

Stochastik — eine interdisziplinäre Wissenschaft. Teil II

L. Rüschemdorf

Institut für Mathematische Stochastik,
Universität Freiburg ¹

In dem zweiten Teil dieser Darstellung sollen einige Anwendungsgebiete etwas detaillierter betrachtet werden, nachdem im ersten Teil historische und grundsätzliche methodische Gesichtspunkte im Vordergrund standen. In einigen Fällen wird mehr auf die methodische Entwicklung, in anderen auf die inhaltlichen Fragestellungen aus den Anwendungen eingegangen.

Neuere Anwendungen der Stochastik basieren auf der methodischen Weiterentwicklung. Es sind eine Reihe von komplexen Modellen konstruiert und analysiert worden (räumliche Modelle, stochastische Prozesse, verallgemeinerte lineare Modelle (GAM), Markovsche Felder, Cox-Regressionsmodelle, ...). Rechenintensive Verfahren haben Einzug gehalten (Bootstrapverfahren, Spline-Wavelet und andere nichtlineare Regressionsmethoden, neuronale Netze, logistische Regression). Umfangreiche Vergleichsstudien der zugehörigen Algorithmen sind erstellt worden. Die folgenden Beispiele aus diesem Bereich sollen erläutert werden:

- A) Stochastik in Algorithmen und Optimierung
- B) Klassifikation und räumliche Modelle
- C) Finanzmathematik und stochastische Prozesse
- D) Biometrie und Expertensysteme
- E) Emissionstomographie

A) Stochastik in Algorithmen und Optimierung Die Entwicklung und Analyse von Algorithmen ist eine der wichtigsten Aufgaben der (angewandten) Mathematik. Schon in den ersten Schulklassen werden Algorithmen gelernt, z.B. wie man, vom kleinen Einmaleins ausgehend, große Zahlen miteinander multipliziert.

Zu den am häufigsten verwendeten Algorithmen gehören der Gauß-Algorithmus zum Lösen linearer Gleichungssysteme oder das Newton-Verfahren zum Auffinden von Nullstellen einer Funktion. Schnelle und leistungsfähige Algorithmen für partielle Differentialgleichungen ermöglichen es, z.B. das aerodynamische Verhalten von Flugzeugen teilweise auf dem Computer zu simulieren. Sie gestatten es, große Datenmengen in statistischen Analysenproblemen oder in der Computertomographie zu verarbeiten. Weiter werden Algorithmen benötigt, etwa zum Erzeugen von Zufallszahlen für stochastische Simulationen.

¹Eckerstraße 1, D-79104 Freiburg

Mit stochastischen Algorithmen lassen sich natürliche Objekte approximativ beschreiben. Einige dieser Konstruktionen sind auch weitgehend in den öffentlichen Medien vorgestellt worden:

Von Mandelbrot stammt die Konstruktion von 'fraktalen Objekten' mit der fraktalen Brownschen Bewegung.



(aus Mandelbrot (1977): The fractal geometry of nature.)

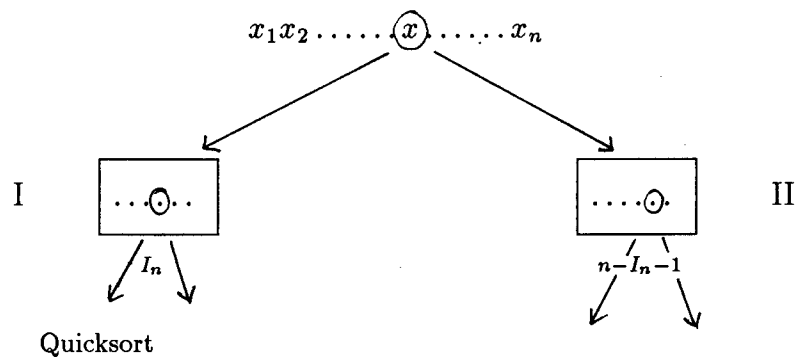


(aus Barnsley (1988): Fractals everywhere.)

Barnley hat die Konstruktion und Datenkompression von natürlichen Objekten mit **iterierten Funktionensystemen** propagiert. Die Bilder sind Repräsentanten der invarianten Maße vom sogenannten **Chaos-Spiel**, bei dem nach einer Verteilung zufällig Funktionen iteriert angewendet werden und die entstehende Punktmenge das Bild darstellt. Die Bedeutung dieser Verfahren zum Zwecke der Datenkompression wird zur Zeit intensiv studiert und weiterentwickelt.

Eine wichtige Aufgabe in der Informatik ist die **Konstruktion und Analyse von Algorithmen** für häufig auftretende Aufgaben wie: Sortieren, Suchen und Speichern von Zahlen, Schlüsselns oder ähnliches, die Speicherverwaltung und vieles mehr.

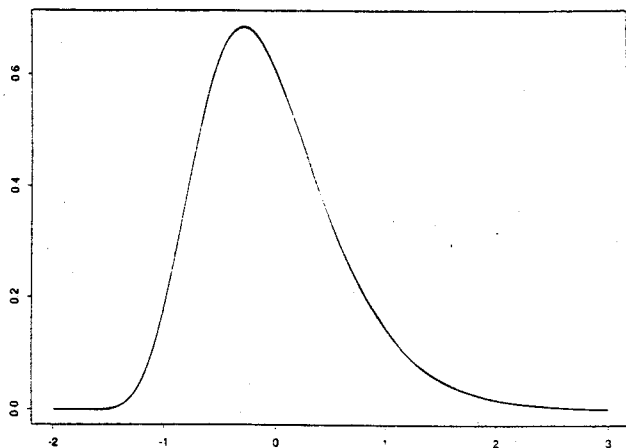
Insbesondere interessiert z.B. das Laufzeitverhalten der Algorithmen und das durchschnittliche Ergebnis der Algorithmen. Es hat in neuerer Zeit erhebliche Fortschritte auf diesem Gebiet gegeben. Betrachten wir z.B. den **Quicksortalgorithmus** zum Sortieren von n reellen Zahlen x_1, \dots, x_n der Größe nach. Dazu wird zufällig ein Element x aus den Zahlen herausgegriffen und die anderen Zahlen werden mit diesem Element verglichen. Es entstehen zwei Gruppen: die Gruppe I der kleineren Elemente vom Umfang I_n und die Gruppe II der größeren Elemente vom Umfang $n - I_n - 1$. Rekursiv werden nun die kleineren Problemgruppen sortiert und dann zusammengesetzt.



Ist L_n die Anzahl der benötigten Rechenschritte zur Herstellung der geordneten Zahlen, so gilt nach Knuth (1973) für den Erwartungswert und für die Varianz von L_n

$$\ell_n = EL_n \approx 2n \log n, \quad \text{Var}L_n \approx \left(7 - \frac{2}{3}\pi^2\right) n^2$$

Ein neues und präziseres Resultat von Cramer (1994) liefert, daß $\frac{L_n - \ell_n}{n}$ angenähert log-normalverteilt ist.



Geglättete empirische Dichte von Quicksort für $n = 5000$ und simultan eine approximierende log-Normalverteilung. Beide Kurven sind kaum zu unterscheiden.

Mit obiger Aussage und weiteren ähnlichen Aussagen läßt sich einsehen, daß nur mit exponentiell kleiner Wahrscheinlichkeit Quicksort nicht von der Größenordnung $n \log n$ Schritte verwendet. Der Algorithmus benötigt im schwierigsten Fall (worst case) n^2 Rechenschritte. Mit obiger Methode läßt sich aber praktisch ausschließen, daß der Quicksortalgorithmus einen großen Rechen- und Zeitaufwand hat.

Es sind für viele Algorithmen sehr präzise Fehlerschranken gefunden worden.

Stochastische Algorithmen Ein Prinzip bei der Konstruktion von stochastischen Algorithmen ist es, in einzelnen Rechenschritten zufällig Fehler in Kauf zu nehmen, dafür aber den global optimalen Wert anzustreben. Üblicherweise sind Algorithmen Berechnungsverfahren mit präzise vorgegebenen Einzelschritten, die ohne weitere Phantasie des Anwenders durch systematisches Befolgen der Regeln das Endergebnis liefern, wobei man versucht, die doch immer wieder auftretenden Fehler durch Entwicklung von fehlerkorrigierenden Codes zu verbessern. Z.B. sind die Chips in Computern so dicht gepackt, daß gelegentliche Störungen durch radioaktive Strahlung auftreten. Die Bilder vom Jupiter von der Voyager-Sonde waren nur auf Grund der Verwendung von Golay Codes so brilliant.

