TIME SERIES, STATISTICS, 
AND INFORMATION 

By 

Emanuel Parzen 

IMA Preprint Series # 663 
July 1990
TIME SERIES, STATISTICS, AND INFORMATION

by

Emanuel Parzen

Department of Statistics, Texas A&M University

Abstract by Image

STATISTICAL CULTURE BY I.O.U.
(INFORMATION OPTIMIZATION UNIFICATION)

INVERSE PROBLEMS WITH
POSITIVITY CONSTRAINTS,
APPLICATIONS IMAGE RECONSTRUCTION
DATA COMPRESSION
PATTERN CLASSIFICATION
CLUSTER ANALYSIS

RENYI
INFORMATION
CHI-SQUARE
DIVERGENCE

STATISTICAL CULTURE MOBILE

ENTROPY

ROAD TO
NEW DIRECTIONS
IN TIME SERIES ANALYSIS
TIME SERIES, STATISTICS, AND INFORMATION

by

Emanuel Parzen

Department of Statistics, Texas A&M University

Abstract

This paper is a broad survey of ideas for the future development of statistical methods of time series analysis based on investigating the many levels of relationships between time series analysis, statistical methods unification, and inverse problems with positivity constraints. It is hoped that developing these relations will: help integrate old and new directions of research in time series analysis; provide research tools for applied and theoretical statisticians in the 1990's and coming era of statistical information; make possible unification of statistical methods and the development of Statistical Culture. New results include a new information divergence between spectral density functions. Topics discussed include:

1. Traditional entropy and cross-entropy,
2. Renyi and Chi-square information divergence,
3. Comparison density functions,
4. Approximation of positive functions (density functions) by minimum information divergence (maximum entropy),
5. Equivalence and Orthogonality of Normal Time Series,
6. Asymptotic Information of Stationary Normal Time Series,
7. Estimation of Finite Parameter Spectral Densities,
8. Minimum information estimation of spectral densities and power index correlations,
9. Tail classification of probability laws and spectral densities,
10. Sample Brownian Bridge exploratory analysis of time series.

1Preliminary version of paper to be presented at Workshop in New Directions in Time Series Analysis at Institute for Mathematics and Its Applications, University of Minnesota on July 16, 1990, a day dedicated to honor John Tukey's contributions to Time Series Analysis. Research supported by U. S. Army Research Office.
0. Introduction

The general level of the current relation (or non-relation) between statistics and time series analysis is: (1) many applied statisticians are ignorant about the theory of time series analysis, (2) many departments of statistics offer almost no courses in time series analysis (often relying on courses taught in economics or engineering departments), (3) theoretical statisticians traditionally have regarded time series analysis as safe to ignore because it is a “technical” subject in which it is difficult to confront basic issues of statistical inference which are the problems about which they want to do research.

Time series models are becoming of research interest to some theoretical statisticians whose primary research areas involve statistical analysis of data obeying the classical model of independent observations. They would like to investigate the extension of their work to data obeying probability models of dependence. This relation of time series analysis and statistics has a possibility of being superficial because it is completely methods-driven, rather than problem-driven. Consequently it may not handle problems that are of real interest to applied users of time series analysis. Questions of asymptotic rate of convergence of parameter estimators are technical problems which fail to treat basic problems (such as model identification and/or non-regular estimation for long-memory or non-Gaussian time series) that are usually the central problems in a time series analysis.

I propose that the narrow reasons why statisticians should learn about the methods of time series analysis (they are important for applications and many potential clients have time series problems) should be supplemented by broad information age reasons; the development of Statistical Culture requires that statisticians should learn about the theory of time series because it will help them improve their mastery of the basic methods of statistical analysis for traditional data consisting of independent observations. The theory of time series analysis needs to become exoteric (belonging to the outer or less initiate circle) as well as esoteric.

My concept of Statistical Culture (Parzen (1990)) proposes that a statistical analysis
should aim to provide not a single answer but a choice of answers (answers by several methods for the same problem). Therefore a framework for comparing answers is required. A framework should also provide ways of thinking for classical problems that extends to as many modern problems as possible. This paper discusses: the basic ideas of a unification of statistical methods in terms of information concepts, the relation between time series and statistics in terms of their relations with information statistics, a framework for time series methods in terms of information concepts, and suggestions for research problems in time series analysis. The references aim to include many influential papers on information methods in statistics; additional references are warmly solicited.

An aim of this paper is to stimulate discussion of the mind-boggling discovery which appears to be emerging in modern statistical research and which I call I. O. U. (Information, Optimization, Unification). It appears that one can find a common type of optimization and approximation problem that provides a link among almost all classical and modern statistical analysis problems! An inner product $< f, g >$ is the integral of the product $fg$. Let the information about an unknown NON-NEGATIVE function $f$ be the values of linear functionals which are inner products of $f$ with specified score functions $J_k$: for $k = 1, \ldots, m$,

$$< f, J_k > = \tau_k$$

for specified constants $\tau_k$ called "moment parameters". Find a NON-NEGATIVE function, denoted

$$f^* \text{ or } f^*(\tau_1, \ldots, \tau_m),$$

which among all functions satisfying the above constraints minimizes an information divergence criterion (which includes as a special case maximum entropy). This problem is called an inverse problem with positivity constraints.

We favor Renyi information criteria which imply that $f^*$ has a representation $f_D$: for
suitable index $\lambda$ and parameters $\theta_k$ (which we call inverse parameters)

$$\left( f^\lambda_{\theta} - 1 \right) / \lambda = \sum_{k=1}^{m} \theta_k J_k$$

The inverse parameters $\theta$ are functions of the moment parameters and obtained by solving

$$< f_{\theta}, J_k > = \tau_k.$$

Uncertainty (probability and statistics) enters the picture because one observes a raw random function (denoted $f^r$) or at least its inner products

$$< f^r, J_k > = \tau_k^r, k = 1, \ldots, m.$$  

One then seeks $f^r$ which is non-negative and minimizes the specified information criterion among all functions satisfying $< f, J_k > = \tau_k^r, k = 1, \ldots, m$. Among new data analytic tools that are open problems for research are “profile functions”, defined as the minimum value of the criterion as a function of the moment parameters.

The problems that need to be solved to apply the foregoing approach to unifying statistical methods include

(1) introducing suitable density function $d(u)$, $0 < u < 1$, whose estimation underlies conventional problems,

(2) determining sufficient statistics $J_k(u)$ whose inner products with $d(u)$ or $d^\lambda(u)$ for some power $\lambda$, are regarded as most significantly different from zero and therefore provide the constraints on the unknown $d$,

(3) determining information measures whose index $\lambda$ provides a parameter formula for $d$ of the form $d^\lambda(u)$ is a linear combination of known functions $J_k(u)$ with coefficients $\theta_k$ to be estimated,

(4) developing and implementing algorithms to compute the solutions of the optimization problems.

