Robust Tests for Change-Points in Time Series
– A Tribute to Bob Burton on the Occasion of his Retirement –

Herold Dehling

RUHR-UNIVERSITÄT BOCHUM

Frontier Probability Days 2018, Oregon State University, Corvallis

Coauthors: Roland Fried (Dortmund), Martin Wendler (Greifswald)

Research supported by DFG-Sonderforschungsbereich 823
Statistical modelling of nonlinear dynamic processes

How it all began: Göttingen 1984/85

Stochastic Processes and their Applications 23 (1986) 301–306 301
North-Holland

AN INVARIANCE PRINCIPLE FOR WEAKLY ASSOCIATED RANDOM VECTORS

Robert M. BURTON*
Department of Mathematics, Oregon State University, Oregon 97331, USA

André Robert DABROWSKI**
Department of Mathematics, University of Ottawa, Ontario, Canada

Herold DEHLING***
Institut für Mathematische Stochastik, Universität Göttingen, Göttingen, FR Germany

Received 28 November 1985
Revised 30 July 1986

The positive dependence notion of association for collections of random variables is generalized to that of weak association for collections of vector valued random elements in such a way as to allow negative dependencies in individual random elements. An invariance principle is stated and proven for a stationary, weakly associated sequence of $\mathbb{R}^d$-valued or separable Hilbert space valued random elements which satisfy a covariance summability condition.

Friendship and Collaboration with Bob

Nine joint papers with Bob, my personal favorites in chronological order:

1. R.M. BURTON, A.R. DABROWSKI, H. DEHLING: An invariance principle for weakly associated random vectors. *Stochastic Processes and their Applications* **23**, 301-306 (1986) [34 citations]
2. R.M. BURTON, H. DEHLING: Large deviations for some weakly dependent random processes. *Statistics and Probability Letters* **9**, 397-401 (1990) [77]
3. J. AARONSON, R. BURTON, H. DEHLING, D. GILAT, T. HILL, B. WEISS: Strong laws for L- and U-statistics. *Transactions of the AMS* **348**, 2845-2866 (1996) [27]
4. R.M.BURTON, H. DEHLING: Universal approximation in p-mean by neural networks. *Neural Networks* **11**, 661-667 (1998) [17]
5. S.BOROVKOVA, R.BURTON, H.DEHLING: Consistency of the Takens estimator for the correlation dimension. *Annals of Appl. Probability* **9**, 376-390 (1999) [21]
6. S. BOROVKOVA, R.M. BURTON, H. DEHLING: Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Transactions of the AMS* **353** (2001), 4261–4318 [54]

All connected to wonderful memories, many concerning visits to Corvallis.

Eike and Malte (Corvallis, summer of 1987)

Gundi, Eike and Malte (Corvallis, summer of 1987)

The Dehlings (Bochum, Christmas 2016)

Borovkova, Burton, Dehling (2001)

TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 353, Number 11, Pages 4261–4318
S 0002-9947(01)02819-7
Article electronically published on June 20, 2001

LIMIT THEOREMS FOR FUNCTIONALS OF MIXING PROCESSES WITH APPLICATIONS TO U-STATISTICS AND DIMENSION ESTIMATION

SVETLANA BOROVKOVA, ROBERT BURTON, AND HEROLD DEHLING

ABSTRACT. In this paper we develop a general approach for investigating the asymptotic distribution of functionals $X_n = f((Z_{n+k})_{k\in\mathbf{Z}})$ of absolutely regular stochastic processes $(Z_n)_{n\in\mathbf{Z}}$. Such functionals occur naturally as orbits of chaotic dynamical systems, and thus our results can be used to study probabilistic aspects of dynamical systems. We first prove some moment inequalities that are analogous to those for mixing sequences. With their help, several limit theorems can be proved in a rather straightforward manner. We illustrate this by re-proving a central limit theorem of Ibragimov and Linnik. Then we apply our techniques to U-statistics

$$U_n(h) = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} h(X_i, X_j)$$

with symmetric kernel $h : \mathbf{R}\times\mathbf{R}\to\mathbf{R}$. We prove a law of large numbers, extending results of Aaronson, Burton, Dehling, Gilat, Hill and Weiss for absolutely regular processes. We also prove a central limit theorem under a different set of conditions than the known results of Denker and Keller. As our main application, we establish an invariance principle for U-processes $(U_n(h))_h$, indexed by some class of functions. We finally apply these results to study the asymptotic distribution of estimators of the fractal dimension of the attractor of a dynamical system.