In stochastischen Algorithmen werden einzelne Rechenschritte dagegen 'zufällig' falsch ausgeführt. Für manche Probleme sind jedoch die Ergebnisse dieser stochastischen Algorithmen besser als mit jedem bekannten deterministischen Algorithmus.

Ein recht universell anwendbarer stochastischer Algorithmus ist der **simulated annealing Algorithmus**. Die Idee dazu kommt aus der Physik von dem Abkühlprozeß bei der Herstellung einer Schmelze. Soll eine Zielfunktion $f : S \rightarrow \mathbb{R}$, S ein Gitter, maximiert werden, so wählt man, ausgehend von einem Gitterpunkt x ,

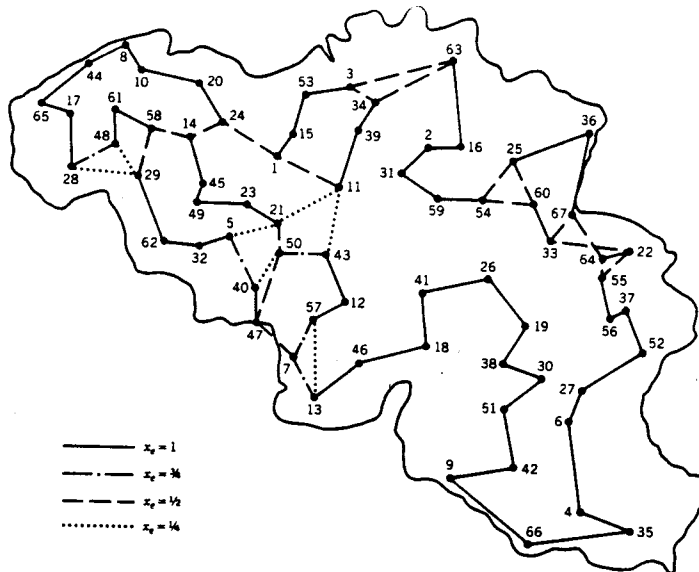
zufällig einen Nachbarn und wechselt mit kleiner Wahrscheinlichkeit zu dem Nachbarn über, auch wenn sich dabei der Zielwert verringert. Die Temperatur entspricht der Größe der Wahrscheinlichkeit. Sie wird in dem fortlaufenden Prozess langsam auf 0 abgesenkt. Für viele Probleme liefert dieser Algorithmus gute, für manche Probleme sehr gute Ergebnisse.

Beispiel 1) Beim Traveling Salesman problem ist simulated annealing (SA) nicht optimal, aber eine Modifikation von SA liefert die zur Zeit besten Ergebnisse für große Probleme. An diesem Beispiel läßt sich gut erläutern, daß eine ganz wichtige Komponente bei der Lösung von Optimierungsproblemen die Darstellung der Daten ist.

Table with 67 columns representing distances between cities in Belgium. The cities listed include BRUSSEL, CHARLEROI, CHIMAY, DENZEE, DENDERMONDE, DINSLUDE, DINANT, DURBUY, EERLO, ENGHEM, EUPEN, GERAARSBERGEN, GENT, HASSELT, HUUY, HOUFFALIZE, JEPER, KORTRIJK, LA ROCHE, LEUVEN, LEUZE, LIEGE, LIER, LUXEMBURG, MAASEIK, MALMEDY, MARCHIE, MECHELEN, MONS, NAMUR, NIEFCHATEAU, NIVELLES, OOSTENDE, OUDENARDE, PHEIPPEVILLE, QLEUVY, ROESELARE, RONSE, BOCHES, ST HUBERT, ST VITH, ST MELAAS, ST TRUIDEN, SPA, STAVELDT, THUEN, and TIEL.

Ausschnitt aus 67 Städtedistanzen in Belgien

Eine Einbettung in die euklidische Ebene ist der Schlüssel für relativ gute intuitive Lösungen.

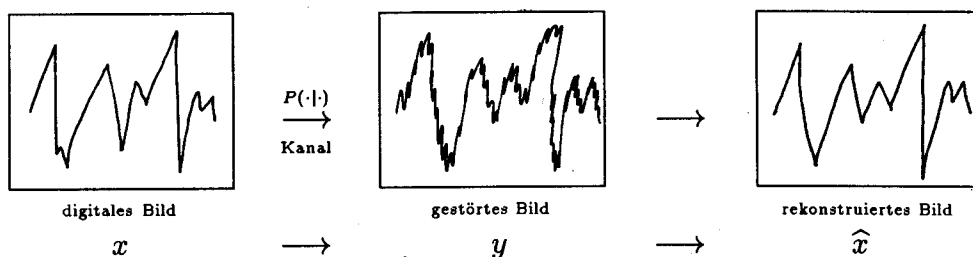


Branch and bound Methode (aus Nemhauser Wolsey), Lösung: 1606,75km

Stochastische Algorithmen sind z.B. die besten für das Problem der Primzahlfaktorzerlegung. Dieses ist von großer praktischer Bedeutung, da solche Zerlegungen eingehen in das einzige als sicher geltende Kodierungsverfahren, den public key code. Für einige als außerordentlich schwierig geltende Berechnungsverfahren (sogenannte NP-Probleme) sind in jüngerer Zeit stochastische Algorithmen gefunden worden, die in polynomialer Zeit operieren und approximativ optimal sind, z.B. zur Bestimmung der Normierungsfunktion im Ising Modell für ferromagnetische Systeme (Jerrum, Sinclair (1989)) oder bei der Approximation konvexer Körper in hohen Dimensionen.

Beispiel 2) Bildrekonstruktion

Die Bayessche Bildrekonstruktion beruht auf folgendem Verfahren. Ein digitales Bild x wird in einem Kanal übertragen und als mit dem Übergangskern $P(\cdot|\cdot)$ gestörtes Bild y empfangen. Aus y soll eine Schätzung \hat{x} für x gewonnen werden.

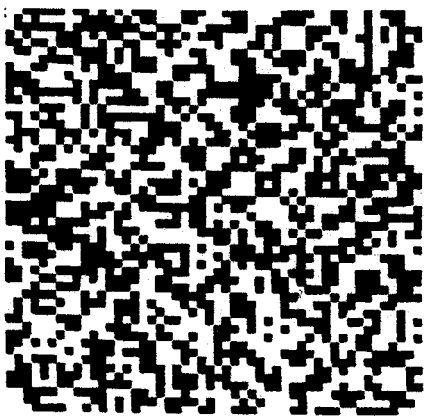


Die Dimension D von x ist extrem hoch, typischerweise etwa 500×500 oder 500×1000 . Sei S die Menge der Farben. Die Bilder sind Elemente aus $\Omega = S^D$. Zur Rekonstruktion verwendet man eine a priori-Verteilung μ_β auf Ω , ein "Markov-Feld" (Gibbs Prozess), $\mu_\beta(x) = \frac{e^{-\beta E(x)}}{\pi(\beta)}$, wobei $E = E_\alpha$, die von einem Parameter α abhängige Energie ist. μ_β enthält die Vorinformation über das Bild. Der Parameter β entspricht der "Temperatur" μ_β beschreibt ein stochastisches Bild und liefert

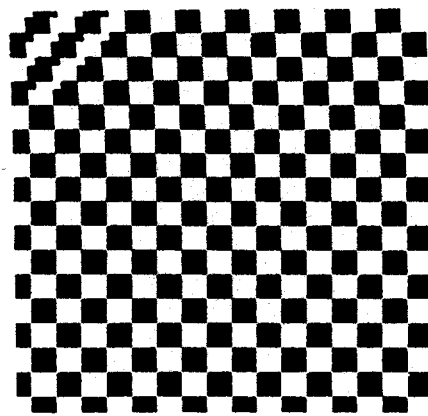
z.B. Informationen über Regularitätseigenschaften der Bilder. Nach der Bayesschen Regel (vgl. Teil I) erhält man:

$$P_{\beta}(x|y) = P(y|x) \frac{\mu_{\beta}(x)}{P_{\beta}(y)}$$

Das a posteriori-Bild ist wieder ein Gibbsches Maß. Die Rekonstruktion \hat{x} gewinnt man z.B. als Maximumstelle von $P_{\beta}(\cdot|y)$ (MAP Rekonstruktion) mit dem simulated annealing algorithmus



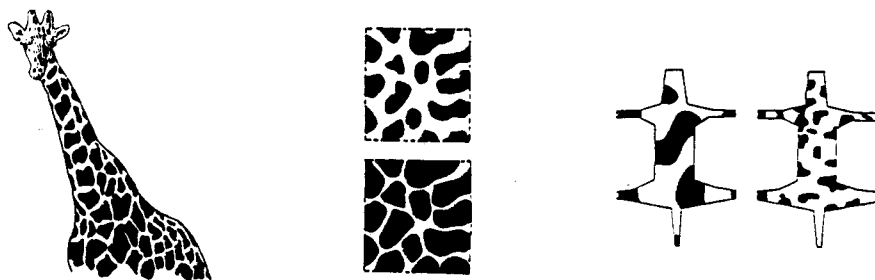
gestörtes Bild mit 30% fliprate



Rekonstruktion des Musters mit der
Bayes-Methode

(aus: Ripley (1988): Statistical inference for spatial processes.)

