MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

T a g u n g s b e r i c h t 52/1989

Asymptotic methods for computer-intensive procedures in statistics

10.12. bis 16.12.1989

Die Tagung fand unter der Leitung von Prof. R.Beran (Berkeley) und Prof. D.W.Müller (Heidelberg) statt.

Die Verfügbarkeit hoher Rechenleistung hat das Gebiet der Statistik zu verändern begonnen. Die Tagung sollte dieser Entwicklung Rechnung tragen und Experten aus dem Gebiet der computer-intensiven Verfahren mit Fachleuten aus der asymptotischen Statistik in Kontakt bringen. Insgesamt wurden 34 Vorträge gehalten, davon einige mit Computer-Demonstrationen.

Breiten Raum nahm naturgemäß das Thema "Bootstrap" ein. Daneben gab es aber auch Beiträge aus den Gebieten: Parallelrechnen, Bildverarbeitung, Mustererkennung, empirische Prozeß-Methoden, Design-Optimierung, stochastische Suchverfahren, Clusteranalyse, Robustheit und Nichtparametrik.

Vortragsauszüge

BERAN, R.

Controlling conditional coverage probability in prediction

Suppose the variable X to be predicted and the learning sample Y_n that was observed are independent, with a joint distribution that depends on an unknown paramter θ. A prediction region D_n for X is a random set, depending on Y_n, that contains X with prescribed probability α. In sufficiently regular models, D_n can be constructed so that overall coverage probability converges to α at rate n^{-r}, where r is any positive integer. This paper shows that the *conditional* coverage probability of D_n, given Y_n, converges in probability to α at a rate which usually cannot exceed $n^{-1/2}$.

BICKEL, P.J.

Second order efficiency and the bootstrap

We consider bootstrap confidence bounds for a parameter $\theta(F)$ when $X_1, \ldots, X_n$ are i.i.d. $F \in \mathcal{F}$ where $\mathcal{F}$ may be parametric or not. If the model is parametric we compare on second order efficiency grounds (using the $n^{-1/2}$ term) bootstrap bounds of various types, "studentized", BCA, etc. If the underlying estimate is efficient, all bounds, whether based on the parametric or nonparametric bootstrap, which are second order correct are also second order efficient. However the bounds differ to order n^{-1}. They match the corresponding "exact" bounds to that order but differ according to what estimate of scale, parametric or nonparametric, is used or whether the parametric or nonparametric bootstrap is used. They also differ depending on which efficient estimate of θ is used. Simulation and theoretical comparisons in performance based on these differences may be worthwhile.

BOWMAN, A.W.

Asymptotic methods and computer-intensive procedures in nonparametric smoothing

Some asymptotic expressions for the bias of adaptive density estimators were examined and found to be misleading in the tails of the density. Numerical integration was used to study the exact theoretical properties of these estimators. (This work is joint with Peter Foster.)
Nonparametric regression was used to incorporate the natural assumption of smoothness over time into a repeated measurements analysis. The quadratic form structure of the test statistic allows highly accurate p-values to be computed in tests for features such as group $\times$ time interaction. (This work is joint with Adelchi Azzalini).

DICICCIO, Th.J.

Approximations to marginal tail probabilities

In many situations, inference about a scalar parameter in the presence of nuisance parameters requires integration of either a joint density of pivotal quantities or a joint posterior density which is known except for a normalizing constant. For such cases, accurate approximations of marginal tail probabilities are useful to avoid highdimensional integrals. Two such approximations are presented. They are based on normal approximations to the distribution of a variable analogous to the signed root log likelihood ratio statistic that arises in parametric inference. The approximations are easy to implement, requiring only first- and second-order partial derivatives of the log joint density. The accuracy of the approximations is illustrated in the conditional analysis of extreme-value regression models for censored data, where they are found to be excellent even for small sample sizes.

DAVIES, P.L.

Some aspects of high breakdown regression

A list of desirable properties of robust regression estimates was given including such properties as high breakdown point, efficiency, differentiability. The following method was proposed for obtaining an estimator which enjoyed at least several of the desirable properties. In the first stage an S-estimator is used to obtain a high breakdown point. This estimator is then smoothed by using a one- or two-step M-estimator with a smooth ψ-function. The problem of leverage points was also considered. This leads to the problem of obtaining a robust measure of dispersion for all nondegenerate probability measures on $\mathbf{R}^p$. Using a modification of a proposal of Donoho and Stahel it was shown that it is possible to construct such a dispersion operator.

DÜMBGEN, L.

On nonparametric changepoint-estimation

Consider a sequence $X_1, X_2, ..., X_n$ of independent random variables, where X_i has distribution F for $i \leq n\theta$ and G otherwise. The changepoint $\theta \in (0,1)$ is an unknown parameter to be estimated, and F and G are two unknown probability distributions.
The nonparametric changepoint-estimators of Darkhovskh (1976) and Carlstein (1988) are described, and rates of consistency are given under some general assumptions on F,G and θ (depending on n).
In a special model the limiting distribution of four particular estimators is presented, and in a mean-shift model the nonparametric estimators are compared with Hinkley's (1970) semiparametric estimator.
Finally two methods for the construction of bootstrap-confidence sets are proposed. One of them is based on the distribution of the estimators, while in the other case bootstrap-tests are inverted.

