İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 1 - 17

STATISTICAL COMPARISON BETWEEN LINEAR AND NONLINEAR PREDICTORS

Yasumasa Matsuda and Mituaki Huzii
Department of Industrial Engineering, Kanagava University and
The National Center for University Entrance Examination, JAPAN

Abstract

In this paper, we construct a type of nonlinear predictor and investigate the relations between the prediction errors of the linear and the nonlinear predictors when the sample size is not large as well as case where the sample size tends to infinity.

Key Words: stationary time series, quadratic predictor, least square estimator, mean square prediction error.

1. Introduction

For predicting the values of future time points of a stationary stochastic process $\{X_t\}$, many predictors are available. Sometimes we use a linear predictor and, sometimes, a nonlinear predictor. As nonlinear predictors, many authors use predictors constructed on the basis of bilinear models (Subba Rao and Gabr (1984)), threshold models (Tong (1990)), and etc.

We assume that the exact model of the time series is not known and evaluate the goodness of a predictor by means of the mean square prediction error. Here, we consider the case when we do not know the values of any moments of $\{X_t\}$.

We suppose that the samples $\{X_t; -(T-1) \leq t \leq 0\}$ are obtained and want to predict the value of X_h for a positive integer h. As a linear predictor, we

usually use the predictor $\hat{X}_h(l)$, which can be expressed as

$$\hat{X}_h(l) = \sum_{i=0}^{p-1} \phi_i X_{-i}, \tag{1}$$

where ϕ_i's are unknown constants and p is a suitably chosen nonnegative integer. For estimation of ϕ_i's we use the least square estimators $\hat{\Phi}_h = (\hat{\phi}_1, \ldots, \hat{\phi}_p)$ constructed by the sample data..

As a nonlinear predictor, we consider a predictor which is expressed by means of a complete system of orthogonal functions of the samples, since we treat the case where the exact model is not known and where the best predictor is systematically approximated as much as possible. For the first step of this study, we take the system of orthogonal polynomial functions, on the grounds that a kind of the best nonlinear predictor can be generally expressed by using the system of orthogonal polynomial functions (Masani and Wiener (1959)) and that such a system includes the linear predictor, which is usually used, as a special case. The adopted nonlinear predictor is

$$\begin{aligned} \hat{X}_h(nl) &= \theta + \sum_{i_1=0}^{p-1} \theta_{i_1} X_{-i_1} + \sum_{i_1=0}^{p-1} \sum_{i_2=0}^{p-1} \theta_{i_1,i_2} X_{-i_1} X_{-i_2} + \cdots \\ &+ \sum_{i_1=0}^{p-1} \sum_{i_2=0}^{p-1} \cdots \sum_{i_K=0}^{p-1} \theta_{i_1,i_2,\ldots,i_K} X_{-i_1} X_{-i_2} \cdots X_{-i_K}, \end{aligned} \tag{2}$$

which is expressed by the K-th order polynomial function of the samples. Usually, we do not know the values of θ, θ_{i_1}'s, θ_{i_1,i_2}'s and etc. We replace $\Theta_h^K = (\theta,\ \theta_{i_1}\text{'s},\ \theta_{i_1,i_2}\text{'s},\ \ldots,\ \theta'_{i_1,\ldots,i_K}\text{s})'$ by the least square estimators $\hat{\Theta}_h^K = (\hat{\theta},\ \hat{\theta}_{i_1}\text{'s},\ \hat{\theta}_{i_1,i_2}\text{'s}, \ldots, \theta_{i_1,\ldots,i_K}\text{'s})$.

When X_t is a Gaussian process, $\hat{X}_h(l)$ is known to be the best predictor for sufficiently large sample size T. When however $\{X_t\}$ is not known to be Gaussian, our interest is whether we should use a nonlinear predictor $\hat{X}_h(nl)$. In general, $\hat{X}_h(nl)$ includes $\hat{X}_h(l)$ as a special case, and the mean square prediction error of $\hat{X}_h(nl)$ is not greater than that of $\hat{X}_h(l)$ for a non-Gaussian process when T is large. But $\hat{X}_h(nl)$ includes higher dimensional parameter than $\hat{X}_h(l)$, and we can suspect that $\hat{X}_h(nl)$ includes less stable estimators of the parameters than $\hat{X}_h(l)$ for especially small sample size. Thus, for especially small-sized samples, we can guess that the variance of $\hat{X}_h(nl)$ is larger than that of $\hat{X}_h(l)$.

STATISTICAL COMPARISON BETWEEN PREDICTORS

In this paper, we discuss which of the predictors $\hat{X}_h(l)$ and $\hat{X}_h(nl)$ is better to use in the relation of sample size T. In section 2, we mathematically evaluate the mean square errors of the linear and nonlinear predictors when the sample size is large and $K = 2$. In section 3, we carry out simulation studies to show their prediction errors when the sample size is small, since it is difficult to evaluate the prediction errors mathematically.

2. Mean square prediction errors for large sample sizes

In this section, we evaluate the mean square prediction errors of the linear and nonlinear predictors.

Put

$$\begin{aligned} \mathbf{V}_t &= (X_t, \ldots, X_{t+1-p})', \\ \mathbf{W^K}_t &= (1, \{X_{t-i_1}\}, \{X_{t-i_1} X_{t-i_2}\}, \ldots, \{X_{t-i_1} \cdots X_{t-i_K}\})', \end{aligned}$$

where $0 \le i_1 \le \cdots \le i_K \le p-1$, and also let $\hat{X}_h(l)^0 = V_0' \Phi_h$ and $\hat{X}_h(nl)^0 = W_0^{K\prime} \Theta_h^K$ be the projections of X_h onto the spaces spanned by V_0 and W_0^K, respectively.

Then, the mean square prediction error of the linear predictor is

$$\begin{aligned} E(X_h - \hat{X}_h(l))^2 &= E(X_h - \hat{X}_h^0(l) + \hat{X}_h^0(l) - \hat{X}_h(l))^2 \\ &= E(X_h - \hat{X}_h^0(l))^2 + E(\hat{X}_h^0(l) - \hat{X}_h(l))^2 \\ &+ 2E(X_h - \hat{X}_h^0(l))(\hat{X}_h^0(l) - \hat{X}_h(l)). \end{aligned}$$

Here, we assume that the predictor $\hat{X}_h(l)$ is constructed by using the LSE $\hat{\Phi}_h$ based on another time series $Y_{-(T-1)}, \ldots, Y_0$ which is independent of $\{X_t\}$ and has the same probability structure with $\{X_t\}$. Then, the third term vanishes, and the prediction error becomes

$$E(X_h - \hat{X}_h(l))^2 = E(X_h - \hat{X}_h^0(l))^2 + E(\hat{X}_h^0(l) - \hat{X}_h(l))^2. \tag{3}$$

By the same arguments, we have

$$E(X_h - \hat{X}_h(nl))^2 = E(X_h - \hat{X}_h^0(nl))^2 + E(\hat{X}_h^0(nl) - \hat{X}_h(nl))^2. \tag{4}$$

From (3) and (4), we find that the mean square prediction errors of the predictors are composed of the mean square errors of the projections, which do not depend on the sample sizes, and the residual terms caused by random fluctuation of the estimators $\hat{\Phi}$ and $\hat{\Theta}_h^K$. Our concern is to find some relations between the bias terms and the sample sizes.