Other aspects of unification of time series analysis methods were discussed in Parzen (1958), Parzen (1961), Parzen (1965), Parzen (1971), and Parzen (1974).
1. Traditional Entropy and Cross-Entropy

The (Kullback-Liebler) information divergence between two probability distributions $F$ and $G$ is defined (Kullback (1959)) by (our definitions differ from usual definitions by a factor of 2)

$$I(F;G) = (-2) \int_{-\infty}^{\infty} \log \{g(x)/f(x)\} f(x) dx$$

when $F$ and $G$ are continuous with probability density functions $f(x)$ and $g(x)$; when $F$ and $G$ are discrete, with probability mass functions $p_F(x)$ and $p_G(x)$, information divergence is defined by

$$I(F;G) = (-2) \sum \log \{p_G(x)/p_F(x)\} p_F(x).$$

An information decomposition of information divergence is

$$I(F;G) = H(F;G) - H(F),$$

in terms of entropy $H(F)$ and cross-entropy $H(F;G)$:

$$H(F) = (-2) \int_{-\infty}^{\infty} \{\log f(x)\} f(x) dx,$$

$$H(F;G) = (-2) \int_{-\infty}^{\infty} \{\log g(x)\} f(x) dx.$$

2. Renyi and Chi Square Information

Adapting the fundamental work of Renyi (1961) this section offers a new definition of
Renyi information of index $\lambda$. For continuous $F$ and $G$: for $\lambda \neq 0, -1$

$$IR_\lambda(F; G) = \frac{2}{\lambda(1 + \lambda)} \log \int \left\{ \frac{g(y)}{f(y)} \right\}^{1+\lambda} f(y) dy$$
$$= \frac{2}{\lambda(1 + \lambda)} \log \int \left\{ \frac{g(y)}{f(y)} \right\}^\lambda g(y) dy$$

$$IR_0(F; G) = 2 \int \left\{ \frac{g(y)}{f(y)} \log \frac{g(y)}{f(y)} \right\} f(y) dy$$
$$= 2 \int \left\{ \frac{g(y)}{f(y)} \log \frac{g(y)}{f(y)} - \frac{g(y)}{f(y)} + 1 \right\} f(y) dy$$

$$IR_{-1}(F; G) = -2 \int \left\{ \log \frac{g(y)}{f(y)} \right\} f(y) dy$$
$$= -2 \int \left\{ \log \frac{g(y)}{f(y)} - \frac{g(y)}{f(y)} + 1 \right\} f(y) dy$$

An analogous definition holds for discrete $F$ and $G$.

The second definition provides: (1) extensions to non-negative functions which are not densities, and also (2) a non-negative integrand which can provide diagnostic measures at each value of $y$.

Renyi information, for $-1 < \lambda < 0$, is equivalent to Bhattacharyya distance (Bhattacharyya (1943)).

In addition to Renyi information divergence (an extension of information statistics) one uses as information divergence between two non-negative functions an extension of chi-square statistics which has been developed by Read and Cressie (1988). For $\lambda \neq 0, -1$, Chi-square divergence of index $\lambda$ is defined for continuous $F$ and $G$ by

$$C_\lambda(F; G) = \int B_\lambda \left( \frac{g(y)}{f(y)} \right) f(y) dy$$

where

$$B_\lambda(d) = \frac{2}{(1 + \lambda)} \left\{ \left[ \lambda \left( \frac{d^\lambda - 1}{\lambda} \right) - d + 1 \right] \right\}$$

$$B_0(d) = 2 \left\{ d \log d - d + 1 \right\}$$

$$B_{-1}(d) = -2 \left\{ \log d - d + 1 \right\}$$
Important properties of $B_\lambda(d)$ are:

\[ B_\lambda(d) \geq 0, B_\lambda(1) = B_\lambda'(1) = 0, \]
\[ B_\lambda'(d) = \frac{2}{\lambda} \left( d^\lambda - 1 \right), B_\lambda''(d) = 2d^{\lambda-1} \]
\[ B_1(d) = (d - 1)^2 \]
\[ B_0(d) = 2(d \log d - d + 1) \]
\[ B_{-0.5}(d) = 4 \left( d^{0.5} - 1 \right)^2 \]
\[ B_{-1}(d) = -2(\log d - d + 1) \]
\[ B_{-2}(d) = d \left( d^{-1} - 1 \right)^2 \]

An analogous definition holds for discrete $F$ and $G$. Axiomatic derivations of information measures similar to $C_\lambda$ are given by Jones and Byrne (1990).

The Renyi information and chi-square divergence measures are related:

\[ IR_0(F; G) = C_0(F; G) \]
\[ IR_{-1}(F; G) = C_{-1}(F; G) \]

For $\lambda \neq 0, -1$,

\[ IR_\lambda(F; G) = \frac{2}{\lambda(1+\lambda)} \log \left\{ 1 + \left( \frac{\lambda(1+\lambda)}{2} \right) C_\lambda(F; G) \right\} \]

Interchange of $F$ and $G$ is provided by the Lemma:

\[ C_\lambda(F; G) = C_{-(1+\lambda)}(G; F) \]
\[ IR_\lambda(F; G) = IR_{-(1+\lambda)}(G; F) \]

Our survey in this paper suggests (in section 6) a new class of information measures:

\[ A_\lambda(F; G) = \int A_{\lambda} \left( \frac{g(y)}{f(y)} \right) f(y) dy \]

For $\lambda \neq 0, -1$, perhaps most usefully $-1 < \lambda < 0$,

\[ A_\lambda(d) = \left\{ \frac{2}{\lambda(1 + \lambda)} \right\} \{(1 + \lambda) \log d - \log \{1 + (1 + \lambda)(d - 1)\}_+\} \]
Note $A_{-1}(F; G) = IR_{-1}(F; G)$,

$$A_0(d) = 2\{\log d - 1 + (1/d)\} = (1/d)B_0(d).$$

3. Comparison Density Functions

Information divergence $I(F; G)$ is a concept that works for both multivariate and univariate distributions. This section shows that the univariate case is distinguished by the fact that we are able to relate $I(F; G)$ to the concept of comparison density $d(u; F, G)$.

Quantile domain concepts introduced in Parzen (1979) play a central role; $Q(u) = F^{-1}(u)$ is the quantile function. When $F$ is continuous, we define the density quantile function $fQ(u) = f(Q(u))$, score function $J(u) = -(fQ(u))'$, and quantile density function

$$q(u) = 1/fQ(u) = Q'(u).$$

When $F$ is discrete, we define $fQ(u) = p_F(Q(u))$, $q(u) = 1/fQ(u)$.

The comparison density $d(u; F, G)$ is defined as follows: when $F$ and $G$ are both continuous,

$$d(u; F, G) = g(F^{-1}(u))/f(F^{-1}(u));$$

when $F$ and $G$ are both discrete

$$d(u : F, G) = p_G(F^{-1}(u))/p_F(F^{-1}(u)).$$

In the continuous case $d(u; F, G)$ is the derivative of

$$D(u; F, G) = G(F^{-1}(u));$$

in the discrete case we define the comparison distribution function

$$D(u; F, G) = \int_0^u d(t; F, G)dt.$$

Let $F$ denote the true distribution function of a continuous random variable $Y$. To test the goodness of fit hypothesis $H_0 : F = G$, one transforms to $W = G(Y)$ whose distribution
function is $F(G^{-1}(u))$ and whose quantile function is $G(F^{-1}(u))$. The comparison density $d(u; F, G)$ and $d(u; G, F)$ are respectively the quantile density and the probability density of $W$.

