Meddelelser, v/Eva B. Vedel Jensen
Matematisk Institut
Aarhus Universitet

BREV
Ukonvoluteret

PP
Danmark

Returneres ved varig adresseændring

Næste nummer af "MEDDELELSER" udkommer i begyndelsen af december.
Bidrag til dette nummer skal være redaktøren i hænde senest **onsdag den 23. november 1994**.
Bidrag bedes sendt til:

Meddelelser, v/ Eva B. Vedel Jensen
Afdeling for Teoretisk Statistik
Matematisk Institut
Ny Munkegade
8000 Århus C
eller med e-mail til: eva@mi.aau.dk

Samme adresse bedes benyttet ved **indmeldelse** i DSTS og ved **adresseændring**.

MEDDELELSER

Dansk Selskab for Teoretisk Statistik

19. aargang nr. 9 november 1994

TODAGESMØDE I SELSKABET

DEN 22. OG 23. NOVEMBER 1994

Afdeling for Teoretisk Statistik, Matematisk Institut, Aarhus Universitet

Efterårets todagesmøde afholdes tirsdag den 22. november og onsdag den 23. november 1994 på Matematisk Institut, Aarhus Universitet, Ny Munkegade, 8000 Århus C.

Middag og frokost: I forbindelse med mødet arrangeres der middag tirsdag aften kl. 19.00 i Matematisk Instituts kantine. Ledsagere er velkomne. Prisen for middagen er 180 kr. (studerende 90 kr.). Om onsdagen arrangeres der frokost i Matematisk Instituts kantine. Prisen for frokosten er 60 kr. (studerende 30 kr.). Tilmelding skal ske til Oddbjørg Wethelund, Afdeling for Teoretisk Statisik, Matematisk Institut, tlf. 89423532, senest onsdag den 16. november.

Program følger på de næste sider sammen med en lokale oversigt.

LOKALE OVERSIGT

Tirsdag 22. november	kl. 15.15-16.30	Auditorium D1
	kl. 17.00-18.15	Auditorium D1
Onsdag 23. november	kl. 10.00-11.15	Auditorium D2
	kl. 11.45-13.00	Auditorium D2
	kl. 14.00-15.15	Auditorium D3

PROGRAM FOR TODAGESMØDET

TORSDAG DEN 22. NOVEMBER

15.15–16.30 Michael Sørensen, Afdeling for Teoretisk Statistik, AU

Statistical inference for discretely observed diffusion models

Diffusion processes are solutions to stochastic differential equations and are often a natural choise when modelling dynamical processes that develop in continuous time. They are used in several fields including physics, biology, environmental sciences, engineering and finance. Observations of the process are typically made at discrete time points. A main problem when drawing statistical inference from discrete time data using a diffusion model is that the likelihood function is only known for very few models. It must therefore be approximated or replaced by a quasi-likelihood. A review will be given of some solutions to this problem which have been developed in the Aarhus research group on statistics for diffusions. First a relatively simple estimating function will be derived, and its properties will be studied analytically as well as in a simulation study. Calculation of the estimating function often requires simulation of the diffusion model. Also more complicated estimating functions will be briefly treated. Ideally, one would base the statistical inference on the likelihood function, and a method for calculating an approximate likelihood function by means of simulations will be presented and its properties will be discussed. Only results and no mathematical detail will be presented. Early in the talk a brief introduction will be given to diffusion processes and stochastic differential equations, including methods for numerical simulations of the stochastic process which solves such an equation.

16.30–17.00 Kaffe

17.00–18.15 Asger Roer Pedersen, Afdeling for Teoretisk Statistik, AU

Statistical analysis of Gaussian diffusion processes based on incomplete discrete observations

Gaussian diffusion processes defined by linear stochastic differential equations of the form

$$dX_t = (A_t X_t + a_t)dt + B_t dW_t, X_0 = \xi_0 \; t \geq 0,$$

where $A : [0,\infty) \to M^{d\times d}$ (the set of $d \times d$ matrices), $a : [0,\infty) \to R^d$ and $B : [0,\infty) \to M^{d\times r}(d \leq r)$ are deterministic functions of time t and W is an $r-$dimensional Wiener process, are important in many sciences from a modeling point of view, and from a statistical point of view they constitute a class of diffusion processes for which an exceptional variety of statistical methods is available.

