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 juni 1996.
Bidrag til dette nummer skal være redaktøren i hænde senest **fredag den 24. maj 1996.**
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

21. aargang nr. 4 maj 1996

25 års jubilæum

Dansk Selskab for Teoretisk Statistik

— se vedlagte hæfte

Selskabets bestyrelse:

Formand: Peter Dalgaard Biostatistisk Afdeling Panum Instituttet Blegdamsvej 3 2200 København N	Tlf: 35 32 79 19 Fax: 35 32 79 07 e-mail: p.dalgaard@biostat.ku.dk
Kasserer: Bendix Carstensen Statens Veterinære Serumlaboratorium Bülowsvej 27 1790 København V	Tlf: 35 30 01 39 Fax: 35 30 01 20 e-mail: bxc@svs.dk
Redaktør: **Eva B. Vedel Jensen** Afdeling for Teoretisk Statistik Matematisk Institut Århus Universitet 8000 Århus C	Tlf: 89 42 31 88 89 42 35 18 *(direkte)* Fax: 86 13 17 69 e-mail: eva@mi.aau.dk
Sekretær: David Edwards Novo Nordisk Krogshøjvej 5 2880 Bagsværd	Tlf: 44 44 88 88 44 42 61 35 *(direkte)* Fax: 44 44 42 10 e-mail: DEd@novo.dk
Jyske anliggender: Susanne Christensen Institut for Elektroniske Systemer Aalborg Universitetscenter Frederik Bajersvej 7 9200 Aalborg Øst	Tlf: 98 15 85 22 =20 98 15 42 11 • 5076 *(direkte)* Fax: 98 15 81 29 e-mail: susanne@iesd.auc.dk
Indkøbschef: Peter Allerup Danmarks Pædagogiske Institut Hermodsgade 28 2200 København N	Tlf: 31 81 01 40 Fax: 31 81 45 51 e-mail: nimmo@allerup.dpi.dk

Symposium on frailty models and their use in epidemiology

'Kommunehospitalet', Entrance 11a,
Øster Farimagsgade 5, 1399 Copenhagen K,
31 May -1 June 1996

Programme:

Friday, 31 May 1996

13.00-14.00	*Frailty models in epidemiology: some promising opportunities* by James W. Vaupel
14.00-15.00	*Multivariate survival data - an additive frailty structure* by Jørgen Holm Petersen
15.00-15.30	Coffee
15.30-16.30	*Genetic models for censored age-of-onset data* by Andrew Pickles
16.30-17.30	*Markov Chain Monte Carlo methods for correlated frailty models* by Duncan C. Thomas

Saturday, 1 June 1996

09.00-10.00	*A frailty model for clustered failure time data with patterned dependence structure* by Michael P. Jones
10.00-10.30	Coffee
10.30-11.30	*Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes* by Phillip Hougaard
11.30-12.30	*Strength and frailty of frailty modelling* by Anatoli I. Yashin

The symposium is organised jointly by Dept. of Biostatistics, Univ. of Copenhagen and Danish Epidemiology Science Centre.

Participation is free of charge but people interested in attending are kindly asked to leave a message at: Jørgen H. Petersen, Department of Biostatistics, University of Copenhagen, fax +45 3532 7907, e-mail: jhp@biostat.ku.dk

Frailty Models in Epidemiology: Some Promising Opportunities

by

James W. Vaupel

The development of frailty models over the past two decades has opened up some exciting possibilities for further research. In particular:

- Frailty was originally conceived as a measure of unobserved genetic factors. Advances in genetics and genetic epidemiology are revealing some of the alleles that underlie frailty. Furthermore, correlated-frailty models are being devised to analyse data on genetically-related individuvals. These two developments permit frailty models that integrate data from epidemiology and genetic epidemiology. These two fields were once distant but may now merge.

- Experiments with large populations of insects, worms, yeast, etc., may lead to deeper understanding of the dynamics of populations with hitten heterogenity and to less-questionable empirical estimates of variation in frailty across individuals.

- Frailty models can be adapted to estimate risk factors on the basis of cross-sectional data on the proportions of individuals at different ages with characteristics related to the risks. Such data, for instance, can be used to estimate the relative mortality risk of different ApoE genotypes.

Multivariate Survival Data - an Additive Frailty Structure

by

Jørgen Holm Petersen

A synthesis is proposed of the idea from traditional variance components models based on normal distributions, and modern survival analysis, with its emphasis on dynamic modelling of the hazard rate of independent individuals. This is done by superimposing an additive variance components type structure on multiplicative gamma frailty models for survival analysis. The additive dependence structure was first and only recently proposed by Yashin et al. \ 95 and Pickles et al. \ 94 in twin set-ups.

It is shown that, apart from consistency with earlier proposed models for independent individuals, the model also has statistical advantages: despite its elaborate nature, it is amenable to more or less straight-forward statistical analysis (non-parametric maximum likelihood estimation) smoothly extending the highly successful established techniques of survival analysis (in particular, the Cox regression model). It is demonstrated how the EM algorithm can be applied to this situation.

