MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

T a g u n g s b e r i c h t 48/1978

Multivariate Statistical Analysis

vom 26.11. bis 02.12.1978

Die Tagung über "Multivariate Statistical Analysis" stand unter der Leitung von D. Plachky (Münster) und S. Schach (Dortmund). Es wurden Probleme im Zusammenhang mit linearen statistischen Modellen, Zeitreihenanalyse, Effizienz- und Klassifikationsfragen sowie Konvergenzaussagen bei asymptotischen Betrachtungen in der Statistik und auch in der Physik erörtert. Die Anwesenheit von überdurchschnittlich vielen ausländischen Experten ermöglichte den Teilnehmern vielseitige Diskussionen über ein Fachgebiet der mathematischen Statistik, das in Deutschland bisher kaum vertreten ist.

Teilnehmer

J.C. Akkerboom, Groningen
H. Baringhaus, Münster
P. Bártfai, Budapest
K. Behnen, Hamburg
St.A. Book, Dominguez Hills
P.L. Davies, Essen
M. Deistler, Wien
K. Doksum, Berkeley
H. Drygas, Kassel
M.L. Eaton, Kopenhagen
P. Groeneboom, Amsterdam
A.K. Gupta, Bowling Green
E.J. Hannan, Canberra
K.-H. Jöckel, Dortmund
J. Jurečková, Prag
H.-P. Kinder, Bremen
H. Klinger, Düsseldorf
P.R. Krishnaiah, Pittsburgh
V. Kurotschka, Berlin
X. Loc, Orsay
V. Mammitzsch, Marburg
O. Moeschlin, Hagen
D. Morgenstern, Hannover
G. Neuhaus, Hamburg
R.A. Olshen, La Jolla
J. Oosterhoff, Amsterdam
R.-D. Reiß, Siegen
G. Rothe, Dortmund
L. Rüschendorf, Aachen
W. Schaafsma, Groningen
N. Schmitz, Münster
P. Shaman, Philadelphia

G.D. Simons, Chapel Hill
E. Sonnemann, Dortmund
W. Stadje, Göttingen
J. Steinebach, Düsseldorf
H. Strasser, Bayreuth
F. Streit, Genf
C.G. Troskie, Göttingen
W. Vogel, Bonn
R. Willers, Dortmund
H. Witting, Freiburg
S. Zabell, New Brunswick
W.R. van Zwet, Leiden

Vortragsauszüge

W. AKKERBOOM: Testing linear hypothesis under restricted alternatives

Let X_i $(i = 1,\dots,n)$ be independent, $\mathcal{N}(\mu_i,\sigma_i^2)$-distributed random variables. Assume that $\mu_{s+1} = \dots = \mu_n = 0$ for some s $(1 \le s \le n)$ and that $\sigma_1^2 = \dots = \sigma_n^2 = \sigma^2$, where σ^2 is either known or unknown. For some r $(1 \le r \le s)$, the null hypothesis that $\mu_1 = \dots = \mu_r = 0$ has to be tested against the alternative that $\mu_i \neq 0$ for some i $(1 \le i \le r)$, under the restriction that $(\mu_1,\dots,\mu_r)$ belongs to some convex polyhedral cone C.
PINCUS (1975) suggested to modify this problem by replacing C by some circular cone. We study the null distribution of the generalized likelihood ratio statistic, both for the modified problem and for the original one. For the modified problem we obtain other tests by exploiting its rotational symmetry.
The multivariate case can be treated along these lines, if the covariance matrix is known up to a multiplicative constant.

P. BÁRTFAI: On the multivariate Chernoff theorem

While the lower estimation for large deviation probabilities of the sample mean of i.i.d. random vectors in R^k is completely solved, the upper estimation is more problematic. Denote by $R(t)$ the moment generating function of X_1, and by $\rho(x) = \inf_t e^{-(t,x)} R(t)$ the Chernoff function of X_1, and let $\rho(A) = \sup_{x\in A} \rho(x)$ $(A \subset R^k)$. The statement

$$(*) \qquad \limsup \sqrt[n]{P\left(\frac{S_n}{n} \in A\right)} \le \rho(A)$$

(and therefore the limit is equal to $\rho(A)$) is true for open A in the following cases: (i) $\rho(A) = 0$, (ii) $\rho(A) = 1$, (iii) A is convex. In the last case the Bernstein-inequality holds: $P(\frac{S_n}{n} \in A) \leq \rho^n(A)$. Further results: (∗) holds true if $R(t) < \infty$ for $|t| < \delta$; in $\mathbb{R}^2$ (∗) is valid if $E(\log^k(1 + |X_1|)) < \infty$ for every k and if the second derivative of the level curve of $\rho(x)$, namely $\{x : \rho(x) = \rho(A)\}$ is monotone for large x values. The main lemma to the proof: if $A = \{x : \rho(x) < c\}$, $A \subset \operatorname{int} \operatorname{dom} \rho$ then $\langle x,t(x)\rangle$ bounded for $x \in A$, where $t(x)$ is the point minimizing the expression $e^{-\langle t,x\rangle}R(t)$.