Bob Burton, Svetlana Borovkova, H.D. (+/- 1996)

S.BOROVKOVA, R.M.BURTON, H.DEHLING (2001): Limit Theorems for Functionals of Mixing Processes with Applications to *U*-Statistics and Dimension Estimation. *Transactions of the AMS* **353**, 4261–4318.

Borovkova, Burton, Dehling (2001)

- Empirical distribution of $g(X_i, X_j), 1 \leq i < j \leq n$,

$$U_n(t) := \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} 1_{\{g(X_i, X_j) \leq t\}}, t \in \mathbb{R},$$

- We proved empirical process CLT, i.e. weak convergence of

$$(\sqrt{n}(U_n(t) - P(g(X, Y) \leq t)))_{t \in \mathbb{R}}$$

  to a Gaussian process.
- Process $(X_i)_{i \geq 1}$ is a functional of an absolutely regular process.
- Motivating example: Correlation integral/dimension

$$\frac{1}{\binom{n}{2}} \#\{1 \leq i < j \leq n : \|X_i - X_j\| \leq t\}.$$

What's Next: Table of Contents

Motivation: Tests for Change-Points in Time Series

Observations are generated by a stochastic process $(X_i)_{i\geq 1}$,

$$X_i = \mu_i + \epsilon_i,$$

where $(\mu_i)_{i\geq 1}$ are the unknown signals, and where $(\epsilon_i)_{i\geq 1}$ is a noise process, satisfying $E(\epsilon_i) = 0$, $\mathrm{Var}(\epsilon_i) < \infty$.

Based on observations $X_1, \ldots, X_n$, we wish to test the hypothesis

$$H\ :\ \mu_1 = \ldots = \mu_n$$

against the alternative

$$A\ :\ \mu_1 = \ldots = \mu_k \neq \mu_{k+1} = \ldots = \mu_n \text{ for some } k \in \{1, \ldots, n-1\},$$

i.e. that there is a level shift somewhere in the process.

CUSUM Test Statistic

- CUSUM test uses the test statistic

$$T_n = \frac{1}{\sqrt{n}} \max_{1\le k\le n} \left(\sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i \right)$$
$$= \sqrt{n} \max_{1\le k\le n-1} \frac{k}{n}\left(1-\frac{k}{n}\right) \left(\frac{1}{k}\sum_{i=1}^{k} X_i - \frac{1}{n-k} \sum_{i=k+1}^{n} X_i \right)$$

- This test statistic is based on the two-sample Gauss test, comparing the means of the two samples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$.
- Asymptotic distribution of the CUSUM test can be derived from a Donsker-type invariance principle for the partial sum process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} (X_i - \mu),\ 0 \le \lambda \le 1.$$

- General principle behind change-point tests: Consider 2-sample test, applied to the samples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$, and max over k.

Robust Change-Point Test

We study two robust change-point tests, motivated by the Wilcoxon and the Hodges-Lehmann tests for the two-sample problem:

- Wilcoxon change-point test statistic

$$W_n = \frac{1}{n^{3/2}} \max_{1 \le k \le n-1} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i \le X_j\}} - \frac{1}{2} \right)$$

- Hodges-Lehmann change-point test statistic

$$T_n = \sqrt{n} \max_{1 \le k \le n-1} \frac{k}{n} \left(1 - \frac{k}{n} \right) \mathrm{med}\{(X_j - X_i) : 1 \le i \le k < j \le n\}.$$

We will show how to derive the asymptotic distribution of a large class of test statistics including the above examples, when the data are dependent.

Three Interesting Processes

The above examples lead to three interesting processes:

1. Two-sample U-statistic process

$$U_{[n\lambda],n-[n\lambda]} := \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j)$$

2. Two-sample empirical U-process

$$U_{[n\lambda],n-[n\lambda]}(t) := \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{g(X_i,X_j)\le t\}}$$

3. Two-sample empirical U-quantile process

$$Q_{[n\lambda],n-[n\lambda]}(p) := \inf\{t : U_{[n\lambda],n-[n\lambda]}(t) \ge p\}$$

We derive their asymptotic distributions when the data $(X_i)_{i\ge 1}$ are dependent.