We consider statistical analysis of such Gaussion diffusion processes based on incomplete observations given by

$$Y_{t_i} = T_{t_i} X_{t_i} + U_{t_i} + e_{t_i}, i = 0, 1, ..., n,$$

where $0 = t_0 < t_1 < \cdots < t_n$ are the discrete observation time-points, $\{T_{t_i}\}_{i=0}^n$ are non-random $m \times d$ matrices $(m \leq d)$ that specify which linear combinations of the coordinates of $\{X_{t_i}\}_{i=0}^n$ that can be observed, $\{U_{t_i}\}_{i=0}^n$ are non-random $m \times 1$ vectors representing additional inputs, and $\{e_{t_i}\}_{i=0}^n$ are random $m \times 1$ vectors that accounts for measurement errors. The statistical analysis consists of maximum likelihood estimation of unknown parameters, validation of the statistical model, reconstruction of the unobservable diffusion vectors from the data, forecasting and prediction.

The presented statistical methods are applied in an example from freshwater ecology, where a three-dimensional Gaussian diffusion process is used to describe the dynamics and interactions of the amounts of three forms of phosphorus in a lake. The statistical analysis is complicated by the fact that only two of the three forms of phosphorus can be observed. As a part of the analysis the corresponding amounts of phosphorus in the unobservable form are reconstructed. The estimated three-dimensional Gaussian diffusion model and the reconstructed amounts of phosphorus in the unobservable form is then used to derive a two-dimensional Gaussian diffusion model that is designed to make predictions about the effects of reducing the external loading of phosphorus to a polluted lake. Such models can be very helpful in the strategic planning of future actions of sewage plants and other environmental efforts.

19.00– ???? Middag i Matematisk Instituts kantine.

ONSDAG DEN 23. NOVEMBER

10.00–11.15 Michael Væth, Institut for Biostatistik, AU

Cancer mortality among the atomic bomb survivors in the Life Span Study

Since 1950 approximately 100.000 A-bomb survivors in Hiroshima and Nagasaki have been followed in the so-called Life Span Study (LLS) to assess the long-term effects of radiation on humans. A review of the mortality data and the statistical methods used for analysis of these data will be given. The talk will focus on the findings for cancer mortality and will discuss statistical problems related to the description and interpretation of the temporal patterns of the excess cancer mortality seen in the LSS cohort.

11.15–11.45 Kaffe

11.45–13.00 Martin Jacobsen, Institut for Matematisk Statistik, KU

Gaussiske diffusioner og autoregressive processer

En oversigt over teorien for tidshomogene Gaussiske diffusioner og sammenhængen til teorien for autoregressive processer, dels via svag konvergens og dels ved at betragte diffusionen til en diskret følge af ækvidistante tidspunkter.

De Gaussiske diffusioner er flerdimensionale processer der løser stokastiske differentialligninger af formen

$$dX_t = (A + BX_t)dt + DdW_t,$$

med W en flerdimensional Browsk bevægelse og A, B og D ikke-stokastiske, tidsuafhængige vektorer og matricer.

Det er specielt muligt at få processer med glatte komponenter (dvs med udfaldsfunktioner der kan differentieres en eller flere gange), og det er sådanne processer, der kan benyttes til at beskrive svage grænser af følger af autoregressive processer af orden ≥ 2 – det er velkendt, at passende følger af autoregressive processer af orden 1 konvergerer svagt mod 'rene' Gaussiske diffusioner uden glatte komponenter.

Blandt de Gaussiske diffusioner iøvrigt fremhæves de roterende Brownske bevægelser og roterende Ornstein-Uhlenbeck processer.

13.00–14.00 Frokost

14.00–15.15 Anders Stockmarr, Institut for Matematisk Statistik, KU

ML-estimation i grænser for autoregressive processer

Der diskuteres Maximum-Likelihood estimation i modeller for stokastiske processer i kontinuert tid med glatte komponenter, som er fremkommet som grænser for $\mathrm{AR}(r+1)$–processer. Disse modeller kan via en sufficient transformation opfattes som delmodeller af modellen

$$dX = (A + \sum_{i=0} B_i S^i(X)) + dW,$$

hvor W er en d–dimensional Brownsk bevægelse, S er den sædvanlige integrationsoperator, medens parametrene $(A, B_0, ..., B_r)$ varierer i $R^d \times \mathrm{Mat}_d(R) \times \cdots \times \mathrm{Mat}_d(R)$.

I tilfældet r=0 (svarende til at X er en Gaussisk diffusion) diskuteres den asymptotiske opførsel af estimatorerne, når tiden går mod uendelig.