Auch Beispiele der Aufarbeitung von Bildern aus der Kernspintomographie (NMR) sind nach dieser Methode untersucht worden. Ein interessanter Aspekt der obigen Methode sind die dazu verwendeten markovschen Felder. Das einfachste dieser Felder ist das Ising Modell für Ferromagnetismus. Es ist sehr ausführlich untersucht worden und beschreibt die Spin-Muster, die bei Magneten auftreten. Deterministische Varianten dieser Interaktionsmodelle sind in jüngerer Zeit in der Theorie zellulärer Automaten und in stetiger Form in Reaktions-Diffusionsgleichungen zur Beschreibung von natürlichen Mustern untersucht worden. Die Beschreibung von Mustern ist ein aktuelles Forschungsgebiet in der Biologie und auch in anderen Disziplinen.

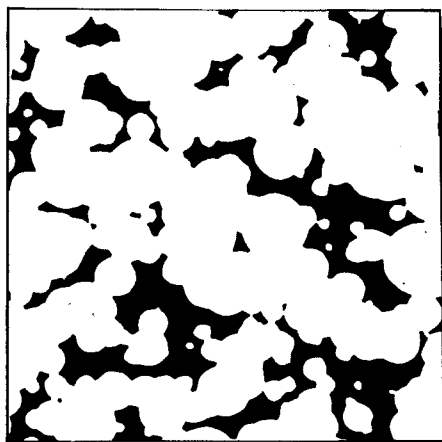


Biologische Muster und Simulationen mit Reaktions-Diffusionsgleichungen
(aus Murray (1993): *Mathematical Biology*.)

Es scheint mir, daß markovsche Felder die Konstruktion von natürlichen Mustern in größerer Vielfalt ermöglichen. Das bleibt aber noch zu untersuchen. Ein großes Problem in diesem Zusammenhang ist es, inwieweit solche einfachen Erklärungsmodelle auch eine biologische Relevanz haben.

B) Klassifikationsprobleme und räumliche Modelle Räumliche stochastische Modelle finden in neueren Arbeiten eine Fülle von essentiellen Anwendungen.

a) Das **Boolsche Modell** ist eines der einfachsten räumlichen Modelle. Es ist eines der klassischen Grundmodelle der stochastischen Geometrie und wurde vor allem in der Geologie untersucht und als Modell für Lagerstätten verwendet. Weitere Bereiche sind die Modellierung von Kolloiden mit Gelstruktur, das Studium der Energieverhältnisse beim Übergang von Flüssigkeiten in Dampf und insbesondere als Testobjekte für automatische Bildanalysegeräte. Es ist erstaunlich, wie mit einfachen geometrischen Objekten (Kreisen) und etwas Zufall (Zentren und Radien) interessante Strukturen entstehen.



Modellierung einer Kalilagerstätte und dem Boolschen Modell
(aus Stoyan, Kendall, Mecke (1985): *Stochastic Geometry and its Applications*.)

Ein Beispiel raumzeitlicher Regressionsmodelle für Luftverschmutzung wurde schon vorher in Teil I behandelt.

b) Tiefenumwandlung seismischer Daten, Öllagermodellierung Öllager befinden sich in der Nordsee in einer Tiefe von etwa 3000 Metern und haben ein Ausmaß von etwa $3,0 \cdot 5,0 \text{ km}^2$. Das Hydrokarbon wandert in den Reservoirs aufwärts bis es unter nichtdurchlässigen geologischen Schichten (typischerweise Schiefer) gefangen wird. Die Abbildung dieser Schichten gibt wichtige Hinweise für die Förderung und das Hydrocarbonvolumen.

Undurchlässige Schichten können weiträumig durch seismische Daten ermittelt werden, d.h. erhalten werden seismische Reflektionszeiten ($t(x), x \in D$), die in Abbildungen ($Z(x), x \in D$) der undurchlässigen Schichten umgewandelt werden sollen. Die Signalgeschwindigkeit jedoch variiert beträchtlich, sowohl vertikal als auch lateral. Daher kann nicht einfach das Produkt aus Zeit mal Geschwindigkeit gebildet werden. Neben den seismischen Daten (Messungen von einem Schiff im Bereich $100 \cdot 100 \text{ m}^2$) liegen weitere Daten aus Bohrlöchern vom Umfang $0,2 \cdot 0,2 \text{ m}^2$, $Z(x_1), \dots, Z(x_n)$ vor. Verwendet wird ein Modell von der Form:

$$\begin{aligned} Z(x) &= V(x)t(x) + \varepsilon(x) \\ &= \left(\beta_1 + \beta_2 t(x) + \beta_3 \frac{x_1 - x_{1,\min}}{x_{1,\max} - x_{1,\min}} \right) t(x) + \varepsilon(x) \end{aligned}$$

mit der Kovarianzfunktion $\text{Cov}(\varepsilon(x)) = \sigma^2 K(\|x - y\|)$. Geophysikalische Information führt zur Annahme

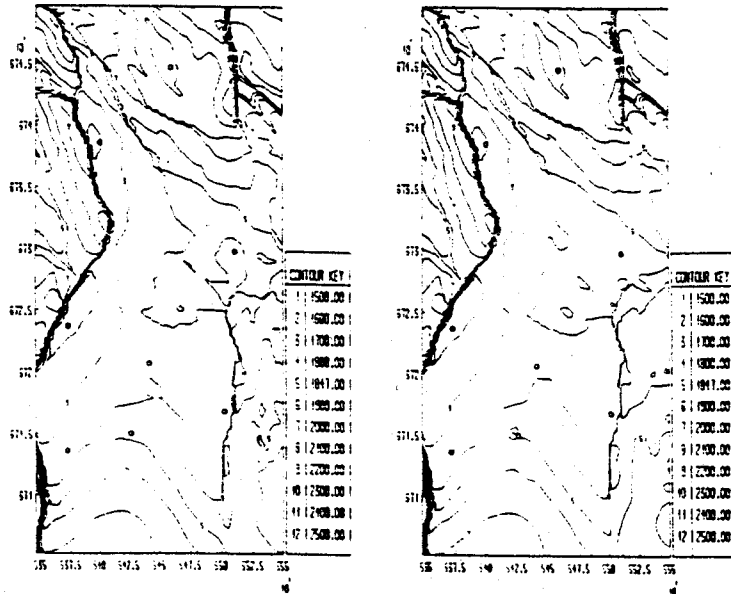
$$K(r) = \begin{cases} 1 - \frac{3}{2} \left(\frac{r}{2.800} \right) + \frac{1}{2} \left(\frac{r}{2.800} \right)^3 & 0 \leq r \leq 2.800 \\ 0 & \text{sonst} \end{cases}$$

und zur a priori-Verteilungsannahme:

$$\beta \sim N \left(\begin{pmatrix} 0,6 \\ 0,0002 \\ 0,07 \end{pmatrix}, \tau^2 \begin{pmatrix} (0,055)^2 & 0 & 0 \\ 0 & (0,0001)^2 & 0 \\ 0 & 0 & (0,022)^2 \end{pmatrix} \right)$$

Der Skalenparameter τ wird variiert. Aus der a posteriori-Verteilung ergeben sich

mit der "Bayes Kriging Methode" folgende Bilder

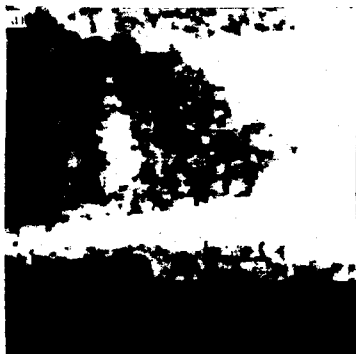


Bayessche Kriging Karten

(aus Hjort, Omre (1994): Topics in spatial statistics. In: Scand. J. Statistics)

Weiter Beispiele von Bedeutung sind die

- Identifikation von Herz-Dysfunktionen basierend auf Echokardiographie-Daten (Ultraschalltechnik). Die Bilder haben geringen lokalen Kontrast und große 'Noise'-Komponenten. Verwendet wird Vorwissen mit 'Pattern recognition'-Techniken. Typische Bilder versprechen wesentliche Reduzierung des 'Noise' und Verbesserung der diagnostischen Möglichkeiten.



