MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

T a g u n g s b e r i c h t 10/1991

Medical Statistics: Statistical Models for Longitudinal Data

24.02. - 02.03.1991

The conference was organized by Norman Breslow (Seattle), Jochen Mau (Düsseldorf) and Martin Schumacher (Freiburg).

There were 38 participants from Germany, Denmark, Finland, Poland, Norway, USA, United Kingdom, France, Switzerland, and Austria. 34 talks were given, covering a wide range within and related to statistical models for longitudinal data.

Most of the talks can be put into one of the following three categories: growth curves and individual time series/repeated measures analysis, survival data analysis, and analysis of repeated observations of various types of data. The areas of application comprised comparative clinical trials, observational therapeutical studies, and followup studies in contexts of anthropometric, epidemiological, oncological, or drug research.

From the viewpoint of statistical methodology, the speakers applied a wide range of mathematical-statistical techniques: marked-point and Markov processes, counting processes, regression models, kernel smoothing algorithms, tests of interval alternatives, bootstrapping, penalized likelihood, multiple and sequential test procedures, and Fréchet-differentiable functionals.

K. ABT:

Planning for descriptive data analysis (DDA) of quantitative and/or qualitative longitudinal observations

The concept of Descriptive Data Analysis (Abt, Meth. Inf. Med., 1987, and Statistics in Medicine, 1991, in press) is discussed in its application to the design and analysis of longitudinal studies, here only to the "marginal" situation of estimating and testing population characteristics through time. Rather than treating the problem by a comprehensive marginal model, the analysis is carried out at each point and for each variable separately. The problem of multiplicity is treated in DDA by considering the N, say, p-values as descriptive characteristics rather then as error probabilities under the N corresponding, prespecified individual null hypotheses. In order that inferences may be drawn, values $p \leq \alpha_D$ (= 0.05, say) should appear in medically plausible, near regular patterns associated with prespecified minimally relevant effect estimates of near-homogeneous signs. For confirmatory statements at a level α_C (= 0.05, say), the Bonferroni-Holm procedure may be applied to a prespecified set of N_B (≤3, say) individual hypotheses. Alternatively, for a prespecified set of $N_{HR} \leq N$ hypotheses of about equal importance the corresponding global hypothesis may be rejected at α_C according to Hailperin-Rüger (Metrika, 1978) if at least a prespecified number $N'_{HR} \leq N_{HR}$ can be rejected each at nominal level $\alpha_{HR} = (N'_{HR}/N_{HR})\alpha_C$.

P. K. ANDERSEN:

Longitudinal analysis of follow-up measurements in clinical trials in chronic diseases

In clinical trials in chronic diseases the main outcome variable is often the survival time of the patients, though much more information on the course of the disease in the individual patients is collected. This consists of clinical, biochemical or histological variables recorded at the time of entry into the trial and at later follow-up times where the patients visit the clinic for further

examinations. We discuss how linear regression models (for continuous variables) and continuous time, discrete state Markov process models (for discrete variables) may be used to evaluate the effect of treatment on the course of the disease, and the use of such models is illustrated with data from clinical trials in liver diseases. The need for multivariate models for the joint distribution of survival time and course of variables is emphasised.

E. ARJAS:

Causality reasoning in longitudinal studies

Most people take it as self-evident that C, in order to be a cause of E, has to temporally precede it. The same holds for different notions of probabilistic causality. Apart from this obvious ordering, little attention has been paid to analysing and modelling explicitly such time progression. We make an attempt in this direction, by modelling the considered sequence of events in time (causal chain) in terms of the general framework of marked point processes, and considering then the associated prediction probabilities. Various dependencies on time, and on time delays, become then explicit. These are illustrated by some examples.
(This is joint work with Mervi Eerola.)

P. BAUER:

The choice of sequential boundaries based on the concept of power spending

It is suggested to choose discrete sequential boundaries for a given overall level α on the basis of the rate at which the power is spent over the inspection times. Three concepts are considered: Spending the power at a constant rate, proportional to the square root of the accumulated sample size and proportional to the accumulated sample size. Concrete two and three stage plans are calculated under the assumption of normally distributed test statistics

with known variance. The impact of the three concepts on bias of the "classical" estimate is investigated.

T. BEDNARSKI:

Robustness problems in Cox's regression model

A given parametric model $\{P_\theta\}_{\theta\in\Theta}$ can be extended to an infinitesimal supermodel, composed of shrinking nonparametric neighbourhoods around each probability distribution of the model. The rate of shrinkage is $1/\sqrt{n}$. An estimator is defined to be robust for the "ideal" model $\{P_\theta\}_{\theta\in\Theta}$ if it remains consistent and asymptotically normal in the infinitesimal supermodel. The robust estimators may be obtained via Fréchet differentiable functionals, Fisher consistent at the ideal model. It is demonstrated that the partial likelihood estimator for the regression parameter β in Cox's model is not robust. In fact the estimator may be arbitrarily inconsistent in the supermodel. This is due to the fact that the score function corresponding to the partial likelihood is unbounded. Modified "score functions" are proposed that lead to smooth (Fréchet differentiable) von Mises functionals and in consequence also to robust estimators in Cox's regression model.