EDDY, W.F.

Asynchronous iteration

A network of computers can be used for calculations which are "embarassingly parallel" (e.g. sampling experiments) by subdividing the calculations among the processors. For more complex calculations (e.g. solving systems of linear equations) complicated algorithms which depend on the ratio of the interprocessor communication speed to the processor computation speed have been developed. In this talk we describe a general iterative algorithm for performing complex calculations on a network of computers. The algorithm only requires that the calculation be expressed as the fixed point of a smooth function from $\mathbf{R}^n$ to $\mathbf{R}^n$ (with spectral radius less than 1). The algorithm has been used to solve linear equations, eigenvector problems, differential equations and integral equations. We discuss an example application.

FALK, M.

On the accuracy of bootstrap estimates of the quantile function of sample quantiles

It is shown that the accuracy of the bootstrap estimate of the quantile function pertaining to the distribution of the sample q-quantile based on n independent identically distributed observations is exactly $O_P(n^{-1/4})$, $q \in (0,1)$ fixed. This rate can be improved considerably by applying smoothed bootstrap estimates. The results are formulated in terms of functional central limit theorems for the corresponding quantile bootstrap processes.

FRANKE, J.

The bootstrap in time series analysis

We discuss two applications of the bootstrap in time series analysis: *1)* to M-estimators of ARMA-parameters, and *2)* to nonparametric kernel spectrum estimates.

Part 1 (with Jens-Peter Kreiss): we describe a bootstrap procedure for ARMA- processes based on resampling from the residuals. Then, we apply it to get approximations of the distribution of centered and scaled M-estimates of the parameters. As the Mallows distance between those laws vanishes, the bootstrap principle holds in theory. In practice, the procedure works quite well for sample sizes N = 30 - 50.

Part 2 (with Wolfgang Härdle): given T data from a linear process with i.i.d. innovations and spectral density $f(\omega)$, we first transform them to the periodogram $I_T(\omega)$. From asymptotics of the periodogram we get the multiplicative regression model $I_T(\omega_j) = f(\omega_j)\,\varepsilon_j$, $j = 1,\ldots,[T/2]$, where $\omega_j = 2\pi j/T$ and the residuals ε_j are "approximately" i.i.d. in a vague sense. We discuss bootstrap approximations of kernel estimators for $f(\omega)$ where we resample from the residuals pretending that the ε_j are really i.i.d. We prove that the Mallows distance between the laws of the rescaled, centered estimate and its bootstrap approximation vanishes asymptotically. However, we have to be careful how to choose the initial estimate for $f(\omega)$ which we need for getting hold of the residuals: if chosen again as a kernel estimate its bandwidth has to be asymptotically larger than the optimal.

GILL,R.

Bootstrapping the multivariate product-limit estimator

The problem of nonparametric estimation of a multivariate survival function $\overline{F}(t) = \Pr(T \geq t)$, $T = (T_1,\ldots,T_k)$, based on randomly censored data has remained a challenge for many years. An efficient estimator is unknown but many competing root-n consistent estimators have been devised. Recently D. Dabrowska (1988, Ann.Stat.) proposed an ad hoc estimator which turns out though usually inefficient to have very attractive properties.

Her estimator can be expressed in terms of the following new representation of a multivariate survival function in terms of conditional multivariate hazard measures:

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

$$\overline{F}(t) = \prod_{C\subseteq\{1,\ldots,k\}} \mathcal{P}_{s_C\leq t_C} \prod_{B\subseteq C}\Big(1+\sum_{\emptyset\subset A\subseteq B}(-1)^{|A|}\Lambda(\mathrm{d}s_A|s_C)\Big)^{(-1)^{|C\setminus A|}}$$

where $\Lambda(\mathrm{d}s_A|s_C) = \Pr(T_A \in \mathrm{d}s_A|T_C \geq t_C)$ can be estimated naively by the ratio of numbers of observations known to lie in a small cell divided by the number known to lie in an appropriate upper quadrant, $\mathcal{P}$ is the product integral.

We discuss asymptotic results, including the correctness of the bootstrap, for the estimator based on compact differentiability. The representation just given is easily interpretable and easy to manipulate by formal algebra to get the right answers. However it involves several "mathematically illegal" operations and a calculus of interpretational rules has to be developed to justify the natural formal manipulations.

GINÉ, E.

