -Wert. Dies geschieht mit den **Perzentil-Tabellen**.

Das Perzentil wird auch als Quantil oder Fraktile bezeichnet.

Unter **Quartilen** versteht man das 25., 50. und 75. Perzentil. Der Median entspricht dem 50. Perzentil.

Es kann sein, daß mehrere Perzentile auf denselben  $x$ -Wert fallen, (z. B. in Tab. 2 fallen das 25. und das 30. Perzentil beide auf den 9. Tag).

Die eindeutige Zuordnung zwischen den  $F(x)$ - und  $x$ -Werten wird umso eher möglich, je höher die Anzahl und je kleiner die Höhe der einzelnen Stufen in der Treppenkurve ist. Dies leitet zu Verteilungen von stetigen Variablen über.

Die Verteilung einer stetigen Variablen geht von unendlich vielen Meßwerten aus und führt zu einer stetigen Verteilungsfunktion ohne Stufen.

In diesem Fall läßt sich jedem  $x$ -Wert genau ein  $F(x)$ -Wert zuordnen und umgekehrt.

## 2. GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITSTHEORIE

Die **Wahrscheinlichkeitstheorie** macht Aussagen über zufällige, d.h. nicht sicher voraussagbare Ereignisse. Ein **Ereignis** A ist dabei das Ergebnis eines Versuches bzw. einer Beobachtung oder die Zusammenfassung von einzelnen Ergebnissen  $e_1, e_2, \dots$ .

### Beispiele:

- Ergebnis eines Münzwurfs, Ergebnis einer Ziehung im Lotto, Wirkung eines Medikaments, Verlauf einer Erkrankung
- Das Messen der "Blutsenkung" hat als mögliche Ergebnisse:  $e_1 = 1, e_2 = 2, \dots$ , wobei die Zahlenangabe in der Maßeinheit "mm" erfolgt.  
Der üblicherweise verwendete Normalbereich für die Blutsenkung nach 1 Stunde ist:  
8 - 12  
Die Ergebnisse  $e_8$  bis  $e_{12}$  bilden zusammengefaßt das Ereignis  
"Blutsenkung normal von 8 - 12"

Die Wahrscheinlichkeit eines Ereignisses A ist eine diesem zugeordnete reelle Zahl  $P(A)$ . Man schreibt oft auch  $\text{Pr}(A)$ , wobei Pr für engl. probability steht. Dies vor allem dann, wenn es Verwechslungen mit der Potenzmenge  $P(M)$  einer Menge M geben könnte.

Für die Wahrscheinlichkeit müssen die **Axiome von Kolmogoroff** erfüllt sein:

- $P(A) \geq 0$  nicht-negativ
- $P(S) = 1$  S ..... Menge aller Ergebnisse
- $P(A \vee B) = P(A) + P(B)$  A, B mit:  $A \cap B = \emptyset$

Hinweis: Ist  $A \cap B \neq \emptyset$  dann lautet der Additionssatz:

$$P(A \vee B) = P(A) + P(B) - P(A \cap B)$$

Man spricht in diesem Fall von vereinbaren Ereignissen.

Ist der Durchschnitt leer, dann heißen die Ereignisse unvereinbar.

Der Begriff Wahrscheinlichkeit bezieht sich auf ein einzelnes Ereignis.

Die Gesamtheit aller möglichen Ereignisse wird durch die Begriffe

- Häufigkeitsverteilung (Häufigkeitsfunktion)
- Wahrscheinlichkeitsverteilung (kurz Verteilung)

wie folgt charakterisiert.

Ein zufallsabhängiges Ereignis wird zu einer **Zufallsvariablen**, indem es zahlenmäßig erfaßt oder verschlüsselt wird.

Zufallsvariable werden üblicherweise mit Großbuchstaben X,Y,... gekennzeichnet, der zugeordnete Wert mit Kleinbuchstaben x,y,....

### Beispiele:

- Dem Merkmal "Körpergröße" wird die Zufallsvariable X zugeordnet.  
 $x = 175$  cm ist eine bestimmte Ausprägung des Merkmals
- Beim Werfen einer Münze hat die Zufallsvariable X als mögliche Ausprägungen:  
Zahl = 0      Wappen = 1

Durch die Zuordnung auf eine reelle Zahl wird ein Wahrscheinlichkeitsmaß definiert, das Verteilung genannt wird.

Wenn die Zufallsvariable nur endlich viele oder abzählbar unendlich viele Punkte  $x_i$ ,  $i=1,2,\dots$  annimmt, spricht man von einer **diskreten Verteilung**.

Die Wahrscheinlichkeit, daß die Zufallsvariable  $X$  den Wert  $x$  annimmt, ist  $P(X=x)$  oder kurz  $P(x)$ . Die Funktion  $P(x)$  heißt **Häufigkeitsfunktion**.

Beispiele:

- **Gleichverteilung**

Jedem möglichen Ergebnis wird die gleiche Wahrscheinlichkeit zugeordnet

$$p(x_1) = p(x_2) = \dots = p(x_n) = \frac{1}{n}$$

Beispiel (Würfel):

Die Zuordnung der 6 verschiedenen Würfeleregebnisse zu den 6 Punkten  $1, 2, \dots, 6$  ist eine Zufallsvariable und erzeugt eine Gleichverteilung auf den Zahlen 1 bis 6.

$$p(1) = p(2) = \dots = p(6) = \frac{1}{6}$$

Die Gleichverteilung auf den ersten  $n$  Zahlen aus  $N$  ist

$$P(x) = \frac{1}{n} \quad (x = 1, \dots, n)$$

- **Geometrische Verteilung**

$$P(x) = q^{x-1} \times p \quad \text{mit } p + q = 1$$

Die Menge der Werte, die die Zufallsvariable annehmen kann, ist abzählbar unendlich.

Die **Verteilungsfunktion**  $F(x) = P(X \leq x)$  gibt die Wahrscheinlichkeit an, daß die Zufallsvariable einen Wert kleiner als  $x$  oder gleich  $x$  annimmt.

Im **diskreten Fall** ist

$$F(x) = \sum_{z \leq x} P(z)$$

Es ist  $F(-\infty) = 0$ ,  $F(\infty) = 1$ .

Beispiel:

- Gleichverteilung: 
$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{i}{n} & i \leq x < i+1 \quad i = 1, \dots, n-1 \\ 1 & x \geq n \end{cases}$$

- geometrische Verteilung: 
$$F(x) = \begin{cases} 0 & x \leq 0 \\ \sum_{j=1}^i q^{j-1} \times p = 1 - q^i & i \leq x < i+1 \quad i = 1, 2, \dots \end{cases}$$

Im **stetigen Fall** nimmt eine Zufallsvariable alle Werte in einem endlichen Intervall, auf einer Halbachse oder der ganzen reellen Achse an. An die Stelle der Häufigkeitsfunktion  $P(x)$  tritt die **Wahrscheinlichkeitsdichte**  $f(x)$  (kurz Dichte).

Zur Dichte und Verteilungsfunktion im stetigen Fall gelangt man nach Einführung differentiell kleiner Intervalle:

$f(x) \Delta x$  ist für kleine  $\Delta x$  angenähert die Wahrscheinlichkeit, daß die Zufallsvariable  $X$  einen Wert zwischen  $x$  und  $x + \Delta x$  annimmt. Somit ist die Verteilungsfunktion nach LEIBNIZ (1. Mittelwertsatz der Integralrechnung)

$$F(x) = \int_{-\infty}^x f(z) dz$$

Die Dichte ist dann gegeben durch  $f(x) = F'(x)$  bis auf endlich viele x.

Beispiel: Gleichverteilung im Intervall  $[a, b]$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Beispiel:

Die Verteilung der Rundungsfehler ist eine Gleichverteilung mit  $a = -1/2$  und  $b = +1/2$

Die **Normalverteilung** (Gaußverteilung) ist eine stetige Verteilung, die auf der ganzen reellen Achse definiert ist. Ihre Wahrscheinlichkeitsdichte ist

$$f(x) = \frac{1}{\sqrt{2 \times \pi \times \sigma}} \times e^{-\frac{(x-\mu)^2}{2 \times \sigma^2}} \quad -\infty < x < \infty$$

Diese Dichte hat die Form einer Glockenkurve mit

- dem Maximum an der Stelle  $\mu$  und
- den Wendepunkten an den Stellen  $\mu - \sigma$  und  $\mu + \sigma$

In der Medizin ist diese Verteilung überaus häufig anzutreffen.

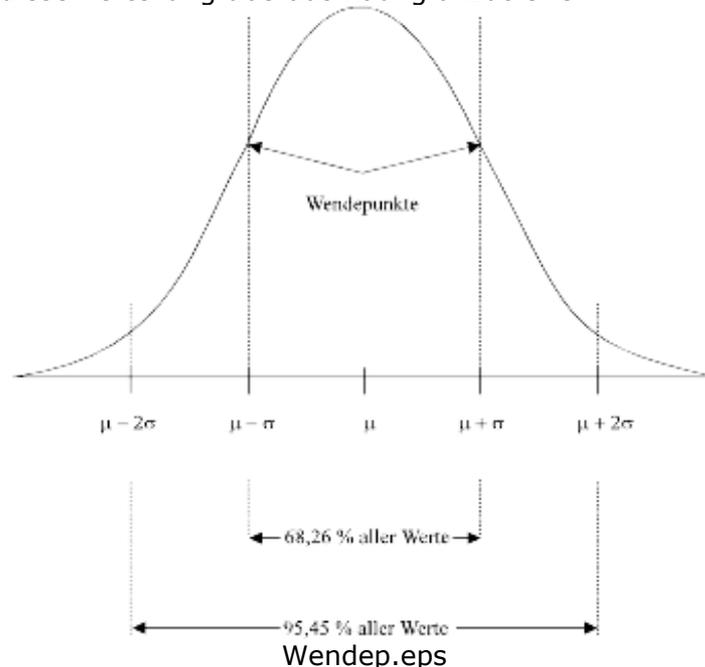


Abb. 3.1: Gauß- oder Normalverteilung

Andere bekannte Verteilungen sind:

- Lognormalverteilung
- Poissonverteilung

Den Zusammenhang zwischen der Wahrscheinlichkeit und den real erfahrbaren relativen Häufigkeiten stellt das **Gesetz der großen Zahlen** her. Danach konvergiert bei genügend großer Anzahl von Wiederholungen eines Versuches die relative Häufigkeit gegen die Wahrscheinlichkeit dieses Ereignisses. Man kann daher die relativen Häufigkeiten als Schätzwerte verwenden.

Die Auswertung von Beobachtungen führt zu einer Häufigkeitsfunktion. Den Forscher interessiert zumeist nicht die untersuchte Stichprobe, wie z.B. die Aufenthaltsdauer bei den 20 Patienten mit Krankheit X, sondern die zugrundeliegende Grundgesamtheit, also die Aufenthaltsdauer von allen Patienten mit der Krankheit X. Die Stichprobe wird dazu benutzt, die Wahrscheinlichkeitsverteilung zu schätzen, welche für die Grundgesamtheit gültig ist.

Die wichtigsten Parameter der Wahrscheinlichkeitsverteilung sind

- der Erwartungswert  $\mu$
- die Varianz  $\sigma^2$ .

Der Erwartungswert  $\mu$  wird durch den arithmetischen Mittelwert  $\bar{x}$ , die Varianz  $\sigma^2$  durch die Varianz  $s^2$  geschätzt.

