MÜNCHENER STOCHASTIK-TAGE 1998 Sekt.-Nr.:15

Survival Analysis

Sektionsleiter: A. Janssen, Düsseldorf

Eichner, Gerrit
Mathematisches Institut der Justus-Liebig-Universität Gießen

Nichtparametrische Schätzung in Survival/Sacrifice-Experimenten

Bei Untersuchungen zur Toxizität von krebserregenden Stoffen ist das Interesse oft darauf gerichtet, wie sich eine Vergiftung damit auswirkt. Es kann dies zum Beispiel die Bildung eines Tumors sein. Hier ist insbesondere die Zeitdauer D von der Vergiftung bis zum Beginn des Tumorwachstums wichtig.
Häufig ist der Wachstumsbeginn äußerlich nicht erkennbar und kann nur durch Gewebeuntersuchungen im Rahmen einer Obduktion festgestellt werden. Dazu ist (in Tierexperimenten) die Tötung, genannt „Sacrificing", des zu untersuchenden Individuums erforderlich. Nach dem Sacrificing zur Zeit Y kann entschieden werden, ob der Tumor bereits zu wachsen begonnen hatte oder nicht, d. h. ob $D \leq Y$ oder $Y < D$.
Stirbt ein Tier ohne äußere Einwirkung nach der Zeit X aufgrund des Tumors, liegt dieser „natürliche" Tod damit vor der Sacrifice-Zeit Y, die für dieses Tier angesetzt war. Es ist also auch klar, daß $D \leq X$. Stirbt ein Tier eines natürlichen Todes, der nichts mit der Krankheit zu tun hat, ist dies mit einem Sacrificing zu eben diesem (Todes-)Zeitpunkt Y gleichzusetzen. Eine Obduktion ergibt dann wiederum, ob der Tumor bereits vorhanden war ($D \leq Y$) oder nicht ($D > Y$).
Das Besondere an diesen Untersuchungen ist, daß D selbst *nie* zu beobachten ist. Stattdessen liegen Daten in zensierter Form vor.
Es ist das Ziel, die Verteilungsfunktion von D zu schätzen. Wir schlagen einen nichtparametrischen Ad-hoc-Schätzer vor und untersuchen seine Konvergenzeigenschaften. Die Tatsache, daß D nie beobachtet wird, führt dazu, daß man es mit einem schlecht-gestellten Inversen Problem zu tun hat. Typischerweise sind Ad-hoc-Schätzer dann nicht monoton, so daß eine Isotonisierung vonnöten ist.

Giard , Nicole
Max Planck Institute for Demographic Research, Rostock
Iachine , Ivan A.
Odense University, Denmark

A multistate model for the aging process

In this paper a 3-state model for the aging process consisting of the states "healthy", "ill" and "dead" and the times of staying in a healthy or in an ill state is suggested. The model is supposed to be applied in the analysis of twin data in order to determine the relative magnitude of genetic and environmental influences on the aging process.
The dependence between the 4 survival times in a twin pair is modelled with the help of some underlying frailty variables which can be interpreted as susceptibility to illness or death. The conditional independence of the survival times given the frailties and a proportional hazard structure are assumed.
Questions of the identifiability of the model with the help of current health-state information of different birth cohorts of twins as well as approaches to estimation methods are discussed.
Some properties of the model and the results of simulation studies are illustrated by graphs.

Mayer , Claus-Dieter
Heinrich-Heine-Universität Düsseldorf

Ein funktionaler Grenzwertsatz für Survivalstatistiken bei hochzensierten Daten

Wir betrachten ein Zweistichprobenproblem mit zufälliger Rechtszensierung. Mit Hilfe von Martingalmethoden läßt sich die asymptotische Verteilung von Survivalstatistiken (z.B. Log-Rank- oder Prentice-Wilcoxon-Statistik) und den von ihnen erzeugten Prozessen herleiten (vgl. Gill (1980)). Man bekommt auf diese Weise sowohl die asymptotische Nullhypothesenverteilung als auch die Verteilung unter lokalen Alternativen in den Griff. In den verwendeten Modellen wird dabei jedoch stets vorausgesetzt, daß der relative Anteil der unzensierten Daten gegen eine von 0 verschiedene Konstante konvergiert. In vielen in der Praxis vorkommenden Fällen ist der Anteil an unzensierten Beobachtungen allerdings so niedrig, daß sich die Frage stellt, wie sich die Verfahren verhalten, wenn man zuläßt, daß dieser Anteil asymptotisch verschwindet. Wir zeigen, daß die genannten funktionalen Grenzwertsätze weiterhin gelten und auch lokale Alternativen erkannt werden können, die nun allerdings mit einer langsameren Rate gegen die Nullhypothese konvergieren. Ferner sind wir in der Lage, analoge Resultate für bedingte Versionen unserer Tests (Permutationstests) zu zeigen. Aus funktionalen Grenzwertsätzen für die bedingten Statistiken folgt die asymptotische Äquivalenz von bedingten und unbedingten Tests auch unter verschiedenen Zensierungsverteilungen, wodurch Ergebnisse von Neuhaus (1993) verallgemeinert werden.