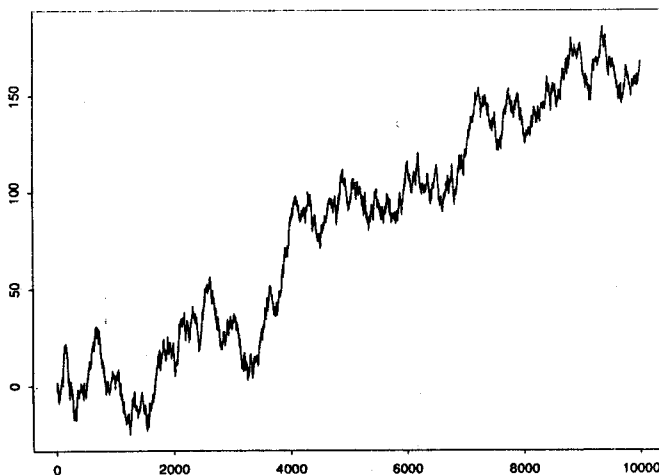
Ultraschallbild von Herzkammeraktivität mit Bayesscher Noise-Reduktion

(aus Hjort, Omre (1994): Topics in spatial statistics. In: Scand. J. Statistics)

- Bestimmung von Fisch-Populationen
- Simulation vom Gesteinsaufbau
- einige erfolgreiche Ziffernerkennungsverfahren basieren auf stochastischen Interaktionsmodellen.

C) Finanzmathematik und stochastische Prozesse Eines der zur Zeit aktivsten Gebiete in der Stochastik ist die Modellierung von Finanzdaten, Aktienkursen, Wechselkursen und Returnverteilung durch stochastische Prozesse, die Ermittlung von Optionspreisen und die Risikosteuerung der Portfolios.

1827 hatte der Botaniker Brown beobachtet, daß Staubpartikel auf einer Wasseroberfläche unregelmäßige Bewegungen durchführen. Einstein und Smoluchowski gelang es 1905 anhand dieser Beobachtungen die atomare Struktur der Materie nachzuweisen und sie leiteten für die zeitliche Dichtefunktion der Brownschen Bewegung B die Diffusionsgleichungen sowie das \sqrt{t} -Gesetz für den Zuwachs der Bewegung her, $B(t+s) - B(s) \sim N(0, \sigma^2 t)$.



Simulation der Brownschen Bewegung

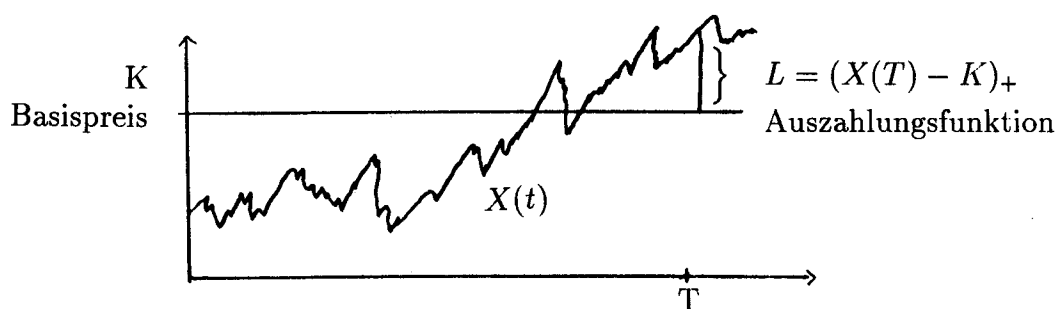
Einstein fand für die Varianz σ^2 die Formel

$$\sigma^2 = \frac{R \cdot T}{3\pi \cdot L \cdot r \cdot \eta} \quad \begin{array}{l} R = \text{universelle Gaskonstante, } L = \text{Loschmidsche Zahl,} \\ r = \text{Radius, } \eta = \text{Reibungszahl, } T = \text{absolute Temperatur} \end{array}$$

Basierend auf dieser Entwicklung wurde von Perrin 1926 experimentell $L \sim 6,02 \cdot 10^{23}$ bzw. die Boltzmannkonstante $k = \frac{R}{L}$ bestimmt. Hierfür erhielt Perrin den Nobelpreis. Wiener 1923 konstruierte als erster ein stochastisches Pfadmodell für die

Brownsche Bewegung, das es ermöglicht, auf einfache Weise eine approximative Simulation der Brownschen Bewegung zu erhalten. Der französische Mathematiker Bachelier hatte schon im Jahre 1900 in seiner Doktorarbeit 'Theorie de la Speculation' das (Brownsche) Diffusionsmodell als Modell für die Aktienpreisentwicklung einiger zu seiner Zeit in Frankreich gehandelter Optionen vorgeschlagen und die berechneten Preise mit den realen Marktpreisen verglichen. Die Bedeutung seiner Arbeit wurde aber in seiner Zeit nicht erkannt und sein Doktorvater, der berühmte Mathematiker Poincaré, schrieb 'das Thema ist in gewisser Weise weit entfernt von den Themen, die üblicherweise von unseren Forschungskandidaten gewählt werden'. Das Bacheliermodell für die Preise enthielt einen zusätzlichen Drift: $dX(t) = \mu(t)dt + \sigma(t)dB(t)$. Samuelson entdeckte 1965 die Bachelier-Arbeit und schlug eine Modifikation vor, die **geometrische Brownsche Bewegung**: $dX(t) = X(t)(\mu dt + \sigma dB(t))$, äquivalent, $X(t) = X(0)e^{\mu t + \sigma B(t) - \frac{1}{2}\sigma^2 t}$. ($X(t)$) beschreibt die zeitliche Entwicklung von Aktienkursen und weiteren ökonomische Zeitreihen. Es vermeidet einige wichtige Nachteile des Bachelierschen Modells.

Mit der Einführung von Optionen, z.B. von europäischen Call Optionen mit Strikepreis K und Auszahlungsfunktion $L = (X(T) - K)_+$ ergibt sich folgendes Bild



Auszahlungsfunktion bei Basispreis K

Für die Frage nach dem Wert einer solchen Option fanden die Mathematiker **Black und Scholes** 1973 einen völlig neuartigen Weg, diesen zu bestimmen. Ihre Methode, basierend auf einem Hedging Argument, wird bis heute an den Börsen angewandt. Sie führt zu dem 'fairen Preis'

$$C_T = S_0 \Phi(d_+) - ke^{-rT} \Phi(d_-)$$

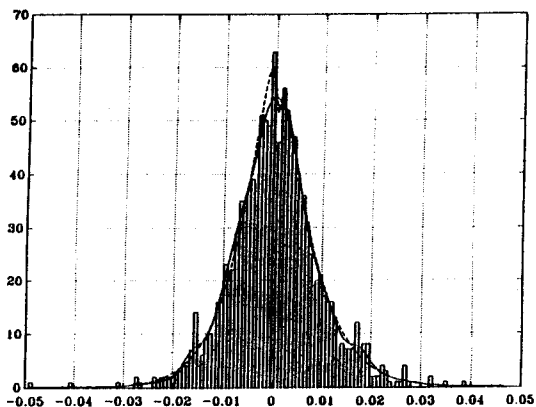
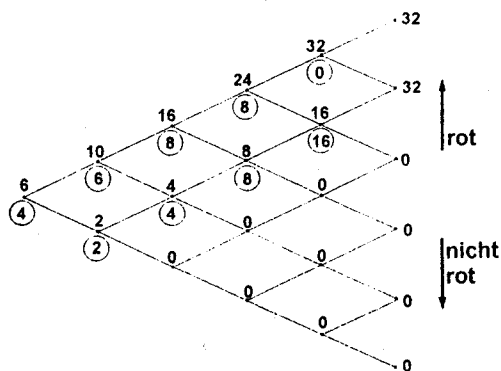
mit $d_{\pm} = \left(\log \frac{X(0)}{K} + T(r \pm \frac{\sigma^2}{2}) \right) / \sigma \sqrt{T}$, $r =$ Zinsfaktor, $\Phi =$ Verteilungsfunktion der Normalverteilung. Das Hedgingargument besteht darin, die Auszahlungsfunktion L durch Verwendung einer geeigneten Anlagestrategie ϑ exakt zu duplizieren