Julemøde den 6. december 1994

Selskabets juleaftenmøde afholdes tirsdag den 6. december kl. 18.15 i Københavns Universitets Hovedbygning, Auditorium 6, Vor Frue Plads.

Foredragsholder er Gabi Schulgen, Freiburg, Tyskland. Efter foredraget vil der være mulighed for at deltage i fællesspining på en nærliggende restaurant. Sted og tilmeldingsprocedure annonceres i næste nummer af Meddelelser. Nedenfor er titel og resumé af foredraget.

Outcome-oriented cutpoints in analysis of quantative exposures: residence near high voltage facilities and the risk of cancer in children

by

Gabi Schulgen[1], Berthold Lausen[2], Jørgen H. Olsen[3] & Martin Schumacher[1]

1. Institut für Medizinische Biometrie und Medizinische Informatik, Albert-Ludwigs-Universität, Freiburg, Germany
2. Forschungsinstitut für Kinderernahrung, Dortmund, Germany
3. Danish Cancer Society, Division for Cancer Epidemiology, Copenhagen, Denmark

In the analysis of epidemiologic data in which exposure has been measured on a continuous scale, cutpoints can be defined to delineate categories or exposure can be modeled as a continuous covariate by assuming a special functional shape of the effect on disease status. Rules for classifying exposure into two or more categories range from a priori selection of cutpoints to data-orientated rules. The risk estimates may vary, however, with the choice of cutpoint. If the cutpoint selected is that for which the most impressive effect of exposure on outcome is observed, the final result must be qualified by adjustment. We will propose a method for adjusting results which are derived by varying the cutpoint on a specified selection interval. Adjustment is derived from the null distribution of the maximally selected test statistic (1, 2). The method should be applied to correct p values if the cutpoint used to define different levels of exposure is selected in such a way that the measure of difference between two risk groups, such as the odds ratio or relative risk, is maximized. No method is yet available for adjusting the resulting risk estimate and the corresponding confidence limits.

The statistical method has been applied in the analysis of the case-control study on association between exposure to magnetic fields and the risk of cancer in children, which was conducted recently in Denmark (3, 4). The objective of the Danish population based case–control study was to investigate whether residence before and after birth near 50 Hz high voltage installations increases a child's risk of cancer and whethter the risk correlates with the strength of the magnetic field. Study subjects are 1707 children under the age of 15 with leukaemia, tumour of the central nervous system, or malignant lymphoma diagnosed in 1968–86 and 4788 control children selected at random from the Danish central population register. A significant association was seen between all major types of childhood cancer combined and exposure to magnetic fields from high voltage installations of more than 0.4 mT (odds ratio 5.6). However, only very few children were exposed to magnetic field strength at or above that level. For a priori chosen cutpoints at lower levels no significant association was seen.

1. Miller, R. & Siegmund, D. (1982): Maximally selected chi-square statistic. *Biometrics* **38**, 1011–6.
2. Lausen, B. & Schumacher, M. (1992): Maximally selected rank statistics. *Biometrics* **48**, 73–85.
3. Olsen, J., Nielsen, A. & Schulgen, G. (1993): Residence near high-voltage facilities and the risk of cancer in children. *British Medical Journal* **307**, 891–5.
4. Schulgen, G., Lausen, B., Olsen, JH. & Schumacher, M. (1994): Outcome-oriented cutpoints in analysis of quantitative exposures. *American Journal of Epidemiology* **140**,172–84.

Statistisk seminar fælles mellem

Institut for Teoretisk Statistik, HHK og

Sociologisk og Statistisk Institut, KU.

Næste seminar bliver

Tid:	**Torsdag d.3. November kl. 15.30**
Sted:	**Større Øvelsessal, Økonomisk Institut**
	Studiestræde 6
Foredragsholder:	**Erling B. Andersen**
Titel:	**Anvendelser af residualer i latent struktur analyse.**

Kalender

SOC:	*Statistisk seminar.* Afholdes paa Økonomisk Institut, Studiestræde 6, København. Arrangeres af Institut for Teoretisk Statistik, Handelshøjskolen i København og Sociologisk og Statistisk Institut, Københavns Universitet.
3. november	Erling B. Andersen (Statistisk Institut, KU): Anvendelser af residualer i latent struktur analyse. **SOC**, Større Øvelsessal, kl. 15.30.
22.-23. november	Todagesmøde, Afdeling for Teoretisk Statistik, Matematisk Institut, Aarhus Universitet.