To evaluate the relations, suppose $\{X_t\}$ is expressed as

$$X_t = \sum_{j=0}^{\infty} \phi_j Z_{t-j}, \tag{5}$$

where $\{Z_t\}$ is a white noise sequence with $EZ_t = 0$, $EZ_{t-i}Z_{t-j} = \sigma^2\delta(i,j)$, where $\delta(i,j)$ is Kronecker's delta. Here we impose the following assumptions.

Assumption 1 *Z_t is a $4K$th order stationary process, and*

$$\sum_{n_1,\ldots,n_{j-1}=-\infty}^{\infty} |Q_j^Z(n_1,\ldots,n_{j-1})| < \infty,$$

for $3 \leq j \leq 4K$, where $Q_j^Z(n_1,\ldots,n_{j-1})$ is a cumulant function of $(Z_t, Z_{t+n_1},\ldots,Z_{t+n_{j-1}})$ defined by a coefficient of the jth order term of a Taylor expansion of the logarithm of their characteristic function.

In many cases, the absolute summability of ϕ_i's is assumed to obtain some theoretical results. Here, we relax this summability condition and study the effects when this summability condition is not satisfied.

Assumption 2 *There exists a constant d with $0 < d \leq 2$ such that*

$$\sum_{j=0}^{\infty} |\phi_j|^d < \infty.$$

Assumption 3 *The distribution function of Z_t has an absolutely continuous component.*

Under the Assumptions 1 and 2, $E|X_t|^j$ for $1 \leq j \leq 4K$ is shown to be finite (see section 4 for a proof) and put

$$R = EV_tV_t', \qquad \hat{R} = \frac{1}{T}\sum_{t=-T+p}^{-h} V_tV_t',$$

$$
\begin{aligned}
r_h &= EX_{t+h}V_t, & \hat{r}_h &= \frac{1}{T}\sum_{t=-T+p}^{-h} X_{t+h}V_t, \\
S_K &= EW_t^K W_t^{K\prime}, & \hat{S}_K &= \frac{1}{T}\sum_{t=-T+p+1}^{-h+1} W_t^K W_t^{K\prime}, \\
s_h^K &= EX_{t+h}W_t^K, & \hat{s}_h^K &= \frac{1}{T}\sum_{t=-T+p}^{-h} X_{t+h}W_t^K.
\end{aligned}
$$

Note that matrix R can be shown to be non-singular only from Assumption 2 (see, for example, Brockwell and Davis, 1990, p167), but S_K is not always non-singular-see Masani and Wiener (1959) for a counter example where they showed that Assumption 3 was a sufficient condition for non-singularity S_K .

The least square estimators $\hat{\Phi}_h$ and $\hat{\Theta}_h^K$ are

$$
\begin{aligned}
\hat{\Phi}_h &= \hat{R}^{-1}\hat{r}_h \\
\hat{\Theta}_h^K &= \hat{S_K}^{-1}\hat{s}_h^K,
\end{aligned}
$$

respectively. And put

$$
\begin{aligned}
\Phi_h &= R^{-1}r_h \\
\Theta_h^K &= S_K^{-1}s_h^K.
\end{aligned}
$$

Hence we have the following theorem:

Theorem 1 *Suppose that Assumptions 1, 2 and 3 are satisfied. Then,* $\hat{\Phi}_h$ *and* $\hat{\Theta}_h^K$ *converge in probability to* Φ_h *and* Θ_h^K*, respectively, as* T *tends to* ∞

For a proof of Theorem 1, see section 4. From this theorem, the variances of the asymptotic distributions for the linear and nonlinear predictors as T tends to infinity are

$$
E\Big(X_h - \hat{X}_h^0(l)\Big)^2 = EX_h^2 - r_h' R^{-1} r_h, \tag{6}
$$

$$
E\Big(X_h - \hat{X}_h^0(nl)\Big)^2 = EX_h^2 - s_h^{K\prime} S_K^{-1} s_h^K. \tag{7}
$$

respectively.

Our next interest lies in the examination of the second terms in (3) and (4) from the standpoints of the tail behavior of $\{\phi_i\}$ in Assumption 2. Define a sequence $\{\alpha_j\}$ such that

$$\alpha_j = \sum_{i=j+1}^{\infty} |\phi_i|^d, \tag{8}$$

which is shown to converge to 0 as j tends to ∞ in consequence of Assumption 2. Let the nonlinear predictor for $K = 2$ be denoted as $\hat{X}_h(q)$, which we call the quadratic predictor in the subsequent discussions. Then, for $K = 2$, the following theorem can be obtained for the linear and the quadratic predictors.

Theorem 2 *Under Assumptions 1, 2 and 3 and for $K = 2$,*

$$\hat{X}_h(l) - \hat{X}_h^0(l) = O_p(l_T), \tag{9}$$

$$\hat{X}_h(q) - \hat{X}_h^0(q) = O_p(q_T), \tag{10}$$

as T tends to ∞, where $\{l_T\}$ and $\{q_T\}$ are sequences which converge both to 0 as T tends to infinity. In particular, the following relations are satisfied:

1. *When $d \leq 1$, we have $l_T = q_T = T^{-\frac{1}{2}}$.*
2. *When $d > 1$, we have $l_T = T^{-\frac{1}{2}}$ and $q_T = o(1)$, if $\sum_{j=1}^{\infty} |\alpha_j|^{\frac{2}{d}} < \infty$. In particular, if $\sum_{j=1}^{\infty} |\alpha_j|^{\frac{1}{d}} < \infty$, we have $l_T = q_T = T^{-\frac{1}{2}}$.*

Proof of Theorem 2 is given in section 4. This theorem shows $\{l_T\}$ and $\{q_T\}$ depend on both d and $\{\alpha_j\}$. We provide some examples below.

Example 1 *When $\alpha_j = O(\mu^j)$ with $|\mu| < 1$, we have $l_T = T^{-\frac{1}{2}}$ and $q_T = T^{-\frac{1}{2}}$, regardless of the value of d.*

Example 2 *When $d > 1$ and $\alpha_j = O(j^{-\mu})$ with $\mu > 0$, we have*

1. *$l_T = T^{-\frac{\mu}{d}}$ and $q_T = T^{-\frac{\mu}{2d}}$, if $0 < \mu < \frac{d}{2}$.*
2. *$l_T = \sqrt{\frac{\log T}{T}}$ and $q_T = T^{-\frac{\mu}{2d}}$, if $\mu = \frac{d}{2}$.*
3. *$l_T = T^{-\frac{1}{2}}$ and $q_T = T^{-\frac{\mu}{2d}}$, if $\frac{d}{2} < \mu < d$.*

4. $l_T = T^{-\frac{1}{2}}$ *and* $q_T = \sqrt{\frac{\log T}{T}}$, *if* $\mu = d$.