Ø. BORGAN:

King-Hardy's method for graduating life-tables and M-estimators for counting process models

The method of King and Hardy (1880) for graduating life-tables reviewed, and its relation to M-estimators for counting process models for life history data is pointed out. General results for M-estimators for parametric counting process models are summarised. Illustration is given based on the survival of insulin-dependent diabetic females in the county of Fyn in Denmark 1973 - 80.

E: BRUNNER:

Rank procedures for repeated measurements with subject-treatment interaction

The test problem of fixed treatment effects is considered in the two-factor mixed model with interaction and unequal cell frequencies when the classical assumptions of normality do not hold. An explicit form of a test statistic is derived using a partial rank transform (ranking all observations within each block) and the asymptotic distribution of the statistic is determined under the assumption that the number of blocks tends to infinity and the cell frequencies are bounded. The statistic reduces to Friedman's statistic if no interactions are involved in the model and all cell frequencies are equal and hence the proposed test can be regarded as a generalisation of Friedman's test for repeated observations. In case of two treatments the exact conditional distribution is determined and estimators and confidence intervals for the shift effect are proposed.

D. CLAYTON:

Hierarchical generalized linear models with intrinsic Gaussian priors (or: Some unexpected uses for a disease mapping program)

The analysis of maps of disease incidence and mortality using Bayesian image analysis techniques suggests some interesting parallel analyses of repeated measures and other time-ordered data.
Typically such methods depend upon prior distributions which can model "smooth" geographical variation. Intrinsic (improper) Gaussian distributions are a convenient choice. Estimation may be carried out using approximation methods, such as penalized likelihood, or by Monte Carlo ("Gibbs sampling") approaches.
Mapping programs may be extended to incorporate covariates both at area ("ecological") and subject level. The resultant models have characteristics of both "multilevel" generalized linear models and of semi-parametric GLMs (or generalized additive models).

An example of an ecological regression analysis was discussed, and the interpretation of parameters indicated.
A possible use of the methodology for modelling intensity on the Lexis diagram was proposed.

A. EKHOLM:

Maximum likelihood combined with empirical Bayes for many short binary time series

Fitting of the logistic normal model for many short binary time series is considered. The covariates may be time dependent. The following stepwise procedure is proposed and motivated. Firstly, we fix the value of the variance of the normally distributed random effects, and find starting values for the regression parameters by maximum likelihood, using the fitting procedure of Ekholm, Green and Palmgren (GLIM Newsletter, 1986). Secondly we "postdict" the random effects by an empirical Bayes procedure. Thirdly, we find estimates for the regression parameters by setting off the postdicted values of the random effects. Finally, we choose as best estimates those obtained for the fixed value of the variance that maximizes the marginal likelihood. Simulation results are reported. A data set on respiratory infection in Ohio children is analyzed. These data have been analyzed before by Laird, Beck and Ware (Preliminary Draft, 1984), and by Zeger, Liang and Albert (Biometrics, 1988).

L. FAHRMEIR:

Dynamic models and conditional mode estimation for categorical longitudinal data (with applications to survival models)

Dynamic (state space) models relate a sequence $\{y_t\}$ of observations to "states" $\{\beta_t\}$ by an observation model. States, which may be stochastic trends, time-varying covariate effects etc., are assumed to obey a Markovian transition model.

To deal with categorical $\{y_t\}$, we replace the linear observation model in the well-known Gaussian situation by a categorical observation model, while generally retaining a linear Gaussian transition model.

To avoid high-dimensional numerical integration, we propose to estimate ("filtering and smoothing") $\{\beta_t\}$ by modes instead of by means of the conditional distribution given by the data $\{y_t\}$. This leads to a penalized likelihood criterion which can be maximised by Kalman-filter and -smoother-type algorithms.

The approach can be easily extended to categorical longitudinal data as well as to discrete time survival models, e.g. to dynamic versions of the grouped Cox model. It allows simultaneous smoothing of baseline hazard rates and estimation of covariate effects. Some illustrations to real data sets are presented.