Empirical processes in connection with the bootstrap

Uniform Donsker classes, i.e. classes of functions where the CLT for the empirical process based on P holds for all P and uniformly in P, can be characterized by a Gaussian property not too difficult to check. These classes include the "Euclidean" ones, but not only them. They have the property that: If $\mathcal{F}$ is uniform Donsker then

$\|R_n - R_0\|_{\mathcal{G}} \to 0$ where $\mathcal{G} = \mathcal{F} \cup \mathcal{F}\cdot\mathcal{F}$ and R_n $n = 0,1,\ldots$ are p.m.'s, implies $n^{1/2}(P_{n,R_n} - R_n) \to_w G_{R_0}$ in $\ell^\infty(\mathcal{F})$, where P_{n,R_n} is the empirical process based on n i.i.d. (R_n) random variables. Taking $R_n = P_{\theta_n}$ this should apply to the "parametric" or semiparametric bootstrap. (Work done with J.Zinn.)

GÖTZE, F.

Discrete nonparametric estimation problems

Consider a discrete estimation problem like the restoration of an n $\times$ n - pixel image from additive white noise of intensity σ^2. Simulated annealing and greedy algorithms approximations to the maximum penalized likelihood estimator are based on an energy function of the image. One can show that the expected time to reach the approximate level of the global maximum of the MPL- function is polynomial in n for the simulated annealing algorithm. The consistency of the image obtained seems to be governed (both for the greedy and the stochastic algorithms) by the requirement of a small noise level σ^2 and that the original image is an approximate fixpoint under the restoration algorithm. Since the efficiency of the MPL - estimators depend crucially on oversmoothing, the exact global maximum point of the MPL- function may be inconsistent.

GRÜBEL, R.

Stochastic models as functionals

A stochastic model relates certain quantities of interest to other (known) quantities and may be regarded as a functional from an to, e.g., a set of distributions. We consider the G/G/1 queueing model and analyze the functional which associates the stationary waiting time distribution with the interarrival and service time distributions; a reformulation of the Spitzer-Baxter identities plays a key role. This analysis leads to an efficient algorithm for computing stationary waiting time distributions. The derivative of the functional is obtained and used to arrive at new approximation formulae.
This approach also leads to non-parametric estimators. We use a different model to show how local linearizations of the functional can be used to obtain asymptotic normality results for such estimators.
A final example explains the use of the FFT algorithm in a simple bootstrap problem.

HÄRDLE, W.

Resampling in curve estimation

We discuss two applications of the Golden Section Bootstrap also called Wild Bootstrap. The first application is with Enno Mammen on the distribution of squared error distance between a parametric and a nonparametric model. The second one is with Steve Marron on constructing simultaneous error bars for nonparametric regression curves. The Golden Section Bootstrap is defined as follows: given $\hat{\varepsilon}_i$ a residual from a nonparametric kernel estimator we choose a two point distribution $G_i = p\,\delta_a + (1-p)\delta_b$ at each X_i so that $EG_iZ = 0$, $EG_iZ^2 = \hat{\varepsilon}_i^2$, $EG_iZ^3 = \hat{\varepsilon}_i^3$. The solution to this set of equations is $G_i = \delta\hat{\varepsilon}_i$ w.p.$(1+\sqrt{5})/10$, $= (1-\delta)\hat{\varepsilon}_i$ w.p.$(1-\sqrt{5})/10$, where $\delta = (1+\sqrt{5})/2$ is the golden ratio number. Using the first terms of the Fibonacci series we can approximate δ by 8/5 and derive efficient algorithms.

HOLM, S.

Abstract bootstrap for linear models

In linear models with i.i.d. error terms, confidence intervals and confidence sets can be generated by a method based in fact only on the (abstract) bootstrap distribution of the true error terms, yet having observable final results. The conditional distribution of the statistic used in this case has the same limit as the one of the statistic used in the ordinary bootstrap method, thus making the abstract method valid. Small simulations indicate that the small sample properties of the abstract method are probably better than those of the ordinary method.

LEVIT, B.Ya.

Second order optimality of estimators in the presence of the nuisance parameters

Recently a second order optimality theory has been developed providing the practitioners with the functional forms of second order minimax and/or admissible estimators. In principle these estimators allow an improvement on the traditional first order optimal ones, however they usually contain some free parameters which should be tuned by numerical computations. Some examples show that performing the computation enlightens the ways in which these estimators should be modified to perform still better for the samples of moderate size.
Some efforts have been made to develop the second order admissibility theory applicable to nonparametric estimators as well. As an example it can be shown that the sample mean is a second order admissible estimator of the population mean iff the distributions of the sample F admit finite exponential moments $\int e^{cx}\, dF$ for any c. The result indicates a clear relation betwen a second order admissibility and a kind of (strong) robustness.

LIU, R.

Robustness and efficiency in resampling

Via a representation theorem we establish that typically the standard delete-1 jackknife and the classical bootstrap are equally efficient for estimating mean-square-error of a statistic in the i.i.d. setting. This equivalence no longer holds as one moves to the linear regression model. It turns out that the bootstrap is more efficient when error variables are homogeneous, and the jackknife is more robust when they are heterogenous. In fact we can divide all the commonly used resampling procedures for linear regression models into two classes: the E-type (the Efficient ones like the bootstrap) and the R-type (the Robust ones like the jackknife). Thus the theory presented here provides a unified view of all the known resampling procedures.