For a density $d(u), 0 < u < 1$, Renyi information (of index $\lambda$), denoted $IR_\lambda(d)$, is non-negative and measures the divergence of $d(u)$ from uniform density $d_0(u) = 1, 0 < u < 1$. It is defined:

$$IR_0(d) = 2 \int_0^1 \{d(u) \log d(u)\} du = 2 \int_0^1 \{d(u) \log d(u) - d(u) + 1\} du$$

$$IR_{-1}(d) = -2 \int_0^1 \{\log d(u)\} du = -2 \int_0^1 \{\log d(u) - d(u) + 1\} du$$

for $\lambda \neq 0$ or $-1$

$$IR_\lambda(d) = \left\{2/\lambda (1 + \lambda)\right\} \log \int_0^1 \{d(u)\}^{1+\lambda} du$$

$$= \left\{2/\lambda (1 + \lambda)\right\} \log \int_0^1 \left(\{d(u)\}^{1+\lambda} - (1 + \lambda) \{d(u) - 1\}\right) du.$$

To relate comparison density to information divergence we use the concept of Renyi information $IR_\lambda$ which yields the important identity (and interpretation of $I(F; G)$!)

$$I(F; G) = (-2) \int_0^1 \log d(u; F, G) du$$

$$= IR_{-1}(d(u; F, G)) = IR_0(d(u; G, F)).$$

For a density $d(u), 0 < u < 1$, define

$$C_\lambda(d) = \int_0^1 B_\lambda(d(u)) du.$$

The comparison density again unifies the continuous and discrete cases. One can show that for univariate $F$ and $G$

$$C_\lambda(F, G) = C_\lambda(d(u; F, G))$$

4. Approximation of positive functions (density functions) by minimum information divergence (maximum entropy)
This section discusses how approximation theory provides models for comparison density functions. To a density \( d(u) \), \( 0 < u < 1 \), approximating functions are defined by constraining (specifying) the inner product between \( d(u) \) and a specified function \( J(u) \), called a score function. We often assume that the integral over \((0,1)\) of \( J(u) \) is zero, and the integral of \( J^2(u) \) is finite. A score function \( J(u) \), \( 0 < u < 1 \), is always defined to have the property that its inner product with \( d(u) \), denoted

\[
[J, d] = [J(u), d(u)] = \int_0^1 J(u)d(u)du,
\]

is finite. The inner product is called a component or linear detector; its value is a measure of the difference between \( d(u) \) and 1.

The question of which distributions to choose as \( F \) and \( G \) is often resolved by the following formula which evaluates the inner product between \( J(u) \) and \( d(u; F, G) \) as a moment with respect to \( G \) if \( J(u) = \varphi(F^{-1}(u)) \):

\[
[\varphi(F^{-1}(u)), d(u; F, G)] = \int_{-\infty}^{\infty} \varphi(y)dG(y) = E_G[\varphi(Y)]
\]

Often \( G \) is a raw sample distribution and \( F \) is a smooth distribution which is a model for \( G \) according to the hypothesis being tested.

We propose that non-parametric statistical inference and density estimation can be based on the same criterion functions used for parametric inference if one uses the minimum Renyi information approach to density estimation (which extends the maximum entropy approach); form functions \( d_{\lambda,m}(u) \) which minimize \( IR_\lambda(d^*(u)) \) among all functions \( d^*(u) \) satisfying the constraints

\[
[J_k, d^*] = [J_k, d] \quad \text{for} \quad k = 1, \ldots, m
\]

where \( J_k(u) \) are specified score functions. One expects \( d_{\lambda,m}(u) \) to converge to \( d(u) \) as \( m \) tends to \( \infty \), and \( IR_\lambda(d_{\lambda,m}) \) to non-decreasingly converge to \( IR_\lambda(d) \).

The case \( \lambda = 1 \) provides approximations in \( L_2 \) norm which are based on a sequence \( J_k(u) \), \( k = 1, 2, \ldots \), which is a complete orthonormal set of functions. If \( d(u) \), \( 0 < u < 1 \),
is square integrable (equivalently, \( IR_1(d) \) is finite) one can represent \( d(u) \) as the limit of

\[
d_m(u) = 1 + \sum_{k=1}^{m} [J_k, d]J_k(u), m = 1, 2, \ldots.
\]

When \( \varphi_k(y), k = 1, 2, \ldots, \) is complete orthonormal set for \( L_2(F) \), a density \( g(y) \) can be approximated by

\[
g_m(y) = f(y) \left\{ 1 + \sum_{k=1}^{m} E_G[\varphi_k(Y)] \varphi_k(y) \right\}
\]

We call \( d_m(u) \) a truncated orthogonal function (generalized Fourier) series.

An important general method of density approximation, called a weighted orthogonal function approximation, is to use suitable weights \( w_k \) to form approximations

\[
d^*(u) = 1 + \sum_{k=1}^{\infty} w_k [J_k, d]J_k(u).
\]

to \( d(u) \). Often \( w_k \) depends on a "truncation point" \( m \), and \( w_k \to 1 \) as \( m \to \infty \).

**Quadratic Detectors.** To test \( H_0 : d(u) = 1, 0 < u < 1 \), many traditional goodness of fit test statistics (such as Cramer-von Mises and Anderson-Darling) can be expressed as quadratic detectors

\[
\sum_{k=1}^{\infty} \{w_k [J_k, d]\}^2 = \int_0^1 \{d^*(u) - 1\}^2 du = C_1(d^*) = -1 + \exp IR_1(d^*).
\]

We propose that these nonadaptive test statistics should be expressed as information measures and compared with minimum Renyi information detectors \( IR_\lambda(d_{\lambda,m}) \); in this way information can provide unification of statistical methods.

Maximum entropy approximators correspond to \( \lambda = 0 \); \( d_{0,m}^\ast(u) \) satisfies an exponential model (whose parameters are denoted \( \theta_1, \ldots, \theta_m \))

\[
\log d_{0,m}^\ast(u) = \sum_{k=1}^{m} \theta_k J_k(u) - \Psi(\theta_1, \ldots, \theta_m)
\]

where \( \Psi \) is the integrating factor that guarantees that \( d_{0,m}^\ast(u), 0 < u < 1 \), integrates to 1:

\[
\Psi(\theta_1, \ldots, \theta_m) = \log \int \exp \left\{ \sum_{j=1}^{m} \theta_j J_j(u) \right\} du
\]
The approximating functions formed in practice are not computed from the true components \([J_k, d]\) but from raw estimators \([J_k, d']\) for suitable raw estimators \(d'(u)\). The approximating functions are interpreted as estimators of a true density. Methods proposed for unification and generalization of statistical methods use minimum Renyi information estimation techniques. Different applications of these methods differ mainly in how they define the raw density \(d'(u)\) which is the starting point of the data analysis.

5. Equivalence and Orthogonality of Normal Time Series

This section formulates in terms of Renyi information some classic results of the theory of time series that should be part of the education of Ph.D.’s in statistics.

To apply Neyman Pearson statistical inference to a time series \(\{Y(t), t \in T\}\) with abstract index set \(T\), we must first define the probability density functional, denoted \(p(Y(\cdot); \theta)\). We assume a family of probability models (for the time series) parametrized by a possibly infinite dimensional parameter \(\theta\).

The model “\(Y(\cdot)\) is normal with known covariance kernel \(K(s, t) = \text{cov}(Y(s), Y(t))\) and unknown mean value function \(m(t) = E[Y(t)]\)” is equivalent to a probability measure \(P_m\) on the function space \(R_T\) of all functions on \(T\). We define \(p(Y(\cdot); m)\) to be the Radon-Nikodym derivative of \(P_m\) with respect to \(P_0\), the probability measure corresponding to \(m(t) = 0\) for all \(t\).