## 2.1 Bedingte Wahrscheinlichkeit

Sei B ein Ereignis mit  $P(B) > 0$ . Dann ist die bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung, dass B eingetreten ist, gegeben durch

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Zwei Ereignisse heißen unabhängig, wenn  $P(A|B) = P(A)$ , d.h. das Wissen über A wird durch B nicht verändert.

Bei unabhängigen Ereignissen gilt der **Multiplikationssatz**:

$$P(A \wedge B) = P(A) \times P(B)$$

Beispiele von bedingten Wahrscheinlichkeiten sind:

- Risikofaktoren
- Nebenwirkungen

Beispiel:

So gilt

$$P(\text{AIDS unter der Bedingung RAUSCHGIFTSÜCHTIGKEIT}) > P(\text{AIDS})$$

### 3. ANWENDUNGEN DER STATISTIK UND WAHRSCHEINLICHKEITSTHEORIE

#### 3.1 Die BAYESsche Formel

Ein Ansatz der computerunterstützten Diagnose ist die BAYESsche Formel. Sie dient dazu auszurechnen, mit welcher bedingten Wahrscheinlichkeit  $P(K_i|S)$  eine Krankheit  $K_i$  vorliegt, wenn das Symptom S festgestellt wurde.

Per Definitionem ist:

$$P(S|K_i) = \frac{P(K_i \wedge S)}{P(K_i)} \Rightarrow P(K_i \wedge S) = P(S|K_i) \times P(K_i) \quad (A)$$

$$P(K_i|S) = \frac{P(K_i \wedge S)}{P(S)} \Rightarrow P(K_i \wedge S) = P(K_i|S) \times P(S) \quad (B)$$

Gleichsetzen von (A) und (B) (=Formel für das gleichzeitige Vorliegen von  $K_i$  und S):

$$P(K_i \wedge S) = P(K_i) \times P(S|K_i) = P(S) \times P(K_i|S) \quad (4.1)$$

Umformen:

$$P(K_i|S) = \frac{P(K_i) \times P(S|K_i)}{P(S)} \quad (4.2)$$

BAYESsche Formel für eine Krankheit

Das Symptom S spricht vor allem dann mit hoher Wahrscheinlichkeit für die Krankheit  $K_i$ , wenn:

- $P(S|K_i)$  groß ist                      d.h. S tritt bei der Krankheit  $K_i$  häufig auf
- $P(K_i)$  groß ist                        d.h. die Krankheit  $K_i$  tritt insgesamt häufig auf
- $P(S)$  klein                              d.h. das Symptom S tritt insgesamt selten auf,  
d.h. S ist relativ spezifisch für die Krankheit

Im Prinzip orientiert sich der Arzt bei der Erstellung einer Differentialdiagnose nach der Formel (4.2), wenn er die Krankheiten  $K_1$  bis  $K_n$  durchgeht.

Wenn

- $K_1, K_2, \dots, K_n$  alle Zustände (inkl. Gesundheit) sind, bei denen das Symptom S vorkommen kann (Vollständigkeit)
- die  $K_i$  paarweise disjunkt sind

dann ist:

$$P(S) = P(K_1 \wedge S) \vee P(K_2 \wedge S) \vee \dots \vee P(K_n \wedge S) = \sum_{i=1}^n P(K_i \wedge S) \quad (4.3)$$

$P(S)$  läßt sich somit (Einsetzen von (A) in (4.3)) schreiben als:

$$P(S) = \sum_{i=1}^n P(K_i) \times P(S|K_i) \quad (4.4)$$

Somit lautet die BAYESsche Formel ((4.4) in (4.2) einsetzen):

$$P(K_i|S) = \frac{P(K_i) \times P(S|K_i)}{\sum_{j=1}^n P(K_j) \times P(S|K_j)}$$

Beispiel:

Mit dem Test S wird nach der Krankheit  $K_0$  gesucht. Bei Patienten, die an  $K_0$  leiden, spricht der Test mit einer Wahrscheinlichkeit von 99% positiv an, d.h.  $P(S|K_0) = 0.99$ . Insgesamt ist der Test in 1% der Fälle (also bei Gesunden und Kranken zusammen) positiv., d.h.  $P(S) = 0,01$ .

Frage:

Mit welcher Wahrscheinlichkeit ist ein Patient an  $K_0$  erkrankt, wenn der Test bei ihm positiv ist?

- 1) Wenn  $P(K_0) = 0.01$  (jeder 100ste ist erkrankt), dann ist (nach (4.2))

$$P(K_0|S) = \frac{0.01 \times 0.99}{0.01} = 0.99$$

Ein Patient mit positivem Testergebnis ist also mit 99%iger Wahrscheinlichkeit erkrankt.

- 2) Wenn  $P(K_0) = 0.0001$  (jeder 10.000ste ist erkrankt), dann ist (nach (4.2))

$$P(K_0|S) = \frac{0.0001 \times 0.99}{0.01} = 0.0099$$

Ein Patient mit positivem Testergebnis ist also nur mit 0,99%iger Wahrscheinlichkeit erkrankt.

Der Unterschied zwischen Fall 1) und 2) :

- Fall 1  
Der Test ist genauso häufig positiv, wie die Krankheit vertreten ist; er spricht also - von ganz wenigen Ausnahmen abgesehen - nur spezifisch auf die Krankheit K an.
- Fall 2  
Der Test ist hundertmal häufiger positiv als die Krankheit K auftritt. Daraus folgt, daß der Test unspezifisch ist, d.h. nicht nur auf K anspricht.

(Quelle: V.HARMS Biomathematik, Statistik und Dokumentation. Harms-Verlag Kiel 1982).

### 3.2 Korrelation

k (skalare) Merkmale  $y_1, y_2, \dots, y_k$  spannen einen k-dimensionalen Merkmalsraum auf.

Die Daten der beobachteten Objekte sind k-dimensionale Punkte (bzw. Vektoren) in diesem Raum.

Bei vielen Verfahren in der medizinischen Anwendung wird versucht, die Dimension des k-dimensionalen Merkmalsraums zu reduzieren.

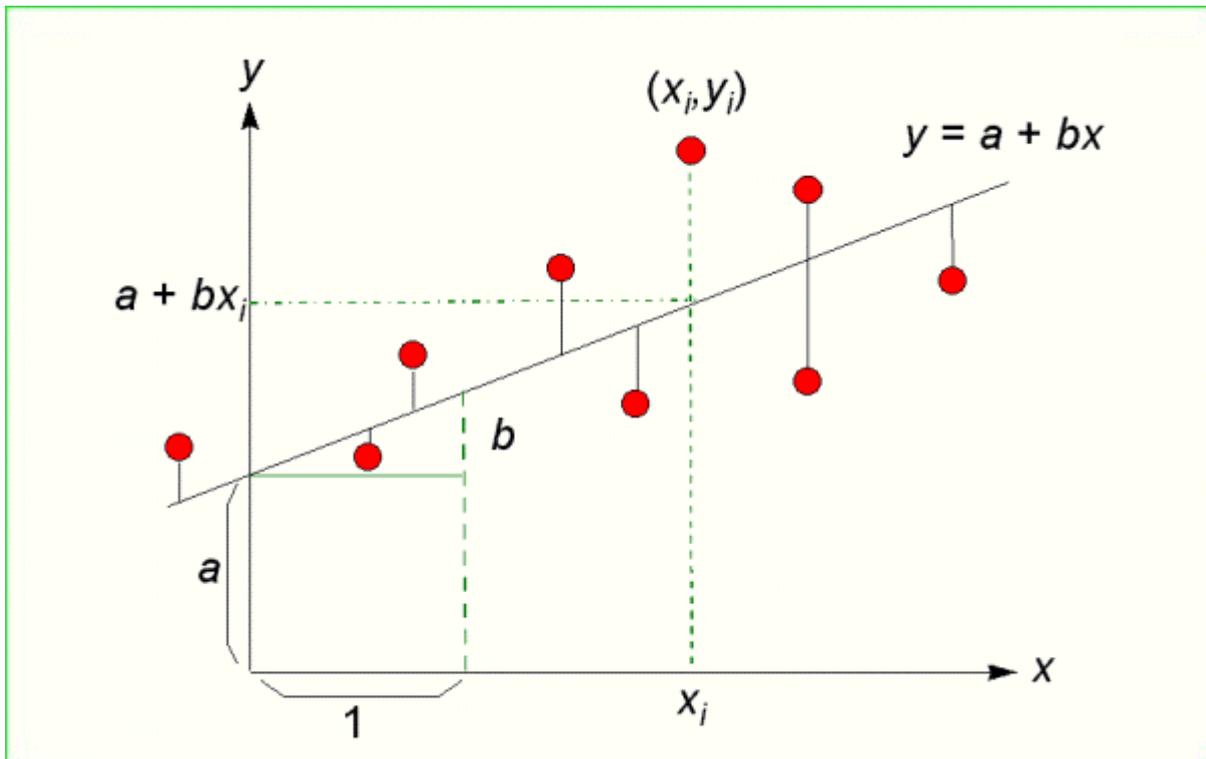
Um die Dimension des Merkmalsraumes zu reduzieren, analysiert man die

Abhängigkeiten der einzelnen Daten zueinander. Wenn z.B. ein Merkmal  $y_i$  ein anderes

Merkmal  $y_j$  vollständig determiniert, dann liefert  $y_i$  keine Information mehr und kann

weggelassen werden. Mathematisch gesehen erfolgt eine Projektion des k-dimensionalen Raumes auf einen Raum mit niedrigerer Dimension.  
Man läßt also einzelne Merkmale weg, die in hohem Maße von anderen Merkmalen abhängig sind.

Ein bekanntes **Maß für die (lineare) Abhängigkeit** von 2 Zufallsvariablen ist der Korrelationskoeffizient. Der Korrelationskoeffizient gibt an, wie stark die Punktwolke der beobachteten Objekte um die Regressionsgerade streut. Die Regressionsgerade entsteht, wenn man eine Schätzung der y-Werte aus den x-Werten vornimmt.



LineareRegression.gif

In dem Diagramm ist das Merkmal X und das Merkmal Y von 9 Personen eingetragen. Die Punktwolke läßt vermuten, daß im Rahmen der zufallsbedingten Streuung ein linearer Zusammenhang zwischen dem Merkmal X und dem Merkmal Y besteht. Man kann daher versuchen, den y-Wert aus dem x-Wert zu schätzen mittels der (Regressions)Geraden

$$y = b_{yx} \times x + a_{yx}$$

Der Regressionskoeffizient  $b_{yx}$  gibt die Steigung an;  $a_{yx}$  ist der Schnitt mit der y-Achse.

Berechnung des Regressionskoeffizientens:

$$b_{yx} = \frac{\text{Kovarianz}}{\text{Varianz der } x\text{-Werte}} = \frac{\frac{1}{n} \times \sum_{i=1}^n (x_i - \bar{x}) \times (y_i - \bar{y})}{\frac{1}{n} \times \sum_{i=1}^n (x_i - \bar{x})^2}$$

Wie gut die Regressionsgerade die Werte von Y wiedergibt misst der **Produktmoment-Korrelationskoeffizient** r.