U. FELDMANN:

Canonical categorical regression

A response variable y with categorical outcome y=i (i=1,..,m) is considered which is linked to a vector of continuous or discrete explanatory variables $\mathbf{x}$ by the posterior probability pr(y=i|$\mathbf{x}$). The canonical model reads:

$$F(\theta_i - \beta_i'\mathbf{x}) = \frac{pr(y=i|\mathbf{x})}{pr(y=i|\mathbf{x}) + pr(y=i+1|\mathbf{x})} \qquad (i=1,..,m-1)$$

$F(\cdot)$ is any continuous cumulative distribution function and θ_i are threshold parameters while β_i denote regression parameter vectors. If all β_i are different from each other, one deals with a totally unordered model. If there is a sequence of at least two identical neighbouring regression vectors, e.g. $\beta_{i-1} = \beta_i$ and if the corresponding thresholds are ordered, e.g. $\theta_{i-1} < \theta_i$, then a partially ordered model is achieved, i.e. the categories y=i-1,i,i+1 are ordered.

If all regression vectors coincide, i.e. $\beta = \beta_i$, and if the thresholds are ordered , i.e. $\theta_{i-1} < \theta_i$, then the totally ordered model is achieved. In this instance the allocation rule

$$y = i \quad \text{if} \quad \theta_{i-1} \leq \beta' x < \theta_i \qquad (i=1,..,m; \quad \theta_0 = -\infty \; ; \; \theta_m = +\infty)$$

corresponds to the Bayes' allocation, provided that $F(0) = 0.5$ holds. The given approach permits a data generated model choice on the basis of maximum likelihood estimation and is applied to medical prognostics.

TH. GASSER:

Approaches to deal with sample of curves

First the single curve regression problem is considered. The nonparametric estimator by the kernel method involves the choice of a bandwidth, which is crucial for the quality of the estimate. A new approach is presented to estimate the optimal bandwidth from the data by using the asymptotic formula. Both in theory and in simulations this outperforms cross-validation type estimators.
The second part is devoted to samples of curves. Interindividual variation occurs typically not only in the intensity but also in the dynamic. First, a true transformation is estimated from the individual data, in order to synchronise individual curves to an average dynamic. After this alignment procedure further statistical procedures can be applied such as averaging in order to get a valid average curve. The method is illustrated by the application to human growth curves.

G. GIANI:

Simultaneous subset selection for comparing several experimental treatments with a control

The problem of comparing several experimental treatments with an unknown control is considered in the context of multiple decision

making. The objective is to identify a subset which contains all treatments being in some sense equivalent to the control to separate them from all the inferior and all the superior treatments. Bhattacharyya (1956) has dealt with this problem for a known standard regarding normally distributed random variables. His proposal for treating the unknown control cases is incorrect. Later Seeger (1972) again picked up this problem in a slight modification. By introducing two controls he tries to overcome the critical points in determining strict lower bounds on the probability of correct selection. But note that Cane (1979) has pointed out an error in Seeger's main result.

Given a distribution function with location parameter and log-concave Lebesgue density a decision rule is proposed determining three disjoint subsets of treatments which should contain all equivalent, all inferior, and all superior ones, respectively. It is shown that the probability of simultaneous correct selection is minimised in one of a finite number of parameter configurations. Finally, the normal case is treated in detail where the rule is based on the sample means. In order to guarantee at least a prespecified probability of correct selection over the entire parameter space, it is clear that a minimum sample size has to be required. This sample size can exactly be calculated and its order of magnitude is demonstrated via some examples.

A. HAMERLE:

<u>On the robustness of covariate effect estimates in parametric hazard rate models</u>

This paper investigates the sensitivity of the (pseudo-) maximum likelihood estimates of the covariate effects to misspecification of the hazard rate to the presence of unmeasured heterogeneity that has not been accounted for. We investigate the wide class of parametric hazard rate models which have a linear model representation in logT. Estimation is based on two classes of models, the Weibull model and the lognormal model. It is shown that in both the Weibull and the lognormal model the (pseudo-) maximum likelihood estimators of the covariate effects are consistent and asymptotically normally

distributed, regardless whether unobserved heterogeneity is present or not, and regardless whether the individual hazard rate is correctly specified or not. It is only assumed that the true model is a linear model in logT, and that the regression is correctly specified. We propose robust variance estimates of the parameter estimates, since the usually calculated standard errors from the (misspecified) likelihood are not the correct ones. It is also shown that the duration dependence parameter cannot be correctly estimated if any misspecification occurs. Finally, the results of some simulation experiments are presented.

R. HILGERS:

A nonparametric simultaneous test for sensitivity and specificity of diagnostic tests

The quality of diagnostic tests not only has impact on general health care but is also basic to all mathematical models having the distinction between healthy and diseased individuals as primary (or implicit) variable.
In this paper a procedure is proposed testing simultaneously whether a specific diagnostic test has at least prespecified values for sensitivity and specificity. The onset is purely nonparametric with respect to the distributions of the (clinical) parameter of interest for healthy and diseased individuals and also distribution-free for test level and power. The test statistic is based on the intersection or non-intersection of the two one-sided nonparametric tolerance regions for sensitivity and specificity, respectively, at an adaquately chosen confidence level.
The approach may be extended to confidence bounds for ROC-curves and even be implemented into a group sequential strategy. A corresponding permanent quality control may also be incorporated.