St.A. BOOK: <u>Large rectangles in the Csörgö-Révész theorem on increments of the two-parameter Wiener Process</u>

A recent theorem of Csörgö and Révész (1978) established a continuous link between the analogues of the law of the iterated logarithm and the Erdös-Rényi law for Brownian motion having two time parameters s and t. For rectangles $R = [s_1,s_2] \times [t_1,t_2]$ in two-dimensional time space, Csörgö and Révész defined $W(R) = W(s_2,t_2) - W(s_2,t_1) - W(s_1,t_2) + W(s_1,t_1)$ and investigated the a.s. behavior at $T \to \infty$ of $M_T = d_T \sup_{R \in L_T} |W(R)|$ where d_T is an appropriate normalizing factor and $L_T = L_T(a_T,b_T)$ is a class of subrectangles of the square $[0,b_T] \times [0,b_T]$ having area a_T. They showed that $\lim_{T\to\infty} M_T = 1$ a.s. if a certain quantity in a_T and b_T tended to ∞ and that $\limsup_{T\to\infty} M_T = 1$ a.s. otherwise. We show here that, if the quantity tends to r, $0 \leq r \leq \infty$, along certain lines, then $\liminf_{T\to\infty} M_T = \sqrt{\frac{r}{r+1}}$ a.s. Brownian motion having several time parameters has appeared in invariance principles for multivariate limit theorems. (This paper is based on joint work with Terence R. Shore.)

K. DOKSUM: <u>Transformations in some multivariate models</u>

Consider the class of (k+1) variate distributions of the random vector $(Y,X_1,\ldots,X_k)$ determined by the following requirements: For some transformation $h_\lambda(y)$ depending on a parameter $\lambda \in \Lambda \subset R^r$, the conditional distribution of $h_\lambda(Y)$ given $X_1,\ldots,X_k$ is symmetric about some point μ_i.

The problem is to draw inferences about λ and the parameters in the conditional and unconditional distributions of Y and $h_\lambda(Y)$ on the basis of a (k+1)-variate random sample $(Y_1, X_{11}, \dots, X_{1k}), \dots (Y_m, X_{m1}, \dots, X_{mk})$. If the shape of the distributions are known, the maximum likelihood estimates of the parameters can be computed. Conditions under which these estimates are consistent and asymptotically normal are given, and the behavior of the asymptotic variances are studied.

H. DRYGAS: <u>Missing observations in multivariate linear models</u>

In the linear multivariate model $Y = X\beta + U_i$ $EU = 0$, $\mathrm{Cov}\ U = I \times \Sigma$, $Y \in \mathbb{R}_{n\times p}$ it is assumed that not all variables y_{ij} are observed. Define $y^*_{ij} = \begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$. If P denotes the block-diagonal transformation which transforms $Y = (y_{ij})$ to $Y^* = (y^*_{ij})$ the model $E\,P\,Y = P\,X\,B$, $\mathrm{Cov}\ PY = P(I \times \Sigma)\,P$ is obtained. It is shown that a BLUE of EY^* exists iff $P_i L_j \subseteq L_i$, where $P = \mathrm{diag}(P_i)$, $L_i = \mathrm{im}(P_i X)$. A BQUE of all estimable functions of Σ exists iff $P_i\ P_j\ P_k$ and $P_i\,X\,e_r e_s'\,X' P_j P_k$ contained in the linear space generated by $\{P_i\,X\,e_r e_s'\,P_j, P_i\ P_k\}$. These conditions simplify if already a BLUE of EY^* exists. A BQUE of all estimable functions of Σ exists for all X iff the model decomposes into a certain number of multivariate models, $Y^* = \begin{bmatrix} Y_1 & & 0 \\ & \ddots & \\ 0 & & Y_r \end{bmatrix}$.

M.L. EATON: <u>The structure of multivariate linear models - a review</u>

If X is a random vector in $(V, (\cdot,\cdot))$ with a regression manifold M and a covariance set γ, $I \in \gamma$, the Gauss-Markov and Least square estimators of $\mu \in M$ are the same when $\Sigma M \subseteq M$, $\Sigma \in \gamma$, and are given by $\hat{\mu} = PX$ where P is the orthogonal projection onto M. When X is normal and $\Sigma M \subseteq M$ for all $\Sigma \in \gamma$, then $\hat{\mu}$ is also the maximum likelihood estimator of μ. A central problem is to describe certain types of covariance sets γ for which maximum likelihood estimation can be carried out explicitly for $\Sigma \in \gamma$. Symmetry models and Jordan algebra models provide a framework in which the covariance models can be decomposed into products of real, complex, quaternion, and Clifford algebra models where maximum likelihood estimation can be carried out in closed form. Further, certain models can be rewritten in a conditional form where rather hard looking estimation problems actually decompose into a conditional model and an unconditional model to which the previous results can be applied.