Berechnung des Produktmoment-Korrelationskoeffizient r:

$$r = \frac{\text{Kovarianz}}{(S \text{ tandardabweichung der } x\text{-Werte}) \times (S \text{ tandardabweichung der } y\text{-Werte})}$$

$$r = \frac{\frac{1}{n} \times \sum_{i=1}^n (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\frac{1}{n} \times \sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\frac{1}{n} \times \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Regressionskoeffizient  $b_{xy}$  und Korrelationskoeffizient r haben stets dasselbe Vorzeichen. Der Korrelationskoeffizient r kann Werte zwischen -1 und +1 annehmen. Bei  $r = -1$  oder  $r = +1$  liegen alle Punkte auf einer Geraden.  $r = 0$  bedeutet, daß zumindest keine lineare Abhängigkeit zwischen x und y besteht.

Der **multiple Korrelationskoeffizient** ist ein Maß für die Abhängigkeit einer Zufallsvariablen Y in bezug auf die Menge  $\{X_1, X_2, \dots, X_n\}$

Im Falle des linearen Modells ist die Abhängigkeit einer Zielgröße von Einflußgrößen allgemein:

$$X = Y^T \beta + Z^T h + E$$

$Y^T, Z^T$  .....die transponierten Matrizen zu Y, Z

X ..... ein Vektor von Zielgrößen

Y .....eine durch das Modell festgelegte Matrix qualitativer deterministischer Einflußgrößen

Z ..... eine Matrix quantitativer Einflußgrößen

$\beta$  ..... ein Vektor von "Effekten"

h ..... ein Vektor von Regressionskoeffizienten

E ..... ein Vektor von zufälligen Fehlern

Voraussetzung: die Fehler sind multivariat normalverteilt.

Ist  $Y = 0$

dann handelt es sich um eine Regressionsanalyse

ist  $Z = 0$

dann handelt es sich um eine Varianzanalyse

ist weder Y noch Z gleich 0,

dann handelt es sich um eine Kovarianzanalyse

Zusammenhänge zwischen Komponenten des Zufallsvektors X werden mit Hilfe der **Hauptkomponentenanalyse** untersucht. Dazu wird der Zufallsvektor X durch eine Lineartransformation so transformiert, daß die Varianz von  $\beta X$  maximal wird. Dies führt auf ein Eigenwertproblem. Geometrisch wird das Koordinatensystem solange rotiert und die Achsen so gewählt, daß sie orthogonal sind und in ihrer Richtung die Varianz maximal wird. Ziel ist dabei die Reduzierung der Dimension des Zufallsvektors. Man versucht die Variation der Daten anhand weniger unabhängiger Linearkombinationen der ursprünglichen Zufallsvariablen zu erklären.

### 3.3 Cluster-Analyse (Mustererkennung)

#### Gegeben:

Stichprobe aus  $n$  Objekten mit einem  $k$ -dim. Merkmalsvektor  $\vec{X} = (x_1, x_2, \dots, x_k)$ .

Es wird vermutet, daß die Stichprobe aus  $m$  verschiedenen Grundgesamtheiten

$V_1, V_2, \dots, V_m$  stammt.

#### Ziel:

Die Stichprobe soll so in Untermengen (**Klassen**) zerlegt werden, daß die Elemente der gleichen Klasse möglichst homogen und Elemente verschiedener Klassen möglichst inhomogen sind. Dabei muß  $m < n$  sein, sonst bildet jedes Objekt eine Klasse.

Die beobachteten Objekte können als Punkte im  $k$ -dimensionalen Raum aufgefaßt werden, der durch den Merkmalsvektor  $X$  aufgespannt wird. Gesucht ist eine Zerlegung des Raumes im obigen Sinne.

Zur Lösung dieser Fragestellung verwendet man Verfahren, die mit Ähnlichkeiten oder Distanzen arbeiten und die unter dem Namen **Cluster-Analyse** bekannt sind.

Darunter versteht man Verfahren, bei denen "feine" Unterschiede der einzelnen Elemente vernachlässigt werden. Dies führt zu einer Klassenbildung, wobei man bewusst in Kauf nimmt, daß damit eine Redundanzvermehrung einhergeht.

Unbedingt erforderlich ist dabei, daß ein Maß für die Distanz oder Ähnlichkeit definiert wird.

Man benötigt:

- Distanzfunktion für Objekte
- Distanzfunktion für Mengen (Klassen)

Die Distanzfunktion für 2 Punkte  $X, Y$  des  $k$ -dimensionalen Raumes ist eine reelle Zahl, für die gilt:

- $f(X, Y) \geq 0$  nicht-negativ
- $f(X, Y) > 0$  für  $X \neq Y$  positiv
- $f(X, Y) = f(Y, X)$  symmetrisch

Gilt zusätzlich die Dreiecksungleichung

$$f(X, Y) + f(Y, Z) \geq f(X, Z)$$

(d.h. ist der Weg von  $X$  über  $Y$  nach  $Z$  mindestens so lang wie der Weg von  $X$  nach  $Z$ ) dann spricht man von einer **Metrik**.

Normiert man die Distanz auf  $[0, 1]$ , dann ist die **Ähnlichkeit** definiert als

$$g(X, Y) = 1 - f(X, Y)$$

$g = 1$  bedeutet maximale Ähnlichkeit,  $g = 0$  keine Ähnlichkeit.

Praktische Schwierigkeiten entstehen dann, wenn

- Daten fehlen (was in der Medizin oft der Fall ist)
- qualitative Merkmale vorliegen

Bei qualitativen Merkmalen kann man versuchen, anstatt der Merkmale Zufallsvariable zu betrachten, was auf eine numerische Codierung hinausführt. Es hängt von der Problemstellung ab, ob die Distanz dann eine sinnvolle Aussage zuläßt.

Beispiel:

- Merkmal = Farbe  $\Rightarrow$  nein
- Merkmal = Schmerz  $\Rightarrow$  ja

Wenn Daten fehlen,

- läßt man die betroffenen Objekte weg oder
- substituiert sie durch den Mittelwert oder ein anderes Maß

Bei quantitativen Merkmalen verwendet man vorwiegend folgende Distanz:

$$L_r(X, Y) = \left( \sum_{i=1}^k |x_i - y_i|^r \right)^{\frac{1}{r}} \quad r > 0$$

2 häufig verwendete Spezialfälle der  $L_r$ -Distanzen sind:

- $r = 1$  : Summe der Abstände der Komponenten
- $r = 2$  : Euklidische Norm

Die Distanzfunktion zwischen Klassen beruht auf der Distanz zwischen Elementen. Es seien  $B_1$  und  $B_2$  zwei Klassen von beobachteten Objekten mit den Umfängen  $n_1$  und  $n_2$ , zwischen denen eine Distanz  $f = f(X, Y)$  mit  $X \in B_1$  und  $Y \in B_2$  definiert ist.

Wichtige **Distanzen zwischen Klassen** sind:

- $\min_{X, Y} f(X, Y)$  ..... Single Linkage
- $\frac{1}{n_1 \times n_2} \times \sum_{X, Y} f(X, Y)$  ..... Average Linkage
- $\max_{X, Y} f(X, Y)$  ..... Complete Linkage

Die "Single Linkage"-Distanz erzeugt langgestreckte Klassen mit brückenartigen Verbindungen zwischen den Klassen.

„Average Linkage“ und „Complete Linkage“ produzieren kompakte "runde" Klassen.

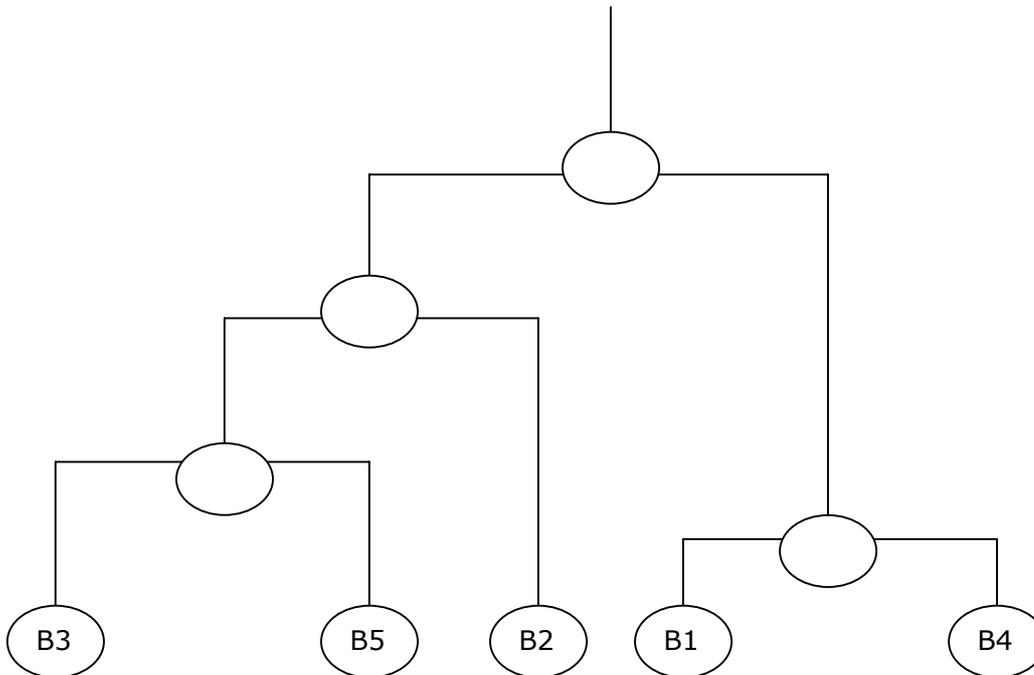
### 3.3.1 Algorithmen zur Bildung von Klassen

#### Hierarchische Klassifikation

Die  $n$  Objekte bilden zunächst 1-elementige Klassen.

Im nächsten Schritt werden jene Klassen vereinigt, deren Distanz am geringsten ist, usw. Auf diese Weise entsteht ein Dendrogramm, das die hierarchische Ordnung widerspiegelt.

Beispiel:



#### Nicht-hierarchische Klassifikation

Eine Zahl  $d \geq 0$  wird vorgegeben.

Man beginnt bei irgendeinem Objekt  $B_i$  und fasst alle Objekte zu einer Klasse

zusammen, die in der  $d$ -Umgebung von  $B_i$  liegen. Dann werden alle Objekte der Klasse untersucht, ob es weitere Objekte  $\leq d$  außerhalb gibt und diese der Klasse hinzugefügt. Die Klasse ist abgeschlossen, wenn kein neues Element mehr hinzukommt.

Von Objekten außerhalb der Klasse wird das Verfahren neu begonnen.

Dieses Verfahren hängt sehr stark von der Wahl von  $d$  ab.

- Für  $d = 0$  entstehen nur 1-elementige Klassen.
- Je größer  $d$  gewählt wird, desto eher fallen alle Objekte in die gleiche Klasse.

Die Cluster-Analyse wird in der Medizin oft zur sogenannten "Hypothesen-Generierung" angewandt.

---

## 4. ROC-ANALYSE

---

Die Abkürzung ROC steht für "Reciever-Operating-Characteristic".