N. KEIDING:

Statistical inference in the Lexis diagram

The Lexis diagram is a (time, age) coordinate system, representing individual lives by line segments of unit slope, joining (time, age) of birth and death. The main theme of this paper is non-parametric continuous-time statistical analysis on the Lexis diagram, where I indicate some possible approaches within modern survival analysis. I also introduce the history of the diagram, point processes on the diagram, and the classical statistical approach based on piecewise constant intensities. The Lexis diagram is also useful for describing morbidity, and the methodology is illustrated by two Danish studies of diabetes incidence (Phil. Trans. R. Soc. Lond. A (1990) **332** , 487-509).

H. KLINGER:

Remarks on statistical models for bioequivalence assessment

In many cases, the assessment of bioequivalence of two drugs, more specifically of a "standard" and a "generic" or "test" drug, is based on estimates of the "area under the curve" (AUC) which are calculated from measurements of serum concentrations of the active compound.
Let (X,Y) denote the pair of random variables of "estimated AUC under standard" and "estimated AUC under test", respectively, observed on the same subject. Standard and test are often defined to be bioequivalent if a suitably chosen parameter θ of the distribution of (X,Y) fulfils the condition $\theta_1 < \theta < \theta_2$ for prescribed values of θ_1 and θ_2. The resulting statistical test problem is given by $H_0 := \{\theta: \theta \leq \theta_1\} \cup \{\theta: \theta \geq \theta_2\}$ and $H_1 := \{\theta: \theta_1 \leq \theta \leq \theta_2\}$.
The consequences of different choices of θ,

$\theta = E(Y/X)$, $\theta = E(Y)/E(X)$, $\theta = \text{median}(Y/X)$, and $\theta = P(Y>X)$

are discussed under the assumption that the distribution of (X,Y) is either bivariate normal, bivariate lognormal, or an element of a suitably chosen nonparametric family of distributions. It is shown that only $\theta = P(Y>X)$ under either a lognormal distribution or a

nonparametric family of distributions leads to satisfactory statistical procedures.

K. KRICKEBERG:

Information systems for primary health care and their applications to longitudinal studies

In developing countries the main tool for planning, managing and monitoring primary health care and for evaluating its impacts is an integrated information system. While supporting the clinical and epidemiological activities of health workers in small health centers it provides at the same time the necessary indicators up to the provincial and national level. It also contributes to the integration of primary health care itself.
The present talk describes a system where a balanced use of the various methods is attempted: the traditional complete routine recording and reporting as well as sampling procedures and epidemiological studies. An essential feature is a provision for correcting errors in indicators due to wrong diagnoses, reporting gaps etc., based on statistical ideas from areas like medical tests. Longitudinal studies incorporated into the system can be used to monitor and improve diagnostic and therapeutic standard schemes of primary health care.

J. KÜBLER:

A nonparametric estimator of the risk function in the proportional hazards model

A widely used regression model for possibly censored survival times is the proportional hazards model suggested by Cox (1972). This model specifies that a covariate X has a proportional effect on the survival time distribution. That means the hazard function can be represented in the form $\lambda(t|x) = \lambda_0(t)R(\eta(x))$, where $\lambda_0(t)$ is a unspecified baseline hazard, R is a known strictly monotone function and η a regression function. This representation of the hazard

function, however, is not unique. This aspect will be taken into account when deriving a new nonparametric estimator of η, which is the main topic of this talk.
After a short repetition of some basic notations, a survey of classical concepts of estimation in the proportional hazards model is given.
When constructing a nonparametric estimator of η, tools of nonparametric regression have to be used. Therefore, two different techniques which are common in this field are introduced in a more general context: the kernel method and the closely related nearest neighbor estimation. The latter has been applied in a slightly generalized form by Tibshirani & Hastie (1987) to the partial likelihood function. Thus, an algorithm can be given to calculate an estimator of η under certain restrictions on the risk function $R(\eta)$ to get a unique representation of the hazard function. This algorithm is based on a simultaneous estimation of the hazard and the regression function. In addition, the third step of the algorithm takes account of the above mentioned restrictions. Important properties of this new estimator of η are discussed. Finally, a simulation study is presented which compares the estimator proposed by Tibshirani & Hastie (1987) with the new estimator of η where the latter is calculated by using different kernels.

N. LAIRD:

Smoothing growth curves using Kriging predictors

Smoothed growth curves for individuals are useful in many contexts, including identifying features of the curves and developing normative standards. Kriging is a method of producing Best Linear Unbiased Predictors developed by geologists for creating contour plots of a random variable distributed over a region. The method requires only specification of the variogram of the process as a function of the distance between the points. We show how the same technique can be modified to produce smoothed growth curves, based on estimating the covariance function of the observations as a function of the time between the observations. Application to a

sample of lung function growth in children of ages 6 to 18 is discussed.