MAMMEN, E.

Bootstrap and wild bootstrap in high dimensional linear models

We consider the case that one observes a data set of n i.i.d. data points $(X_1,Y_1),\dots,(X_n,Y_n)$ with $X_i \in \mathbf{R}^p$ and $Y_i \in \mathbf{R}$ and that one wants to estimate the least squares linear model defined by the parameter $\beta = \arg\min_b E(Y_i - X_i^T b)^2$. An estimate for β is given by the least squares estimator $\hat{\beta} = (\Sigma X_i X_i^T)^{-1} \Sigma X_i Y_i$. In this talk we compare different estimates for the distribution of $(\hat{\beta} - \beta)$: the normal approximation with mean 0 and estimated covariance matrix, the bootstrap estimate (based on resampling from $\{(X_1,Y_1),\dots,(X_n,Y_n)\}$), and the wild bootstrap estimate (based on the "wild estimation" of the conditional distribution $\mathcal{L}(Y_i - X_i^T\beta | X_i)$ for every $1 \le i \le n$ by an arbitrary distribution with mean 0, variance $(Y_i - X_i^T\hat{\beta})^2$, and third moment $(Y_i - X_i^T\hat{\beta})^3$). In an asymptotic approach where everything (especially also the dimension of the fitted linear model) may depend on n we will show that bootstrap works under weaker assumptions than wild bootstrap but that - if the linear model is true (i.e. $E(Y_i|X_i) = X_i^T\beta$) wild bootstrap is strictly more accurate than bootstrap for the distribution of the studentized estimator.

MILLAR, P.W.

Bootstrap, stochastic search, and the logistic model

Let $(X_1,\dots,X_n)$ be i.i.d. random variables, $X_i = (Y_i,Z_i)$, $Y_i = 0$ or 1, Z_i with values in $\mathbf{R}^d$. Let $\hat{P}_n$ be the empirical measure of $\{X_i\}$, indexed by the V-C class $\mathcal{V}$ consisting of sets of the form $\{i\} \times K$, $i = 0,1$, and K a lower left "octant" of $\mathbf{R}^d$. The logistic model is parametrized by $\theta = (\beta,F)$, $\eta \in \mathbf{R}^{d+1}$, F an unknown probability; the joint distribution P_θ then satisfies $P_\theta\{Z_i \in A\} = F(A)$, $P_\theta\{Y_i = 1 | Z_i = z\} = p(\beta,z)$ where $\log p(\beta,z)[1 - p(\beta,z)]^{-1} = \beta_0 + \underset{\sim}{\beta}_1{}' z$, $z \in \mathbf{R}^d$, $\beta_0 \in \mathbf{R}^1$, $\underset{\sim}{\beta}_1 = (\beta_1,\dots,\beta_d)$, $\beta = (\beta_0,\beta_1,\dots,\beta_d)$. The parameter set Θ then consists of all such θ. The goodness of fit statistic $M_n \equiv \inf_{\theta\in\Theta} \sqrt{n} \sup_{v\in\mathcal{V}} |\hat{P}_n - P_\theta|$ is shown to have the asymptotic limit

$\inf_{u \in \text{span}\Theta} | W - l(u)|$ where W is a Gaussian process in $L_\infty(\mathcal{V})$, l is a linear operator from Θ to $L_\infty(\mathcal{V})$ and $| \, |$ is the $L_\infty(\mathcal{V})$ norm. A computationally feasible variant of M_n is proposed, wherin $\inf_\Theta$ is replaced by $\inf_{\Theta_n}$; here Θ_n is a random subset of Θ consisting of j_n bootstrap replicas of an appropriate preliminary estimate of θ. This new stochastic GOF statistic is shown to have the same limit as M_n (under regularity); asymptotically valid critical values are shown to be obtainable by a special "conditional bootstrap" method.

MÜLLER, D.W.

Excess mass estimates and the modality of a distribution

A method for investigating the number of modes of a distribution is being proposed and studied. The method uses the excess mass functional as a tool for exhibiting sets of excessive empirical mass in comparison with multiples of uniform measure. By this approach one separates the investigation about the number of modes from questions concerning their location. For distributions on the line, the excess mass functional can be estimated at a square root rate, a rate typically not found for classical methods. The asymptotic behavior of estimators is analyzed, and tests for multimodality based on the excess mass are derived. (Joint work with G.Sawitzki.)

NIEMANN, H.

Iterative learning of concepts

A "concept" in our approach is a data structure representing an object or event in the real world. It has as substructures parts, specializations, concretezations, relations, and attributes. Automatic learning of a concept requires determination of those substructures. The learning process is iterative; it uses concept-schemas representing a priori knowledge, an observation, and the concept acquired so far. The learning algorithm consists of the three main steps of observation description, concept formation, and generalization.

OLSHEN, R.