Der Ursprung der ROC-Analyse liegt in der Nachrichtentechnik. Bei der Übertragung von digitalen Signalen über einen Kanal kommt es zu Verzerrungen, d.h. ein Signal wird nicht mit dem gleichen Pegel (0 oder 1) beim Empfänger ankommen, mit dem es abgesendet wurde. Der Empfänger muß daher einen (Schwell)Wert festlegen,

- oberhalb dessen er ein empfangenes Signal als 1-Signal entschlüsselt
- unterhalb dessen er Eingangssignale als 0-Signal klassifiziert

Wird die Schwelle hoch angesetzt, dann werden viele 1-Signale fälschlicherweise als 0 erkannt (da sie bei der Übertragung verändert wurden).

Wird die Schwelle niedrig festgelegt, dann werden viele 0-Signale als 1 eingestuft.

ROC-Kurven stellen den Zusammenhang der Genauigkeit bei der Übertragung von 0- und 1-Signalen in Abhängigkeit vom gewählten Schwellenwert dar. Die ROC-Kurve kann deshalb als **Annahmeline** verstanden werden (DE DOMBAL).

Dieses Verfahren wird nicht nur in der Nachrichtentechnik und Radartechnologie, sondern auch in der Medizin und Psychologie häufig verwendet.

### 4.1 ROC-Kurven in der Medizin

ROC-Kurven dienen in der Medizin zur Darstellung des Leistungsvermögens von diagnostischen Testverfahren bzw. von diagnostischen Entscheidungen.

Unter den Begriff diagnostischer Test fallen

- Tests in klinisch-chemischen Labors  
(z.B. die Bestimmung verschiedener Laborwerte aus einer Blutprobe und die daraus resultierende Entscheidung normaler/abnormaler Befund)
- die Klassifikation von Röntgen-, CT-Bilder etc. auf einer Rangskala

Rangskalen sind in der Medizin ein bewährtes Instrument zur Bewertung von Objekten. Radiologische Aufnahmen werden meistens auf 5-stufige Rangskalen klassiert

(es bedeutet 1: sicher normaler Befund, ..., 5: sicher abnormaler Befund).

In der Psychologie werden manchmal bis zu 40-stufige Rangskalen (von -20 bis +20) eingesetzt.

Eine Möglichkeit, die Güte eines Tests zu beschreiben, ist die Angabe der Maßzahl Genauigkeit (accuracy). Diese ist definiert als Quotient aus der Anzahl der richtigen Entscheidungen (der richtig diagnostizierten Fälle) und der Anzahl der insgesamt durchgeführten Entscheidungen (der insgesamt diagnostizierten Fälle).

Um den Anteil der richtigen Entscheidungen angeben zu können, muß der tatsächlich vorhandene Zustand bekannt sein; es muß also eine Bezugsdiagnose ("gold standard diagnosis") bekannt sein.

Die Angabe der Genauigkeit eines Tests ist jedoch aus 2 Gründen nicht sehr aussagekräftig:

- Die Genauigkeit hängt von der Verbreitung oder Prävalenz (prevalence) der untersuchten Krankheit ab.  
Angenommen ein Entscheidungsträger entscheidet unabhängig vom Testergebnis immer "Befund negativ".  
Bei einer Verbreitung der Krankheit von 5%, d.h. 5% der gesamten Population weist diese Krankheit auf, erreicht der Entscheider im Durchschnitt eine Genauigkeit von

95%.

Beträgt die Verbreitung der Krankheit 50%, dann erreicht die verfolgte Entscheidungsstrategie eine durchschnittliche Genauigkeit von nur 50%.

- Die Angabe der Genauigkeit läßt keinen Schluß zu, wie genau das Testverfahren, bei tatsächlich positiven bzw. tatsächlich negativen Fällen ist. Diese Unterscheidung ist dann von Bedeutung, wenn die Folgen von falsch-positiven und falsch-negativen Entscheidungen unterschiedlich sind (was in der Praxis meistens der Fall ist).

Die ROC-Analyse basiert auf den beiden Maßzahlen Sensitivität (sensitivity) und Spezifität (specificity).

- Die Sensitivität ist definiert als Quotient aus der Anzahl der wahr-positiven TR-Entscheidungen und der Anzahl der tatsächlich positiven Fälle:

$$\text{Sensitivität} = \text{TPR} = \frac{\text{Anzahl der wahr\_positiven Entscheidungen}}{\text{Anzahl der tatsächlich positiven Fälle}} \% \quad \text{Gl.(1)}$$

Die Sensitivität wird in üblicherweise % angegeben.

Diese Maßzahl (TPR steht für True Positive Ratio) kann Werte zwischen 0% (wenn die Anzahl der wahr-positiven Entscheidungen gleich 0 ist) und 100% (wenn die Anzahl der wahr-positiven Entscheidungen gleich der Anzahl der tatsächlich positiven Fälle ist) annehmen.

Die Anzahl der positiven Entscheidungen kann größer sein als die Anzahl der tatsächlich positiven Fälle; ein Teil davon ist dann jedoch falsch-positiv.

- Die Spezifität ist definiert als Quotient aus der Anzahl der wahr-negativen TN-Entscheidungen und der Anzahl der tatsächlich negativen Fälle:

$$\text{Spezifität} = \text{TNR} = \frac{\text{Anzahl der wahr\_negativen Entscheidungen}}{\text{Anzahl der tatsächlich negativen Fälle}} \% \quad \text{Gl.(2)}$$

Auch die Maßzahl Spezifität (TNR steht für True Negative-Ratio) kann Werte zwischen 0% und 100% annehmen.

Sensitivität und Spezifität sind also spezielle Genauigkeiten:

Die Sensitivität gibt die Genauigkeit des Testverfahrens in bezug auf tatsächlich positive Fälle an und die Spezifität gibt die Genauigkeit in bezug auf tatsächlich negative Fälle an.

Zwischen der "einfachen" Genauigkeit und der Sensitivität bzw. Spezifität besteht folgender Zusammenhang:

$\text{Genauigkeit} = \text{Sensitivität} \times \frac{\text{Anteil tatsächlich Positive}}{\text{Anzahl der Fälle}} + \text{Spezifität} \times \frac{\text{Anteil tatsächlich Negative}}{\text{Anzahl der Fälle}}$
--

Gl (3)

Beweis:

Die Genauigkeit ist wie folgt definiert:

$$\text{Genauigkeit} = \frac{\text{Anzahl der korrekten Entscheidungen}}{\text{Anzahl der Fälle}}$$

Also gilt:

$$\text{Genauigkeit} = \frac{\text{Anzahl der wahr\_positiven Entscheidungen}}{\text{Anzahl der Fälle}} + \frac{\text{Anzahl der wahr\_negativen Entscheidungen}}{\text{Anzahl der Fälle}}$$

Erweitert man mit der Anzahl der tatsächlich positiven bzw. negativen Fälle dann ergibt sich:

$$\text{Genauigkeit} = \frac{\text{Anzahl der wahr\_positiven Entscheidungen}}{\text{Anzahl der tatsächlich positiven Fälle}} \times \frac{\text{Anzahl der tatsächlich positiven Fälle}}{\text{Anzahl der Fälle}} + \frac{\text{Anzahl der wahr\_negativen Entscheidungen}}{\text{Anzahl der tatsächlich negativen Fälle}} \times \frac{\text{Anzahl der tatsächlich negativen Fälle}}{\text{Anzahl der Fälle}}$$

Dieser Ausdruck entspricht Gl.(3). Qed.

Korrespondierend zu den Größen TP und FP werden die Größen

- FP für die Anzahl der falsch-positiven Entscheidungen bzw.
  - FN für die Anzahl der falsch-negativen Entscheidungen
- verwendet. Damit definieren sich die Größen FNR und FPR wie folgt:

$$\text{FNR} = \frac{\text{Anzahl der falsch\_negativen Entscheidungen}}{\text{Anzahl der tatsächlich positiven Fälle}}$$

$$\text{FPR} = \frac{\text{Anzahl der falsch\_positiven Entscheidungen}}{\text{Anzahl der tatsächlich negativen Fälle}}$$

Es gilt:

- $\text{TPR} + \text{FNR} = 1$
- $\text{TNR} + \text{FPR} = 1$

Diese Maßzahlen haben auch einen Zusammenhang mit der Wahrscheinlichkeitsrechnung und zwar stehen sie im Zusammenhang mit bedingten Wahrscheinlichkeiten. Die Größen TPR, TNR, FPR, FNR sind relative Häufigkeiten, die als Schätzwerte für Wahrscheinlichkeiten herangezogen werden.

Zusammenhänge:

$$\text{TPR} = \text{Sensitivität} = P(T+ \mid D+)$$

$$\text{TNR} = \text{Spezifität} = P(T- \mid D-)$$

$$\text{FPR} = 1 - \text{Spezifität} = P(T+ \mid D-)$$

$$\text{FNR} = 1 - \text{Sensitivität} = P(T- \mid D+)$$

Verbreitung (Prävalenz) der Krankheit =  $P(D+)$

$$P(D-) = 1 - P(D+)$$

$$\text{TPR} + \text{FNR} = P(T+ \mid D+) + P(T- \mid D+) = 1$$

$$\text{TNR} + \text{FPR} = P(T- \mid D-) + P(T+ \mid D-) = 1$$

$$\begin{aligned} \text{Genauigkeit} &= \text{Sensitivität} \times p(D+) + \text{Spezifität} \times p(D-) \\ &= \text{TPR} \times p(D+) + \text{TNR} \times p(D-) = P(T+ \leq \mid D+) \times P(D+) + P(T- \leq \mid D-) \times P(D-) \end{aligned}$$

Dabei bedeutet

- T+ ..... positives Testergebnis
- T- ..... negatives Testergebnis
- D+ ..... Vorhandensein der untersuchten Krankheit
- D- ..... Absenz der Krankheit

$P(D+)$  bezeichnet die Wahrscheinlichkeit, dass ein zufällig aus der betrachteten Population ausgewähltes Individuum von der Krankheit betroffen ist,

$P(D-)$  steht für die Wahrscheinlichkeit, daß ein zufällig ermitteltes Individuum gesund ist.

$P(T+ | D+)$  bezeichnet die Wahrscheinlichkeit, daß ein positives Testergebnis erhalten wird unter der Voraussetzung, daß die Krankheit bei diesem Individuum vorliegt. Analog sind die anderen bedingten Wahrscheinlichkeiten  $P(T- | D-)$ ,  $P(T+ | D-)$ ,  $P(T- | D+)$  zu interpretieren.

Beispiel:

Die bei einem diagnostischen Test T erhaltenen Ergebnisse können in Form einer Entscheidungsmatrix dargestellt werden. Angenommen bei einem Test werden insgesamt 1.200 Personen untersucht und man kennt auf Grund anderer Untersuchungen den wirklichen Krankheits- bzw. Gesundheitszustand D der Personen; ein mögliches Ergebnis ist dann beispielsweise das folgende:

	D+	D-	Summe
T+	140 (TP)	100 (FP)	240
T-	60 (FN)	900 (TN)	960
Summe	200	1.000	1.200

Tab. 1

Aus dieser Entscheidungsmatrix ergeben sich die folgenden Maßzahlen:

Direkte Berechnung

abgeleitete Größe

$$TPR = \frac{140}{200} = 0.70$$

$$FNR = 1 - TPR = 1 - 0.70 = 0.30$$

$$TNR = \frac{900}{1000} = 0.90$$

$$FPR = 1 - TNR = 1 - 0.90 = 0.10$$

$$FPR = \frac{100}{1000} = 0.10$$

$$TNR = 1 - FPR = 1 - 0.10 = 0.90$$

$$FNR = \frac{60}{200} = 0.30$$

$$TPR = 1 - FNR = 1 - 0.30 = 0.70$$

$$P(D+) = \frac{200}{1200} = 0.17$$

$$P(D-) = 1 - P(D+) = 1 - 0.17 = 0.83$$

$$\text{Genauigkeit} = TPR \times P(D+) + TNR \times P(D-) = 0.70 \times 0.17 + 0.9 \times 0.83 = 0.866$$

Die Genauigkeit dieses Tests ist also bei negativen Fällen größer als bei positiven, d.h.  $TNR > TPR$ .