5. $l_T = T^{-\frac{1}{2}}$ *and* $q_T = T^{-\frac{1}{2}}$, *if* $\mu > d$.

Note that a stationary causal ARMA process is a special case of the Example 1 and a stationary fractional ARIMA process is a special case of the Example 2 (Brockwell and Davis (1990, p522)).

As it is shown in Theorem 2 $l_T^{-1}(\hat{X}_h(l) - \hat{X}_h^0(l))$ and $q_T^{-1}(\hat{X}_h(q) - \hat{X}_h^0(q))$ converge in probability, and hence, in distribution, to some variables with variances C_1 and C_2, say, we approximate the variances of $\hat{X}_h(l) - \hat{X}_h^0(l)$ and $\hat{X}_h(q) - \hat{X}_h^0(q)$ for large sample sizes by $C_1 l_T^2$ and $C_2 q_T^2$, respectively. The latter are called asymptotic variances. This shows the relation between the mean square prediction errors and the sample size.

Theorem 3 *Under Assumptions 1,2 and 3 for* $K = 2$,

$$E(X_h - \hat{X}_h(l))^2 \simeq E(X_h - \hat{X}_h^0(l))^2 + C_1 l_T^2, \tag{11}$$

$$E(X_h - \hat{X}_h(q))^2 \simeq E(X_h - \hat{X}_h^0(q))^2 + C_2 q_T^2, \tag{12}$$

where "$\simeq$*" means the variances of* $\hat{X}_h(l) - \hat{X}_h^0(l)$ *and* $\hat{X}_h(q) - \hat{X}_h^0(q)$ *are replaced by their asymptotic variances.*

Here we examine Theorem 3 by concrete examples. We generate time series by

$$X_t = 1.5X_{t-1} - 0.56X_{t-2} + f_t - 1.3f_{t-1}, \tag{13}$$

where f_t is a stochastic process with mean 0. To examine the effects of $\{\alpha_k\}$ defined in (8) as it converges to 0 slowly, let f_t be generate by

$$f_t = (1 - B)^{\eta_1} Z_t,$$

where B is a backward shift operator and η_1 is a constant with $0 \leq \eta_1 < 0.5$, and Z_t's are independent standard normal random variables. Note that Assumptions 1-3 are satisfied. When $\eta_1 = 0$, X_t is an ordinary ARMA process and the Assumption 2 is satisfied with $d = 1$. When $0 < \eta_1 < 0.5$, X_t is a stationary long memory process and the Assumption 2 is satisfied with $d = 2$.

We generate time series by (13) and repeat the evaluation of the prediction error 1000 times and calculate

$$MSE(l,p) = \frac{1}{1000}\sum_{i=1}^{1000}(X_h^{(i)} - \hat{X}_h(l)^{(i)})^2, \tag{14}$$

$$MSE(q,p) = \frac{1}{1000}\sum_{i=1}^{1000}(X_h^{(i)} - \hat{X}_h(q)^{(i)})^2, \tag{15}$$

where $\hat{X}_h(l)^{(i)}, \hat{X}_h(q)^{(i)}$ and $X_h^{(i)}$ mean the values of $\hat{X}_h(l), \hat{X}_h(q)$ and X_h in the ith simulation study, respectively.

To examine (11) and (12), we express MSE(L,p) and MSE(Q,p) as

$$MSE(l,p) = E(X_h - \hat{X}_h^0(l))^2 + v(l), \tag{16}$$

$$MSE(q,p) = E(X_h - \hat{X}_h^0(q))^2 + v(q). \tag{17}$$

and evaluate $v(l)$ and $v(q)$ by calculating $E(X_h - \hat{X}_h^0(l))^2$ and $E(X_h - \hat{X}_h^0(q))^2$ mathematically by (6) and (7), respectively.

In Figure 1, we show the values of $v(l)$ and $C_1 l_T^2$ and those of $v(q)$ and $C_2 q_T^2$ for $h = 1$, $p = 4$ and η_1 being equal to 0.2, since the results for other values have same tendencies. From Figure 1, we find that (11) holds with $C_1 = 7.0$ and $l_T = T^{-0.5}$ and (12) holds with $C_2 = 2.0$ and $q_T = T^{-0.3}$. (11) holds when the sample size is larger than 50, while (12) holds when the sample size is larger than 400.

3. Mean square errors for small sample sizes

In this section, we show the comparison between linear and nonlinear predictors especially when the sample size is small. When sample size is not large, it is very difficult to evaluate the prediction errors of $\hat{X}_h(l)$ and $\hat{X}_h(nl)$ mathematically and, so, we carry out simulation studies for the case $K = 2$.

For our purpose, we generate X_t by (13) for f_t given by

$$\text{Case 1.} \quad f_t = (1-B)^{\eta_1} Z_t,$$

$$\text{Case 2.} \quad f_t = (1-B)^{\eta_1} \varepsilon_t,$$

$$\text{Case 3.} \quad f_t = \sum_{j=1}^{\infty} j^{-\eta_2} \varepsilon_{t-j},$$

where Z_t and ε_t are random variables with standard normal and exponential distributions, respectively. Here the exponential distribution is adopted as a typical example of non-Gaussian distributions.

When $\eta_2 > 1.0$, Assumption 2 is satisfied with $d = 1$ and, when $\eta_2 \leq 1.0$, the Assumption 2 is satisfied with $d = 2$. By (14), (15), (16) and (17) we evaluate MSE(l,p), MSE(q,p), $v(l)$ and $v(q)$, respectively, for Cases 1-3. In the following, $E(X_h - \hat{X}_h^0(l))^2$ and $E(X_h - \hat{X}_h^0(q))^2$ are denoted as AMSE(l,p) and AMSE(q,p), respectively.

In Figure 2, we show results for Case 1 with η_1 being 0.2, Case 2 with η_1 being 0.4 and Case 3 with η_2 being 0.6, when $p = 4$ and $h = 1$, since other cases give similar results with the above cases. ¿From Figure 2, we find that AMSE(q,4) is smaller than AMSE(l,4) for Cases 2-3 but is equal to AMSE(l,4) for Case 1. This is clear since time series in Case 1 is Gaussian. For each Case, $v(q)$ is much larger than $v(l)$ when the sample size is less than 100 and MSE(q,4) is larger than MSE(l,4) when the sample size is less than 100.