<u>Gait analysis and the bootstrap</u>

The talk was a report on bootstrap-based prediction in models that arise in gait analysis. By "gait analysis" was meant the study of free speed human walking on a level surface. Because walking is nearly periodic, a suitable model for motion data for the i^{th} in a group of N learning sample subjects is $y^{(i)}(\theta) = \alpha_0^{(i)} + \Sigma_{j=1..J}[\alpha_j^{(i)} \cos j\theta + \beta_j^{(i)} \sin j\theta]$; the zeroth order term (α_0) is studied separately from the sum of harmonic terms. The focus of the talk was the latter. We observe $y_h^{(i)}(\theta)$ at evenly spaced points (indexed by k), with errors ε_{ik}. The vectors $((\alpha_1^{(i)}, \ldots, \beta_J^{(i)}))_{i=1...N}$ are assumed iid, and the mean 0 $\{\varepsilon_{ik}\}$ are, also; $\sigma^2(\theta)$ is the variance of $y_h^{(i)}(\theta)$. We wish to determine if a test $\tilde{y}_h = \tilde{y}_h(\theta)$ differs from the learning sample. The $\alpha_j^{(i)}$'s and $\beta_j^{(i)}$'s are estimated by least squares, and from these estimates $\sigma^2(\theta)$ is estimated. Simultaneously for every $m > 0$, in an obvious notation, $P\left\{\max_\theta |(\tilde{y}_h(\theta) - \hat{\bar{y}}(\theta))/\hat{\sigma}(\theta)| > m\right\}$ is estimated by a bootstrap process. It has been found that (subject to moment and smoothness assumptions) while critical values cannot be estimated hyperaccurately, coverage probabilities can be ($O(N^{-3/4+\gamma}) \;\forall\; \gamma > 0$). A $\sqrt{\log N/N}$ almost sure rate of convergence of certain bootstrap conditional probabilities to their true values has been established. Material discussed during the talk involves joint work with many others: C.Bai, P.Bickel, E.Biden, D. Sutherland, and M. Wyatt.

RASCH, D.

<u>The use of the asymptotic covariance matrix for small sample inference and experimental design in nonlinear regression</u>

Let us consider the model (random variables underlined)

$\underline{y}_i = f(x_i,\theta) + \underline{e}_i$, $i = 1,\ldots n$, $\theta \in \Omega$, $\dim(\Omega) = p \leq n$, $\theta^T = (\theta_1,\ldots,\theta_p)$ and the least squares estimator $\hat{\underline{\theta}} = \arg\inf_{\theta \in \Omega} [\Sigma_{i=1...n}(y_i - f(x_i,\theta))^2]$. Let further - with $u_j(x,\theta) = \partial/\partial\theta_j\, f(x,\theta)$ and $x^T = (x_1,\ldots,x_n)$ - the $n \times p$ matrix F be given by $F(x,\theta) = (u_j(x_i,\theta))$. Then Jennrich (1969) showed for

normally and independently distributed $\underline{e}_i$ that under mild conditions $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically $N(0,\Sigma)$ distributed where Σ is the limes of $n\ V(\hat{\theta},\theta,x)$ with $V = V(\hat{\theta},\theta,x) = \sigma^2\ [F^T(x,\theta)F(x,\theta)]^{-1}$.
The author presents results of his research group concerning tests and confidence estimations valid for small $n \geq n_0$ based on V and gives n_0 - values for special functions f. Some theorems concerning D-optimum exact experimental designs are also given. Further the expert system CADEMO is mentioned which includes all recent results.

REISS, R.-D.

Conditional curves, Poisson processes, bootstrap

Consider functionals of the conditional distribution $F(\bullet \mid x)$ of Y given X = x like the mean and the median getting in that particular cases, as a function of x, the mean and the median regression function. Using a Poisson process approximation it was proved by Falk and Reiss (1989) that conditional statistical functionals are asymptotically normal if the asymptotic normality holds for the pertaining unconditional procedure. Extending the framework to functionals having their values in the space of distribution functions one is able to reduce conditional bootstrap problems to unconditional ones.

RÖSLER, U.

Fireflies in a black box

Assume you observe the number of active fireflies in a black box as a function of time. Assuming independence and stationarity, what can you extract from the data on the underlying structure. The superposition of independent processes complicates the matter. We tried different methods, Markov processes, Semimarkov processes, alternating renewal processes. All these are ϕ mixing processes with exponential rate. We present some central limt theorems and discuss some estimators derived from these.
This problem shows up in biochemistry observing ion channels in a cell membrane.

ROMANO, J.P.

Bootstrap choice of tuning parmeters

Consider the problem of estimating $\theta = \theta(P)$ based on data x_n from an unknown distribution P. Given a family of estimators $T_{n,\beta}$ of $\theta(P)$, the goal is to choose β among $\beta \in I$ so that the resulting estimator is as good as possible. Typically, β can be regarded as a tuning or smoothing parameter, and proper choice of β is essential for good performance of $T_{n,\beta}$. In this paper, we discuss the theory of β being chosen by the bootstrap. Specifically, the bootstrap estimate of β, $\hat{\beta}_n$, is chosen to minimize an empirical bootstrap estimate of risk. A general theory is presented to establish the consistency and weak convergence properties of these estimators. Confidence intervals for $\theta(P)$ based on $T_{n,\hat{\beta}_n}$ are also asymptotically valid. Several applications of the theory are presented, including optimal choice of trimming proportion, bandwidth selection in density estimation, and optimal combinations of estimates.