W. LEHMACHER:

Multiple testing with two groups of repeated measures

There are several multivariate tests existing for the comparison of two independent groups of repeated measures with L time points, like Hotellings's T^2 or O'Brien's sum tests. It is shown, that these tests are suitable for the application of the closed testing principle of Marcus, Peritz and Gabriel. It facilitates the construction of multiple testing procedures controlling the experimentwise error rate. After the rejection of the global hypothesis all lower dimensional marginal hypotheses are tested in a stepwise manner until finally the local hypotheses at the L timepoints are tested. A hypothesis is then rejected if its related test as well as all tests for higher dimensional hypotheses containing this hypothesis are significant at level α. This procedure is just as powerful as the multivariate test, and, in addition, it is possible to detect differences in single time points or sets of time points. Further strategies, like a modified Holm procedure, are discussed.

K.-Y. LIANG:

Random effects estimation in generalized linear models

We propose a method for estimating the random effects in the generalized linear mixed model for either discrete or continuous data. The main driving force is the use of estimating function which is essential for repeated discrete responses.
We suggest a simple bootstrap method to estimate the uncertainty of random effect estimates. An application to the estimation of trends in AIDS incidence across risk groups and regions is presented.

J. MAU:

The Lexis diagram for multivariate survival times

Meaningful graphical methods highlight selected aspects of multivariate data. Separating intra- from inter-individual variation while regarding both real and experimental time is a common objective in the analysis of long-term studies of survival patterns. These goals are achieved by the constellation graph - a scatter diagram in a $(1, \pi)$-semicircle where each vector of multivariate survival times is represented in polar coordinates - and a generalization of the Lexis diagram - a polygon representation of each life history in a total-time-on-test vs. real-time coordinate system. In some analogy to univariate survival times in a Lexis diagram, one may define a marked point process associated with each vector of survival times to obtain a statistical model as a basis for estimating some underlying intensities of interest. In the context of the Lexis diagram these are typically conditioned on the number of components still at risk.

K. MIESCKE:

Statistical analysis of periodic time series with application to circadian rhythms

Periodicity of time series is observed in various situations of real life. One for example is biorhythms in man. Hereby, the daily cycle deviates from its normal 24 hour value due to works in shifts, long distance flights, or some other lack of synchronization with the natural day and night cycle, whereas it is highly desired to stabilize the same to keep a person at the usual daytime performance level.
Besides the problem of detecting and estimating a period, there are also situations where the period is known and a technique is needed to measure the closeness to it. Likewise, two periods may exist simultaneously and a decision has to be made which of the two is dominating.

Several existing methods are described and discussed. New models and methods including nonparametric approaches and smoothing techniques are proposed. Some applications to a winter depression study, where the clients lived through several 26-hour days, are given using body temperature data.

CH. E. MINDER:

A time series model with random coefficients

In many time series applications, one observes simultaneously the responses of several similar observational entities over time. In such situations, one often is not interested solely in the parameters governing the responses of each individual entity, but also in their distribution.
Under the assumption that subsystems are independent and first order autoregressive, two methods for estimating parameters of the super-distribution of entity-parameters will be presented. One is based on the method of moments, and thus distribution-free, the other is based on a parametrisation of both error distribution (normal) and super-distribution (truncated normal).
Apart from simulations, and an example, connexions and implications for hierarchical models in general are discussed.

H.-G. MÜLLER:

Estimating direction fields in autonomous equation models, with an application to T4 cell counts in a HIV-seropositive cohort

Assume pairs $X_i = f(T_i)$, $Z_i = {}^{d}/_{dt}\, f(t) \mid_{t=Ti} + \eta_i$ are observed, where f is strictly monotone and smooth, and T_i are unobservable i.i.d. r.v.'s, η_i is noise.
It is of interest to recover f, which satisfies the autonomous equation $f' = g(f)$, $f(t_o) = g_o$, for some unknown g which satisfies $g(x) = E\,(Z \mid X=x)$. It is proposed to estimate g by nonparametric regression, and f by integration of the autonomous equation with $\hat{g}$ in lieu of g. Consistency is discussed.

The method is applied to obtain an estimated time trajectory f(t) of T4 lymphocyte marker cells in a HIV-seropositive cohort, where each subject is observed for a short time period. From these measurements, X_i and Z_i are obtained. The time origin t=0 corresponds to seroconversion, T_i to the unknown time between seroconversion and appearance of the subject in the study.

M. OLSCHEWSKI:

Methodological problems in the analysis of survival data incorporating quality of life measurements

Standard endpoints used in survival analysis are death, relapse or other disease-specific endpoints. Nowadays, additional endpoints describing the subjective well-being of patients under therapies are introduced for an evaluation of therapeutic regimens. The measurement of this so-called quality of life parameters is discussed. A combined analysis of the classical survival endpoint with these quality of life measurements is proposed by introducing quality of life adjusted survival times (QUALYs). A major drawback when applying methods of survival analysis by QUALYs is that in case of censored observations a dependency of the censoring distribution with the QUALY distribution is induced leading to serious bias in the estimation of the QUALY distribution. Some proposals of overcoming this problem are presented. Simulation results are used to illustrate the amount of bias.