P. GROENEBOOM: Efficiencies of test statistics in testing the multivariate linear hypothesis

Consider the multivariate linear hypothesis in its canonical form. Let each row of the matrices $Y_1 = Y_1(n_1 \times m)$, $Y_2 = Y_2(n_2 \times m)$ and $Y_3 = Y_3(n \times m)$, $n = N - n_1 - n_2$, be distributed independently according to an m-variate normal distribution with covariance matrix $\sum$ and expectations given by

$$\varepsilon\, Y_1 = B_1 \ , \quad \varepsilon\, Y_2 = B_2 \quad \text{and} \quad \varepsilon\, Y_3 = 0.$$

The matrix B_2 consists of nuisance parameters and N denotes sample size. The multivariate linear hypothesis is defined by $H_0 : B_1 = 0$. Let $S_{h,N} = Y'_1 Y_1$ and $S_{e,N} = Y_3'Y$. Pillai's test (perhaps more correctly to be called Bartlett's test) for testing H_0 is based on $t_N^{(1)} = \operatorname{tr} S_{h,N}\,(S_{h,N} - S_{e,N})^{-1}$. Hotelling's T_0^2 is based on $t_N^{(2)} = \operatorname{tr} S_{h,N}\, S_{e,N}^{-1}$, Roy's largest root test is based on $t_N^{(3)}$ = largest root of $S_{h,N}\, S_{e,N}^{-1}$ and the likelihood ratio (LR) test is based on $|S_{e,N}(S_{e,N} + S_{h,N})^{-1}|$. The power of these tests is a function of the matrix $N\Theta$, $N \operatorname{diag}(\theta_1,\ldots,\theta_m)$, where $\theta_1 \geq \ldots \geq \theta_m$ are the latent roots of the matrix $\sum^{-1/2} B_1'B \sum^{-1/2}$. Therefore a (fixed) alternative to H_0 can be represented by Θ. The exact Bahadur slopes of the sequences of test statistics $\{t_N^{(1)}\}_{N\in\mathbb{N}}$, to $\{t_N^{(1)}\}_{N\in\mathbb{N}}$ at Θ are given by $c_1(\Theta) = -m \log \{\frac{1}{m} \operatorname{tr}(I + \Theta)^{-1}\}$, $c_2(\Theta) = \log(1 + \operatorname{tr}\Theta)$, $c_3(\Theta) = \log(1 + \theta_1)$ and $c_4(\Theta) = \log|I + \Theta|$, respectively. This implies inadmissibility in the sense of Bahadur efficiency for the first three tests. However, looking at next order terms (and using the concept of Bahadur deficiency) it can be shown that Pillai's test is superior against Θ with $\theta_1 = \ldots = \theta_m$ and that Hotelling's T_0^2 and Roy's test are superior against Θ with $\theta_1 > 0$ and $\theta_2 = \ldots = \theta_m$. After a natural change of scale (keeping the largest root θ_1 fixed) these "halflines of deficiency" of the LR test are enlarged to regions of positive Lebesgue measure for alternatives Θ_N such that $\Theta_N \to 0$.

A.K. GUPTA: Statistical classification: A progress report

Statistical classification procedures are reviewed and a classification procedure, based on the maximum-likelihood criterion, for classification into one of two multivariate populations when multiple observations are available on the same variable for each individual, has been studied in the present paper. The distribution of the classification statistics is derived and the probability of misclassification, exact and/or approximate, is given. These procedures are then extended to more than two populations.

E.J. HANNAN: Some asymptotic results for time series and their applications

A great deal of the asymptotic theory of time series analysis depends on theory for expressions of the form

(1) $\sum_1^N x(n)\ \varepsilon(n)'$

where $x(n)$ is generated by a vector random process and $\varepsilon(n)$ by a vector, stationary, random process and $x(n)$ is measurable F_{n-1} where F_n is the σ-algebra generated by $\varepsilon(m)$, $m \leq n$. The central limit theorem and the strong law for such sequences are related to a more general time series analysis context and in particular to the theory of linear systems, namely to the estimation of the true parameter point for the (analytic) manifold of all ARMAX systems of a given McMillan degree, m.
The estimation of the McMillan degree relates to the law of the iterated logarithm for special cases of (1). An approach to this estimation problem would be the minimisation with respect to m of an expression of the form

$$\log \hat{\Sigma} + \{c d(m)\ \ln\ln N\}/N,\ c > 1\ ,\quad \Sigma = \mathcal{E}\{\varepsilon(n)\ \varepsilon(n)'\}$$

where $d(m)$ is the dimension of the manifold. The treatment of this problem appears to demand the proof of the law of the iterated logarithm uniformly for an infinite sequence of statistics that are special cases of (1), though in some cases the treatment is simpler.