Rahnenführer , Jörg
Heinrich-Heine-Universität Düsseldorf

Testen in bivariaten Überlebenszeitmodellen

Interessiert man sich für ein eindimensionales parametrisches Testproblem, so beruhen lokal optimale Tests auf der Linearisierung $\dot{L}$ des Dichtequotienten in der Nullhypothese (Tangente, score Funktion). Unter geeigneten Regularitätsvoraussetzungen kann man LAN-Theorie betreiben und mit Hilfe der Tangente $\dot{L}$ asymptotische relative Effizienzen von Tests berechnen.
Bei der Behandlung von zensierten Daten geht man über zur Modellbildung über Hazardraten λ_ϑ und deren entsprechende Entwicklung in der Nullhypothese γ. Mit Hilfe einer Isometrie zwischen $\dot{L}$ und γ (Efron/Johnstone '90, Ritov/Wellner '88) können so auch auf der Seite der Hazardraten optimale Teststatistiken angegeben und Effizienzen berechnet werden. Dies deckt z.B. auch den Fall proportionaler Hazards ab mit $\gamma \equiv 1$. Schätz- und Testtheorie sind in dieser Situation weit entwickelt (vgl. Gill '80).
Wir wollen hier das bivariate Analogon betrachten, das in vielen Anwendungen eine Rolle spielt, z.B. wenn man zwei Komponenten eines Systems beobachtet, die ausfallen können oder die Zeitpunkte des Eintretens eines Patienten in zwei Stadien einer Krankheit. Man interessiert sich dafür, wie diese beiden Lebenszeiten zusammenhängen, etwa ob sie unabhängig sind. Dazu geht man also von einem zweidimensionalen parametrischen Modell mit Dichten f_ϑ und Hazardraten λ_ϑ aus. In der zweidimensionalen Welt ist das Leben allerdings komplizierter. Die Abbildung der Tangente $\dot{L}$ auf die Hazardratenableitung γ ist nun keine Isometrie mehr, sie ist nicht einmal mehr bijektiv.
Die Modellbildung über zweidimensionale Hazardraten reicht also nicht mehr aus. In der Schätztheorie gibt es neuere Verfahren, die diesen Mangel beheben (Dabrowska '89, van der Laan '93, '96). Ziel des Vortrags ist es, Auswege anzugeben, wie man in der zweidimensionalen Situation auch zu optimalen Tests kommen kann.

Sen , Arusharka
University of Hyderabad, Andhra Pradesh, India

Limit Distributions of Kaplan-Meier U-Statistics

Suppose $X_i, i \geq 1$, are i.i.d with distribution function (d.f.) F, subject to *right random censoring* by another i.i.d sequence $Y_i, i \geq 1$, having d.f G, so that we can observe only $V_i := (Z_i, \delta_i)$, where $Z_i := \min(X_i, Y_i), \delta_i := I(X_i \leq Y_i), i \geq 1$. Let $\hat{F}_n$ denote the *Kaplan-Meier estimator* of F based on $V_i, 1 \leq i \leq n$. For $\varphi : R \to R$, Stute (1995) proved a CLT for $\int \varphi d\hat{F}_n$ which is the random censoring analogue of the sample mean $n^{-1} \sum_{i=1}^{n} \varphi(X_i)$. Earlier, Stute and Wang (1993) established almost sure convergence of the same.
In this article, we study the limit distributions of the *Kaplan-Meier U-statistics* $\int \int_{x<y} \varphi(x,y) \hat{F}_n(dx) \hat{F}_n(dy)$, introduced in Bose and Sen (1996), for a symmetric $\varphi : R^2 \to R$. We treat both the 'non-degenerate' (normal limit) and the 'degenerate' (*double Wiener integral* limit) cases. The nature of degeneracy in this case is different from that in the classical case and is a fall-out of random censoring (examples are included). As for the method, we follow Stute (1995) and obtain weak representations of our U-statistic in terms of (classical) U-statistics based on $V_i, 1 \leq i \leq n$. We use appropriate analogues of Stute's (1995) conditions in our case.

Stute, Winfried **Hauptvortrag**
Mathematisches Institut, Universität Gießen

Gut Und Schlecht Gestellte Probleme In Der Survival Analysis

Ein Charakteristikum der Survival Analysis besteht darin, daß die interessierenden Daten häufig nur unvollständig bekannt sind. Eine unkritische Anwendung von Standardverfahren der Statistik würde unweigerlich zu Verzerrungen und damit zu falschen Schlußfolgerungen führen. An die Stelle der empirischen Verteilungsfunktion treten Schätzer, die häufig nicht mehr explizit angebbar sind und deren asymptotisches Verhalten ungleich komplizierter ist. Der Vortrag soll an mehreren Beispielen die Problematik deutlich machen und aufzeigen, daß viele Fragestellungen im Sinne von Hadamard schlecht gestellt sind. Dabei erfordern selbst die formal gut gestellten Probleme neue Lösungsmethoden.

Yashin , Anatoli I. ; Begun , Alexander Z.
Max Planck Institute for Demographic Research, Rostock
Iachine , Ivan A.
Odense University, Denmark

An Extension of Gamma- and Inverse Gaussian Frailty Models in the Statistical Analysis of Twin Survival Data

The correlated gamma-frailty models were extensively used in the the statistical analysis of data on dependent durations. Recent applications of these models to the analysis of genetic aspects of aging and longevity open a new avenue in merging methods of quantitative genetics and survival analysis. In this paper we extend methods of gamma-frailty to more general three-parameter frailty distribution. In addition to gamma-frailty this model includes the inverse-Gaussian bivariate distribution of frailty as a particular case. The semiparametric representation of the bivariate survival function with such frailty distribution is derived. Several methods of statistical analysis of data are suggested. They include methods of parametric as well as semiparametric analysis of dependent durations. The new version of the EM-algorithm is developed. The estimation procedures are tested on simulated survival data on related individuals with and without observed covariates. Then these methods are applied to the survival data on Danish twins born in 1870-1900. Several genetic models of frailty are tested. The result of this analysis are compared with the results of earlier genetic studies of human longevity.