A. NEISS:

Bone narrow transplantation (BMT) versus chemotherapy (CHEMO) in leukemia patient: a non-randomized comparison.

484 CHEMO- and 243 BMT-patients were compared with regard to disease free survival time (DFS). Using the original data there is a bias in the result because of unbalanced distributions of the covariate, and the waiting time for transplantation. Two procedures

to eliminate this bias are discussed and applied to the data. The first one is to give a penalty to the BMT-group by using a stochastic process approach which leads to a modified Kaplan Meier estimator for the adjusted survival function. The second one is based on ideas in the analysis of truncated data giving the CHEMO group an award. The unbalanced distributions of concomitant variables are adjusted by using the proportional hazard model.
The results show that no therapy offers a substantial advantage over the other.

R. REPGES:

Adverse drug reactions after short time exposure: a counting process approach

The usual assessments of adverse drug reactions assume a constant exposure. So do also epidemiologic studies for risk assessments. To model the situation of short time exposure, exposures of varying doses and various covariates, we consider a counting process $N_{ij}(t)$, counting the transitions from state i to state j up to time t. The states are a list of all occurrences of exposure and reactions. By means of the Aalen-Nelson theory the compensator A_{ij} of N_{ij} can be estimated, such that $N_{ij}-A_{ij}$ is a martingale M_{ij}. Furthermore $<M_{ij}>$, the variance process, is also estimated by A_{ij}. Under mild conditions on N_{ij}, the A_{ij}-process is

continuous, $A_{ij}(t) = \int_0^t \lambda_{ij}(s)ds$, and $\lambda_{ij}\,dt = P(dN_{ij}(t)=1 \mid F_t)$.

This relates to a Markov chain with continuous time. It can be decomposed in Poisson point processes with parameters $q_{ij} = \lambda'_{ij}(0)$ and are thus special cases of the counting processes $N_{ij}(t)$ considered.

B. SCHNEIDER:

Neuronal networks and biometry

Neuronal networks are defined as following:

a) There are N units, each being characterized by its state $a_i(t)$ at time t, where $a_i \in \mathbb{R}_1$, $(i = 1,\ldots,N)$.
b) The state $a_i(t)$ produces an output $o_i(t)$ which may be transferred to other units $(o_i(t) = f(a_i(t)))$.
c) There are weights $w_{ij} \in \mathbb{R}_1$ $(i, j = 1, \ldots N)$. If the unit j has output $o_j(t)$ it influences the unit i with the partial net input: $net_{ij}(t) = w_{ij}o_j(t)$. The total net input which unit i receives from all other units j is: $net_i(t) = \Sigma_j\, net_{ij}(t) = \Sigma_j\, w_{ij}\, o_j(t)$.
d) (updating rule): It is assumed that the net works in discrete time intervals t= 0,1,2, ... synchronously. If the units are at time t in states $a_i(t)$, they are at time t + 1 (without external excitation) in the states: $a_i(t+1) = F(net_i(t))$. Suitable updating functions F (.) are: the linear threshold $(F(.) = 0$ for $net_i < \theta$ and $= 1$ for $net_i \geq \theta)$ or the stochastic logistic function: $P(a_i(t+1)=1) = (1+ \exp(- net_i(t)/T))^{-1}$.

In some examples the construction of networks for classification and clustering are shown. An important feature is "learning" of the network, i.e. the change of the weights w_{ij} by giving a learning pattern a_{pj} to unit j and a_{pi} to unit i. Learning rules are: the Hebbian rule: $w_{ij} = a_{pi}a_{pj}$ and the Delta rule: $w_{ij} = (a_{pi}-a_i)a_{pj}$, where a_i is line real activation state of unit i produced by the learning input pattern a_{pj}.

K. ULM:

Isotonic regression for dose-response relations

The proof of a dose-response relationship is an important criterion to establish causality. The usual way to prove this relationship is to define subgroups with respect to the exposure and to calculate

the risk (incidence or prevalence) in these subgroups.
The methods described in the literature to analyze this relationship are a test on trend or parametric models (e.g. Poisson regression). In all these methods the levels of the exposure in the subgroups have to be quantified and the results are depending on these levels. Therefore a nonparametric approach, the isotonic regression, is proposed. The only assumption is monotonicity, which means, the response cannot decrease. The result is independent of the definition of the exposure-levels in the subgroups.
For analyzing the association an exact permutation test is recommended. The approximation based on normal distribution agrees very well with the exact test.