Next we consider the relations between the prediction error of the linear predictor and that of the quadratic predictor for time series which do not necessarily satisfy the Assumptions 1-3. To seek the relations, data is obtained for Case 1 with $\eta_1 = 0.2$, Case 2 with $\eta_1 = 0.4$ and Case 3 with $\eta_2 = 0.6$, and by the following nonlinear models:

$$\text{Case 4.} \quad X_t = Z_t + 0.2Z_{t-1}^2 + 0.3Z_{t-2}^2,$$

$$\text{Case 5.} \quad X_t = Z_t + 0.2Z_{t-1}^3 + 0.3Z_{t-2}^3,$$

$$\text{Case 6.} \quad X_t = 0.4X_{t-1} - 0.3X_{t-2} + 0.4X_{t-1}Z_{t-1} + Z_t,$$

$$\text{Case 7.} \quad X_t = 0.4X_{t-1} - 0.3X_{t-2} + 0.4X_{t-2}Z_{t-1} + Z_t,$$

$$\text{Case 8.} \quad X_t = 0.4X_{t-1} - 0.3X_{t-2} + 0.4X_{t-2}Z_{t-2} + Z_t,$$

$$\text{Case 9.} \quad X_t = \begin{cases} -2.0X_{t-1} + 1.0 + Z_t, & X_{t-1} \geq -1.0 \\ 0.4X_{t-1} + 0.5 + Z_t, & X_{t-1} < -1.0, \end{cases}$$

$$\text{Case 10.} \quad X_t = \begin{cases} -0.5X_{t-1} + 0.4X_{t-2} - 1.0 + Z_t, & X_{t-2} \geq 1.0 \\ 0.4X_{t-1} + 1.0 + Z_t, & X_{t-2} < 1.0, \end{cases}$$

$$\text{Case 11.} \quad X_t = \{0.5 + \exp(-0.1X_{t-1}^2)\}X_{t-1} + Z_t,$$

$$\text{Case 12.} \quad X_t = \{0.5 + 0.4\cos(0.1X_{t-1})\}X_{t-1} + Z_t,$$

where Z_t's are independent random variables which have standard normal distributions. Cases 4-12 are concerned as nonlinear models which are often examined in many literatures, since we want to investigate the forecasting performances of the linear and the quadratic predictors for wide varieties of time series.

By simulation study, we evaluate MSE(l,p) and MSE(q,p) as in (14) and (15), respectively, for sample sizes 50, 100, 200, 500 and 900. And we calculate the ratio

$$r = \frac{MSE(q,p)}{MSE(l,p)}.$$

The results are presented in Table 1 for $p = 4$, and $h = 1, 2, 5$ and 10.

It depends on properties of time series how many sample sizes we need for the quadratic predictor being more accurate than the linear predictor, but we find it would be enough to take the sample size 100 for many cases we examined. However, for Case 9, which is the threshold model, the quadratic predictor has smaller prediction errors than the linear predictor even when the sample size is 50. For a longer forecasting lead time, it needs more sample sizes for the quadratic predictor to be more accurate than the linear predictor.

Proofs of Theorems 1 and 2

In this section, we give the proof of Theorem 2 given in section 2. Let $Q_j^X(n_1, \ldots, n_{j-1})$ be a j-th order cumurant of $(X_t, X_{t+n_1}, \ldots, X_{t+n_{j-1}})$ and put $b(\lambda) = \sum_{j=0}^{\infty} \phi_j \exp(-ij\lambda)$. For a proof of Theorem 2, two lemma are necessary.

Lemma 1 *Under Assumptions 1 and 2, for* $3 \leq j \leq 4K$,

$$\sum_{|k|<T} (1 - \frac{|k|}{T}) Q_j^X(p_1 k + q_1, \ldots, p_{j-1} k + q_{j-1})$$

converges to a constant as T *tends to* ∞, *where* p_i*s take the values 0 or 1 and are not all 0, and* q_i*s are integers.*

Proof. We show the proof only for $j = 4K$, since other cases can be proved by essentially the same arguments. Put $m = 4K - 1$ for simplicity. Now the process $\{X_t\}$ has a $(m+1)$th order spectral density $f_{m+1}^X(\lambda_1, \ldots, \lambda_m)$ such that

$$f_{m+1}^X(\lambda_1, \ldots, \lambda_m) = b(-\lambda_1) \cdots b(-\lambda_m) b(\lambda_1 + \cdots + \lambda_m) f_{m+1}^Z(\lambda_1 + \cdots + \lambda_m, \lambda_2, \ldots, \lambda_m),$$

where $f^Z_{m+1}(\cdot)$ is a $(m+1)$th order spectral density of $\{Z_t\}$. See Brillinger (1974). Denote by $K_T(\lambda)$ the Fejer kernel and we put p_1 to be equal to 1 without loss of generality. Then

$$\begin{aligned}
&\sum_{|k|<T}(1-\frac{|k|}{T})Q^X_{m+1}(k+q_1,\ldots,p_mk+q_m) \qquad (18)\\
= \quad &\int\ldots\int\{\int_{-\pi}^{\pi}K_T(\lambda_1+\ldots+p_m\lambda_m)\cdot 2\pi\exp(iq_1\lambda_1)b(-\lambda_1)b(\lambda_1+\cdots\lambda_m)\\
&\times f^Z_{m+1}(\lambda_1+\cdots+\lambda_m,\lambda_2,\ldots,\lambda_m)d\lambda_1\}b(-\lambda_2)\cdots b(-\lambda_m)\\
&\times\exp\{i(q_2\lambda_2+\cdots+q_m\lambda_m)\}d\lambda_2\cdots d\lambda_m.
\end{aligned}$$

Let

$$H(\lambda_1,\ldots,\lambda_m)=2\pi\exp(iq_1\lambda_1)b(-\lambda_1)b(\lambda_1+\cdots\lambda_m)f^Z_{m+1}(\lambda_1+\cdots+\lambda_m,\lambda_2,\ldots,\lambda_m).$$

Note that for each $\lambda_1\in(-\pi,\pi)$,

$$\int\ldots\int|H(\lambda_1,\ldots,\lambda_m)|^2d\lambda_2\cdots d\lambda_m<\infty.$$

Then,

$$\begin{aligned}
&|\int\ldots\int\int\{K_T(\lambda_1+\ldots+p_m\lambda_m)H(\lambda_1,\ldots,\lambda_m)\\
&\quad -H(-p_2\lambda_2-\ldots-p_m\lambda_m,\lambda_2,\ldots,\lambda_m)\}d\lambda_1\\
&\quad \times b(-\lambda_2)\cdots b(-\lambda_m)\exp\{i(q_2\lambda_2+\cdots+q_m\lambda_m)d\lambda_2\cdots d\lambda_m|\\
&\leq\{\int\ldots\int|\int K_T(\lambda_1+\ldots+p_m\lambda_m)H(\lambda_1,\ldots,\lambda_m)d\lambda_1\\
&\quad -H(-p_2\lambda_2-\ldots-p_m\lambda_m,\lambda_2,\ldots,\lambda_m)|^2d\lambda_2\ldots d\lambda_m\}^{\frac{1}{2}}\\
&\quad \times\{\int\ldots\int|b(-\lambda_2)\ldots b(-\lambda_m)|^2d\lambda_2\ldots\lambda_m\}^{\frac{1}{2}}. \qquad (19)
\end{aligned}$$