$$L = L_0 + \int_0^T \vartheta(t) dX(t).$$

Der Erwartungswert von L_0 bzgl. eines 'äquivalenten Martingalmaßes' ist dann der faire Preis. Die Möglichkeit einer solchen Darstellung kann für sogenannte vollständige Preismodelle nachgewiesen werden. Das Hedgingargument läßt sich an folgendem (von Hipp vorgeschlagen) Spielmodell einfach erläutern. Die Auszahlungsfunktion bei einer Spielfolge vom Roulettetyp mit Ergebnissen rot und schwarz sei

$$L = \begin{cases} 32 & \text{wenn } Z(6) \geq 4, \\ 0 & \text{sonst,} \end{cases}$$

d.h. nach 5 Spielen zahle 32, falls mindestens 4 mal rot auftritt, usw. $Z(t)$ sei die Anzahl der Spiele mit Ergebnis 'rot' bis zum Zeitpunkt t . Sie läßt sich exakt mit einer Hedgingstrategie reproduzieren, nämlich $L = 6 + \sum_{t=1}^5 \vartheta(t)(Z(t+1) - Z(t))$. Der faire Preis ist demnach 6.



Modellierung von Finanzdaten

Anpassung von S&P500 Daten 1982-1986 durch eine Weibullverteilung.

— $\hat{=}$ Kernschätzer, --- $\hat{=}$ Weibull

In jüngerer Zeit wurden viele 'Verbesserungen' der geometrischen Brownschen Bewegung vorgeschlagen und untersucht sowie alternative Modelle für den logarithmischen Zuwachs (in der geometrischen Brownschen Bewegung normalverteilt ist), wie z.B. Weibullverteilung, stabile Verteilung und hyperbolische Verteilung. Dabei wurden sehr gute Übereinstimmungen mit realen Daten erzielt.

Stochastische Prozesse sind auch bei weiteren Anwendungen zentral. Klassische Modelle für ökonomische Zeitreihen sind z.B. ARMA-Prozesse.

In neueren Studien zu lokalen Kommunikationsnetzwerken hat sich herausgestellt, daß der 'Ethernet Traffic' long range dependence Eigenschaften hat und durch selbstähnliche Prozesse gut beschrieben wird.

D) Biometrie und Expertensysteme Ein aktuelles und auch in der historischen Entwicklung wichtiges Feld der Statistik ist die Biometrie. Im Jahre 1978 waren von den 15000 Statistikern, die im Direktorium der Amerikanischen Statistischen Gesellschaft geführt wurden, etwa ein Fünftel mit Problemen der Biostatistik befaßt. (Die obigen Zahlen sind inzwischen auf über 20000 angestiegen.)

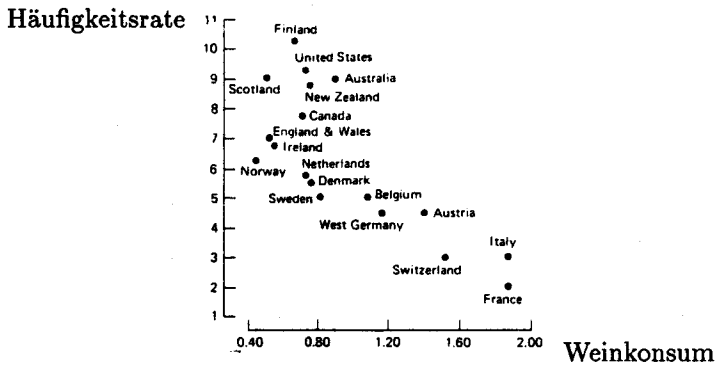
1849 sammelte Snow in London Cholera-Daten und stellte Zusammenhänge mit der Herkunft des verwendeten Trinkwassers in der jeweilig untersuchten Region fest. Dieses war eine der ersten systematischen epidemiologischen Studien. Die Ätiologie der Erkrankung war zu jener Zeit noch weitgehend unbekannt.

Große Probleme bei der Planung und Auswertung von epidemiologischen und klinischen Studien bereiten insbesondere die große Anzahl von Faktoren, die eine Erkrankung beeinflussen können sowie die sogenannten Niedrig-Dosen-Effekte. Zu den größten Herausforderungen der Biostatistik zählen zur Zeit die Beschreibung der Ätiologie von Herz-Kreislauf- und Tumorerkrankungen. Beide Erkrankungen haben erblich bedingte Anteile und auch der Umwelteinfluß in Form von gefährlichen Substanzen wie z.B. bei Lungentumoren von Zigarettenrauchen und Asbestinhalation steht außer Frage.

	Ages 55-64								Age 25-34	
	Hypertensive Disease		Ischaemic Heart Disease		Cerebro-vascular Disease		Bronchitis		Road Accidents	
	M	F	M	F	M	F	M	F	M	F
Doctors	-0.38	0.03	-0.51	-0.46	-0.69	-0.36	-0.42	-0.42	0.42	0.15
Nurses	0.10	-0.17	0.65	0.64	-0.20	0.11	0.14	0.25	-0.15	0.06
G.N.P.	-0.62	-0.22	-0.17	-0.26	-0.69	-0.77	-0.67	-0.65	0.10	0.19
Population density	-0.05	0.04	-0.44	-0.45	-0.02	-0.16	0.13	-0.08	0.00	-0.08
Cigarettes	0.26	0.23	0.28	0.44	0.08	0.22	0.44	0.35	0.47	0.34
Alcohol										
Total	-0.09	0.18	-0.70	-0.58	0.15	-0.16	-0.29	-0.28	0.38	0.21
Wine	-0.01	0.23	-0.70	-0.61	0.13	-0.14	-0.32	-0.27	0.13	0.02
Beer	-0.03	-0.09	0.23	0.31	0.14	0.17	0.35	0.22	0.37	0.25
Spirits	-0.38	-0.15	-0.26	-0.32	-0.35	-0.51	-0.57	-0.47	0.30	0.37
Calories	0.43	0.06	0.51	0.61	-0.02	0.36	0.57	0.67	-0.20	-0.03
Total fat	-0.16	-0.40	0.45	0.46	-0.45	-0.16	0.11	0.17	-0.18	0.01
Saturated fat	0.10	-0.17	0.64	0.62	-0.16	0.15	0.30	0.35	-0.08	0.05
Mono-unsaturated fat	-0.11	-0.42	0.60	0.60	-0.35	-0.05	0.11	0.15	-0.28	-0.17
Poly-unsaturated fat	-0.45	-0.28	-0.48	-0.47	-0.51	-0.60	0.33	-0.30	-0.05	-0.11
Keys' prediction*	0.04	-0.19	0.70	0.69	-0.10	0.19	0.24	0.27	0.06	0.14

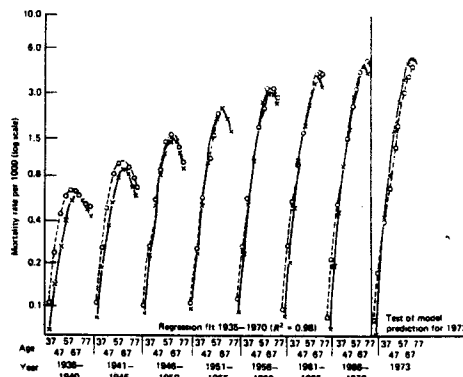
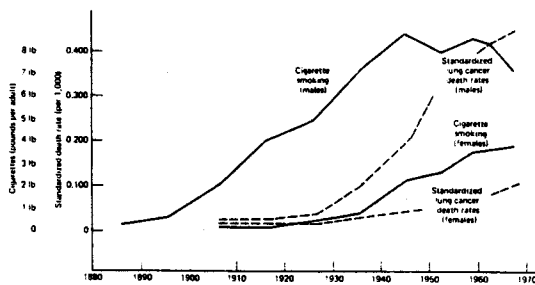
Korrelation verschiedener Erkrankungen und Variablen
(aus Kotz & Johnson: Encyclopedia of statistical science.)

Darüber hinaus sind aber für eine quantitative Modellierung die Interaktion dieser Faktoren untereinander sowie mit anderen Formen der Luftverschmutzung, mit Stress, Schlafverhalten, Alkohol und psychosozialen Faktoren zu berücksichtigen.