M. VÆTH:

Poisson regression methods for grouped survival data: The impact of measurement errors in covariates and other sources of excess variability

Grouped survival data is often analyzed using Poisson regression methods. It is shown that overdispersion induced by heterogeneity on the individual level leads to a variance-mean relation of the form variance/mean = $1 + \phi^2$. Moreover ϕ^2 is essentially equal to zero when the association between the random component and the time at risk is negligible. A reasonable solution in measurement error problems is then to replace the hazard function $h(\beta x)$ by $E(h(\beta x)|z)$ where x denotes the true (but unobserved) covariate and z is the measured covariate. Some numerical results will be presented to evaluate the impact of measurement errors and other sources of excess variability. The present investigation was initiated in connection with work on the analysis of the cancer mortality among the A-bomb survivors in Hiroshima and Nagasaki.

J. WAHRENDORF:

Repeated longitudinal determination of HPV-status of pregnant and non-pregnant women

The Human Papilloma Virus family (HPV) seems to play an important role in the aetiology of cervical cancer. Yet, a detailed understanding of the mechanism by which this role is exhibited is still lacking. This is not at least due to the fact that the determination of HPV-status in individuals seems to depend on various circumstances and covariates such as pregnancy, smoking and sexual habits, among others.

To investigate the role of such covariates, a longitudinal study was set up assessing HPV-status of pregnant and non-pregnant women in three-monthly intervals at six time points. Covariates were assessed initially as well as for some covariate at all later inspection points. The chance of having a positive HPV-test in material collected from a cervical smear depends on whether at all the individual is infected and whether a suitable amount of viral DNA is sampled.

Following preliminary descriptive analyses of these longitudinal data an autoregressive linear model is proposed as comprehensive approach.

Let y_{it} be the HPV-status (0 or 1) of an individual i (i=1,...,n) at time t (t=1,...,T) and $\underline{x}_{it}$ a vector of covariates measured (or considered) at time t, then the model formulates

$$l(y_{it}) = y_{i,t-1} + \underline{\beta}_t \underline{x}_{it}$$

with a logit link function $l(.)$ and binomial error, α and $\underline{\beta}_t$ being parameters.

This model proves to be very useful to analyse the data at hand showing that HPV-positivity is highly increased by positivity at the previous examination, that pregnant women have higher odds for positivity compared to non-pregnant women, but not so in the period around delivery, and that current cigarette smoking influences the HPV-status as well.

S. WELLEK:

Testing for equivalence of survivor functions under proportional hazards assumptions

The direct approach to decision making in equivalence assessment by means of classical hypotheses testing methods is adopted for some common settings of survival analysis. It is shown that under proportional hazards assumptions the problem of testing for equivalence of survivor functions both in the one- and the two-sample situations can be reduced to testing H : $\theta \notin (\theta_1, \theta_2)$ versus K : $\theta \in (\theta_1, \theta_2)$ either in the family $(\mathcal{E}(\theta))_{\theta>0}$ of exponential distributions, or in the Gaussian location family $(N(\theta,1))_{\theta \in \mathbb{R}}$ standardized with respect to the variance. In either case, the reduced problem admits of an UMP solution which can be made explicit enough in order to provide fully practicable procedures. For the original problem of testing for equivalence of a single survivor function to a continuous hypothetical one in absence of censoring, an UMP test is obtained as well. For the one-sample problem with censoring, and the problem of testing for equivalence of two survivor functions, asymptotic solutions are given.

Berichterstatter: B. Haastert

Tagungsteilnehmer

Prof.Dr. Klaus Abt
Abteilung Biomathematik
Klinikum
Universität Frankfurt
Theodor Stern Kai 7

6000 Frankfurt 70

Dr. Ornulf Borgan
Institute of Mathematics
University of Oslo
P. O. Box 1053 - Blindern

N-0316 Oslo 3

Prof.Dr. Per Kragh Andersen
Statistical Research Unit
University of Copenhagen
Blegdamsvej 3

DK-2200 Kobenhavn N

Prof.Dr. Norman Breslow
Dept. of Biostatistics SC-32
University of Washington

Seattle , WA 98195
USA

Prof.Dr. Elja Arjas
Dept. of Mathematics
University of Oulu
Linnanmaa

SF-90570 Oulu

Prof.Dr. Edgar Brunner
Abteilung Medizinische Statistik
Fachbereich Medizin
Universität Göttingen
Humboldtallee 32

3400 Göttingen

Prof.Dr. Peter Bauer
Institut für Medizinische
Dokumentation und Statistik
Universität Köln
Josef Stelzmann Straße 9

5000 Köln 41

Dr. David Clayton
Medical Research Council Centre
Biostatistics Unit
5, Shaftesbury Road