Genetic Models for Censored Age-of-onset Data

by

Andrew Pickles, London

A variety of model constructions are considered that enable components of variance associated with genetic and environmental influences to be estimated from censored age-of-onset data. Constraints on the marginal distributions and their implication for model construction are explained.

For proportional hazards models the relative importance of the distributional form of the frailty components, how their effects combine and the form of the conditional hazard function are described. As an alternative approach, the direct modelling of the age-of-onset density function is considered. This approach proves particularly attractive where the data are obtained by retrospective recall and may be subject to particular forms of measurement error.

The methods are illustrated using data on the age-of-onset of different aspects of puberty in a large US sample of adolescent twins and begins to examine the extent to which synchronicity in the timing of these different aspects is due to their sharing of a common genetic mechanism.

Markov Chain Monte Carlo Methods for Correlated Frailty Models

by

Duncan C. Thomas

Family studies of disease incidence are aimed at estimating familial relative risks and comparing alternative hypotheses for such dependency. One of the simplest models for dependent survival data is the shared frailty model, in which all members of a family are assumed to have the same value of an unobservable "frailty", conditional upon which survival times assumed to be independent and determined by a proportional hazards model.

The frailties are often assumed to have a gamma prior distribution, since it is conjugate to the Poisson form of the survival likelihood. Standard methods of inference on the parameters of this model are based on the marginal likelihood, obtained by integrating over the distribution of frailties, although Gibbs sampling methods have also been developed. In nonparametric estimation, the frailty distribution is assumed to consist of an unknown number of support points, with locations and masses (the relative risks and population prevalences, respectively) estimated by the E-M algorithm.

While reasonable for analysis of elementary structures like sibships or parent-offspring pairs, this model does not allow for differences in the degrees of relationship in larger pedigrees. Recent attention has therefore been directed towards correlated frailty models, in which each individual has a separate frailty with some known correlational structure.

One way to generate such a correlated frailty model is to postulate that each individual carries two or more frailties, which are shared exactly with immediate relatives. Likelihood methods now entail multiple integration over the joint distribution of the frailties, and have been implemented for the gamma frailty model.

An extension of the nonparametric approach to correlated frailties using Markov chain Monte Carlo (MCMC) methods will be discussed. This model can be interpreted as a single genetic locus with an unknown number of alleles, each conferring a different relative risk. Each individual is assumed to carry two frailties, one inherited from each parent with equal prior probability, with one or both contributing to the relative risk depending upon what is assumed about dominance.

The basic Gibbs sampler entails iterative selection of a frailty for each subject by sampling from their full conditional distributions, given the current assignments of relatives' frailties and the subject's own survival time. Parameter inference can be based either on its posterior distribution estimated by Gibbs sampling over both parameters and frailties in a fully Bayesian treatment or on a Monte Carlo estimate of the likelihood ratio obtained by Gibbs sampling only over frailties at a fixed parameter value.

A small simulation study has shown that multimodality in the frailty distribution is detactable using such nonparametric methods. However, although the resulting Markov chain can be shown to be irreducible, we have found that person-by-person Gibbs sampling of the frailties leads to unacceptably slow mixing and instead prefer to sample gentic descent states (i.e., the ancestral origins of each frailty). This method allows simultaneous changes to the frailties for all subjects who inherited the same allelee, leading to larger movements through the space of possible frailties.

We are currently exploring the use of "reversible jump MCMC" methods for modelling the population distribution of frailties with an unknown number of support points. Specifically, at each cycle, a proposal is made either to increase the number of support points by randomly splitting an existing point into two adjacent points in the reverse fashion. The proposal is then accepted with probability given by the ratio of the posterior probabilities of the two states divided by the ratio of the corresponding proposal probabilities. Conditional on the number of support points, their locations and masses are then randomly updated and the individuals' frailties are reassigned, using standard Gibbs sampling methods. We have found that this approach provides a good estimator of the underlying population distribution of frailties in simulated data, although it appears to overestimate the true number of support points.

Some extension of the MCMC methods to major gene segregation and linkage models, including polygenic components, environmental covariates, and their interactions, as implemented in the Genetic Analysis Package, will be briefly discussed, with applications to a study of breast cancer in 70 multiple-case families.

A Frailty Model for Clustered Failure Time Data with Patterned Dependence Structure

by

Michael P. Jones and Steven G. Self

Nielsen et al. (1992) introduced counting process models for multivariate survival data in which the hazard function for a cluster depends on an unobservable quantity called a frailty. They proposed use of the EM algorithm in obtaining maximum likelihood estimates of the parameters. We extend their work by proposing a more complex frailty structure which takes into account the known interrelationships among cluster members. Inference procedures for regression coefficients and the inter-cluster variance are derived from a pseudo-likelihood. Some simulation results are available.

Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes

by

Philip Hougaard, M.L.T. Lee and G.A. Whitmore

Count data often show overdispersion compared to the Poisson distribution. Overdispersion is typically modeled by a random effect for the mean, based on the gamma distribution, leading to the negative binomial distribution for the count. This paper considers a larger family of mixture distributions, including the inverse Gaussian mixture distribution. It is demonstrated that it gives a significantly better fit for several data sets, including one on the frequency of epileptic seizures. The better fit is due to the suggested distributions having a longer right tail. The same approach can be used to generate counting processes from Poisson processes, where the rate or the time is random. It can also be generalized to multivariate count data genrerated by several random effects.

Strength and Frailty Modelling

by

Anatoli I. Yashin

Which models of bivariate survival are most appropriate for the analysis of duration data on related individuals? In this paper several random effect models used in the analysis of related durations are reviewed. They include liability models used in genetic epidemiology, and frailty models exploited in demography and biostatistics. Properties of these models are compared. Role of frailty models in evaluation of the effects of omitted covariates on regression parameters in survival models is considered. Role of multivariate frailty models in the improvement of the efficiency of the regression parameters in survival models with observed covariates is discussed. It is shown that the correlated frailty model allows for combining the ideas of quantitative genetics and genetic epidemiology with methods and approaches used in demography and biostatistics to address interdisciplinary problems. Properties of this model with and without observed covariances are analysed. Several ad hoc estimation strategies are suggested. Methods for evaluation of the lower bounds for the biological limits of human longevity in the presence of observed covariates are discussed.

Nyt om navne

Fra *RSS NEWS* under **1996 medal awards** har vi følgende: "The **Guy Medal in Silver** is awarded to **Steffen Lauritzen** for his innovative contributions to statistical theory and practice, and especially for his work on graphical models including the two papers read to the Society 'Local computations on graphical structures and their application to expert systems' (*JRSS B*, **50**, 157–224, 1988, with D J Spiegelhalter) and 'Substantive research Hypothesis, conditional independence graphs and graphical chain models' (*JRSS B*, **52**, 21–72, 1990, with N Vermuth)."

SENIORFORSKER I STATISTIK

Ved Afdeling for Biometri og Informatik, Forskningscenter Foulum er en stilling som seniorforsker i statistik ledig til besættelse pr. 1. august 1996 eller efter aftale.

Se opslag i Magisterbladet.

Forskningslektor i statistik

Ved Det teknisk-naturvidenskabelige fakultet, Institut for Elektroniske Systemer, er en tidsbegrænset stilling som forskningslektor tilknyttet DINA Aalborg ledig til besættelse i perioden fra 1. september 1996 til 30. juni 1997 med mulighed for forlængelse. (**Stilling nr. 96303**).

DINA (Danish Informatics Network in Agriculture) er et netværk mellem forskningsmiljøer i jordbrugsforskning, matematik, datalogi og statistik. DINA er hovedsagelig finansieret af forskningsrådenes informatikprogram (PIFT) og har til formål at højne niveauet for anvendelse af informatik i jordbruget. DINA har forskningsgrupper ved Aalborg Universitet, Forskningscenter Foulum, Landbohøjskolen og Danmarks Tekniske Universitet. DINA Aalborg er placeret ved Afdeling for Matematik og Datalogi. Forskningen ved DINA Aalborg omhandler komplekse stokastiske systemer, og ansøgeren skal deltage aktivt i denne forskning.

Yderligere oplysninger kan fås ved henvendelse til Professor Steffen L. Lauritzen, tlf.nr. 98 15 42 11 - 5073, e-mail: steffen@iesd.auc.dk.

Notat om stillingsstruktur for videnskabeligt personale med forskningsopgaver ved de højere uddannelsesinstitutioner er gældende for stillingen og kan rekvireres ved fakultetet, tlf.nr. 98 15 85 22, lokal 7471. Her kan også indhentes oplysninger om kvalifikationskrav, ansættelsesvilkår og regler for behandling af ansættelsessager.

Ansættelse i henhold til overenskomst mellem Staten og AC.

Ansøgeren skal i sin ansøgning beskrive det grundlag, som den pågældende ønskes bedømt på for såvidt angår videnskabelige kvalifikationer.

Ansøgningen skal desuden indeholde dokumentation for:

Uddannelse.

Videnskabelige kvalifikationer. Der skal vedlægges publikationsliste med angivelse af de arbejder, ansøgeren ønsker at påberåbe sig. Bedømmelsesudvalget kan kræve tilvejebragt yderligere materiale.

Ansøgningen skal herudover indeholde oplysninger om personlige data og hidtidige arbejdsforhold m.m.

Bedømmelsen af ansøgere til stillingen sker ved fagkyndigt udvalg.

Ansøgningen mrk. stillingsnummer, bilagt publikationsliste, påberåbte publikationer og andet materiale i 3 eksemplarer samt liste over det medsendte materiale, skal være universitetet i hænde senest d. 17. maj 1996 med morgenposten.

Ansøgningen sendes til:

Aalborg Universitet
Det teknisk-naturvidenskabelig fakultet
Postbox 159
9100 Aalborg