J. JUREČKOVÁ: Asymptotic relations of M-estimates and R-estimates in multivariate regression model

Let $\underset{\sim}{X}_i = (X_{i1},\dots,X_{ip})'$, $i = 1,\dots,n$ $(p \geq 1)$ be n independent random vectors with continuous cdf's $F^{(i)}(\underset{\sim}{x}) = F(\underset{\sim}{x} - e_i\ \underset{\sim}{\beta})$, $i = 1,\dots,n$. The problem is that of estimating $\underset{\sim}{\beta}$ without knowledge of F. The

following estimates are established: (i) Hodges-Lehmann's type estimator $\hat{\beta}^R$; (ii) one-step version $\hat{\beta}^2$ of $\hat{\beta}^R$ with the least-squares estimate $\hat{\beta}^{(1)}$ as an initial estimator; (iii) Huber's type M-estimator $\tilde{\beta}^M$ and (iv) one step version $\hat{\beta}^{(3)}$ of $\hat{\beta}^M$ with $\hat{\beta}^{(1)}$ as an initial estimate.
Under some conditions on F, e_i β and on the functions generating the estimates, the asymptotic distribution of $n^{1/2}(\hat{\beta}^M - \hat{\beta}^R)$ is derived. It then implies the conditions under which some pairs of estimates become asymptotically equivalent in the weak sense as $n \to \infty$. The multi-parameter case is mentioned; the conditions for the asymptotic equivalence of estimates turn out to be the same as in one-parameter case.

P.R. KRISHNAIAH: Distributions of functions of Eigenvalues of Wishart matrix and their applications

Let $x_1, \ldots, x_n$ be distributed independently with mean vectors $\mu_1, \ldots, \mu_n$ and covariance matrix Σ. Then the distribution of $S = \sum_{j=1}^{n} x_j x_j'$ is known to be Wishart distribution with n degrees of freedom; it is central Wishart if $\mu_i = 0$ $(i = 1, \ldots, n)$ and noncentral Wishart otherwise. Let $\ell_1 > \ldots > \ell_p$ be the eigen-values of S and let $\sum = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. A review of the literature on the distributions of various functions of the roots and their applications is given. We also discussed the recent results of C. Fang and P.R. Krishnaiah as the asymptotic joint distribution of certain analytic functions of the eigenvalues of the noncetral Wishart matrix when the roots have multiplicity.

V. KUROTSCHKA: Optimale Planung multivariater Experimente

Multivariate Lokationsexperimente lassen sich bei endlicher Menge von Versuchsbedingungen als allgemeine MANOVA-Experimente charakterisieren bei unendlicher Menge von Versuchsbedingungen (in den meisten praktischen Anwendungen eine konvexe Teilmenge des $\mathbb{R}_K$) durch allgemeine multivariate lineare Experimente approximieren (Entwicklung nach Tschebytscheffsystemen oder Approximation durch Systeme von Splinefunktionen). Betrachtet man die statistischen Verfahren zur Analyse solcher linearen Beobachtungsmodelle in Abhängigkeit von Versuchsplänen, so lassen sich anhand der Beurteilungsgrößen der verschiedenen parametrischen und nichtparametrischen Verfahren Optimalitätskriterien für Versuchspläne ableiten, die unmittelbar praktisch interpretierbar sind. Dabei besitzen die durch die verschiedenen Optimalitätskriterien definierten Optimierungsprobleme Zielfunktionale, die von der jeweiligen (unbekannten!) Kovarianzstruktur des Beobachtungsmodells abhängen. Mann kann jedoch zeigen, daß in vielen Fällen durch Umrechnung bzw. Modifikation diese Abhängigkeit von der Kovarianzstruktur eliminiert werden kann und die Planungsprobleme bei multivariaten Beobachtungsmodellen sogar auf solche bei univariaten Beobachtungsmodellen zurückgeführt werden können, so daß für diese Fälle alle Ergebnisse über Existenz, Eindeutigkeit und die Form der optimalen Versuchspläne direkt aus der univariaten Theorie übernommen werden können.

Für die Fälle, bei denen die multivariaten Planungsprobleme nicht direkt auf univariate zurückgeführt werden können, lassen sich zumindest verschiedene Sätze aus der Äquivalenztheorie der Versuchsplanung übertragen und damit die iterativen Konstruktionsmethoden vom univariaten auf den multivariaten Fall ausdehnen.

X. LOC: Invariant absolutely minimal designs for polynomial regressions on a triangle

We study the invariant subspaces of the 2mth-moment space for an 2-dimensional simplex in order to apply it to the allocation of D-optimal designs for polynomial regressions. The main result is a theorem on the existence and uniqueness of the invariant absolutly minimal design (IAMD) for a regular triangle in $\mathbb{R}^2$. This result answers some conjectures of Farrell-Kiefer-Walbran (Fifth Berkeley Symposium on Prob. and Stat., Vol. 1, 113-138 (1967)). Some aspects of the higher dimensional cases will be discussed.