Bei Kenntnis der TPR- und TNR-Werte kann (bei sonst gleichbleibenden Bedingungen) die Änderung der Genauigkeit bei einer anderen Verbreitung der Krankheit berechnet werden.

Beispiel:

Steigt  $P(D+)$  von 0.17 auf 0.6, dann ergibt sich eine Genauigkeit von

$$0.7 \times 0.6 + 0.9 \times 0.4 = 0.78$$

Dieses Absinken der Genauigkeit beruht auf dem Umstand, daß mit steigender Verbreitung der Krankheit die geringere Genauigkeit bei positiven Fällen stärker zur Gesamtgenauigkeit beiträgt.

Die bis hierher angestellten Betrachtungen stellen eine punktuelle Darstellung eines diagnostischen Tests dar. Es ist möglich, das Kriterium zu verändern, das als Trennung zwischen negativem und positivem Befund herangezogen wird. ROC-Kurven stellen die daraus resultierenden Veränderungen der Sensitivität und Spezifität dar.

## 4.2 Herleitung einer ROC-Kurve

ROC-Kurven können auf 2 Arten gewonnen werden.

- Man betrachtet die zugrunde liegende Verteilungen bzw. Wahrscheinlichkeitsdichten; diese können entweder diskret oder stetig sein.
- Man kann für die ROC-Kurve eine mathematische Funktion angeben. (wird hier nicht betrachtet.)

### 4.2.1 Aus diskreten Daten

Ausgangspunkt: ein quantitativer diagnostischer Test, der für jede untersuchte Person eine Testgröße  $x$  (z.B. die Titerhöhe eines Tumormarkers) liefert.

Diese Testgröße  $x$  wird mit einem fixen Schwellenwert  $\varepsilon$  verglichen:

- $x < \varepsilon \Rightarrow$  die getestete Person leidet nicht an der untersuchten Krankheit
- $x \geq \varepsilon \Rightarrow$  die getestete Person leidet an der Krankheit

Beispiel:

Gegeben:

Ergebnisse eines fiktiven Tests von untersuchten 1.000 Personen, wobei die Testergebnisse in klassierter Form vorliegen (Tab.2).

Klasse	Klassengrenze	Gesunde	Kranke	Summe
1	$X_1 - X_2$	10	0	10
2	$X_2 - X_3$	50	0	50
3	$X_3 - X_4$	150	0	150
4	$X_4 - X_5$	315	30	345
5	$X_5 - X_6$	105	60	165
6	$X_6 - X_7$	56	75	131
7	$X_7 - X_8$	14	45	59
8	$X_8 - X_9$	0	40	40
9	$X_9 - X_{10}$	0	30	30
10	$X_{10} - X_{11}$	0	20	20
<b>Summe</b>		<b>700</b>	<b>300</b>	<b>1.000</b>

Tab. 2:

In jeder Klasse ist eine Unterscheidung in "Gesunde" und "Kranke" angegeben, d.h. es wurde der wirkliche Zustand der untersuchten Person anhand eines anderen diagnostischen Verfahren bestimmt.

Abbildung 7.1 stellt die Verteilung der Testergebnisse (Titerhöhe eines Tumormarkers) auf die einzelnen Klassen in graphischer Form als Histogramm dar.

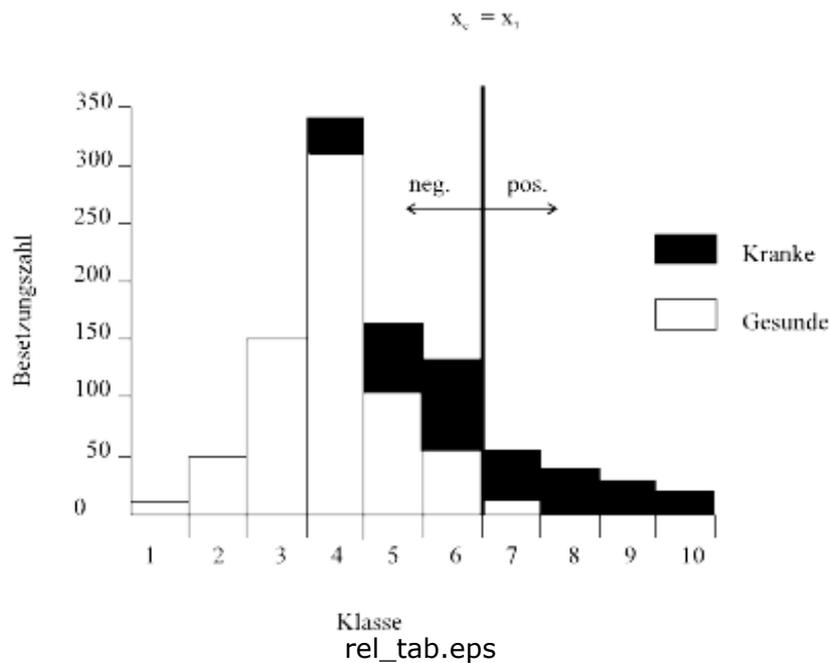


Abb. 7.1: Histogramm: Verteilung der Testergebnisse auf die einzelnen Klassen

Diese Testgröße wird mit einem fixen Schwellenwert  $\epsilon$  verglichen;

- $x < \epsilon \Rightarrow$  die getestete Person leidet nicht an der untersuchten Krankheit
- $x \geq \epsilon \Rightarrow$  die getestete Person leidet an der Krankheit

Der Schwellenwert  $x_c$  ist beliebig wählbar, (in Abb. 7.1 ist  $\epsilon = x_7$ ).

Als Bezeichnungen für diesen Schwellenwert haben sich die englischen Fachausdrücke "Threshold-level" bzw. "**Cut-off-point**" eingebürgert.

Für den cut-off-point  $\epsilon = x_7$  kann die Vierfeldertafel berechnet werden (Tab. 3).

		Tatsächlicher Zustand		Summe
		Krank (D+)	Gesund (D-)	
Test- ergebnis	Abnormal (T+)	135	14	149
	Normal (T-)	165	686	851
Summe		300	700	1000

Tab. 3:

Allgemein besitzt eine Vierfeldertafel das in Tab. 4 dargestellte Aussehen.

		Tatsächlicher Zustand		Summe
		X	$\neg X$	
Test- ergebnis	X	TP	FP	
	$\neg X$	FN	TN	
Summe				

Tab. 4

$x$  bezeichnet das Vorhandensein des Merkmals (bzw. die Entscheidung für das Vorhandensein),  $\neg x$  bedeutet die Absenz des Merkmals.

Durch die Wahl des Cut-off-points  $\varepsilon = x_7$  ergibt sich

- eine Sensitivität von  $\frac{135}{300} = 0.45$
- eine Spezifität von  $\frac{686}{700} = 0.98$

Der Cut-off-point kann aber auch mit jedem anderen Wert angesetzt werden. Tab. 5 stellt die Werte für die Sensitivität und Spezifität dar, die sich durch eine unterschiedliche Wahl der Schranke  $\varepsilon$  ergeben.

Klasse	Sensitivität (TPR * 100%)	Spezifität (TNR * 100%)	FPR
$X_1$	$300/300 = 100$	$0/700 = 0$	1
$X_2$	$300/300 = 100$	$10/700 = 1.43$	0.9857
$X_3$	$300/300 = 100$	$60/700 = 8.57$	0.9143
$X_4$	$300/300 = 100$	$210/700 = 30$	0.7
$X_5$	$270/300 = 90$	$525/700 = 75$	0.25
$X_6$	$210/300 = 70$	$630/700 = 90$	0.1
$X_7$	$135/300 = 45$	$686/700 = 98$	0.02
$X_8$	$90/300 = 30$	$700/700 = 100$	0
$X_9$	$50/300 = 16.6$	$700/700 = 100$	0
$X_{10}$	$20/300 = 6.6$	$700/700 = 100$	0
$X_{11}$	$0/300 = 0$	$700/700 = 100$	0

Tab. 5

Trägt man in einem kartesischen Koordinationssystem

- auf der Ordinate die Sensitivität TPR
  - auf der Abszisse 1- Spezifität (= das Komplement der Spezifität = FPR)
- auf, dann erhält man die in den Abb. 7.2 dargestellten Punkte.

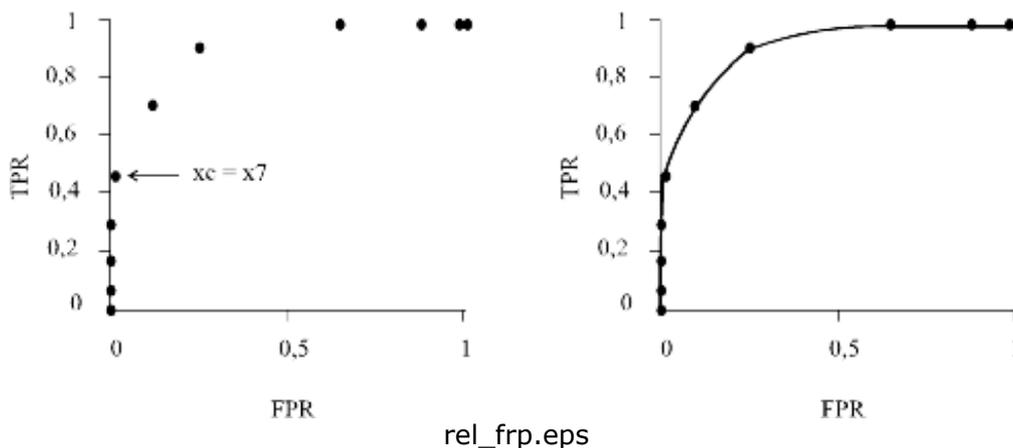


Abb. 7.2 und 7.3: Herleitung der ROC - Kurve

Legt man durch die Punkte einen geschlossenen Kurvenzug, so ergibt sich die zu diesem Testverfahren gehörende ROC-Kurve (Abb.7.3).

Ein vorsichtiger Entscheider wird den Cut-off-point eher niedrig ansetzen. Einer solchen Schranke entspricht ein Punkt im rechten oberen Teil der ROC-Kurve. Es werden dadurch sehr viele positive Fälle als solche erkannt, andererseits werden dadurch aber auch viele negative Personen als positiv klassifiziert.

Wird der Cut-off-point hoch angesetzt - was einem Punkt nahe dem Koordinatenursprung entspricht -, dann werden falsch-positive Ergebnisse selten sein, es wird aber auch zu vielen falsch-negativen Entscheidungen kommen.

Die ROC-Kurve stellt somit den Zusammenhang zwischen der Sensitivität und der Spezifität dar. Eine Zunahme der Sensitivität (also der Genauigkeit bei tatsächlich positiven Fällen) kann nur auf Kosten einer Abnahme der Spezifität (also einer Zunahme der Größe FPR; d.h. einer Abnahme der Genauigkeit bei tatsächlich negativen Fällen) erfolgen.