Theorem (Parzen (1958)): In order that \(P_m\) be absolutely continuous with respect to \(P_0\) it is necessary and sufficient that \(m\) is in \(H(K)\), the reproducing kernel Hilbert space of functions on \(T\) with reproducing kernel \(K\) and inner product between functions \(f\) and \(g\) in \(H(K)\) denoted \(<f, g>_H(K)\) and satisfying

\[<f, K(\cdot, t)>_H(K) = f(t)\]

for every \(f\) in \(H(K)\) and \(t\) in \(T\).

The probability measures \(P_m\) and \(P_0\) of normal time series are either equivalent or orthogonal; when they have the same covariance kernel \(K\) and mean value functions dif-
fering by \( m \) they are orthogonal if and only if \( \| m \|_{H(K)} = \infty \), and they are equivalent if and only if \( \| m \|_{H(K)} \) is finite [which is equivalent to \( m \) is a member of \( H(K) \)] and

\[
\log p(Y(.); m) = \langle Y, m \rangle_{H(K)} - .5\| m \|_{H(K)}^2
\]

The random variable \( \langle Y, m \rangle_{H(K)} \) is a “congruence inner product” (Parzen 1970); it is the linear combination of \( \{Y(t), t \text{ in } T\} \) corresponding to \( m \) under the congruence which maps \( K(\cdot, t) \) into \( Y(t) \).

To compute the Renyi information of index \( \lambda \) let \( \sigma = \| m \|_{H(K)}, \langle Y, m \rangle_{H(K)} = \sigma Z \), where \( Z \) denotes a Normal \((0,1)\) random variable. Then

\[
IR_\lambda(P_0, P_m) = \{2/\lambda(1 + \lambda)\} \log \text{Int},
\]

where

\[
\text{Int} = \int_{RT} \{p(Y(.); m)\}^{1+\lambda} dP_0
= E[\exp\{(1 + \lambda)(\sigma Z - .5\sigma^2)\}]
= \exp\{.5(1 + \lambda)^2\sigma^2 - .5(1 + \lambda)\sigma^2\}
= \exp\{.5\lambda(1 + \lambda)\sigma^2\}
\]

**Theorem:** Renyi information of two common covariance normal time series:

\[
IR_\lambda(P_0; P_m) = \sigma^2 = \| m \|_{H(K)}^2
\]

This beautiful formula illustrates that our definition of Renyi information has been adjusted to be equivalent to a chi-squared statistic.

Method of proof uses limits of information numbers: To prove results about equivalence and orthogonality one studies the limit of information measures of finite dimensional restrictions \( P_m^{(n)} \) representing probability measures of \( Y(t), t \) in a finite set \( T^{(n)} \) of points in \( T \) converging monotonely to a set \( T^{(\infty)} \) dense in \( T \). The norm of the restriction of \( m \) to \( T^{(n)} \), denoted \( \| m \|_{H(K, T^{(n)})} \), may be shown to converge to \( \| m \|_{H(K)} \). For a time
series \( \{Y(t), t \in T\} \) with abstract index set \( T \), the finite dimensional Renyi information is denoted

\[
IR^{(n)}_\lambda = IR_\lambda \left( P^{(n)}_0; P^{(n)}_m \right)
\]

**Theorem.** General martingale theory can be used to show that

1. \( P_0 \) and \( P_m \) are orthogonal if and only if \( IR^{(n)}_{-5} \) converges to \( \infty \) as \( n \) increases,
2. \( P_m \) is absolutely continuous with respect to \( P_0 \) if \( IR^{(n)}_{1} \) has a finite limit as \( n \) increases.

**Proposition:** Renyi information divergence of two zero mean univariate normal distributions. Let \( P_j \) be the distribution on the real line corresponding to Normal \((0, K_j)\) with variance \( K_j \). Let \( p(y) \) denote the probability density of \( P_1 \) with respect to \( P_2 \). Let \( \kappa = \frac{K_2}{K_1} \). Then

\[
p(y) = \kappa^5 \exp \left\{ -\frac{.5(\kappa - 1)}{K_2} \frac{y^2}{2} \right\}
\]

\[
IR_{-1}(P_2; P_1) = \kappa - 1 - \log \kappa,
\]

\[
IR_\lambda(P_2; P_1) = (1/\lambda) \{ \log \kappa - (1 + \lambda)^{-1} \log \{ 1 + (1 + \lambda)(\kappa - 1) \} \}
\]

\[
C_\lambda(P_2; P_1) = \{ 2/\lambda (1 + \lambda) \} \kappa^{5(1+\lambda)} \{ 1 + (1 + \lambda)(\kappa - 1) \}^{-.5}
\]

Asymptotic information can be computed from the fact (compare Hannan (1970), p. 429)

\[
\lim_{n \to \infty} || m ||^2_{H(K_n)} / \sum_{t=1}^{n} m^2(t) = \int_{0}^{1} \frac{1}{R(0)f(\omega)} dM(\omega)
\]

where \( M(\omega) \) is the asymptotic spectral distribution function of \( m(\cdot) \) assuming that \( m(\cdot) \) obeys Grenander’s conditions, or is “persistently exciting” in the language of control engineers (Bohlin (1971)):

\[
\rho_m(v) = \int_{0}^{1} e^{2\pi i \omega v} dM(\omega),
\]

where \( \rho_m(v) \) is limit of sample autocorrelations of \( m(\cdot) \).
6. Asymptotic Information of Stationary Normal Time Series

This section discusses unification of information measures of stationary normal time series and information measures of non-negative functions which are spectral density functions.

When a time series \( \{Y(t), t = 1, 2, \ldots\} \) is modeled by probability measures which are orthogonal over the infinite sequence but equivalent for any finite sample, we define asymptotic information divergence (or rate of information divergence)

\[
\text{AsymIR}_\lambda(P_2; P_1) = \lim_{n \to \infty} (1/n)\text{IR}_\lambda(P_2^{(n)}; P_1^{(n)}).
\]

Let \( Y(.) \) be zero mean normal stationary with covariance function

\[
R(v) = E[Y(t)Y(t - v)].
\]

The correlation function is defined

\[
\rho(v) = R(v)/R(0).
\]

We prefer to analyze the time series after first subtracting sample mean and dividing by its sample standard deviation; its covariance function asymptotically equals \( \rho(v) \).

An important classification of time series is by memory type: no memory, short memory, long memory according as \( I_\infty = 0; 0 < I_\infty < \infty, I_\infty = \infty \) where

\[
I_\infty = I(Y|Y_1, Y_2, \ldots) = E_{Y_1Y_2, \ldots} I(f_{Y}|Y_1, Y_2, \ldots; f_Y)
\]

is the information about \( Y(t) \) in \( Y(t - 1), Y(t - 2), \ldots \), its infinite past (see Parzen (1981), (1983)).

Assume that \( Y(.) \) is short memory and satisfies

\[
\sum_{v=-\infty}^{\infty} |R(v)| \text{ finite.}
\]

The spectral density function \( f(\omega), 0 \leq \omega < 1 \), is defined as the Fourier transform of the correlation function:

\[
f(\omega) = \sum_{v=-\infty}^{\infty} \exp(-2\pi i v \omega) \rho(v)
\]
We call a time series *bounded* memory if the spectral density is bounded above and below:

\[ 0 < c_1 \leq f(\omega) \leq c_2 < \infty. \]

Let \( P_f \) denote the probability measure on the space of infinite sequences \( R_\infty \) corresponding to a normal zero mean stationary time series with spectral density function \( f(\omega) \).