Korrelation von Weinkonsum und Herzinfarkt

Im allgemeinen werden die Modelle vom Biometer zusammen mit dem Spezialisten des Anwendungsgebietes etwa aus der Epidemiologie, Toxikologie, Pharmakologie, Radiologie, Genetik, etc. erstellt. Vom Biostatistiker wird schließlich auch erwartet, daß er wesentlich zur Entscheidung der Frage, ob eine Beziehung kausal ist oder ob nur eine Assoziation oder Korrelation vorliegt, beiträgt. Von einer kausalen Erklärung kann dann ausgegangen werden, wenn eine **Dosis-Response Relation** zwischen der Substanz und dem Tumor nachgewiesen wird. Es gibt aber auch andere Ansätze zur Erfassung von kausalen Beziehungen durch Betrachtung der zeitlichen Abhängigkeiten, Ausschluß des Einflusses von Faktoren in Kontrollstudien oder ähnliches. Eine kürzliche Studie ergab ein Nord-Süd-Gefälle bei Rauchern; im Norden wird mehr geraucht. (regionaler Einfluß)



Sterberaten für Lungentumor:

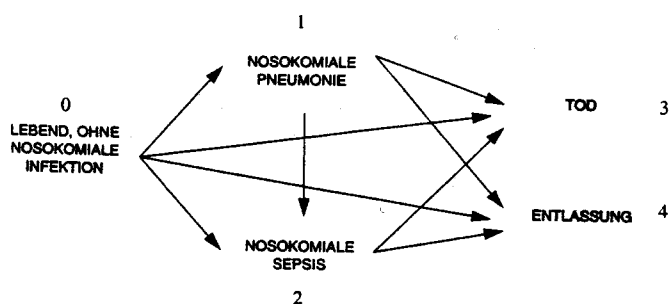
Regressionsfunktion

(aus Kotz & Johnson: Encyclopedia of statistical science.)

männlich - weiblich

Für eine Reihe von Studien (Verläufen) wurden explizite modellmäßige Zusammenhänge gefunden, z.B. 'power laws' der Form $y = \vartheta nrk^n(t - \lambda)^{r^n-1}$ für die Abhängigkeit der Lungentumormortalität von $t = \text{Alter}$, $\vartheta, k = \text{Populationsparameter}$. In neuerer Zeit sind auch zunehmend detailliertere Modelle und Analysetechniken angewendet worden. Erwähnen möchte ich die Verwendung von Punktprozessmodellen und Cox-Regressionsmodellen bei 'missing data' Problem, Verlaufsstudien, etc.

Mit Punktprozessmodellen wurden in einer Studie der Abteilung Biometrie in Freiburg neue Erkenntnisse über die Risikofaktoren für in der Klinik erworbene Infektionen (nosokomiale Infektionen) gewonnen. Insbesondere konnte eine durch einen typischen 'length bias' verursachte Überschätzung der durch nosokomiale Infektion verursachten zusätzlichen Aufenthaltsdauer in Intensivstationen, auf realistische Weise reduziert werden. Es können in diesem Modell auch dynamische Effekte beschrieben werden.



Potentielle Stadien und Übergänge zur Modellierung des Auftretens und der Auswirkungen nosokomialer Infektionen

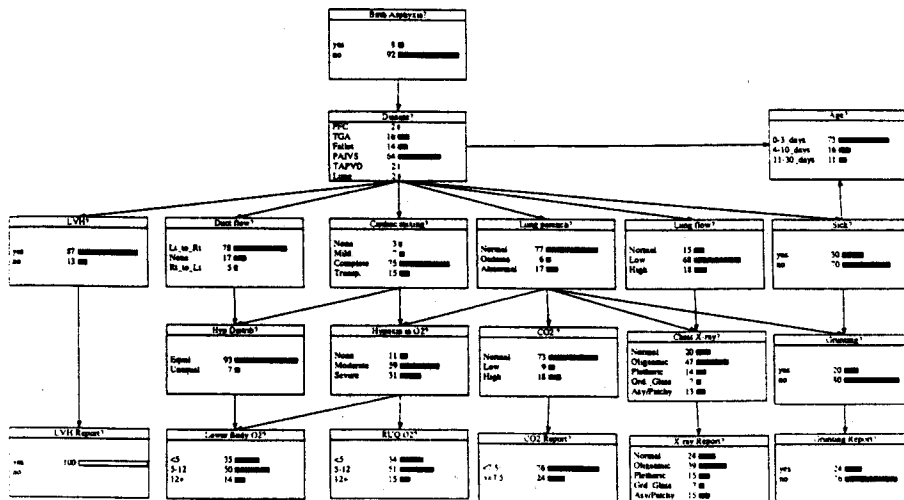
Ansatz	Postoperative Wundinfektion		Nosokomiale Pneumonie		Nosokomiale Beatmungspneumonie	
	Geschätzte zusätzliche Liegedauer	95%-KI	Geschätzte zusätzliche Liegedauer	95%-KI	Geschätzte zusätzliche Liegedauer	95%-KI
„Naiver“ Ansatz	20.7	(18.4, 23.0)	14.4	(10.7, 18.2)	16.6	(12.2, 21.0)
Matching-Ansatz	11.4	(7.1, 15.7)	8.2	(5.9, 10.5)	9.2	(5.9, 12.5)
Ansatz A	9.8	(5.7, 13.8)	3.4	(0.8, 6.0)	5.1	(1.5, 8.9)
Ansatz B	9.9	(5.5, 15.6)	3.4	(1.4, 5.4)	4.9	(2.5, 7.6)
Kontrafaktischer Ansatz	11.5	(8.9, 14.0)	4.0	(1.5, 6.1)	6.5	(4.2, 8.4)

Schätzwerte im Punktprozessmodell

(aus Schulgen (1995): Mehrstudienmodelle für Ereigniszeitdaten: Modellierung des Auftretens und der Auswirkungen nosokomialer Infektionen. Dissertation Freiburg/Dortmund)

Eine interessante und wichtige Entwicklung ist in der Konstruktion von graphischen Interaktionsmodellen und kausalen Modellen zu sehen. In diesen Modellen wird Expertenwissen über a priori-Verteilungen und geschätzten Übergangsverteilungen modellmäßig erfasst. Insbesondere bedingte Unabhängigkeiten verschiedener

Faktoren erlauben eine übersichtliche Modellierung. Als Ergebnis erhält man ein **wahrscheinlichkeitstheoretisches Expertensystem**, in dem medizinische Behandlungen vorgeschlagen werden und Untersuchungsergebnisse zu einer 'Propagation' (Revision) der Wahrscheinlichkeiten für bestimmte Erkrankungen führt.



Propagation der Wahrscheinlichkeiten nach positivem LVH-Report.

(aus: Spiegelhalter, Dawid, Lauritzen und Cowell (1993): Bayesian analysis in expert systems. In: Statistical Science)

Die bisherigen Ergebnisse dieser Systeme sind sehr positiv. Der Anwendungsbereich dieser Systeme geht über die medizinischen Anwendungen hinaus und ist auch z.B. für technische und wirtschaftliche Interdependenzmodellierung von großer Bedeutung.

E) Emissionstomographie Mit der Entwicklung immer leistungsfähigerer Computersysteme gewinnen sowohl in der biologischen Forschung als auch in der medizinischen Diagnose Abbildungstechniken an Bedeutung, die noninvasive Aufnahmen des menschlichen Körpers ermöglichen. Neben bekannten Verfahren wie der Kernspintomographie (NMR) und der Sonographie bietet vor allem die Computertomographie die Möglichkeit, nicht nur anatomische Strukturen sondern auch physiologische und biochemische Prozesse in vivo sichtbar zu machen.

Die Computertomographie läßt sich grundsätzlich in zwei Hauptgruppen unterteilen, die Transmissionstomographie und die Emissionstomographie.

Im klinischen Einsatz sehr weit verbreitet ist die gemeinhin als CT bekannte Transmissionstomographie. Bei CT-Untersuchungen wird der Patient von einer rotierenden Strahlungsquelle aus mit einem gerichteten Photonenstrahl beschossen, dessen Intensität beim Verlassen aus dem Körper gemessen wird. Der Konzentrationsabfall des Photonenstrahls ist ein Maß für das integrierte Rückhaltevermögen

entlang seiner Durchgangsbahn. Mit Aufnahmen von verschiedenen Seiten läßt sich z.B. über inverse Radon-Transformationen die Verteilung des Rückhaltevermögens und damit der Gewebestruktur im gesamten Körper rekonstruieren. Infolgedessen bietet die CT eine Möglichkeit, anatomische Strukturen des Gewebes sowie Verteilungen von Kontrastmitteln sichtbar zu machen.