Ein Testverfahren ist nur dann sinnvoll, wenn eine ROC-Kurve oberhalb der Diagonale des durch die Koordinaten (0,0) und (1,1) definierten Einheitsquadrates befindet. Eine ROC-Kurve, die ident mit der Diagonalen ist, bedeutet, daß der Test nicht besser ist als eine zufällige Entscheidung, ob das untersuchte Individuum krank ist oder nicht.

### 4.2.2 Aus stetigen Daten

Bis jetzt wurde davon ausgegangen, daß die der ROC-Kurve zugrunde liegenden Verteilungen auf diskreten Daten beruhen. Es besteht jedoch auch die Möglichkeit, daß von stetigen Daten ausgegangen wird. Die Verteilung der positiven und der negativen Fälle wird in diesem Fall durch Wahrscheinlichkeitsdichtefunktion dargestellt (Abb. 7.4).

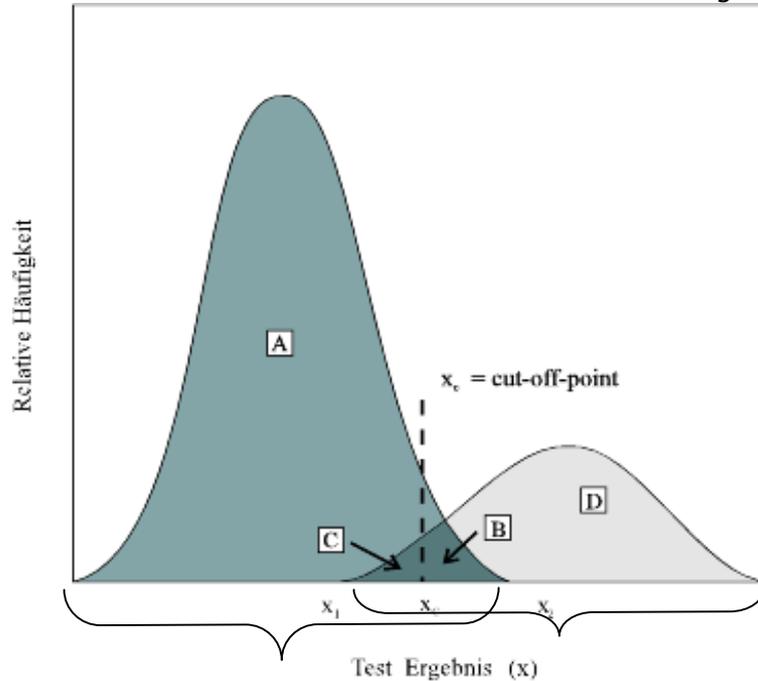


Abb. 7.4 Verteilung Gesunde vs. Kranke  
si schraff.eps

Die linke Kurve stellt die Verteilung der Testergebnisse der gesunden Individuen, die rechte Kurve stellt die Verteilung der Testergebnisse der Kranken dar. Für die weiteren Ausführungen bezeichnet

- $\phi_{D+}$  ..... Wahrscheinlichkeitsdichte der kranken Personen  
 $\phi_{D-}$  ..... Wahrscheinlichkeitsdichte der gesunden Personen

Es gilt:

- $\int_{-\infty}^{\infty} \phi_{D+} dx = 1$
- $\int_{-\infty}^{\infty} \phi_{D-} dx = 1$

Bei der in Abb. 7.4 getroffenen Wahl des Cut-off-points  $\varepsilon = x_7$  ergibt sich folgende Vierfeldertafel

	D+	D-
T+	D (TPR)	B (FPR)
T-	C (FNR)	A (TNR)

Die Flächen A, B, C und D lassen sich durch Integration berechnen, wobei die Integrale mit  $P(D+)$  bzw.  $P(D-)$  zu gewichten sind:

$$A(x) = P(D-) \times \int_{-\infty}^{\varepsilon} \phi_{D-} dx$$

$$B(x) = P(D-) \times \int_{\varepsilon}^{\infty} \phi_{D-} dx$$

$$C(x) = P(D+) \times \int_{-\infty}^{\varepsilon} \phi_{D+} dx$$

$$D(x) = P(D+) \times \int_{\varepsilon}^{\infty} \phi_{D+} dx$$

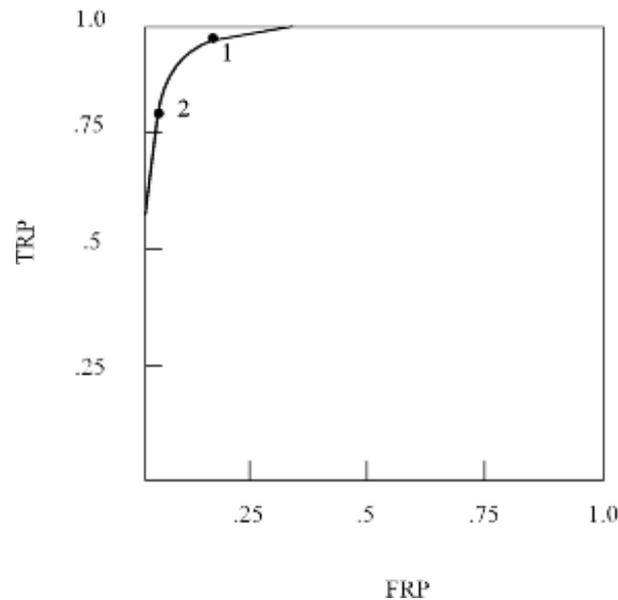
Sensitivität, Spezifität und Genauigkeit lassen sich entweder aus der Vierfeldertafel ableiten oder mit Hilfe von Integralen berechnen.

$$\text{Sensitivität}(x) = \frac{D}{C+D} = \int_{\varepsilon}^{\infty} \phi_{D+} dx$$

$$\text{Spezifität}(x) = \frac{A}{A+B} = \int_{-\infty}^{\varepsilon} \phi_{D-} dx$$

$$\text{Genauigkeit}(x) = \frac{A+D}{A+B+C+D} = P(D-) \times \int_{-\infty}^{\varepsilon} \phi_{D-} dx + P(D+) \times \int_{\varepsilon}^{\infty} \phi_{D+} dx$$

Aus in den Abb. 7.4 dargestellten Dichten ergibt sich die folgende ROC-Kurve (Abb. 7.5).



kl\_kurv.eps  
Abb. 7.5: ROC-Kurve im stetigen Fall

Da es sich um stetige Daten handelt, erhält man in diesem Fall im Gegensatz zu diskreten Daten einen geschlossenen Kurvenzug und nicht einzelne Punkte.

Der auf der ROC-Kurve mit "1" markierte Punkt entspricht dem Cut-off-point bei  $x_1$ , der mit "2" markierte Punkt entspricht dem Cut-off-point  $x_2$  (vgl. Abb. 7.4).

Aus der graphischen Darstellung der Verteilungen können 2 weitere Maßzahlen abgeleitet werden:

- **Vorhersagewert eines positiven Testergebnisses**  $P(D+ | T+)$   
(predictive value of a positive test)
- **Vorhersagewert eines negativen Testergebnisses**  $P(D- | T-)$   
(predictive value of a negative test)

Der Vorhersagewert eines positiven Ergebnisses<sup>3</sup>  $P(D+ | T+)$  ist die bedingte Wahrscheinlichkeit, daß die Krankheit vorliegt unter der Voraussetzung, daß bei dem betreffenden Individuum ein positives Testergebnis erhalten wurde.

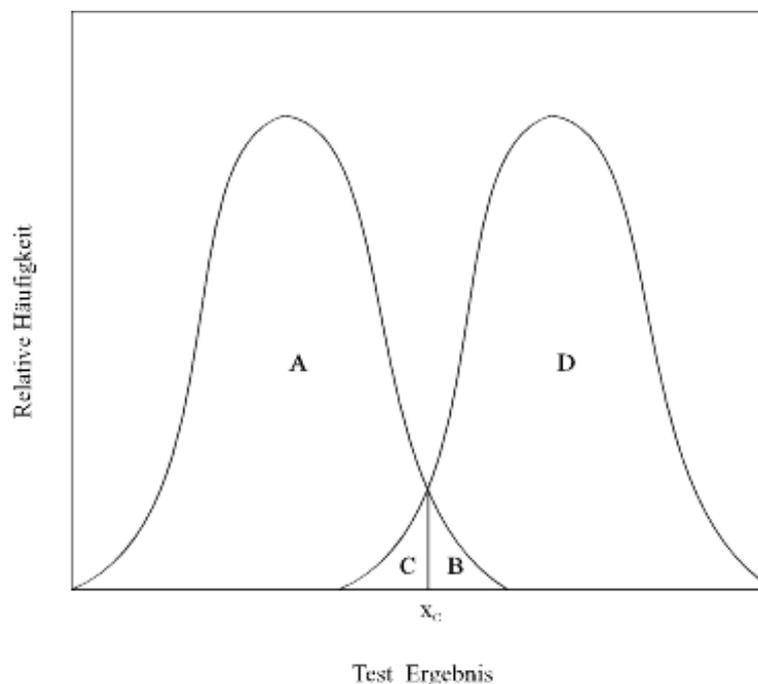
Berechnung von  $P(D+ | T+)$  und  $P(D- | T-)$ :

$$P(D+ | T+) = \frac{D}{B+D} = \frac{TPR}{TPR+FPR} = \frac{P(D+) \times \int_{\epsilon}^{\infty} \phi_{D+} dx}{P(D+) \times \int_{\epsilon}^{\infty} \phi_{D+} dx + P(D-) \times \int_{\epsilon}^{\infty} \phi_{D-} dx}$$

$$P(D- | T-) = \frac{A}{A+C} = \frac{TNR}{TNR+FNR} = \frac{P(D-) \times \int_{-\infty}^{\epsilon} \phi_{D-} dx}{P(D-) \times \int_{-\infty}^{\epsilon} \phi_{D-} dx + P(D+) \times \int_{-\infty}^{\epsilon} \phi_{D+} dx}$$

Die Sensitivität und die Spezifität sind unabhängig von der Verbreitung der Krankheit, der Vorhersagewert eines Testergebnisses jedoch schon. Dieser Sachverhalt kann graphisch veranschaulicht werden.

Abb. 7.6 repräsentiert eine Population, in der 50% die untersuchte Krankheit aufweisen:  
 $P(D+) = P(D-) = 0.5$ .



<sup>3</sup>  $P(D+ | T+)$  ist von  $(T+ | D+)$  strikt zu unterscheiden:

$P(T+ | D+)$  ist die bedingte Wahrscheinlichkeit, daß ein positives Testergebnis erhalten wird unter der Voraussetzung, daß die Krankheit vorliegt.