A result of Pinsker [(1964), p. 196] can be interpreted as providing a formula for asymptotic information divergence between two zero mean stationary time series with respective rational spectral density functions \( f(\omega) \) and \( g(\omega) \). Write \( \text{AsymIR}_\lambda(f, g) \) for \( \text{AsymIR}_\lambda(P_f; P_g) \). Adapting Pinsker (1964) one can prove that

\[ \text{AsymIR}_{-1}(f, g) = \int_0^1 \{ (f(\omega)/g(\omega)) - 1 - \log(f(\omega)/g(\omega)) \} \, d\omega \]

The definition of Renyi information can be extended to non-negative functions \( d(u) \) which do not necessarily integrate to 1. Because spectral densities are even functions we take the integral to be over \( 0 \leq \omega < .5 \). One obtains the following important theorem.

**Theorem:** Unification of information measures of Pinsker (1964) and Itakura-Saito (1970).

\[ \text{AsymIR}_{-1}(f, g) = IR_{-1}(f(\omega)/g(\omega))_{0,.5} \]

The validity of this information measure can be extended to non-normal asymptotically stationary time series (Ephraim, Lev-Ari, Gray (1988)).

One can heuristically motivate Pinsker's information theoretic justification of the Itakura-Saito distortion measure by the formula (at the end of section 5) for the information divergence between two univariate normal distributions with zero means and different variances. Motivated by this formula we propose a formula for bounded memory time series (whose proof is given by Kazakos and Kazakos (1980)) which motivates a new distortion measure: \( \text{AsymIR}_\lambda(f, g) = \)

\[ (1/\lambda) \int_0^1 \{ \log(f(\omega)/g(\omega)) - (1/(1 + \lambda)) \log \{ 1 + (1 + \lambda)((f(\omega)/g(\omega)) - 1) \} \} \, dw \]
The properties of the integrand are the same as those of $B_\lambda(d)$.

Kazakos and Kazakos (1980) also give formulas for asymptotic information of multiple stationary time series.

7. Estimation of Finite Parameter Spectral Densities

This section formulates in terms of Renyi information the classic asymptotic maximum likelihood Whittle theory of spectral estimation.

For a random sample of a random variable with unknown probability density $f$, maximum likelihood estimators $\theta^*$ of the parameters of a finite parameter model $f_{\theta}$ of the probability density $f$ can be shown to be equivalent to minimizing

$$IR_{-1}(f^*, f_{\theta})$$

where $f^*$ is a raw estimator of $f$ (initially, a symbolic sample probability density formed from the sample distribution function $F^*$). A similar result, called Whittle's estimator (Whittle (1953)), holds for estimation of spectral densities of a bounded memory zero mean stationary time series for which one assumes a finite parametric model $f_\theta(\omega)$ for the true unknown spectral density $f(\omega)$.

A raw fully nonparametric estimator of $f(\omega)$ from a time series sample $Y(t)$, $t = 1, \ldots, n$, is the sample spectral density (or periodogram)

$$f^*(\omega) = \left| \sum_{t=1}^{n} Y(t) \exp(-2\pi i \omega t) \right|^2 \div \sum_{t=1}^{n} |Y(t)|^2$$

Note that $f^*(\omega)$ is not a consistent estimator of $f(\omega)$; nevertheless,

$$E[f^*(\omega)] \text{ converges to } f(\omega),$$

a fact which can be taken as the definition of $f(\omega)$.

An estimator $\theta^*$ which is asymptotically equivalent to maximum likelihood estimator is obtained by minimizing AsymIR_{-1}(f^*; f_\theta) = IR_{-1}(f^*, f_{\theta})_{0.5} =

$$\int_{0}^{1} \{(f^*(\omega)/f_\theta(\omega)) - 1 - \log(f^*(\omega)/f_\theta(\omega))\} \, d\omega$$

17
which can be interpreted as choosing $\theta$ to make $\frac{f^-(\omega)}{f_\theta(\omega)}$ as flat or constant as possible.

We usually use the representation

$$f_\theta(\omega) = \sigma^2/\gamma_\theta(\omega)$$

where $\gamma_\theta(\omega)$ is the square modulus of the transfer function of the whitening filter represented by the spectral density model $f_\theta$, constructed so that

$$\int_0^1 \log f_\theta(\omega) d\omega = \log \sigma^2.$$

Minimizing $\text{AsymIR}_{-1}(f^-, f_\theta)$ is equivalent to minimizing

$$\left(1/\sigma^2\right) \int_0^1 \{f^-(\omega)\gamma_\theta(\omega)\} d\omega + \log(\sigma^2)$$

which is equivalent to minimizing over $\theta$

$$\sigma^2_\theta = \int_0^1 \gamma_\theta(\omega) f^-(\omega) d\omega$$

and setting

$$\sigma^{-2} = \int_0^1 \gamma_{\theta^-}(\omega) f^-(\omega) d\omega = \sigma^2_{\theta^-}$$

The information divergence between the data and the fitted model is given

$$\text{IR}_{-1}(f^-, f_{\theta^-}) = \log \sigma^2_{\theta^-} - \log \sigma^{-2} = I_{\infty^-} - I_{\infty}^-$$

defining $-I_{\infty}^- = \log \sigma^{-2}$,

$$-I_{\infty}^- = \log \sigma^{-2} = \int_0^1 \log f^-(\omega) d\omega$$

This criterion (however, corrected for bias in $I_{\infty}^-$) arises from information approaches to model identification (Parzen (1983)). A model fitting criterion (but not a parameter estimation criterion) is provided by the information increment

$$I(Y \mid \text{all past } Y; \ Y \text{ values in model } \theta)$$

$$= \int_0^1 -\log \{f^-(\omega)/f_{\theta^-}(\omega)\} = \text{IR}_{-1}(f^-/f_{\theta^-})_{0, -.5}$$
One can regard it as a measure of the distance of the whitening spectral density

\[ f^*(\omega) = f^-(\omega)/f_{\vartheta}^-(\omega) \]

from a constant function; note that \( f^*(\omega) \) is constructed to integrate to 1. When one accepts that the optimal smoother of \( f^*(\omega) \) is a constant, a "parameter-free" non-parametric estimator of the spectral density \( f(\omega) \) by a smoother of \( f^-(\omega) \) is given by the parametric estimator \( f_{\vartheta}^- \). By "parameter-free" we mean that we are free to choose the parameters to make the data (raw estimator) shape up to a smooth estimator. The parameters are not regarded as having any significance or interpretation; they are merely coefficients of a representation of \( f(\omega) \).

Portmanteau statistics to test goodness of fit of a model to the time series use sums of squares of correlations of residuals; an analogous statistic is

\[ IR_1(f^-/f_{\vartheta}^-)_0,5 = \log \int_0^5 \{f^-(\omega)/f_{\vartheta}^-(\omega)\}^2 d\omega \]

Goodness of fit of the model to the data (as measured by how close \( f^*(\omega) \) is to the spectral density of white noise) is the ultimate model identification criterion to decide between competing parametric models.