R.A. OLSHEN: Asymptotically efficient classification with applications to Gait analysis and non-linear regression

Louis Gordon and I have studied classes of decision rules for classification which are based on an adaptive partitioning of an Euclidean observation space. The class of partitions has a conputationally attractive form, and the related decision rules are invariant under strictly monotone transformations of the coordinate axes. We provide sufficient conditions that a sequence of decision rules be asymptotically Bayes risk efficient as sample size increases. The sufficient conditions involve no regularity assumptions on the underlying parent distributions. Variations of rules of J. Friedman, and of Morgan and Sonquist are among those rules to which our results apply.
Our arguments rest primarily upon three mathematical tools: a martingale-like convergence theorem for non-nested sequences of sigma-fields, a large deviation result of Kiefer, and an optional stopping inequality for the expected marginal probability content of certain projections of components of a partition.
The Friedman schemes have been applied successfully to a study of children's gait. Moreover, the cited theoretical developments can be extended to cover general non-linear regressions.

J. OOSTERHOFF: Large deviations of multivariate L-estimators with monotone weight functions

Let $X_1, X_2, \ldots$ be i.i.d. random p-vectors with cdf F, let $\mathfrak{D}^p$ be the space of all cdf's on $\mathbb{R}^p$ with compct support and let $K(G,F)$ be the Kullback-Leibler information of $G \in \mathfrak{D}^p$ w.r.t. F, and $K(\Omega,F) = \inf\{K(G,F) = G \in \Omega\}$, $\Omega \subset \mathfrak{D}^p$.
Consider the linear combination of order statistics

$$T_n = \left(\sum_{i=1}^{n} c_{i,n}^{(i)} X_{i:n}^{(i)}, \ldots, \sum_{i=1}^{n} c_{i,n}^{(p)} X_{i:n}^{(p)}\right), \quad c_{i,n}^{(d)} = \int_{(i-1)/n}^{i/n} J_d(u)\,du,$$

$d = 1,\ldots,p$; $i = 1,\ldots,n$, where $X_{1:n}^{(d)},\ldots,X_{n:n}^{(d)}$ are the order statistics of the d-th coordinates of $X_1,\ldots,X_n$, and $J_1,\ldots,J_p \in \mathcal{L}$ are weight functions.
Define the map $T_J : \mathfrak{D}^p \to \mathbb{R}^p$ by $T_J^{(d)}(G) = \int_0^1 J_d(u)\, G^{(d)^{-1}}(u)\,du$, $d = 1,\ldots,p$, and let $\Omega_J(A) = T_J^{-1}(A)$ for $A \subset \mathbb{R}^p$.

The following extension of Chernoff's large deviation theorem is shown to hold:

$$\lim_{n\to\infty} n^{-1} \log P = \{T_n \in A\} = -K(\Omega_J(A),F)$$

if $A \subset \mathbb{R}^p$ is convex and increasing and open, and $J_1,\dots,J_p$ are nondecreasing.

G. ROTHE: Some properties of the asymptotic relative Pitman efficiency

For $(\Omega,\mathfrak{A},P_\theta)$, $\theta \in \Theta \in \mathbb{R}$ interval, $0 \in \Theta$, let $\{\phi_n\}$ be a sequence of level-α-tests for $H: \theta = 0$ against $K: \theta \neq 0$. Assume further

[A] there is a bijective, strictly increasing $H: [0,\infty[\to [\alpha,1[$, s.th. $\theta_n{}^2 n \to \eta \Rightarrow E\theta_n(\phi_n) \to H(\eta)$.

[B] $\psi_n: \theta \to E_\theta(\phi_n)$ is continous from above at $\theta = 0$ for every n.

[C] $\theta_n{}^2 \eta \to \infty \Rightarrow E\theta_n(\phi_n) \to 1$

Define, for $\beta \in]\alpha,1[$ and $\theta \in \Theta - \{0\}$,

$$\underline{N}_\beta(\theta) = \min\{n : E_\theta(\phi_n) \geq \beta\}; \quad \overline{N}_\beta(\theta) = \min\{n : E_\theta(\phi_m) \geq \beta \ \forall\ m \geq n\} .$$

Then we have the following results: Under cond. A,

$\theta_n{}^2 \underline{N}_\beta(\theta) \to H^{-1}(\beta) \iff$ B is stf. ; $\theta_n{}^2 \overline{N}_\beta(\theta) \to H^{-1}(\beta) \iff$ C is satisfied.

If $\{\phi_n{}^{(i)}\}$, $i = 1,2$ are sequences satisfying A (with fct. H_i), B,C; then the asymptotic Pitman efficiency exists and equals $e_{12}^P(\beta) = H_2^{-1}(\beta)/H_1^{-1}(\beta)$. The result, e.g., allows the comparison of tests based of test statistics with asymptotic χ^2-distribution but different degrees of freedom by Pitman efficiency (for example the Pitman "comparison" of Friedman test against Anderson (Kannemann)-test).