ROUSSEEUW, P.J.

Asymptotics of the remedian

The remedian with base b proceeds by computing medians of groups of b observations, and then medians of these medians, until only a single estimate remains. This method merely needs k arrays of size b (where $n = b^k$), so the total storage is O(log n) for fixed b, or alternatively $O(n^{1/k})$ for fixed k. Its storage economy makes it useful for robust estimation in large data bases, for real-time engineering applications in which the data themselves are not stored, and for resistant "averaging" of curves or images. The method is equivariant for monotone transformations. Optimal choices of b with respect to storage and finite-sample breakdown are derived. The remedian is a consistent estimator of the population median, and it converges at a nonstandard rate to a median-stable distribution.

SAWITZKI, G.

Distributed computing: the NetWork implementation

NetWork is an implementation model for distributed computing in an environment with random availability. The idea is to make use of 'idle time' on computer networks, while guaranteeing absolute priority of the 'home user' of any station. As an example, for an iterative problem defined by $F: \mathbf{R}^N \to \mathbf{R}^N$, define the restriction to a 'slice' $S \subset \{1, \dots, N\}$ as

$$F(x) := x_i \qquad (i \notin S), \quad F(x) := F(x)_i \qquad (i \in S).$$

Assign slices to (random) processors. Instead of the original iteration $F_n x$ you get random results defining a process Z_n with $Z_0 = x_0$, $Z_n = c(Z_{n-1}, F(x_{n-\tau}))$ for a random delay $\tau > 0$. Random assignment of tasks to processors can be optimized for minimal net interference and maximum net performance. The same applies to the choice of c to guarantee $Z_n x_0 \to \lim F_n x_0$ for contractions F. The implementation model is used for pattern processing with a neural net as a demonstration example.

SEILLIER - MOISEIWITSCH, F.

Quasilikelihood based prediction intervals

Prediction intervals for a new observable X are constructed from a pivotal quantity P (with distribution function F). It is assumed that X is generated from some generalized linear model, as are the data $\underline{X}^{(n)}$ which allow the parameters $\underline{\beta}$ and ϕ to be estimated. By plugging in these estimates one introduces an error in the coverage probability of order $O_p(n^{-1/2})$ in the conditional probability and of order $O_p(n^{-1})$ in the overall probability. The distribution of the former is normal asymptotically. Two ways of getting rid of the bias $O(n^{-1})$ are considered. The first introduces a perturbation of order $O(n^{-1})$ in the nominal coverage probability. The order proposes to use, instead of α, $F_n^{-1}(\alpha;\hat{\underline{\beta}}_n,\hat{\phi}_n)$ in the construction of the interval, i.e. the largest α th quantile of the (overall) distribution of $F(P(X;\hat{\underline{\beta}}_n,\hat{\phi}_n))$. This critical value is estimated via resampling procedures. If a series of such intervals have been constructed, tests based on scoring rules and a martingale central limit theorem are proposed to check the adequacy of the model.

SHORACK, G.R.

Limiting behavior of L-statistics

Consider $L_n \equiv n^{-1}\Sigma_{1...n}\, c_{ni}\, g(\xi_{n:i})$ for Uniform(0,1) order statistics $0 \le \xi_{n:1} \le ... \le \xi_{n:n} \le 1$. Use integrated scores c_{ni} obtained from some J function that is "nice" (essentially Lipschitz in the middle and regularly varying in the tails). Suppose $J \ge 0$ and g $\uparrow$. Then appropriately normalized L_n is asymptotically normal if and only if the quantile function $K(t) \equiv \int_{1/2...t} J(s)\, dg(s)$ is in the domain of attraction of the normal distribution. Moreover, all possible subsequential limits are obtained, as well as a condition determining when and only when there is stochastic compactness. The same sort of solutions are obtained when k_n and k_n' observations are trimmed from the two tails. (The case $k_n \rightarrow \infty$, but $k_n/n \rightarrow 0$ and the case $\sqrt{n}(k_n/n - a) \rightarrow 0$ for $0 < a < 1$ are considered.) This is joint work with David Mason.

STEIGER,W.