By L_2 convergence of the Cesaro sum, (19) converges to 0 and (18) converges to

$$\begin{aligned}
&\int\ldots\int H(-p_2\lambda_2-\ldots-p_m\lambda_m,\lambda_2,\ldots,\lambda_m)\\
&\quad \times b(-\lambda_2)\cdots b(-\lambda_m)\exp\{i(q_2\lambda_2+\cdots+q_m\lambda_m)\}d\lambda_2\ldots d\lambda_m,
\end{aligned}$$

as T tends to ∞. This completes the proof.

Proof of Theorem 1. Since the spectral density function of X_t is square integrable from Assumption 2, EX_tX_{t-k} tends to 0 as k tends to be infinity. By this fact and Lemma 1, we can show easily that $Var(\hat{R}), Var(\hat{r}_h), Var(\hat{S}_k)$ and

$Var(\hat{s}_h^k)$ converge to 0 as T tends to infinity, and $\hat{R}, \hat{r}_h, \hat{S}_k$ and $\hat{s}_h^k$ converge in probability to R, r_h, S_k and s_h^k, respectively by the law of large numbers. This completes the proof of Theorem 1.

Put

$$\begin{aligned}
\bar{X} &= \frac{1}{T}\sum_{t=-(T-1)}^{0} X_t, \\
\gamma_2(p) &= EX_tX_{t+p}, \\
\hat{\gamma}_2(p) &= \frac{1}{T}\sum_{t=-(T-1)}^{0} X_tX_{t+p}, \\
\gamma_3(p_1,p_2) &= EX_tX_{t+p_1}X_{t+p_2}, \\
\hat{\gamma}_3(p_1,p_2) &= \frac{1}{T}\sum_{t=-(T-1)}^{0} X_tX_{t+p_1}X_{t+p_2}, \\
\gamma_4(p_1,p_2,p_3) &= EX_tX_{t+p_1}X_{t+p_2}X_{t+p_3}, \\
\hat{\gamma}_4(p_1,p_2,p_3) &= \frac{1}{T}\sum_{t=-(T-1)}^{0} X_tX_{t+p_1}X_{t+p_2}X_{t+p_3}.
\end{aligned}$$

By using the result of Lemma 1, we can show the following lemma.

Lemma 2 *Under Assumptions 1 and 2,*

$$\bar{X} = O_p(q_T), \tag{20}$$

$$\hat{\gamma}_2(p) - \gamma_2(p) = O_p(l_T), \tag{21}$$

$$\hat{\gamma}_3(p_1,p_2) - \gamma_3(p_1,p_2) = O_p(q_T), \tag{22}$$

$$\hat{\gamma}_4(p_1,p_2,p_3) - \gamma_4(p_1,p_2,p_3) = O_p(q_T), \tag{23}$$

as T tends to ∞, where l_T and q_T are given in Theorem 2.

Proof. We give the proof only for (21) when $p_1 = p_2 = 0$ and (23) when $p_1 = p_2 = p_3 = 0$. First, we show (21). Put $Q_j(\cdot) = Q_j^X(\cdot)$ for simplicity. Then,

$$\begin{aligned}
Var(\hat{\gamma}_2(0)) &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} Cov(X_s^2, X_t^2) \\
&= \frac{1}{T}\sum_{|k|<T}\left(1 - \frac{|k|}{T}\right) Cov(X_t^2, X_{t+k}^2),
\end{aligned}$$

and

$$Cov(X_t^2, X_{t+k}^2) = Q_4(0,k,k) + 2\gamma_2(k)^2.$$

Since $Q_4(0,k,k)$ is Cesaro summable by Lemma 1, we only have to evaluate the order of $\gamma_2(k)^2$. If $d \leq 1$, $\gamma_2(k)^2$ is absolutely summable and $TVar(\hat{\gamma}_2(0))$ converges to a Constant. If $d > 1$, $\gamma_2(k)^2 = O(\alpha_k^{\frac{2}{d}})$ by Holder inequality. Therefore if $\sum_{k=-\infty}^{\infty} \alpha_k^{\frac{2}{d}} < \infty$, $TVar(\hat{\gamma}_2(0))$ converges to a constant as T tends to ∞. If $\sum_{k=-\infty}^{\infty} \alpha_k^{\frac{2}{d}}$ diverges to ∞, we can say only that $Var(\hat{\gamma}_2(0)) = o(1)$. This completes the proof of (21).

Next we show (23) when $p_1 = p_2 = p_3 = 0$.

$$\begin{aligned} Var(\hat{\gamma}_4(0,0,0)) &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} Cov(X_s^4, X_t^4) \\ &= \frac{1}{T}\sum_{|k|<T}(1-\frac{|k|}{T})Cov(X_t^4, X_{t+k}^4), \end{aligned}$$

and

$$\begin{aligned} &Cov(X_t^4, X_{t+k}^4) = \\ &\quad Q_8(0,0,0,k,k,k,k) + 6Q_6(0,0,0,k,k)\gamma_2(0) + 16Q_6(0,0,k,k,k)\gamma_2(k) \\ &\quad +6Q_6(0,k,k,k,k)\gamma_2(0) + 4Q_5(0,0,0,k)Q_3(0,0) + 24Q_5(0,0,k,k)Q_3(k,k) \\ &\quad +24Q_5(0,k,k,k)Q_3(0,k) + 4Q_5(k,k,k,k)Q_3(0,0) + 16Q_4(k,k,k)Q_4(0,0,k) \\ &\quad +18Q_4(0,k,k)Q_4(0,k,k) + 48Q_4(k,k,k)\gamma_2(k)\gamma_2(0) + 72Q_4(0,k,k)\gamma_2(k)\gamma_2(k) \\ &\quad +36Q_4(0,k,k)\gamma_2(0)\gamma_2(0) + 48Q_4(0,0,k)\gamma_2(k)\gamma_2(0) + 24Q_3(0,0)Q_3(0,k)\gamma_2(0) \\ &\quad +16Q_3(0,0)Q_3(0,0)\gamma_2(k) + 36Q_3(k,k)Q_3(k,k)\gamma_2(0) + 144Q_3(k,k)Q_3(0,k)\gamma_2(k) \\ &\quad +24Q_3(k,k)Q_3(0,0)\gamma_2(0) + 36Q_3(0,k)Q_3(0,k)\gamma_2(0) + 24\gamma_2(k)^4 \\ &\quad +72\gamma_2(k)^2\gamma_2(0)^2. \end{aligned}$$

The term which converges most slowly among the above terms is $16Q_3(0,0)Q_3(0,0)\gamma_2(k)$, since Cesaro summability of the other terms is guaranteed by Lemma 1. Note that $\gamma_2(k)$ is $O(\alpha_k^{\frac{1}{d}})$ by Holder inequality. If $\sum_{k=-\infty}^{\infty} \alpha_k^{\frac{1}{d}} < \infty$, $TVar(\hat{\gamma}_2(0,0,0))$ converges to a constant as T tends to ∞. If $\sum_{k=-\infty}^{\infty} \alpha_k^{\frac{1}{d}}$ diverges to ∞, $Var(\hat{\gamma}_2(0,0,0)) = o(1)$. This completes the proof of (23).