Tabelle: Vergleich medizinischer Abbildungstechniken:

Methoden	gemessener Parameter	medizinische Anwendung
Digitale Subtraktions-Angiographie (DSA)	Verteilung von Kontrastmitteln	Darstellung des Gefäßsystems
Ultraschall - Sonographie - Doppler Verfahren	Zeit- und Frequenzänderungen von Ultraschallwellen	Anatomische Aufnahmen, Strömungseigenschaften in Gefäßen
Transmissionstomographie (CT)	Dichteunterschiede des Gewebes	Anatomische Strukturen, Kontrastmittelverteilungen
Kernspintomographie (NMR)	Relaxationszeiten, Kernspinresonanzen	Anatomische Aufnahmen, lokale Stoffwechselverteilungen
Emissionstomographie (ET)	Verteilung von radioaktiven Isotopen	Stoffwechselaktivitäten, lokale biochemische Prozesse

Im Gegensatz dazu können mit der Emissionstomographie erstmalig einzelne Körperfunktionen und spezielle Krankheitssymptome beobachtet werden, die nicht durch anatomische Strukturen, sondern nur in spezifischen lokalen Stoffwechselaktivitäten zum Ausdruck kommen. Dem Patienten wird eine mit einem radioaktiven Isotop gelabelte Substanz, ein sogenanntes Radiopharmazeutikum, injiziert. Das Radiopharmazeutikum lagert sich im Rahmen der normalen Stoffwechselaktivitäten der Trägersubstanz in einem Körperteil an. Entsprechend der Halbwertszeit des Isotopes finden dann Emissionen statt, die von außerhalb des Körpers angeordneten Detektorsystemen aufgezeichnet werden.

Das mathematische Problem besteht in der Rekonstruktion der Emissionsorte auf Grundlage der Beobachtungen in den Detektorelementen. Da die Verteilung der Emissionsorte auch die Verteilung der Trägersubstanz widerspiegelt, versucht man, aus medizinischer Sicht Aufschluß über lokale Stoffwechselforgänge der Trägersubstanz zu erlangen und spezifische Aktivitätsmuster mit Krankheiten oder funktionalen Phänomenen zu identifizieren. Die Verteilung der Emissionsorte wird durch ein Poissonprozessmodell beschrieben. Die Schätzung der Intensitätsfunktion dieses

Poissonprozesses ist das zentrale statistische Problem: es handelt sich von der Methodik her um ein Problem mit unvollständigen Daten. Der EM-Algorithmus, eine in der Statistik entwickelte Modifikation des Maximum-Likelihood-Algorithmus, ist das adäquate Verfahren zur Lösung des Problems.

Durch die Art der verwandten Isotope und durch die unterschiedlichen Detektorsysteme kann man die Emissionstomographie in drei Haupttypen unterscheiden.

1) **Positronen Emissionstomographie (PET)**. Die Positronen Emissionstomographie macht sich die Tatsache zunutze, daß nahezu alle biochemischen Systeme des menschlichen Körpers die Elemente Kohlenstoff, Sauerstoff, Stickstoff und Wasser enthalten. Mit C^{11} , N^{13} und O^{15} besitzen sie radioaktive Isotope mit Halbwertzeiten zwischen 2 und 20 Minuten, die sich mit den entsprechenden Träger-substanzen natürlicherweise in fast alle Stoffwechselsysteme einbringen lassen. Zum Labeln von Wasser wird das Fluorid F^{18} mit einer Halbwertzeit von 110 Minuten benutzt.

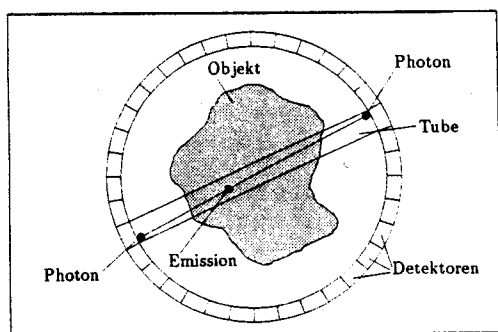
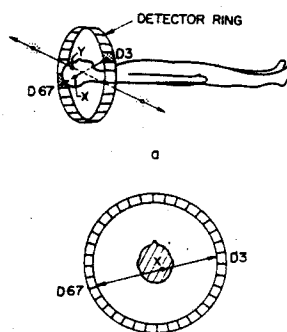


Abbildung 1:
Aufbau eines PET-Tomographen



Versuchsaufbau: Photonenstrahlung,
Detektorring

Entsprechend ihrer Halbwertzeiten emittieren die radioaktiven Isotope Positronen, die mit in der Nähe liegenden Elektronen zusammenprallen. Diese Annihilation setzt 2 Photonen frei, die in einem Winkel von 180° zueinander – aber in zufälliger Richtung vom Annihilationspunkt – auseinanderfliegen. Als γ -Strahlung durchdringen die Photone den Körper des Patienten und werden von Detektorsystemen außerhalb des Körpers aufgezeichnet, deren einzelne Elemente ringförmig um den Patienten angeordnet sind (Abbildung 1). Das nahezu gleichzeitige Auftreffen zweier Photonen läßt den Rückschluß zu, daß die Emission innerhalb des durch die beiden Photonen definierten Detektorstreifens, der sogenannten Tube, stattgefunden hat. Dabei wird die Zeit des Auftreffens zweier Photonen einer Emission als klein im Vergleich zur zeitlichen Differenz zwischen verschiedenen Emissionen angenommen.

Da zur Erzeugung der radioaktiven Isotope ein Cyclotron notwendig ist, sind die Kosten der PET Anwendungen enorm. Das erklärt, daß trotz überzeugender

Ergebnisse in der noninvasiven medizinischen Forschung der Einsatz in der klinischen Diagnose weiterhin beschränkt bleibt.

2) Time of Flight Positronen Emissionstomographie (ToF PET) Eine Weiterentwicklung der PET ist die Time of Flight Positronen Emissionstomographie (ToF PET). Sie ermöglicht bei gleicher Anordnung der Detektorsysteme und gleichen radioaktiven Isotopen neben der Identifikation von zwei Photonen zu einer Emission und der damit verbundenen Lokalisation des Emissionsortes in einer Tube auch die Beobachtung der zeitlichen Differenz zwischen den beiden Photonen. Mit der zeitlichen Differenz kann der Emissionsort innerhalb der Tube noch einer bestimmten Subtube (Abbildung 2) zugeordnet werden.

Ein wesentlicher Nachteil der ToF PET liegt in den zur Messung der zeitlichen Differenz notwendigen größeren Detektorelementen. So muß die Anzahl der Detektorelemente auf einem Detektorring gegenüber der PET reduziert werden. Eine Erhöhung der Auflösung durch die Unterteilung der Tubes in Subtubes wird somit durch die Verbreiterung der Tubes teilweise relativiert.

Insgesamt hat sich die ToF PET jedoch neben der PET als eigenständiger Emissionstyp durchgesetzt.

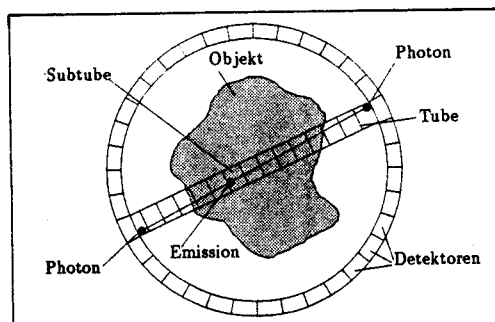


Abbildung 2: Aufbau eines ToF PET-Tomographen

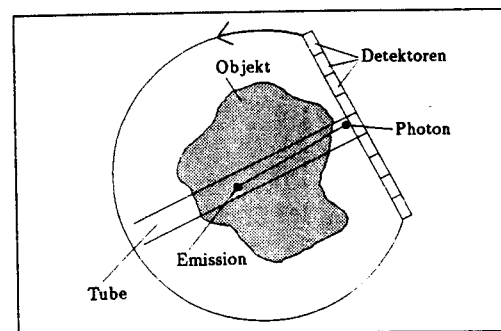


Abbildung 3: Aufbau eines SPECT Tomographen

3) Single Photon Emission Computed Tomographie (SPECT) Eine Emissionstomographieart, die sich aufgrund der geringeren Geräte- und Betriebskosten auch im klinischen Bereich weiter durchgesetzt hat, ist die Single Photon Emission Computed Tomographie (SPECT), die auch als Schichtszintigraphie (vgl. Pschyrembel) bekannt ist.