8. Minimum information estimation of spectral densities

This section provides a perspective on maximum entropy spectral estimation from the point of view of minimum Renyi information approximation.

The maximum entropy approach to the problem of spectral estimation of a stationary time series was originated by Burg (1967). It derives a parametric model \( f_{\vartheta} \) for the true \( f \) by imposing constraints on linear functionals of \( f \) of the form: for \( v = 0,1,\ldots,m \)

\[ \rho(v) = \langle \exp(2\pi i \nu \omega), f(\omega) \rangle_{L_2(0,1)} = \rho^-(v) \]

where \( \rho^-(v) \) are estimators (from a sample) of autocorrelations. Note that the remarkable properties of Burg's estimators derive from the fact that he first estimates in a novel
way the partial correlations and should not be interpreted as proof of the superiority of
maximum entropy philosophy.

Let $f^*_{-1}$ denote the function (among all functions $f$ satisfying these constraints) which
minimizes the neg-entropy (of order -1)

$$IR_{-1}(f)_{0.5} = \int_0^1 \{-\log f(\omega)\} d\omega$$

The solution $f^*_{-1}$ has the following parametric form:

$$\{1/f^*_{-1}\} \text{ is linear combination of } \exp(2\pi iv\omega), v = -m, \ldots, m$$

The non-negativity of $f^*$ then guarantees that $f^*$ is an autoregressive spectral density:

$$f^*_{-1}(\omega) = \sigma^2_m \sum_{j=0}^m a(j) \exp(2\pi i j \omega)|^{-2}$$

A negative opinion about applying autoregressive spectral estimates is expressed by
Diggle (1990), p. 112: “A final method, which we mention only briefly, is to fit an $AR(p)$
process to the data $\{y_t\}$ and to use the fitted autoregressive spectrum as the estimate of
$f(\omega)$. The motivation for this is threefold: fitting an autoregressive process is computa-
tionally easy, autoregressive spectra can assume a wide variety of shapes, and automatic
criteria are available for choosing the value of $p$. Nevertheless, the method seems to fit
uneasily into a discussion of what is essentially a non-parametric estimation problem. It is
analogous to the use of polynomial regression for data smoothing, and is open to the same
basic objection, namely that it imposes global assumptions which can lead to artefacts in
the estimated spectrum.”

If one’s criterion is to minimize

$$IR_0(f)_{0.5} = \int_0^1 \{f(\omega) \log f(\omega)\} d\omega,$$

the neg-entropy of order 0, the solution $f^*$ obeys an exponential model (Bloomfield (1973)):

$$\{\log f^*\} \text{ is linear combination of } \exp(2\pi iv\omega), v = -m, \ldots, m$$

20
These optimization problems are related to the problem: subject to the constraints, with specified score functions \( J_k(\omega) \),

\[
< f, J_k > = \text{specified constant for } k = 1, \ldots, m
\]

find a density \( f(\omega), 0 \leq \omega < 1 \), minimizing the \( L_p \) norm, with \( p = 1 + \lambda \),

\[
\int_0^1 \{f(\omega)\}^{1+\lambda} \, d\omega
\]

Theorems about this problem are given by Chui, Deutsch, and Ward (1990).

*Power correlations, inverse and cepstral correlations:* Parametric models for the spectral density \( f \) can be obtained from various maximum entropy criteria. To check which model is parsimonious, one requires goodness of fit procedures which check the significant difference from zero of the Fourier transforms of various functions of \( f \) such as \( 1/f \), \( \log f \), or \( f^\lambda \).

Let \( e_v(\omega) = \exp(2\pi ij\omega) \), and interpret inner products as \( L_2(0,1) \). Define:

inverse correlations

\[
\rho^{(-1)}(v) = < e_v, 1/f >
\]

cepstral correlations

\[
\rho^{(0)}(v) = < e_v, \log f >
\]

ordinary correlations

\[
\rho^{(1)}(v) = < e_v, f >
\]

power correlations of power \( \lambda \)

\[
\rho^{(\lambda)}(v) = < e_v, f^\lambda >
\]

Identification of a parametric model for \( f \) should include routine estimation and interpretation of these various correlations.

In general if one expands \( f^\lambda(\omega) \) as a linear combination of orthogonal functions \( J_k(u) \), \( 0 \leq \omega < 1 \), one forms the transform (called power orthogonal series coefficients)

\[
\rho^{(J,\lambda)}(k) = < J_k, f^\lambda >
\]
An open research problem is identification of appropriate orthogonal functions $J_k(u)$, $0 \leq \omega < 1$.


This section discusses models for spectral density functions which are based on their analogy with quantile density functions.

From extreme value theory, statisticians have long realized that it is useful to classify distributions according to their tail behavior (behavior of $F(x)$ as $x$ tends to $\pm \infty$). It is usual to distinguish three main types of distributions, called (1) limited, (2) exponential, and (3) algebraic. Parzen (1979) proposes that this classification be expressed in terms of the density quantile function $fQ(u)$; we call the types short, medium, and long tail.

A reasonable assumption about the distributions that occur in practice is that their density-quantile functions are regularly varying in the sense that there exist tail exponents $\alpha_0$ and $\alpha_1$ such that, as $u \to 0$,

$$fQ(u) = u^{\alpha_0}L_0(u), \quad fQ(1-u) = u^{\alpha_1}L_1(u)$$

where $L_j(u)$ for $j = 0, 1$ is a slowly varying function.

A function $L(u)$, $0 < u < 1$ is usually defined to be slowly varying as $u \to 0$ if, for every $y$ in $0 < y < 1$, $L(uy)/L(u) \to 1$ or $\log L(uy) - \log L(u) \to 0$. For estimation of tail exponents we will require further that, as $u \to 0$,

$$\int_0^1 \{\log L(uy) - \log L(u)\} \, dy \to 0$$

which we call integrally slowly varying. An example of a slowly varying function is $L(u) = \{\log u^{-1}\}^\beta$.

Classification of tail behavior of probability laws. A probability law has a left tail type and a right tail type depending on the value of $\alpha_0$ and $\alpha_1$. If $\alpha$ is the tail exponent, we
define:

\[ \alpha < 0 \quad \text{super short tail} \]
\[ 0 \leq \alpha < 1 \quad \text{short tail} \]
\[ \alpha = 1 \quad \text{medium tail} \]
\[ \alpha > 1 \quad \text{long tail} \]

Medium tailed distributions are further classified by the value of \( J^* = \lim J(u) \):

\[ \alpha = 1, J^* = 0 \quad \text{medium long tail} \]
\[ \alpha = 1, 0 < J^* < \infty \quad \text{medium-medium tail} \]
\[ \alpha = 1, J^* = \infty \quad \text{medium-short tail} \]

One immediate insight into the meaning of tail behavior is provided by the hazard function \( h(x) = f(x) \div \{1 - F(x)\} \) with hazard quantile function \( hQ(u) = fQ(u) \div 1 - u \). The convergence behavior of \( h(x) \) as \( x \to \infty \) is the same as that of \( hQ(u) \) as \( u \to 1 \). From the definitions one sees that \( h^* = \lim_{x \to \infty} h(x) \) satisfies

\[ h^* = \infty \text{(increasing hazard rate) Short or medium-short tail} \]
\[ 0 < h^* < \infty \text{(constant hazard rate) Medium-medium tail} \]
\[ h^* = 0 \text{(decreasing hazard rate) Long or medium-long tail} \]

*Formulas for computing tail exponents.* The representation of \( fQ(u) \) suggests a formula for computation of tail exponents \( \alpha_0 \) and \( \alpha_1 \) (which may be adapted to provide estimators from data):

\[ -\alpha_0 = \lim_{u \to 0} \int_0^1 \{ \log fQ(uy) - \log fQ(u) \} \, dy \]
\[ \alpha_1 = \lim_{u \to 0} \int_0^1 \{ \log fQ(1 - yu) - \log fQ(1 - u) \} \, dy \]

*Memory classification of spectral densities:* Spectral densities with no poles or zeroes represent time series with bounded memory. We regard spectral density functions as
analogous to quantile density functions. A model for a spectral density with a pole or zero at zero frequency (a similar representation holds for an arbitrary frequency \( \omega_0 \)) is (Parzen (1986))

\[
f(\omega) = \omega^{-\delta} L(\omega)
\]

where \( L \) is a slowly varying function at \( \omega = 0 \) and \( L(0) > 0 \). An important role is played by \( f(1/n) = n^{\delta} L(1/n) \).