Proof of Theorem 2. First observe that

$$\begin{aligned} \hat{\Phi}_h - \Phi_h &= \hat{R}^{-1}\hat{r}_h - \theta_h \\ &= \hat{R}^{-1}\{\hat{r}_h - r_h - (\hat{R} - R)\Phi_h\}. \end{aligned}$$

Since we have shown in Lemma 2 that $\hat{r}_h - r_h$ and $\hat{R} - R$ are both $O_p(l_T)$ and that $\hat{R}^{-1}$ is $O_p(1)$, $\hat{\Phi}_h - \Phi_h = O_p(l_T)$ holds. By the same argument,

$$\begin{aligned}\hat{\Theta}_h^2 - \Theta_h^2 &= \hat{S}_2^{-1}\hat{s}_h^2 - \Theta_h^2 \\ &= \hat{S}_2^{-1}\{\hat{s}_h^2 - s_h^2 - (\hat{S}_2 - S_2)\Theta_h^2\}.\end{aligned}$$

By noting that $\hat{S}_2^{-1}$ is $O_p(1)$ and that $\hat{s}_h^2 - s_h^2$ and $\hat{S}_2 - S_2$ are both $O_p(q_T)$ by (20), (22) and (23) in Lemma 2, $\hat{\Theta}_h^2 - \Theta_h^2 = O_p(q_T)$ holds.

References

[1] Brockwell, P.J. and Davis, R.A. (1990), *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag.

[2] Brillinger, D.R. (1974), *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, Inc.

[3] Hannan, E. J. (1976), The asymptotic distribution of serial covariances. *Ann. of Statist.*,4, 396-99.

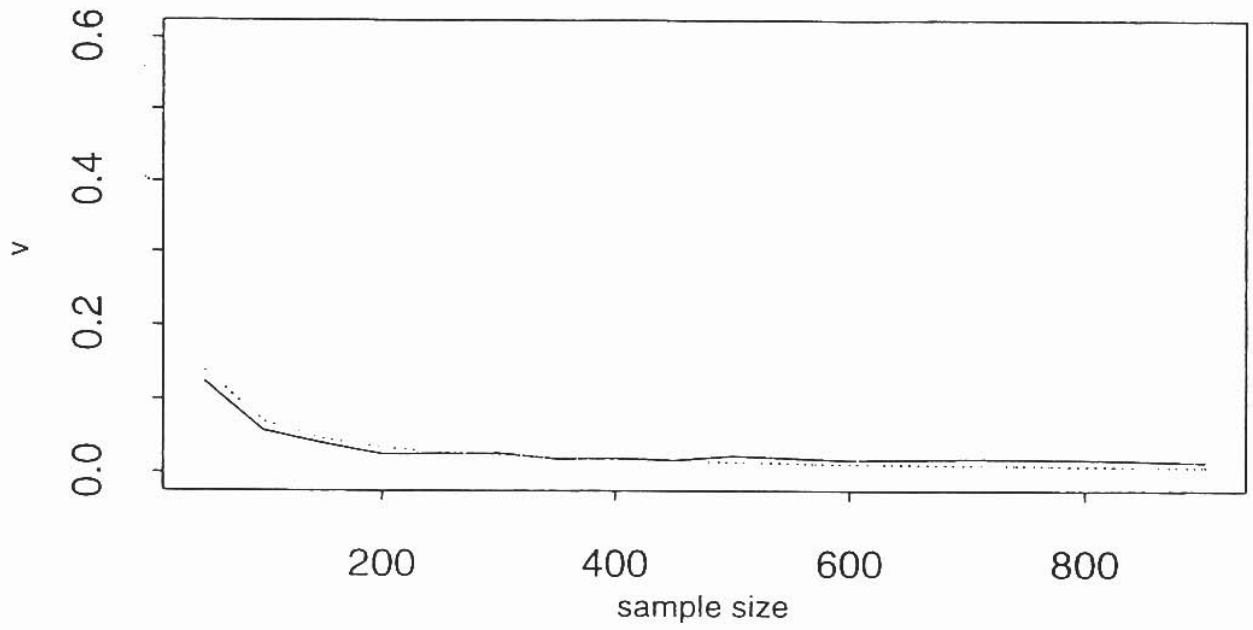
[4] Masani, P and Wiener, N. (1959), Non-linear prediction, *The H. Cramer volume*, John Wiley & Sons, New York, 190-212.

[5] Subba Rao, T. and Gabr, M.M. (1984), *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Lecture Notes in Statistics, Berlin: Springer-Verlag.

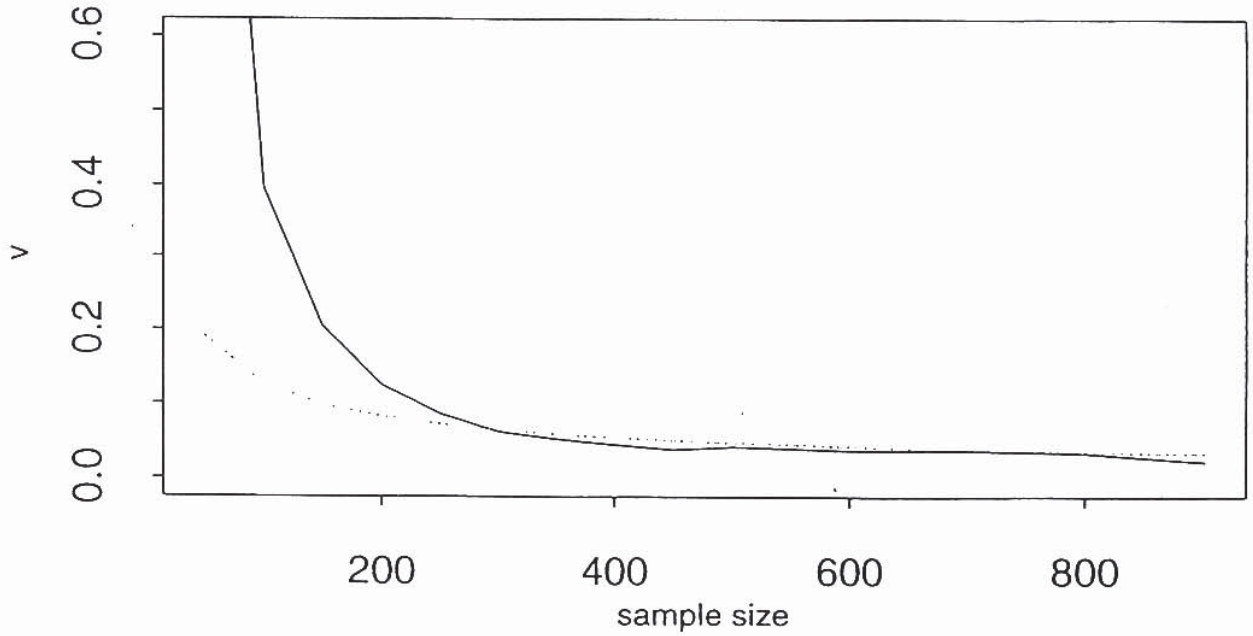
[6] Tong, H. (1990), *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press.