Short Range Dependence I: Mixing Processes

Rosenblatt (1956) initiated a systematic study of short range dependent processes.

Definition (Rosenblatt 1956)

$(X_n)_{n\geq 1}$ is called α-*mixing* with mixing coefficients $(\alpha_k)_{k\geq 0}$, $\alpha_k \to 0$, if

$$\begin{aligned}&|P((X_1,\ldots,X_k)\in A, (X_{k+m},\ldots,X_{k+m+l})\in B)\\ &\quad -P((X_1,\ldots,X_k)\in A)P((X_{k+m},\ldots,X_{k+m+l})\in B)|\leq \alpha_m\end{aligned}$$

holds for all $k, l \geq 1$ and all measurable subsets $A\subset \mathbb{R}^k, B\subset \mathbb{R}^l$.

The process is called ϕ-*mixing*, if

$$\begin{aligned}&|P\left((X_{k+m},\ldots,X_{k+m+l})\in B\,|\,(X_1,\ldots,X_k)\in A\right)\\ &\quad -P((X_{k+m},\ldots,X_{k+m+l})\in B)|\leq \phi_m\end{aligned}$$

Related notions: *absolutely regular*, ψ-*mixing*

Short Range Dependence II: Functionals of Mixing Processes

A more general concept, covering large classes of examples, is provided by functionals of mixing processes, i.e.

$$X_k = f(Z_k, Z_{k-1}, Z_{k-2}, \ldots),$$

where $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ should e.g. be Lipschitz-continuous and $(Z_k)_{k\in\mathbb{Z}}$ a stationary mixing process. Examples include:

- ARMA processes: $X_n = \sum_{k=0}^{\infty} \alpha_k Z_{n-k}$
- Remainders in dyadic expansion: $X_n = \sum_{k=1}^{\infty} \frac{Z_{n+k}}{10^k}$
- Remainders in continued fraction expansion; the digits are ψ-mixing
- Expanding dynamical systems; e.g. piecewise monotone expanding maps, Hofbauer & Keller 1984.

Two-Sample U-Statistics Process

$$U_{[n\lambda],n-[n\lambda]} = \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j), \; 0 \le \lambda \le 1.$$

The analysis of the asymptotic behavior of the two-sample U-statistic process uses the Hoeffding decomposition of the kernel

$$h(x,y) = \theta + h_1(x) + h_2(y) + \psi(x,y),$$

where

$$\begin{aligned} \theta &= Eh(X,Y) \\ h_1(x) &= Eh(x,Y) - \theta \\ h_2(y) &= Eh(X,y) - \theta \\ \psi(x,y) &= h(x,y) - h_1(x) - h_2(y) - \theta. \end{aligned}$$

Two-Sample U-Statistics Functional CLT

Theorem (D., Fried, Garcia, Wendler 2015)

Let $(X_i)_{i \geq 1}$ be a functional of an absolutely regular process satisfying some technical assumptions. Then the 2-sample U-statistic process

$$\sqrt{n}\left(\lambda(1-\lambda)(U_{[n\lambda],n-[n\lambda]} - \theta)\right)_{0 \leq \lambda \leq 1}$$

converges in distribution to $((1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda))_{0 \leq \lambda \leq 1}$, *where* $(W_1(\lambda), W_2(\lambda))$ *denotes 2-dimensional Brownian motion with covariance function*

$$E(W_i(\lambda)\, W_j(\mu)) = (\lambda \wedge \mu) \sum_{k \in \mathbb{Z}} \mathrm{Cov}(h_i(X_1), h_j(X_k)).$$

- Csörgő and Horváth (1988) proved this for IID data.

Idea of Proof I: Hoeffding decomposition

Using Hoeffding decomposition $h(x,y) = \theta + h_1(x) + h_2(y) + \psi(x,y)$, we obtain

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (h(X_i, X_j) - \theta)$$

$$= (n - [n\lambda]) \sum_{i=1}^{[n\lambda]} h_1(X_i) + [n\lambda] \sum_{j=[n\lambda]+1}^{n} h_2(X_j) + \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \psi(X_i, X_j)$$

- The first two terms can be treated using the functional CLT for partial sums of vectors $\sum_{i=1}^{[n\lambda]} (h_1(X_i), h_2(X_i))$.
- Need to show that $\sup_{0\leq\lambda\leq 1} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \psi(X_i, X_j) = o_P(n^{3/2})$. Note that the kernel $\psi(x,y)$ is degenerate, i.e. that $E\psi(x,Y) = E\psi(X,y) = 0$.