GB- Cambridge CB2 2BW

Prof.Dr. Tadeusz Bednarski
Inst. Mat.
PAN
Kopernika 18

51-617 Wroclaw
POLAND

Dr. Anders Ekholm
Department of Statistics
University of Helsinki
Aleksi 7

SF-00100 Helsinki

Prof.Dr. Ludwig Fahrmeir
Lehrstuhl für Statistik
Universität Regensburg
Universitätsstr. 31

8400 Regensburg

Prof.Dr. Alfred Hamerle
Taegermoosstr. 9

7750 Konstanz

Prof.Dr. Uwe Feldmann
Medizinische Statistik, Biomathematik und Informationsverarbeitung
Theodor-Kutzer-Ufer, Klinikum

6800 Mannheim 1

Dr. Reinhard Hilgers
Abteilung Medizinische Statistik
Fachbereich Medizin
Universität Göttingen
Humboldtallee 32

3400 Göttingen

Prof. Dr. Theo Gasser
Zentralinstitut für
Seelische Gesundheit Mannheim
Postfach 5970

6800 Mannheim 1

Prof.Dr. Niels Keiding
Statistical Research Unit
University of Copenhagen
Blegdamsvej 3

DK-2200 Kobenhavn N

Dr. Guido Giani
Abteilung für Medizinische
Statistik und Dokumentation
RWTH Aachen
Pauwelsstraße 30

5100 Aachen

Prof.Dr. Hanns Klinger
Institut für Statistik und
Dokumentation
Universität Düsseldorf
Universitätsstr. 1

4000 Düsseldorf 1

Dr. Burkhard Haastert
Institut für Statistik in der
Medizin
Moorenstraße 5

4000 Düsseldorf 1

Prof.Dr.Dr.h.c. Klaus Krickeberg
U.F.R. de Mathematiques,
Statistique et Informatique
Universite Rene Descartes
12 rue Cujas

F-75005 Paris

Jürgen Kübler
Fachbereich Statistik
Universität Dortmund
Postfach 50 05 00

4600 Dortmund 50

Prof.Dr. Klaus J. Miescke
Dept. of Mathematics, Statistics
and Computer Science
University of Illinois at Chicago
P. O. Box 4348

Chicago , IL 60680
USA

Prof.Dr. Nan M. Laird
Dept. of Biostatistics
Harvard School of Public Health
677 Huntington Ave.

Boston , MA 02115
USA

Dr. Christoph E. Minder
Institut für Sozial- und
Präventivmedizin
Universität Bern
Finkenhubelweg 11

CH-3012 Bern

Prof.Dr. Walter Lehmacher
Institut für Biometrie und
Epidemiologie
Tierärztliche Hochschule Hannover
Bischofsholer Damm 15

3000 Hannover

Prof.Dr. Hans-Georg Mueller
Division of Statistics
University of California
469 Kerr Hall

Davis , CA 95616
USA

Prof.Dr. Kung-Yee Liang
Dept. of Biostatistics
Johns Hopkins University
School of Hygiene and Public Health
615 N. Wolfe St.

Baltimore , MD 21205
USA

Prof.Dr. Albrecht Neiss
Institut für Biostatistik und
Dokumentation
Universität Innsburck
Müllerstraße 59

A-6020 Innsbruck

Prof.Dr. Jochen Mau
Institut für Statistik in der
Medizin
Universität Düsseldorf
Moorenstr. 5

4000 Düsseldorf 1

Manfred Olschewski
Institut für Medizinische Biometrie
und Medizinische Informatik
Abt. Mediz. Biometrie u. Statistik
Stefan-Meier-Straße 26

7800 Freiburg

Prof.Dr. Rudolf Repges
Abteilung für Medizinische
Statistik und Dokumentation
RWTH Aachen
Pauwelsstraße 30

5100 Aachen

Dr. K. Ulm
Institut für Medizinische Statistik
und Epidemiologie
Technische Universität
Ismaninger Str. 22

8000 München 80

Claudia Schmoor
Institut für Medizinische Biometrie
und Medizinische Informatik
Klinikum der Universität
Stefan-Meier-Str. 26

7800 Freiburg

Prof.Dr. Michael Vaeth
Afd. for Teoretisk Statistik
Matematisk Institut
Aarhus Universitet
Ny Munkegade

DK-8000 Aarhus C

Prof.Dr. Berthold Schneider
Institut für Biometrie
Medizinische Hochschule Hannover
Postfach 61 01 80

3000 Hannover 61

Prof.Dr. Jürgen Wahrendorf
Institut für Epidemiologie und
Biometrie
Deutsches Krebsforschungszentrum
Postfach 10 19 49

6900 Heidelberg 1

Prof.Dr. Martin Schumacher
Institut für Medizinische Biometrie
und Medizinische Informatik
Klinikum der Universität
Stefan-Meier-Str. 26

7800 Freiburg

Stefan Wellek
Institut für Medizinische Statistik
und Dokumentation
Universität Mainz
Langenbeckstr. 1

6500 Mainz 1