2kurven.eps

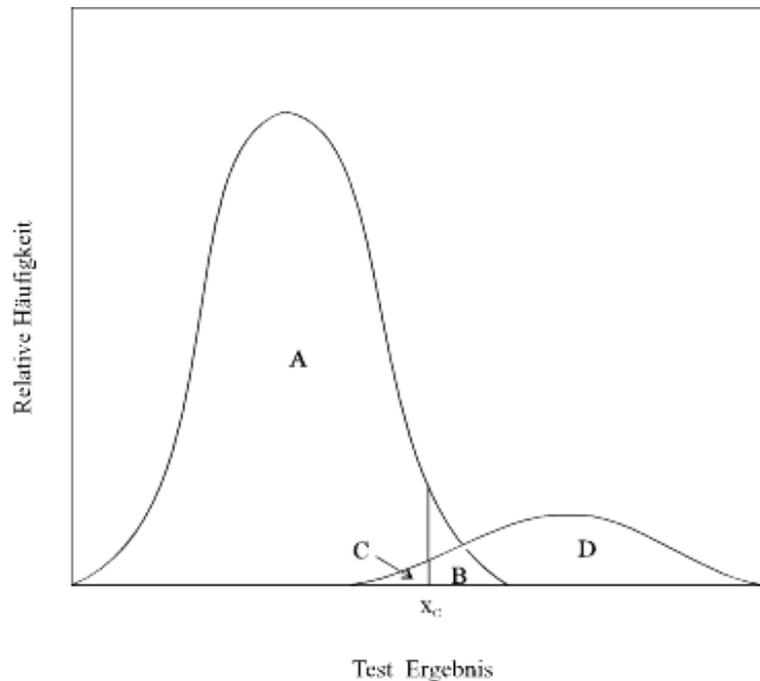
Abb. 7.6: Gleichverteilung Gesunde vs. Kranke

Wird der Cut-off-point wie dargestellt gewählt, dann ergibt sich eine Sensitivität und eine Spezifität von ca. 93% (da Symmetrie vorliegt) und ein Vorhersagewert eines positiven Tests von ebenfalls 93%.

Sinkt die Verbreitung der Krankheit (Abb. 7.7;  $P(D+) = 0.1$ ,  $P(D-) = 0.9$ ) und bleiben sonst alle Bedingungen gleich (insbesondere die Wahl des Cut-off-points), dann hat das zwar keinen Einfluß auf die Sensitivität und die Spezifität, der Vorhersagewert eines positiven Ergebnisses beträgt jetzt allerdings nur noch knapp 61%, d.h. 40% aller positiven Testergebnisse sind falsch-positiv.

Dieser Umstand ist darin begründet, daß das Verhältnis der Flächen D und B immer

kleiner wird und damit der Ausdruck  $\frac{D}{B+D}$  gleichfalls sinkt.



gra\_kld.eps

Abb. 7.7: Gesunde vs. seltene Krankheit

Bei einer Verbreitung der Krankheit von  $P(D+) = 0.01$  ergibt sich bei unverändertem Cut-off-point ein Vorhersagewert von ca. 12% (d.h. 7 von 8 positiven Tests sind falsch-positiv !!).

---

## 5. ENTSCHEIDUNGSMODELLE UND -STRATEGIEN

---

**Gründe**, warum man in der MI Algorithmen zur Entscheidungsunterstützung sucht:

- Algorithmen setzen operationale Regeln für Entscheidungen voraus. Auf diese Weise wird in der MI Grundlagenforschung betrieben.
- Im Bereich Diagnostik und Therapie ist das Wissen des erfahrenen Mediziners (Experten) entscheidend.  
Über Expertensysteme wird versucht, das Expertenwissen im Computer zu speichern und auf diese Weise Hilfestellung anzubieten.
- Für den Arzt und den Studenten ergibt sich die Möglichkeit der Ausbildung.

In diesem Abschnitt werden Verfahren besprochen, die eine Zuordnung zu einer Grundgesamtheit ermöglichen. Es handelt sich dabei um statistische und prädikatenlogische Entscheidungsmodelle.

## 5.1 Statistische Verfahren

Gegeben:

2 Grundgesamtheiten  $G_1$  und  $G_2$ . Es sind dann 4 Fälle bei der Entscheidung möglich.

		Wirklichkeit	
		$G_1$	$G_2$
Entscheidung	$G_1$	0	$K(1/2)$
	$G_2$	$K(2/1)$	0

Tab. 8.1: Kosten einer Entscheidung

Beispiel:

$G_1$ ..... Gesunde

$G_2$ ..... Patienten mit Diabetes

Beide Grundgesamtheiten werden durch die jeweilige Verteilung des stetigen Merkmals  $X = \text{Glucose}$  beschrieben.

$X=x$  wird an einem Menschen bestimmt und es ist zu entscheiden, ob er gesund oder krank ist.

Die Strategie  $R = (R_1, R_2)$  ist die folgende:

- Ist  $x \in R_1$ , dann ist der Mensch gesund
- Ist  $x \in R_2$ , dann hat der Mensch Diabetes

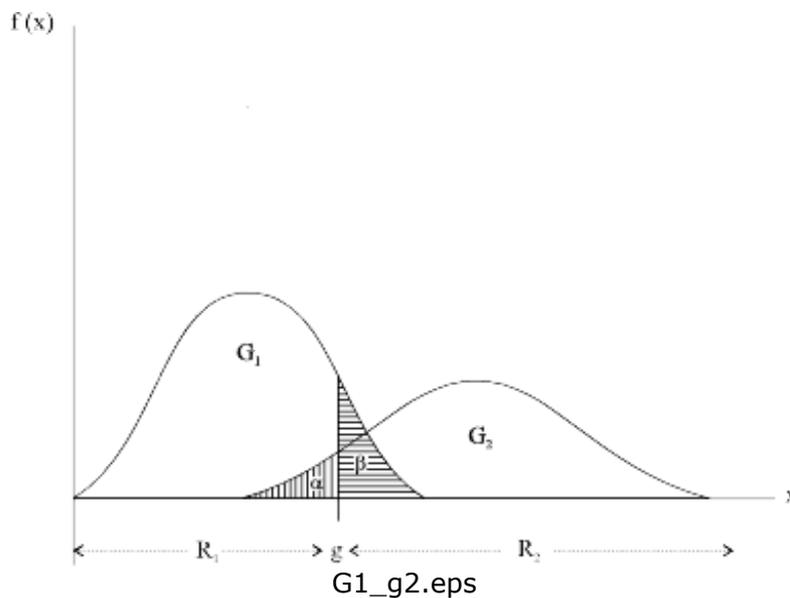


Abb. 8.1: Strategie bei Gesunden vs. Kranke

Abhängig von der Grenze  $g$  wird ein Gesunder mit der Wahrscheinlichkeit  $\beta$  als krank und ein Kranker mit der Wahrscheinlichkeit  $\alpha$  als gesund eingeordnet.

Die  $x$ -Achse wird durch  $g$  in die Bereiche  $R_1$  und  $R_2$  eingeteilt. Die Kosten der Zuordnung eines Kranken zu den Gesunden seien  $K(1/2)$ , die eines Gesunden zu den Kranken  $K(2/1)$ . Ferner sei  $p$  die Wahrscheinlichkeit, daß ein Gesunder untersucht wird.

Die Kosten der Strategie  $R$  sind:

$$p \times K(2/1) \times \beta + (1 - p) \times K(1/2) \times \alpha$$

Die Strategie R ist so zu wählen, dass die erwarteten Kosten minimal werden.  
Sind  $f_1(x)$  und  $f_2(x)$  die beiden Wahrscheinlichkeitsdichten, dann ist g so zu wählen, daß

$$p \times K(2/1) \times \int_g^{\infty} f_1(x) dx + (1-p) \times K(1/2) \times \int_{-\infty}^g f_2(x) dx = \text{Minimum}$$

Man spricht in diesem Fall (Minimierung der zu erwartenden Gesamtkosten) von einer **BAYES-Prozedur**. (Voraussetzung: die a priori Wahrscheinlichkeit p ist bekannt.)

Ist die Wahrscheinlichkeit p unbekannt, kann das **Minimax-Prinzip** verwendet werden.  
Die gesuchte Prozedur soll derart sein, daß gilt:

$$\min_R \max (\text{Kosten})$$

Bei der **Diskriminanzanalyse** erfolgt die Trennung mit Hilfe des Quotienten, der aus der Varianz innerhalb und der Varianz zwischen den Gruppen gebildet wird. Dieser Quotient wird minimiert. Geometrisch wird eine Hyperebene im n-dimensionalen Raum gesucht, der die Gruppen "am besten" trennt.

Voraussetzung:

- multivariate Normalverteilung der Merkmale
- identische Kovarianzmatrix

## 5.2 Prädikatenlogische Verfahren

Ein mathematischer Beweis erfolgt üblicherweise in der Form, daß man von Axiomen und Voraussetzungen ausgeht und nach gültigen Regeln andere Aussagen ableitet.

Die gleiche Vorgangsweise hat man in der medizinischen Diagnostik:

Man geht von Aussagen (Befunden) aus, wendet gültige Regeln (medizinisches Wissen) an und leitet andere Aussagen (Diagnosen) her.

Man kann daraus folgern, daß die medizinische Diagnostik formal darstellbar ist. Dazu wird die Prädikatenlogik herangezogen.

Die **Prädikatenlogik** verwendet eine eigene Sprache. Diese besteht aus

- Elementen zur Formulierung von Fakten
- Regeln zur Ableitung neuer Fakten aus alten

Die **Elemente zur Formulierung von Fakten** sind:

Konstante sind Objekte mit fixen Namen, z.B. LEBER.

Variable werden verwendet, um Objekte im zeitlichen Verlauf zu beschreiben.

Prädikate bezeichnen Eigenschaften.

So ist die Eigenschaft der Leber „*Verdauungsorgan zu sein*“ gegeben durch  
VERDAUUNGSORGAN(LEBER)

Der Wert eines Prädikats ist "wahr" oder "falsch".

Beispiel:

Ein zweistelliges Prädikat ist z.B.:

RISIKO(LEBER, ALKOHOL)

Löst man ein Prädikat auf, dann erhält man eine **Aussage**, z.B. Alkohol ist ein Risiko für die Leber.

**Funktionen** haben - wie Prädikate - Argumente, als Ergebnis jedoch nur einen Wert (keinen Wahrheitswert).

Beispiel:

Die Funktion KONSISTENZ mit dem Argument LEBER hat z.B. die Werte DERB, WEICH, usw.

Funktionen können Argument von Prädikaten sein, jedoch nicht Prädikate Argument von Funktionen.

Der Operator "Negation" wirkt auf 1 Prädikat, die Operatoren "Konjunktion", "Disjunktion", "Implikation" wirken auf 2 Prädikate. Damit werden Ausdrücke formuliert.

Die Formulierung,

*„die Leber ist in ihrer Konsistenz derb und vergrößert“*

erfolgt in folgender Weise:

DERB(KONSISTENZ(LEBER)) UND VERGRÖSSERT(LEBER)

Die **Quantoren** "∀ für alle" und "∃ es gibt" werden ebenfalls zur Formulierung von Ausdrücken verwendet.

Auf diese Weise lassen sich medizinische Entscheidungen "*prädikatenlogisch*" formulieren.

### 5.3 Sequentielle Strategien

In der medizinischen Diagnostik führen die zahlreichen komplexen Abhängigkeiten unter den Merkmalen dazu, daß die Beobachtung aller Merkmale (Symptome) zur Entscheidungsfindung

- zu teuer kommt oder
- aus Rücksicht auf den Patienten nicht durchgeführt werden kann.

Die Strategie des Arztes besteht in einem sequentiellen Vorgehen, d.h. zuerst werden leicht zu erhebende Daten gesammelt usw.

Man sucht also eine optimale Strategie, sodaß die Diagnosegesamtkosten minimal werden. Die dafür verwendeten Algorithmen können der **Spieltheorie** entnommen werden.