Idea of Proof II: Generalized Correlation Inequalities

In order to bound $\sup_{0\leq\lambda\leq 1}\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n}\psi(X_i,X_j)$, we study moments of increments of

$$\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n}\psi(X_i,X_j),$$

- Note that, for IID data, $\psi(X_i,X_j)$ are mutually uncorrelated.
- For dependent data, we establish generalized correlation inequalities, controlling

$$\left|Eg(X_{i_1},\ldots,X_{i_k},X_{i_{k+1}},\ldots,X_{i_n})-Eg(X_{i_1},\ldots,X_{i_k},X'_{i_{k+1}},\ldots,X'_{i_n})\right|,$$

where $(X'_i)_{i\geq 1}$ is an independent copy of $(X_i)_{i\geq 1}$; see Borovkova, Burton, D. (2001) for functionals of absolutely regular processes.

Two Sample Empirical U-Process

$$U_{[n\lambda],n-[n\lambda]}(t) = \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{g(X_i,X_j)\leq t\}}$$

- Under the hypothesis of stationarity, $U_{[n\lambda],n-[n\lambda]}(t)$ is an estimator of $U(t) := P(g(X,Y) \leq t)$, where X, Y are independent copies of X_1.
- Note that $U_{[n\lambda],n-[n\lambda]}(t)$ is a two-sample U-statistic process for fixed t, with kernel $h(x,y;t) := 1_{\{g(x,y)\leq t\}}$
- Previous results for empirical processes of 1-sample U-statistics, i.e. of the data $g(X_i, X_j)$, $1 \leq i < j \leq n$:
  - Serfling (1984) for IID data,
  - Arcones and Yu (1994) for absolutely regular data,
  - Borovkova, Burton, D. (2001) for functionals of a.r. data
  - Lévy-Leduc, Boistard, Moulines and Taqqu (2012) for LRD data

Two Sample Empirical U-Process CLT

As a corollary to the two-sample U-statistic process CLT we obtain finite-dimensional convergence of the two-sample empirical U-process.

Theorem (D., Fried, Wendler 2018+)

Let $(X_i)_{i\geq 1}$ be a functional of an absolutely regular process satisfying some technical assumptions. Then

$$\left(\sqrt{n}\lambda(1-\lambda)(U_{[n\lambda],n-[n\lambda]}(t)-U(t))\right)_{0\leq\lambda\leq 1} \xrightarrow{\mathcal{D}} (W(\lambda))_{0\leq\lambda\leq 1},$$

where $W(\lambda)=(1-\lambda)W_1(\lambda)+\lambda(W_2(1)-W_2(\lambda)$, and where $(W_1(\lambda),W_2(\lambda))$ is 2-dimensional Brownian motion with covariance

$$E(W_i(\lambda)\,W_j(\mu))=(\lambda\wedge\mu)\sum_{k\in\mathbb{Z}}\mathrm{Cov}(h_i(X_1;t),h_j(X_k;t)).$$

Here, $h_1(x;t)$ and $h_2(y;t)$ denote the terms in the Hoeffding decomposition of $h(x,y,t)=1_{\{g(x,y)\leq t\}}$.

- Open problem: Process convergence, i.e. as process in (λ,t).

Bahadur-Kiefer representation

Proposition (D., Fried, Wendler 2018+)

Under some technical assumptions concerning short range dependence of the process $(X_i)_{i\geq 1}$, continuity of the kernel $g(x,y)$ and differentiability of $U(t) = P(g(X,Y) \leq t)$, we obtain

$$\sup_{0\leq\lambda\leq 1} \lambda(1-\lambda)\left(Q_{[n\lambda],n-[n\lambda]}(p) - Q(p) + \frac{U_{[n\lambda],n-[n\lambda]}(Q(p)) - p}{u(Q(p))}\right) = O(n^{-\frac{5}{9}}),$$

almost surely. Here $u(t) = \frac{d}{dt}U(t)$.