Table 1. The ratios MSE(Q,4)/MSE(L,4) for Cases 1-12.

T	Case	h=1	h=2	h=5	h=10	Case	h=1	h=2	h=5	h=10
50		1.740	1.796	2.358	3.579		1.408	1.922	1.825	3.050
100		1.207	1.139	1.237	1.443		1.038	1.200	1.189	1.255
200	1	1.062	1.055	1.068	1.110	7	0.928	1.068	1.103	1.120
500		1.012	1.023	1.015	1.014		0.877	1.009	1.029	1.011
900		1.006	1.014	1.002	1.000		0.857	1.004	1.015	1.011
50		2.243	2.225	2.555	5.653		1.386	1.798	2.902	3.103
100		1.167	1.088	1.127	1.822		0.976	1.130	1.293	1.329
200	2	0.981	0.882	0.876	1.079	8	0.893	0.889	1.179	1.092
500		0.903	0.822	0.766	0.908		0.821	0.879	1.035	1.016
900		0.883	0.791	0.695	0.852		0.810	0.873	1.022	1.018
50		1.913	2.106	3.518	9.408		0.956	1.835	2.302	2.541
100		1.170	1.121	1.073	1.542		0.613	1.172	1.295	1.143
200	3	0.904	0.826	0.749	0.958	9	0.539	1.053	1.084	1.061
500		0.865	0.770	0.697	0.822		0.502	1.001	1.025	1.018
900		0.867	0.782	0.691	0.815		0.504	0.995	1.021	1.010
50		1.557	1.575	1.855	2.236		1.381	1.711	1.803	2.043
100		1.037	1.142	1.207	1.228		1.023	1.133	1.149	1.232
200	4	0.931	0.983	1.076	1.094	10	0.907	0.968	1.071	1.082
500		0.892	0.950	1.013	1.017		0.853	0.932	1.037	1.039
900		0.891	0.937	1.010	1.009		0.847	0.928	1.024	1.015
50		5.220	8.792	17.208	12.155		1.963	2.686	3.313	5.620
100		2.237	2.278	2.503	2.451		1.284	1.460	1.285	1.353
200	5	1.532	1.351	1.225	1.312	11	1.095	1.127	1.098	1.118
500		1.159	1.087	1.048	1.066		1.042	1.054	1.049	1.025
900		1.037	1.067	1.026	1.017		1.024	1.033	1.035	1.017
50		2.568	2.123	3.140	3.753		1.655	1.768	2.675	7.218
100		1.219	1.239	1.272	1.490		1.201	1.209	1.243	1.391
200	6	0.945	1.107	1.108	1.144	12	1.064	1.075	1.068	1.087
500		0.877	1.015	1.028	1.013		1.025	1.031	1.026	1.023
900		0.844	0.986	1.015	0.999		1.011	1.017	1.017	1.016



(a): v(l) for h=1



(b): v(q) for h=1

Figure 1. The values of $v(l)$ and $C_1 l(T)^2$ and those of $v(q)$ and $C_2 q(T)^2$ for $\eta_1 = 0.2$. In (a), —— indicates $v(l)$ and ····· indicates $7.0 l_T^2$, where $l_T = T^{-0.5}$. In (b), —— indicates $v(q)$ and ····· indicates $2.0 q_T^2$, where $q_T = T^{-0.3}$.