Computation of multivariate medians

Let $S = \{X_1,...,X_n\}$ be a given set of points in $\mathbf{R}^d$. The goal is to generalize the usual median to the case $d > 1$. One generalization peels off convex hulls as far as possible: points on the convex hull $C(S_1)$, are assigned depth 1, where $S_1 = S$. Thereafter, $S_{i+1} = S_i \setminus C(S_i)$ is obtained, $i \ge 1$, and points on $C(S_i)$ are assigned a depth of i. A second generalization uses Tukey's directional depth, and the last is based on Regina Liu's simplicial depth, where depth(X_i) counts the number of simplices $\Delta[X_{j_1},...,X_{j_{d+1}}]$ that contain X_i, $i \ne j_s$. For each notion of depth, a median is a point of maximal depth. All three medians were shown to have 0 asymptotic breakdown point. The first and third are invariant under affine transformations. Computational aspects were discussed. An $O(n^d)$ algorithm for the simplicial median was described, $d \leq 3$, along with a lower bound of n(log n). The problem of determining the exact complexity of the simplicial median in the plane seems to involve deep combinatorial properties of configurations of points. It would be very interesting to know whether it is necessary to find the depth of every point in order to determine the point of maximal depth.

STREITBERG, B.

Exact distributions for two problems from multivariate nonparametrics

Under the assumption of independent p-variate gaussian observations, the usual T^2 statistic is UMP-invariant for the standard multivariate one-sample and two-sample testing problems. While the assumption of gaussianity is not crucial for large samples (T^2 is asymptotically distribution free), it is not well known that the independence assumption might also not be justified in many applications. The argument is as follows: consider for instance a randomized clinical trial conducted for the sake of comparing two drugs A,B and assume the H_0 situation that A and B are identical. The patients $i = 1,...,n$ are recruited haphazardly and their potential reaction to the drugs (A,B) can be described by pairs (F_i,G_i) of p-dimensional measures, where $H_0: F_i = G_i$ for $i = 1,...,n$. In a carefully planned design, independence over patients can be justified. There is, however, no good reason for assuming homogenity $F_i = F_j$ for $i \neq j$ (why should two different patients have exactly the same probability of, say, recovering from an illness?). A well-planned experiment is randomized, e.g. a permutation $\sigma : \{1,...,n\} \rightarrow \{1,...,n\}$ is chosen uniformly from the symmetric group S_n and patients with $\sigma(i) \leq n_1$ are treated with A. The conditional H_0-distribution of the observations $y = (y_1,...,y_n)$, given σ, factorizes, but is not homogeneous. The unconditional H_0-distribution of y is a mixture over S_n and, therefore, permutation invariant, but does not, in general, factorize.

Much weaker assumptions are possible for permutation tests: the H_0-distribution of y is invariant under the action of a group G, where $G = S_2{}^n$ for the one-sample case (sign invariance) and $G = S_n$ for the two-sample case (permutation invariance), both with the obvious actions. Given a statistic V, where large values of V serve to indicate a possible break of symmetry, the p-value $|\{g \in G: V(g(y)) \leq V(y)\}| / |G|$ is conditionally distribution free. If one uses $V = T^2$, the conditional H_0-distribution is asymptotically again $\chi^2[p]$, given a mild Lindeberg (one-sample case) or Noether condition. More interesting is the fact that the exact distribution can be computed for reasonable p,n (in general the problem is NP-hard) rather easily using a formal generating function for $s = gy_1 + gy_2 + \ldots + gy_m$ where m = n in the one-sample case and $m = n_1$ in the two-sample case. The basic idea can, for a special univariate case, already be found in Euler´s Introductio.

VAN ZWET, W.R.

Hoeffding´s decomposition and the bootstrap

We discuss Hoeffding´s decomposition and its relation to the bootstrap. It is shown that the naive bootstrap works only for asymptotically normal statistics. In more complicated cases one needs detailed knowledge of the structure of the statistic to be able to make an appropriate version of the bootstrap work. However, such knowledge also enables one to determine the distribution of the statistic to the required order by other methods. All such methods are asymptotically equivalent to the bootstrap, and massive computation will be needed to determine the most promising procedures.

YOUNG, G.A.

Saddlepoint approximation to Sudent´s t, with application to bootstrapping the studentized mean

An approximation to the distribution of $\lambda = \bar{x}/s$, $s^2 = n^{-1} \sum_{i=1..n} (x_i - \bar{x})^2$, can be obtained by (i) saddlepoint approximation of the joint distribution of $(\bar{x}, \overline{x^2})$, (ii) transformation to obtain the joint distribution of (λ, s), (iii) Laplace approximation to obtain the marginal distribution of λ. The method is described, illustrated, and its deficiencies discussed. Application to analytic approximation of the bootstrap distribution of the studentized mean is considered, shadowing Davison & Hinkley (1988).

Berichterstatter: D.W. Müller

Tagungsteilnehmer

Prof. Dr. R. J. Beran
Department of Statistics
University of California
367 Evans Hall

Berkeley, CA 94720
USA

Dr. L. Dümbgen
Institut für Angewandte Mathematik
der Universität Heidelberg
Im Neuenheimer Feld 294

6900 Heidelberg 1

Prof. Dr. P. J. Bickel
Department of Statistics
University of California
367 Evans Hall

Berkeley, CA 94720
USA

Prof. Dr. W. F. Eddy
Dept. of Statistics
Carnegie Mellon University

Pittsburgh, PA 15213
USA

Dr. A. W. Bowman
Dept. of Statistics
University of Glasgow

GB- Glasgow G 12 8QW

Prof. Dr. M. Falk
Fachbereich 6 Mathematik
Universität Siegen
Hölderlinstr. 3

5900 Siegen

Prof. Dr. P.L. Davies
FB 6 - Mathematik
Universität-GH Essen
Universitätsstr. 1-3
Postfach 10 37 64