L. RÜSCHENDORF: Comparison of random variables

Let M be a set of measurable functions on $(E,\mathfrak{A})$ and define for two probability measures P_1,P_2 on $(E,\mathfrak{A})$: $P_1 \leq_M P_2$ if $\int f\,d P_1 \leq \int f\,d P_2$ for all $f \in M$. We consider characterizations of $\leq_M$ for a series of sets M.

Let $\mathcal{H}(F_1,\dots,F_n)$ denote the set of all n-dim. distributions with marginals $F_1,\dots,F_n$. We consider the question of existence of elements $H_1,H_2 \in \mathcal{H}(F_1,\dots F_n)$ such that $\int f\,d H_1 \leq \int f\,d F \leq \int f\,d H_2$,

$\forall\ f \in M$ and all $F \in \mathcal{H}(F_1,\ldots,F_n)$. In this way we get a band caused by dependence for $\int f\,d\,F$.

W. SCHAAFSMA: Selection of variables in the two-group discriminant analysis

Let $X_1,\ldots,X_m$ be an i.r.s. from $N_p(\mu_1,\Sigma)$ and $X_{m_1+1},\ldots,X_{m_1+m_2}$ from $N_p(\mu_2,\Sigma)$. The usual unbiased estimators for μ_1,μ_2 and Σ are denoted by $X^{(1)}$, $X^{(2)}$, S. Let $W = S^{-1}(X^{(2)} - X^{(1)})$ denote the plug-in estimator for the vector $\omega = \Sigma^{-1}(\mu_2-\mu_1)$ of weights of Fisher's linear discriminant. The performance of this estimator is characterized by the expectation $D = W^T(\mu_2-\mu_1)/(W^T\,\Sigma\,W)^{1/2}$ of W. The asymptotic distribution of D is determined: $\mathcal{L}\,m(\Delta-D) \to \frac{1}{2}(x\ \Delta^{-1} + \Delta)\ \chi^2_{p-1}$ if $m \to \infty$ and $x = m^2/m_1m_2$ is fixed. In the formula $\Delta = \{(\mu_2-\mu_1)^T\ \Sigma^{-1}(\mu_2-\mu_1)\}^{1/2}$ denotes the Mahalanobis distance. If one starts from a given ordering of the "variables" (or "characteristics" or "features") then the dependency of $E\ D(p) \approx \Delta(p) - \frac{1}{2}(x\ \Delta(p)^{-1} + \Delta(p))\ (p-1)/m$ upon the number of variables can be considered. It may happen that $E\ D(p)$ is decreasing for $p > p^*(m_1,m_2)$ which means that is would be bad to take more than $p^*(m_1,m_2)$ variables. Selecting variables corresponds with the determination of this "optimal measurement complexity" $p^*(m_1,m_2)$.

P. SHAMAN: The inverted complex Wishart distribution and its application to spectral estimation

The inverted complex Wishart distribution is studied and its use for the construction of spectral estimates is illustrated. The density, some marginals of the distribution, and the first- and second-order moments are given. For a vector-valued time series, estimation of the spectral density at a collection of frequencies and estimation of the increments of the spectral distribution function in each of a set of frequency bands are considered. A formal procedure applies Bayes theorem, where the complex Wishart is used to represent the distribution of an average of adjacent periodogram values. A conjugate prior distribution for each parameter vector is a product of inverted complex Wishart distributions.

G.D. SIMONS: Expectation inequalities for functions of one or more variables

Many papers have described conditions which guarantee various probability inequalities of the form $P(Z \in A) \geq P(Z' \in A)$, and more generally expectation inequalities of the form $Ek(Z) \geq Ek(Z')$, where Z and Z' are random vectors that are stochastically ordered in some sense. This paper considers the further generalization $E\ell(Z) \geq Ek(Z')$, involving two functions k and ℓ, and establishes a mathematical framework within which a wide variety of known and new inequalities can be viewed from a common perspective. Specific attention is given to three types of stochastic orderings. Many new inequalities are obtained for elliptically contoured distributions, including multivariate normal distributions as a special case.

W. STADJE: The asymptotic number of observations of sequential and non-sequential tests for prescribed error sums

Let $X_1, X_2, \ldots$ i.i.d. distributed according to P or Q. Then define $K(x) := \inf\{m \in \mathbb{N} \mid \exists \phi: \mathcal{X}^m \to [0,1] \text{ test}, \ \alpha(\phi) + \beta(\phi) \leq x\}$ ($\mathcal{X}$ sample space, $\alpha(\phi), \beta(\phi)$ error probabilities),
$s(x) := \inf\{E_P(\tau) \vee E_Q(\tau) \mid (\phi,\tau) \text{ seq. test}, \ \alpha(\phi,\tau) + \beta(\phi,\tau) \leq x\}$ for $x \in (0,1)$. The following inequalities hold:

$$\frac{1-x}{\|P-Q\|} \leq s(x) \leq K(x) \leq \left(\frac{\sigma_P^2}{\mu_P^2} + \frac{\sigma_Q^2}{\mu_Q^2}\right)\frac{1}{x} \ ,$$

$$K(x) \geq \frac{2x-1+(2-x)\log(2-x) - x\log x}{\mu_Q - \mu_P} \geq \frac{2\log 2 - 1}{\mu_Q - \mu_P} \ .$$