The spectral density is integrable if \( \delta < 1 \), which is the condition for stationarity. The spectral density of a non-stationary time series needs careful definition. The case \( \delta = 1 \) is of particular interest; it corresponds to “\( 1/f \)” noise. The case \( \delta > 1 \) could be called “fractal noise”. A time series whose first difference is stationary has \( \delta = 2 \). Heuristically, \( \delta \) is interpreted for a zero mean time series \( Y(t) \) by

\[
E \left[ \frac{1}{n} \sum_{t=1}^{n} Y(t)^2 \right] \text{ grows as } n^{\delta} L(1/n).
\]

The index \( \delta \) associated with frequency \( \omega \) is interpreted:

\[
E \left[ \frac{1}{n} \sum_{t=1}^{n} \exp(2\pi i \omega t) Y(t)^2 \right] \text{ grows as } n^{\delta} L(1/n),
\]

and, when \( \delta = 0 \), converges to \( R(0) f(\omega) \) if it is finite.

This approach provides definitions of spectral density for asymptotically stationary time series (Parzen (1962)).

Note that a finite dynamic range (bounded memory) spectral density has \( \delta = 0 \), but \( \delta = 0 \) does not imply finite dynamic range since \( f(\omega) \) can tend to \( \infty \) as \( \omega \to 0 \); an example is \( f(\omega) \sim (\log \omega)^2 \), \( \rho(v) \sim (\log v)/v \) as \( v \to \infty \).

A traditional parametrization of stationary long memory time series is \( \delta = 2H - 1 \), where \( H \) is the Hurst index satisfying \( .5 < H < 1 \); Hurst estimated \( H = .7 \) for the Nile water level time series. The covariance function has the asymptotic representation

\[
R(v) \text{ decays slowly like } v^{2H-2} = v^{\delta-1}.
\]
The memory index delta plays an important theoretical role. In many time series theorems the asymptotic behavior of a statistic is expressed in terms of $f(0)$, the value at zero frequency of the spectral density function. These results often have analogies for long memory time series if one replaces $f(0)$ by $f(1/n) = n^\delta$ asymptotically; compare Samarov and Taqqu (1988).

Estimation of delta can be considered estimating a "fractal dimension", the exponent of the rate of growth of the mean of the sample spectral density. Values of delta are used to describe music and how the brain works! *U. S. News and World Report*, June 11, 1990, p. 62 writes: “Surprisingly, the same mathematical formula that characterizes the ebb and flow of music has been discovered to exist widely in nature, from the flow of the Nile to the beating of the human heart to the wobbling of the earth’s axis. Remarkably, this equation is closely related to other mathematical formulas used by computer experts to generate amazingly realistic pictures of coastlines, clouds and mountain ranges and other natural scenery."

Estimating delta from data has many of the same difficulties as estimating the tail index of a probability distribution. Since delta is a property of a long memory time series, it undoubtedly can not be estimated with great accuracy from relatively short lengths of observed time series.

We would appreciate references to research about delta, especially the conjectures in this section.

10. Sample Brownian Bridge Exploratory analysis of time series.

To a time series sample \( \{Y(t), t = 1, 2, \ldots, n\} \) one can associate functions $d^\gamma(u)$, $0 \leq u \leq 1$, and

\[
D^\gamma(u) = \int_0^u d^\gamma(t) dt
\]

satisfying $D^\gamma(1) = 0$. Let $\mu_n$ and $\sigma_n$ denote respectively the sample mean and sample
standard deviation. Define, for \( j = 1, \ldots, n \),

\[
d^*(u) = \sqrt{n} \frac{(Y(j) - \mu_n)}{\sigma_n}, \quad \frac{j-1}{n} < u < \frac{j}{n}
\]

Note that for \( k = 1, \ldots, n \)

\[
D^*\left(\frac{k}{n}\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{k} \frac{(Y(t) - \mu_n)}{\sigma_n}
\]

We call \( D^*(\cdot) \) the sample Brownian Bridge of an observed time series. We propose that a time series analysis should routinely examine the graph of \( D^*(u), 0 \leq u < 1 \); one can show by examples that it provides graphical tools of identification of various types of long memory time series.

**Theorem:** For a stationary time series with bounded memory

\[
\{D^*(u), 0 \leq u \leq 1\} \text{ converges in distribution to } \left\{ f^{-5}(0)B(u), 0 \leq u \leq 1 \right\},
\]

where \( B(u), 0 \leq u \leq 1 \), is a Brownian Bridge stochastic process and \( f(0) \) is the spectral density at zero frequency.

For long memory time series we would like to understand the asymptotic behavior of

\[
\left\{ f(1/n) \right\}^{-5} D^*(u), \quad 0 \leq u \leq 1.
\]

**Simulation and Time Series:** Note that similar processes are studied by researchers (Schruben, Iglehart) in operations research departments who study simulation methods of forming confidence intervals for \( \mu \), the true mean of \( Y(\cdot) \); they standardize \( D^*(u) \) by its maximum minus its minimum.

**Quality Control and Time Series:** The process \( D^*(\cdot) \) also has applications to quality control problems of identifying departures from the null hypothesis that \( Y(\cdot) \) is white noise. Components (linear functionals) of \( d^*(\cdot) \) are related to accumulation analysis methods of Taguchi.
REFERENCES


\[ A_\lambda(d) = \left\{ \frac{2}{\lambda(1 + \lambda)} \right\} \left\{ (1 + \lambda) \log d - \log \left\{ 1 + (1 + \lambda)(d - 1) \right\}_+ \right\} \]

\[ B_{-1}(d) = -2 \left\{ \log d - d + 1 \right\} \]
# Recent IMA Preprints

**#**

**Author/s**

Mark J. Friedman and Eusebius J. Doedel, Numerical computation and continuation of invariant manifolds connecting fixed points

Scott J. Spector, Linear Deformations as Global Minimizers in Nonlinear Elasticity

Denis Serre, Richness and the classification of quasilinear hyperbolic systems

L. Preziosi and F. Rosso, On the stability of the shearing flow between pipes

Avner Friedman and Wenxiong Liu, A system of partial differential equations arising in electrophotography

Jonathan Bell, Avner Friedman, and Andrew A. Lacey, On solutions to a quasilinear diffusion problem from the study of soft tissue