4300 Essen 1

Prof. Dr. J. Franke
Fachbereich Mathematik
der Universität Kaiserslautern
Erwin-Schrödinger-Straße
Postfach 3049

6750 Kaiserslautern

Prof. Dr. T. J. DiCiccio
Department of Statistics
Stanford University
Sequoia Hall

Stanford, CA 94305-4065
USA

Prof. Dr. R. D. Gill
Mathematisch Instituut
Rijksuniversiteit te Utrecht
P. O. Box 80.010

NL-3508 TA Utrecht

Prof. Dr. E. Gine
Dept. of Mathematics
CUNY College of Staten Island
St. George Campus
130 Stuyvesant Place

Staten Island , NY 10301
USA

Prof. Dr. Fr. Götze
Fakultät für Mathematik
der Universität Bielefeld
Postfach 8640

4800 Bielefeld 1

Prof. Dr. R. Grübel
Faculty of Mathematics and
Informatics
Delft University of Technology
P. O. Box 356

NL-2628 BL Delft

Dr. W. Härdle
Institut für Gesellschafts- und
Wirtschaftswissenschaften
der Universität Bonn
Adenauerallee 24-26

5300 Bonn 1

Prof. Dr. S. Holm
Statistika Institutionen
Göteborgs Universitet

S-411 25 Göteborg

Dr. J. P. Kreiss
Institut für Mathematische
Stochastik
der Universität Hamburg
Bundesstr. 55

2000 Hamburg 13

Prof. Dr. B. Ya. Levit
Fakultät für Mathematik
der Universität Bielefeld
Postfach 8640

4800 Bielefeld 1

Prof. Dr. R. Y. Liu
Dept. of Statistics
Rutgers University
Hill Center, Busch Campus

New Brunswick , NJ 08903
USA

Dr. E. Mammen
Institut für Angewandte Mathematik
der Universität Heidelberg
Im Neuenheimer Feld 294

6900 Heidelberg 1

Prof. Dr. P.W. Millar
Department of Statistics
University of California
367 Evans Hall

Berkeley , CA 94720
USA

Prof. Dr. D.W. Müller
Institut für Angewandte Mathematik
der Universität Heidelberg
Im Neuenheimer Feld 294

6900 Heidelberg 1

Prof. Dr. R.-D. Reiß
Lehrstuhl für Mathematik VI
FB 6 - Mathematik
der Universität Siegen - GH
Hölderlinstr. 3

5900 Siegen

Prof. Dr. H. Niemann
Institut für Mathematische
Maschinen- und Datenverarbeitung
Universität Erlangen
Martensstr. 3

8520 Erlangen

Prof. Dr. J. P. Romano
Department of Statistics
Stanford University
Sequoia Hall

Stanford , CA 94305-4065
USA

Prof. Dr. R. A. Olshen
Department of Statistics
Stanford University
Sequoia Hall

Stanford , CA 94305-4065
USA

Prof. Dr. U. Rösler
Institut für Mathematische
Stochastik
der Universität Göttingen
Lotzestr. 13

3400 Göttingen

Dr. H. Putter
Fakultät für Mathematik
der Universität Bielefeld
Postfach 8640

4800 Bielefeld 1

Prof. Dr. P. J. Rousseeuw
Vesaliuslaan 24

B-2520 Edegem

Prof. Dr. D. Rasch
Akad. d. Landwirtschaftswissensch.
Forschungszentrum für Tierprod.
Dummerstorf-Rostock
Wilhelm-Stahl-Allee

DDR-2551 Dummerstorf

Dr. G. Sawitzki
Institut für Angewandte Mathematik
der Universität Heidelberg
Im Neuenheimer Feld 294

6900 Heidelberg 1

Prof. Dr. F. Seillier
Department of Statistics
University of California
367 Evans Hall

Berkeley , CA 94720
USA

Prof. Dr. B. Streitberg
Institut für Statistik und
Ökonometrie
Universität Hamburg
von Melle Park 5

2000 Hamburg

Prof. Dr. R. J. Serfling
Dept. of Mathematical Sciences
Johns Hopkins University

Baltimore , MD 21218
USA

Prof. Dr. G. A. Young
Statistical Laboratory
Cambridge University

GB- Cambridge , CB2 1SB

Prof. Dr. G. Shorack
Dept. of Mathematics
University of Washington
C138 Padelford Hall, GN-50

Seattle , WA 98195
USA

Prof. Dr. W. R. van Zwet
Mathematisch Instituut
Rijksuniversiteit Leiden
Postbus 9512

NL-2300 RA Leiden

Prof. Dr. W. L. Steiger
Department of Computer Science
Rutgers University
Hill Center, Busch Campus

New Brunswick , NJ 08903
USA