Here μ_P, σ_P^2 resp. μ_Q, σ_Q^2 are the expectation and variance of $\log \frac{t_Q}{t_P}(X_1)$ under P resp. Q. Further, $K_n(x)$ and $s_n(x)$ are computed asymptotically for a sequence of distributions P_n, Q_n with the property

$$\exists R_1 \neq \varepsilon_0 \text{ prob.meas.} \ \exists k(n)\uparrow : \mathcal{L}_{P_n}(s^n_{k(n)}) \to R_1 \quad \text{weakly.}$$

Here $s^n_m := \sum_{i=1}^m \log \frac{f_{Q_n}}{f_{P_n}}(X_i)$.

Essentially, the following theorems are given:

(1) $\lim\limits_{n\to\infty} \frac{K_n(x)}{k(x)} = f^{-1}(1-x)$, where $f(\lambda) := \int (1-\ell^x)^+ \, d R_\lambda(x)$

$(R_\lambda)_{\lambda>0}$ the continuous convolution semi-group, into which we can imbed the infinitely divisible R_1. As a curious by-product we obtain that using two additional observations always diminish the minimal possible error sum, whereas one more obs. does not.

(2) $\lim_{n\to\infty} \frac{s_n(x)}{k(n)} = s(x)$

if some additional condition (not concerning moments!) are satisfied. This is established using an invariance principle for $D[0,\infty)$.

In the second part of the talk, we give asymptotic formulae for $\tau_X(b) := \inf\{t \geq 0 \mid X(t) \geq b\}$ and $f(a,b) = \inf\{t \geq 0 \mid X(t) \notin (a,b)\}$ where $X = (X(t))_{t\geq 0}$ is a process with stationary and independent increments: Let $\mu_1 = E(X(1)) > 0$; $\mu_2 = E(X(1)^2) < \infty$. Then $X(\tau(b)) - b \xrightarrow{\mathfrak{D}}$, $X(f(a,b)) - b \xrightarrow{\mathfrak{D}}$ $(b \to \infty,\ a \to -\infty)$ $b^{-1} E(f(a,b)) \to \mu_1^{-1}$. If additionally $\varlimsup_{t\to 0} E(X^-(t)^3) < \infty$,

$\forall \varepsilon > 0$: $\lim_{b\to\infty,\ a\to-\infty,\ |ab^{-1}|\geq a} \left(E(f(a,b)) - \frac{b}{\mu_1}\right) = \frac{m}{\mu_1}$,

where m is a constant computable in terms of the X-process.

The proofs use generalisations of renewal arguments and many factorization identities.

J. STEINEBACH: <u>On the increments of partial sum processes with multidimensional indices</u>

Let $\mathbb{N}^d$, where $d > 1$ is an integer, denote the positive integer d-dimensional lattice points and let $\{X_n\}_{n\in\mathbb{N}^d}$ be a set of i.i.d. random variables with $EX_n = 0$, $\mathrm{Var}(X_n) = 1$ and finite moment-generating function $\phi(t) = E\, e^{tX_n} < \infty$, $t \in (0,\tau)$. The notation $n < m$ $(n \leq m)$, where $n = (n_1,\ldots,n_d)$ and $m = (m_1,\ldots,m_d) \in \mathbb{N}^d$ means that $n_i < m_i$ $(n_i \leq m_i)$ for all $i = 1,\ldots,d$. $J = (n,m]$ denotes 'half-open intervals' in the lattice, i.e. $J = \{\ell \in \mathbb{N}^d : n < \ell \leq m\}$, and $|J|$ is used for $\Pi_{i=1}^d (m_i - n_i)$, the cardinality of J. For $N \in \mathbb{N}$ put $\underline{N} = (N,\ldots,N) \in \mathbb{N}^d$.
We consider 'increments' $S(J) = S(n,m) = \sum_{n<\ell\leq m} X_\ell$ of the partial sum process over intervals J, and prove that

$$\lim_{N\to\infty} \max_{\substack{J\subset(\underline{0},\underline{N}] \\ |J|\leq K_n}} \frac{S(J)}{\sqrt{K_N[C \log N]}} = a$$

almost surely for certain constants a,C and integer sequences $\{K_N\}_{N\in\mathbb{N}}$. Here (i) K_N is non-decreasing ($\uparrow\infty$), (ii) K_N/N is non-increasing, (iii) $K_N/N^\delta \to 0$ ($N\to\infty$, $\forall\ t>0$) and the cases (I) $K_N/\log N\to\infty$ ('large' intervals, but 'not too large' in view of (iii)), (II) $K_N/C\log N\to C>0$ (ERDÖS-RÉNYI intervals) (III) $K_N/\log N\to 0$ ('small' intervals) yield essentially different results.