- The Bahadur-Kiefer representation makes it possible to derive the limit distribution of the 2-sample U-quantiles from corresponding results for the 2-sample empirical U-process.

Limit Distribution of 2-Sample Empirical U-Quantiles

Theorem (D., Fried, Wendler 2018+)

Under the same assumptions as above, the process

$$\sqrt{n}\left(\lambda(1-\lambda)\left(Q_{[n\lambda],n-[n\lambda]}(p)-Q(p)\right)\right)_{0\le\lambda\le 1}$$

converges in distribution to the process

$$((1-\lambda)W_1(\lambda)+\lambda(W_2(1)-W_2(\lambda)))_{0\le\lambda\le 1},$$

where $(W_1(\lambda), W_2(\lambda))$ *is a two-dimensional Brownian motion with*

$$\mathrm{Cov}(W_i(\mu), W_j(\lambda)) = (\mu\wedge\lambda)\frac{1}{u^2(Q(p))}\sum_{k\in\mathbb{Z}} E(h_i(X_0;Q(p)), h_j(X_k;Q(p))).$$

Here $u(t)=\frac{d}{dt}U(t)$.

Asymptotics of Hodges-Lehmann Process

Corollary (D., Fried, Wendler 2018+)

Under some technical assumption (short range dependent observations, properties of the marginal distribution), the process

$$\left(\sqrt{n}\lambda(1-\lambda)\,\mathrm{med}\{(X_j - X_i) : 1 \leq i \leq [n\lambda], [n\lambda]+1 \leq j \leq n\}\right)_{0\leq\lambda\leq 1}$$

converges in distribution towards the Gaussian process

$$\left(\frac{\sigma}{u(0)} W^{(0)}(\lambda)\right)_{0\leq\lambda\leq 1},$$

where $W^(0)$ is a Brownian bridge, $\sigma^2 = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(F(X_0), F(X_k))$ and $u(t)$ is the density of $Y - X$.

Both σ^2 and $u(0)$ can be consistently estimated from the data.

Recent Progress

1. Long-range dependent data
   - D., Rooch, Taqqu (2013, 2017) study the Wilcoxon test for long-range dependent data, i.e. when the correlations are non-summable.
   - Tewes (2017) studies tests for change-points in the distribution, i.e. the hypothesis that the process $(X_n)_{n\geq 1}$ is stationary against the alternative of a change in the distribution function, using the Cramér-von Mises test statistic $\max_{1\leq k\leq n}\int (F_{1:k}(x) - F_{k+1:n}(x))^2 dx$, and the Kolmogorov-Smirnov test.
2. Tests for changes in the dependence structure of a bivariate time series:
   - D., Vogel, Wendler, Wied (2017) study robust tests for changes in the dependence structure of two time series using Kendall's tau.
   - Schnurr, D. (2017) address the same problem using the concept of ordinal pattern dependence.

Some References

1. H.DEHLING, R.FRIED, I.GARCIA, M.WENDLER (2015): Change-Point Detection under Dependence Based on Two-Sample U- Statistics. In: *Asymptotic Laws and Methods in Stochastics*, Fields Institute Communication **76**, 195–220.
2. H.DEHLING, R.FRIED, M.WENDLER (2018+): A Robust Method for Shift Detection in Time Series. *arXiv:1506.03345 [math.ST]*.
3. H.DEHLING, A.ROOCH, M.S.TAQQU (2017): Power of Change-Point Tests for Long-Range Dependent Data. *Electronic Journal of Statistics* **11**, 2168–2198.
4. H.DEHLING, A.ROOCH, M.S.TAQQU (2013) Nonparametric Change-Point Tests for Long-Range Dependent Data. *Scandinavian J. of Statistics* **40**, 153–173.
5. J.TEWES (2017): Change-point tests under local alternatives for long-range dependent processes. *Electronic Journal of Statistics* **11**, 2461-2498.
6. A.SCHNURR, H.DEHLING (2017): Testing for Structural Breaks via Ordinal Pattern Dependence. *J. of the American Statistical Association* **112**, 706–720.
7. H.DEHLING, D.VOGEL, M.WENDLER, D.WIED (2017): Testing for Changes in Kendall's Tau. *Econometric Theory* **33**, 1352–1386.

Bob Burton (1997)