Bei der SPECT wird zum Labeln der Trägersubstanz die metastabile Version von Technikum 99, ^{99m}Tc , verwendet. ^{99m}Tc hat eine Halbwertszeit von 6 Stunden. Die beim Zerfall freigesetzte Energie besteht zu 90% aus γ -Strahlung.

Das im Gegensatz zur PET einzelne Photon fliegt in zufälliger Richtung vom Emissionspunkt weg und wird von einem planaren Detektorarray außerhalb des Patienten aufgezeichnet (Abbildung 3). Dem Detektorarray ist ein Gitter vorgelagert, so daß nur senkrecht zum Array auftreffende Photonen registriert werden. Das Auftreffen eines Photons in einem Detektorelement ermöglicht also den Rückschluß, daß die Emission innerhalb des senkrecht auf dem Detektorelement liegenden Streifens (Tube) stattgefunden hat. Mit Aufnahmen von verschiedenen Seiten des Körpers, die durch die relativ lange Halbwertszeit von ^{99m}Tc möglich sind, wird die Gesamtverteilung des Radiopharmazeutikums rekonstruiert.

Insgesamt liegt die Auflösung von SPECT-Bildern jedoch unter der von PET- oder ToF PET-Aufnahmen.

4) Anwendungen der Emissionstomographie In der Kardiologie, aber vor allem in der Gehirnforschung, gewinnen die Emissionstomographietechniken aufgrund ihrer gänzlich neuen Möglichkeiten immer mehr an Bedeutung.

Die Ursachen vieler Fehlfunktionen des Gehirns sind in Veränderungen unterliegender biochemischer Prozesse zu suchen. Eine Beobachtung dieser Reaktionen war bislang nur über Blutbilder, Analysen der Gehirnflüssigkeit oder des Urins und gelegentlich über Biopsien möglich. Exakte Messungen lokaler Verteilungen einzelner biologischer Substrate konnten an lebenden Patienten nicht durchgeführt werden. Auch lassen sich die molekulare Diffusion, die Proteinsynthese oder unterschiedliche Rezeptorsysteme darstellen. Die genaue Kenntnis dieser Prozesse ermöglicht im weiteren eine Analyse der Struktur, Organisation und chemischen Basis der normalen Gehirnfunktionen. Auf Arbeiten zur Lokalisation und Aktivität der verschiedenen Gehirnzentren weist Phelps et al. 1985 hin (siehe auch Abbildung 4).

Das Wissen um die Organisation des Gehirns und dessen strukturentwicklungsbedingte Zusammenhänge ermöglicht aber auch eine schnellere und genauere Diagnose einzelner Fehlfunktionen und Ausfälle. Therapeutische Maßnahmen zur Stabilisierung der unterliegenden Prozesse können gezielt eingeleitet werden. Gleichzeitig ist eine direkte Überwachung und Erfolgskontrolle möglich.

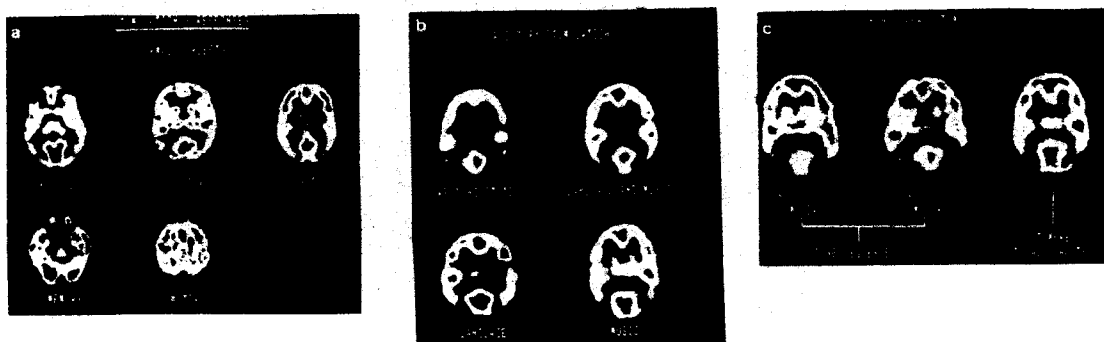


Abbildung 4: PET-Aufnahmen verschiedener Gehirnaktivitäten

(aus Phelps, Maziotto (1985): Positron Emission Tomography: Human brain function and biochemistry. In: Science.)

Phelps et al. (1985) beschreiben in einem Übersichtsartikel Anwendungen aus vielen Bereichen der Gehirnforschung wie den Anfallsleiden, bei Demenz Prozessen, neurodegenerativen Krankheiten oder akuten Gehirnfarkten aber auch bei psychiatrischen Syndromen, verschiedenen Formen der Epilepsie oder Depressionen auf. Beispielfhaft sei an dieser Stelle eine Untersuchung von Chorea Huntington (erblicher Veitstanz) erwähnt. Die Chorea Huntington ist eine vererbare Geisteskrankheit, die über erste Symptome wie Ausfälle, anormale Bewegungen, einen langsamen körperlichen Verfall und zunehmende psychische Veränderungen zwischen dem 30. und 45. Lebensjahr diagnostiziert werden kann und im folgenden einen chronisch progredienten Verlauf annimmt (vgl. Pschyrembel).

Abbildung 5 zeigt die anatomischen und funktionalen Veränderungen des Gehirns im Krankheitsverlauf einer Chorea. Bei den drei oberen Aufnahmen handelt es sich um CT-Aufnahmen, die unteren Aufnahmen stammen von einem Emissionstomographen. In der ersten Spalte ist der normale anatomische Aufbau sowie der normale Glucoseverbrauch des caudate nucleus (Pfeil) dargestellt. Die zweite Spalte zeigt Aufnahmen aus einem sehr frühen Stadium der Chorea Huntington. Im CT-Bild sind noch keine strukturellen anatomischen Veränderungen zu erkennen. Im Gegensatz dazu tritt in der ET-Aufnahme ein stark verminderter Glukoseverbrauch im Bereich des caudate nucleus hervor. Diese Glukoseunterversorgung führt dann zu den aus dem fortgeschrittenen Stadium der Chorea bekannten atrophischen Veränderungen der Ganglienzellen vor allem des Striatums und der Hirnrinde (Pschyrembel), wie sie in der dritten Spalte auch bei der CT-Aufnahme deutlich werden. Somit ermöglicht die Emissionstomographie durch die Beobachtung von Veränderungen im Glucosestoffwechsel eine Diagnose der Chorea Huntington Jahre vor der Erstmanifestation von klinischen Krankheitssymptomen.

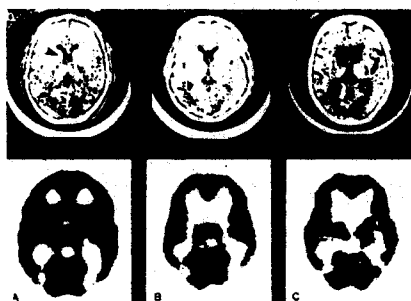


Abbildung 5: Vergleich von CT- und PET-Aufnahmen einer Chorea Huntington (aus Phelps, Maziotta (1985): Positron Emission Tomography: Human brain function and biochemistry. In: Science.)

Zusammenfassung Im Gesamtbild ergeben sich die folgenden Gesichtspunkte:

- Stochastik liefert eine wesentliche Erweiterung und Ergänzung der deterministischen Modell- und Theoriebildung.

- Es gibt in den unterschiedlichsten Gebieten eine Fülle von nur – oder zumindest vorwiegend – stochastisch beschreibbaren Phänomenen. Stochastik hat einen hohen Grad an Interdisziplinarität.
- Die Anwendungen der Stochastik liegen in den unterschiedlichsten Bereichen und fordern eine Zusammenarbeit über verschiedener Fakultäten hinweg.

Die Stochastik liefert Ordnungsprinzipien und Erklärungsmodelle für sinnfällige Phänomene (im weitesten Sinne) und trägt somit zum Verstehen der Welt bei.

Anmerkung: Den Kollegen C. Hipp (Karlsruhe), M. Schumacher und insbesondere H. Witting (Freiburg) danke ich für Hinweise und die Bereitstellung von Materialien, die in diesem Artikel verwendet wurden, sowie für zahlreiche Vorschläge zur Organisation und Darstellung.