The theorem extends the ERDÖS-RÉNYI (1970) law of large numbers (case (II), $d=1$) to sequences of i.i.d. random variables indexed by $\mathbb{N}^d$ and generalizes a result of CHAN (1976) concerning the d-parameter Wiener process.

C.G. TROSKIE: The distributions of the ratios of latent roots of a covariance matrix with applications to principal component and Ridge regression

Assume a linear model $Y = X\beta + e$, where the variables in the model, say $(y, x_1,\dots,x_p)$ have a joint multivariate normal distribution. If $\lambda_1 > \lambda_2 >\dots> \lambda_p$ are the latent roots of X'X and $\mu_1 >\dots> \mu_{p+1}$ are the roots of $\begin{pmatrix} X'X & X'Y \\ Y'X & Y'Y \end{pmatrix}$ then the small roots of these matrices are of considerable interest in Principal Component or Ridge Regression. In both these procedures adjustments are made to the least squares estimator in order to eliminate the M-effects of these small roots. VINED (1975) suggested that attention should be given to the "condition number" i.e. the ratio of the largest root to the smallest root, λ_1/λ_p. The distribution of this ratio is considered and also the distribution of $\lambda_1/\sum \lambda_i$. Asymptotic results are also given.

R. WILLERS: Weak consistency on the least squares estimator in linear models with stochastic regessors

A sufficient condition for the weak consistency of the least squares estimator in linear models with stochastic regressors is given and applied to the (inhomogeneous) autoregressive process $y(t) = \alpha_1\ y(t-1) +\dots+ \alpha_p\ y(t-p) + z'(t)\gamma + e(t)$, $e(t)$, $y(t)$ real random variables, $z(t)\in\mathbb{R}^k$ fixed, $t=1,2,3,\dots$. If in the case of the autoregressive model the following conditions are satisfied

(A) $E(e(t)) = 0, \quad g_* \le E(e(t)^2), \quad E(e(t)^4) \le g^*$ for all t,

$$E\left[\prod_{i=1}^{4} e(t_i)^{n_i}\right] = \prod_{i=1}^{4} E\left[e(t_i)^{n_i}\right], \text{ if } n_i > 0, \quad \sum_{i=1}^{4} n_i \le 4 ,$$

t_i distinct $i = 1,2,3,4$, (quasi-independence of 4th order)

(B) $z^p - \alpha_1 z^{p-1} - \ldots - \alpha_p = 0, \; z \in \mathbb{C} \Rightarrow |z| < 1$,
(stability-condition) ,

(C) $\lambda_{min}\, z_n' z_n \to \infty$,

then the least squares estimator of the parameters α_i and γ_i of the autoregressive process are weakly consistent.

S. ZABELL: Rates of convergence for conditional expectations

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with bounded continuous density or lattice-valued probability mass function $f(x)$. If $E[\exp \alpha|X_1|] < \infty$ for some $\alpha > 0$, $\mu = E[X_1]$, $c_n = o(n)$, and h is a measurable function such that $E[|h(X_1)|]$, $E[|h(X_1)| \exp \alpha|X_1|] \equiv M$ are finite, then

$$E[h(X_1) \mid X_1 + \ldots + X_n = n\mu + c_n] = E[h(X_1)] + M \cdot O\left(\frac{1 + |c_n|}{n}\right)$$

uniformly in h. It follows that

$$\|\mathcal{L}(X_1 \mid X_1 + \ldots + X_n = n\mu + c_n) - \mathcal{L}(X_1)\|_{var} = O\left(\frac{1 + |c_n|}{n}\right).$$

Applications of these results include rates of convergence for the binomial Poisson convergence theorem, spacings, and random allocation.

W.R. VAN ZWET: The asymptotic equilibrium of point-charges on the sphere

If one puts n point-charges (elections) on a conducting sphere, these charges will move to point $x_1, \ldots, x_n$ on the sphere for which the potential energy $\sum_{i \neq j} \sum \frac{1}{d(x_i, x_j)}$ is a minimum (d = Euclidean distance).

Let μ_n be the probability measure that assigns mass $1/n$ to each of the points $x_1, \ldots, x_n$. Robbins and also Konvaar raised the problem of providing a rigorous proof that μ_n converges weakly to the uniform distribution on the sphere. Such a proof is provided. More generally, it is shown that if the sphere is replaced by an arbitrary compact set $K \subset \mathbb{R}^3$ with positive capacity, then μ_n converges weakly to the so-called equilibrium measures on K. A proof was also given independently by Konvaar.

G. Rothe (Dortmund)