David G. Schaeffer and Michael Shearer, Loss of hyperbolicity in yield vertex plasticity models under nonproportional loading

Herbert C. Kranzer and Barbara Lee Keyfitz, A strictly hyperbolic system of conservation laws admitting singular shocks

S. Laederich and M. Levi, Qualitative dynamics of planar chains

Milan Miklavčič, A sharp condition for existence of an inertial manifold

Charles Collins, David Kinderlehrer, and Mitchell Luskin, Numerical approximation of the solution of a variational problem with a double well potential

Todd Arbogast, Two-phase incompressible flow in a porous medium with various nonhomogeneous boundary conditions

Peter Poláček, Complicated dynamics in scalar semilinear parabolic equations in higher space dimension

Bei Hu, Diffusion of penetrant in a polymer: a free boundary problem

Mohamed Sami ElBialy, On the smoothness of the linearization of vector fields near resonant hyperbolic rest points

Max Jodeit, Jr. and Peter J. Olver, On the equation $\nabla f = M \nabla g$

Shui-Nee Chow, Kening Lu, and Yun-Qiu Shen, Normal form and linearization for quasiperiodic systems

Prabir Daripa, Theory of one dimensional adaptive grid generation

Michael C. Mackey and John G. Milton, Feedback, delays and the origin of blood cell dynamics

D.G. Aronson and S. Kamin, Disappearance of phase in the Stefan problem: one space dimension

Martin Krupa, Bifurcations of relative equilibria

D.D. Joseph, P. Singh, and K. Chen, Couette flows, rollers, emulsions, tall Taylor cells, phase separation and inversion, and a chaotic bubble in Taylor-Couette flow of two immiscible liquids

Artemio González-López, Niky Kamran, and Peter J. Olver, Lie algebras of differential operators in two complex variables

L.E. Fraenkel, On a linear, partly hyperbolic model of viscoelastic flow past a plate

Stephen Schecter and Michael Shearer, Undercompressive shocks for nonstrictly hyperbolic conservation laws

Xinfu Chen, Axially symmetric jets of compressible fluid

J. David Logan, Wave propagation in a qualitative model of combustion under equilibrium conditions

M.L. Zeeman, Hopf bifurcations in competitive three-dimensional Lotka-Volterra Systems

Allan P. Fordy, Isospectral flows: their Hamiltonian structures, Miura maps and master symmetries

Daniel D. Joseph, John Nelson, Michael Renardy, and Yuriko Renardy, Two-Dimensional cusped interfaces

Avner Friedman and Bei Hu, A free boundary problem arising in electrophotography

Hamid Bellout, Avner Friedman and Victor Isakov, Stability for an inverse problem in potential theory

Barbara Lee Keyfitz, Shocks near the sonic line: A comparison between steady and unsteady models for change of type

Barbara Lee Keyfitz and Gerald G. Warnecke, The existence of viscous profiles and admissibility for transonic shocks

P. Szmolyan, Transversal heteroclinic and homoclinic orbits in singular perturbation problems

Philip Boyland, Rotation sets and monotone periodic orbits for annulus homeomorphisms

Kenneth R. Meyer, Apollonius coordinates, the N-body problem and continuation of periodic solutions

Chjan C. Lim, On the Poincare–Whitney circuitspace and other properties of an

Stanley Minkowitz and Matthew Witten, Periodicity in cell proliferation using an asynchronous cell population

M. Chipot and G. Dal Maso, Relaxed shape optimization: The case of nonnegative data for the Dirichlet problem

Jeffery M. Franke and Harlan W. Stech, Extensions of an algorithm for the analysis of nongeneric Hopf bifurcations, with applications to delay-difference equations

Xinfu Chen, Generation and propagation of the interface for reaction–diffusion equations

Philip Korman, Dynamics of the Lotka–Volterra systems with diffusion

Harlan W. Stech, Generic Hopf bifurcation in a class of integro-differential equations

Stephane Laederich, Periodic solutions of non linear differential difference equations

Peter J. Olver, Canonical Forms and Integrability of BiHamiltonian Systems

S.A. van Gils, M.P. Krupa and W.F. Langford, Hopf bifurcation with nonsemisimple 1:1 Resonance

R.D. James and D. Kinderlehrer, Frustration in ferromagnetic materials

Carlos Rocha, Properties of the attractor of a scalar parabolic P.D.E.

Debra Lewis, Lagrangian block diagonalization

Richard C. Churchill and David L. Rod, On the determination of Ziglin monodromy groups

Xinfu Chen and Avner Friedman, A nonlocal diffusion equation arising in terminally attached polymer chains

Peter Gritzmann and Victor Klee, Inner and outer j- Radii of convex bodies in finite-dimensional normed spaces

P. Szmolyan, Analysis of a singularly perturbed traveling wave problem

Stanley Reiter and Carl P. Simon, Decentralized dynamic processes for finding equilibrium

Fernando Reitich, Singular solutions of a transmission problem in plane linear elasticity for wedge-shaped regions

Russell A. Johnson, Cantor spectrum for the quasi-periodic Schrödinger equation

Wenxiong Liu, Singular solutions for a convection diffusion equation with absorption

Deborah Brandon and William J. Hrusa, Global existence of smooth shearing motions of a nonlinear viscoelastic fluid

James F. Reineck, The connection matrix in Morse–Smale flows II

Claude Baesens, John Guckenheimer, Seungwan Kim and Robert Mackay, Simple resonance regions of torus diffeomorphisms

Willard Miller, Jr., Lecture notes in radar/sonar: Topics in Harmonic analysis with applications to radar and sonar

Calvin H. Wilcox, Lecture notes in radar/sonar: Sonar and Radar Echo Structure

Richard E. Blahut, Lecture notes in radar/sonar: Theory of remote surveillance algorithms

D.V. Anosov, Hilbert’s 21st problem (according to Bolibruch)

Stephane Laederich, Ray–Singer torsion for complex manifolds and the adiabatic limit

Geneviève Raugel and George R. Sell, Navier-Stokes equations in thin 3d domains: Global regularity of solutions I

Emanuel Parzen, Time series, statistics, and information

Andrew Majda and Kevin Lamb, Simplified equations for low Mach number combustion with strong heat release

Ju. S. Il’yashenko, Global analysis of the phase portrait for the Kuramoto–Sivashinsky equation

James F. Reineck, Continuation to gradient flows

Mohamed Sami Elbialy, Simultaneous binary collisions in the collinear N-body problem

John A. Jacquez and Carl P. Simon, AIDS: The epidemiological significance of two different mean rates of partner-change

Carl P. Simon and John A. Jacquez, Reproduction numbers and the stability of equilibria of SI models for heterogeneous populations

Matthew Stafford, Markov partitions for expanding maps of the circle

Ciprian Foias and Edriss S. Titi, Determining nodes, finite difference schemes and inertial manifolds

M.W. Smiley, Global attractors and approximate inertial manifolds for abstract dissipative equations

M.W. Smiley, On the existence of smooth breathers for nonlinear wave equations

Hitay Özbay and Janos Turi, Robust stabilization of systems governed by singular integro-differential equations

Mary Silber and Edgar Knobloch, Hopf bifurcation on a square lattice

Christophe Golé, Ghost circles for twist maps

Christophe Golé, Ghost tori for monotone maps

Christophe Golé, Monotone maps of $T^n \times R^n$ and their periodic orbits