Spieltheoretische Modelle lassen sich auf die aus der Algebra bekannten **linearen Programme** zurückführen. Darunter versteht man lineare Funktionen mehrerer Variablen  $m$ , die maximal (oder minimal) gemacht werden, wobei Nebenbedingungen gegeben sind, die ebenfalls linearen Charakter haben.

In der Regel handelt es sich dabei um eine große Zahl  $n$  von Ungleichungen, wobei  $n > m$  ist. Die Lösung derartiger Probleme erfolgt z.B. mit dem Simplex-Algorithmus.

#### Hinweis:

Die **lineare Programmierung** ist die wichtigste mathematische Grundlage der Verfahrensforschung (Operations Research).

Die optimale Lösung kann mittels **dynamischer Programmierung** gefunden werden. Dazu wird die Diagnose-Symptom-Matrix in der Form eines Entscheidungsbaumes dargestellt. Dieser enthält 3 Typen von Knoten:

- Terminale Knoten, in denen eine Diagnose gestellt wird
- Entscheidungsknoten, in denen ein Merkmal ausgewählt wird
- Merkmalsknoten, in denen ein Merkmal beobachtet wird

Die sequentielle Entscheidungsstrategie ist der Prototyp eines rekursiven Problems. Der bei jedem Entscheidungsknoten beginnende Unterbaum kann wie der ganze Baum behandelt werden. Die optimale Strategie ist dann eine Folge optimaler Teilstrategien. Sie ergibt sich, wenn man beginnend mit dem ersten Entscheidungsknoten stets zu einem Merkmalsknoten verzweigt, dessen erwartete Kosten gleich jenen im Entscheidungsknoten sind. Die Kosten in einem terminalen Knoten (Diagnose steht fest) sind gleich 0. Zur Berechnung der Kosten benötigt man a priori Wahrscheinlichkeiten.

Der Entscheidungsbaum kann sehr groß sein und dann für den Computer unlösbar werden. Gibt es  $k$  Merkmale mit jeweils  $m$  Ausprägungen, dann sind auf der letzten Stufe bis zu  $k! \times m^{k-1}$  Merkmalsknoten und  $k! \times m$  terminale Knoten vorhanden.

Eine Verminderung der Schritte des Entscheidungsprozesses ist nur möglich, wenn

- vorzeitig abgebrochen wird oder
- Schätzwerte der Entscheidungsstrategie anstatt der Kosten verwendet werden. (Zur Berechnung der Kosten müssen die Kosten für die folgenden Teilstrategien bekannt sein). Man verwendet dann a posteriori Wahrscheinlichkeiten oder deren Schätzungen nach dem BAYES Theorem.

Die Modelle zur Entscheidungsunterstützung verlangen auch **Vollständigkeit**, die in der Praxis nur schwer zu erreichen ist. Der Mensch ist im Vergleich zum Computer in der Lage auch aus unpräzisen und unvollständigen Daten vernünftige Schlüsse zu ziehen.

Deshalb versucht man in der AI-Forschung den menschlichen Entscheidungsprozeß zu simulieren.

Die AI verwendet vorwiegend Methoden zur Problemlösung mittels Computer auf der Basis symbolischer Folgerung ("symbolic reasoning") anstelle arithmetischer Berechnungen. Dabei unterscheidet man zwischen

- Modellbildung  
Die Modellbildung besteht darin, daß Merkmale und deren Beziehungen für ein bestimmtes Wissensgebiet (z.B. Rheuma-Krankheiten) ausgewählt werden und das vorhandene Wissen eingebunden wird.
- Folgern mit Problemlösung  
Folgern mit Problemlösung besteht in der Entwicklung und Durchführung einer Strategie, die auf der Basis des gespeicherten Wissens und unter Verwendung der Daten eines Objekts zu Entscheidungen führt.
- heuristischer Suche  
Die heuristische Suche besteht in der Anwendung von Methoden zur Verringerung der Anzahl von Alternativen.

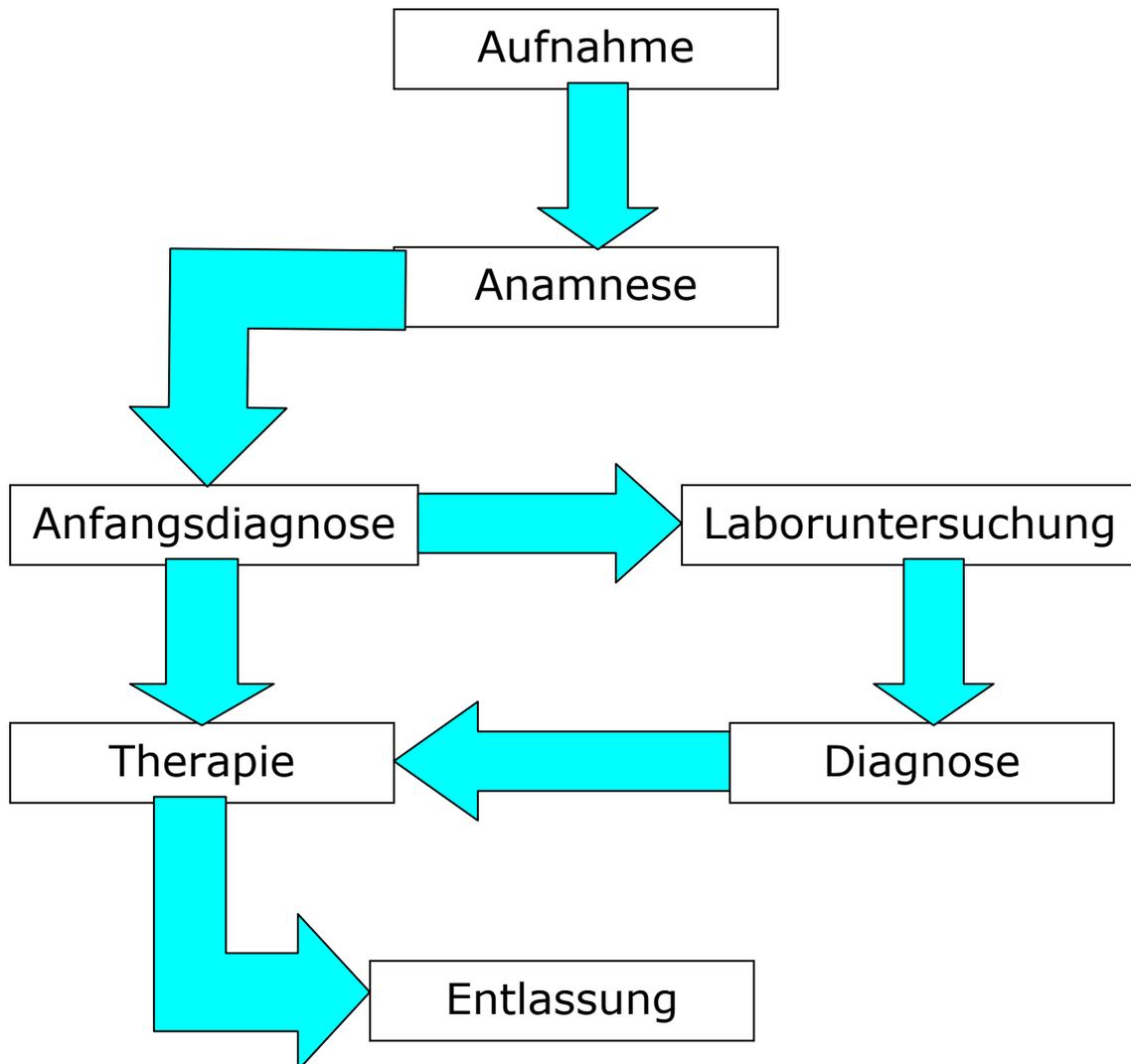
Beispiel zur Entscheidungsunterstützung:

MYCIN (Entscheidungsunterstützung bei der Identifikation von Keimen und der Auswahl einer geeigneten Therapie)

## 5.4 Computerunterstützte Diagnose

Bei den Verfahren zur computerunterstützten medizinischen Diagnostik wird in der Regel das gespeicherte medizinische Wissen dazu verwendet, aus konkreten Informationen über einen bestimmten Patienten Unterstützung bei der allgemeinen und speziellen Diagnostik anzubieten.

Ein diagnostischer Prozess in der Medizin, der alle wesentlichen Elemente enthält, lässt sich wie folgt darstellen:



### Gründe für computerunterstützte Diagnose:

- Der Diagnoseprozeß ist also vorwiegend ein Mengenproblem.  
In der Medizin zählt man an
  - 60.000 Krankheiten und phänomenologisch definierte Syndrome
  - 80.000 bis 100.000 Symptome, also anamnestische Angaben, Beschwerden, Laboruntersuchungen, Biopsie-, Histologie-, Röntgen-, US-, Biosignalbefunde usw. Die Symptome stehen in der Regel in einer bestimmten, für die diagnostische Entscheidung des Arztes wichtigen Beziehung zu einer oder mehreren Krankheiten. Meist müssen Symptomenkomplexe berücksichtigt werden, um eine Krankheit zu diagnostizieren.

Erschwerend dabei ist, daß viele Patienten an mehreren Krankheiten gleichzeitig leiden, d.h. es liegt oft ein sich überlagerndes Symptomenbild vor.

- In Österreich gibt es über 7.000 zugelassene Medikamenten, die der Arzt einschließlich ihrer Kontraindikationen kennen muß.
- Der Arzt wird gezwungen, zu allen für eine Krankheit typischen Symptomen Stellung zu nehmen. Er muß die Diagnose auch gegenüber sehr selten auftretenden Krankheiten abgrenzen. Ein Diagnosesystem kann dem Arzt hierbei helfen, indem es Vorschläge zur weiteren Untersuchung des Patienten ausgibt. Diese Vorschläge kann man auch ordnen nach
  - den Kosten
  - ihrer Gefährlichkeit
  - Schmerzhaftigkeit für den Patienten.
- Eine umfangreiche Sammlung medizinischen Wissens, die in einem Computersystem verfügbar ist, bildet die Grundlage für ein diagnostisches Lehrsystem. Medizinstudenten und junge Ärzte sind in der Lage, einen diagnostischen und therapeutischen Prozeß nachzuvollziehen.
- Es werden nicht nur diagnostische und therapeutische Entscheidungen des Arztes unterstützt, sondern es findet zwangsläufig eine Überprüfung, ein Überdenken, eine Erweiterung und Systematisierung des medizinischen Wissens selbst statt.

Im Zusammenhang mit computerunterstützter Diagnostik spricht man vielfach auch von Expertensystemen. Diese Expertensysteme erhalten ihre Berechtigung dadurch, daß sie die praktische ärztliche Erfahrung, Kreativität und Intuition durch

- (im Idealfall vollständig) gespeicherte große Datenmengen
- die Möglichkeit komplexer logischer Schlußfolgerungen ergänzen.

#### **Einsatzmöglichkeiten:**

- Screening-Programme, deren Aufgabe es ist, große Populationen diagnostisch zu erfassen (Hilfestellung im Bereich der Gesundheitsvorsorge und der allgemeinpraktischen ärztlichen Tätigkeit).
- Diagnostische Programme, die sich auf bestimmte Gebiete der Medizin, wo Spezialwissen vorhanden ist, beschränken (Hilfestellung für den Facharzt).