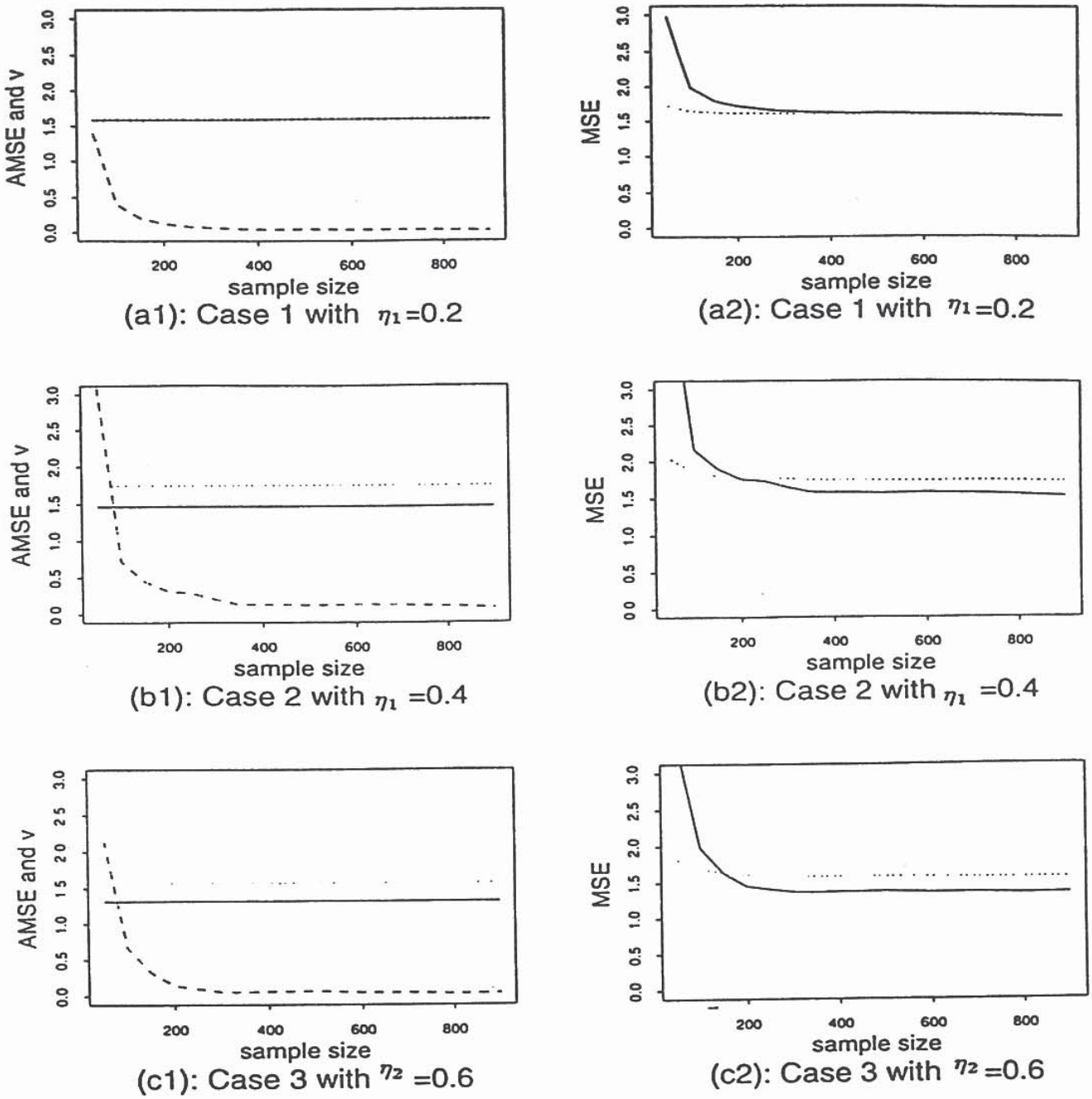


Figure 2. The values of AMSE, v and MSE. In (a1), (b1) and (c1), —— indicates AMSE(q,4), ... indicates AMSE(l,4), ----- indicates $v(q)$ and – – – indicates $v(l)$. In(a2), (b2) and (c2), —— indicates MSE(q,4) and ..... indicates MSE(i,4).

ÖZET

Bu çalışmada, doğrusal olmayan bir kestirici elde edilmiş doğrusal ve doğrusal olmayan kestiricilerin kestirim hataları arasındaki ilişki örneklem hacminin büyüklüğüne göre incelenmiştir.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 19 - 29

TEST OF HETEROGENEITY OF GENERAL POPULATION

Dmitry A. Klyushin, Yuriy I.Petunin
Kiev Shevchenko University, Department of Cybernetics
252017, Vladimirskaya, 64, Kiev, Ukraine

Ismihan G. Bairamov
Ankara University, Faculty of Science, Department of Statistics
06100, Tandogan, Ankara, Turkey

Abstract

The concept of homogeneous and heterogeneous general populations and also the concept of combined and mixed heterogeneous ones are introduced. The algorithms of analysis of homogeneity and heterogeneity of general population with the help of Quetelet curves and spacing curves are considered. The methods stated are used for investigation of heterogeneity of the population of tumour radioresistant cells in different generations of Guerin's carcinoma in rates.

Key words: General population, homogeneous and heterogeneous population, Quetelet curve, spacing curve.

1. Introduction

Consider new concept of the general population. Let $G = \{X_i, i \in I\}$be any set of the elements (individuals); we shall be interested in values of some numerical characteristics only, but not individuals X_i by itself. It is supposed that there is no any ordered structure in the set G (the indexation X_i, which we introduced, is defined only for distinguishing of the elements, but not indication of the order of their arranging). In various applied sciences it is naturally to accept the following hypothesis: numerical characteristics x_i may be considered as a value of some random variable x, however, the set G by itself is not the sample (because it is not ordered), but it is only the set of the sample values of some unknown sample G^*, which is generated by the Nature. There is ordered structure in the sample G^*, but it is unknown to experimenter. Suppose, that the set G^*, and then G, is finite, and *card*

$G^* = n$. Let $G^* = \{x_1^*, x_2^*, ..., x_n^*\}$, denote by $F^*(u_1, u_2, ..., u_n)$ the joint distribution of the random variables $x_1^*, x_2^*, ..., x_n^*$, which may be dependent among themselves, and by $F_1^*(u)$, $F_2^*(u)$, ..., $F_n^*(u)$ - marginal distributions of these random values. Put

$$F(u) = F^*(u) = \frac{1}{n} \sum_{k=1}^{n} F_k^*(u).$$

We shall call the distribution $F(u)$ the basis distribution of the sample G^*. The joint distribution $F^*(u_1, u_2, ..., u_n)$, and therefore the marginal distributions $F_k^*(u)$, and also the basis distribution $F(u)$, are not known to us, however, $F(u)$ may be approximately estimated on the basis of the set G of the sample values of G^*. Indeed, let $\widetilde{F_n}(u)$ be empirical distribution function constructed on the basis of the variational series of the sample G^* or G. By virtue of the well-known results of the theory of probabilities, $\widetilde{F_n}(u) = F(u)$, so that $\widetilde{F_n}(u)$ may be considered as an approximation of the distribution $F(u)$. It should be noted that $F(u)$ is the probability of the following event: the value, which is chosen at random from the set G, does not exceed u. As far as there is no ordered structure in the set G then the unique probabilistic characteristics of this set is one-dimensional distribution $F(u) = p(x < u)$. We shall call the set G with the basis distribution $F(u)$ the general population, thus in our sense the general population is unordered set of dependent or independent sample values.

Let G be a general population and S be some method of sampling of the elements from the set G. As a result of using n times this method of sampling S we obtain the ordered set of numbers (vector) $(x_1, x_2, ..., x_n)$, consisting of the values of the characteristics x of the elements $X_1, X_2, ..., X_n$. We shall call the vector $(x_1, x_2, ..., x_n)$ the sample x obtained with the help of the method S from the set G. For the general population G and given method S for every natural n the joint distribution function $F(u_1, u_2, ..., u_n)$ of the random variables $(u_1, u_2, ..., u_n)$ is defined:

$$F(u_1, u_2, ..., u_n) = p(x_1 < u_1, ..., x_n < u_n).$$

As far as we have not the sample G^* with joint distribution $F^*(u_1, u_2, ..., u_n)$ then the following problem is arising: to renew the sample G^* on the basis of the set of sample values G, i.e. by using the set G to construct n random variables which have joint distribution coinciding with $F^*(u_1, u_2, ..., u_n)$. To solve this problem we proposed two methods [1]. The first method is based on the urn model. We shell assume that all the set of the sample values G is pertaining in the urn; take at random one of the elements γ_1 from the set G, then from the set $G \setminus \{\gamma_1\}$ of the other elements we take at random the following γ_2 and so on. Thus, we have the serie of trainings consisting of random samplings without replacement of the elements of the set G containing in the urn. Received multidimensional random variable $\gamma^{(n)} = (\gamma_1, \gamma_2, ..., \gamma_n)$is called *induced sample* obtaining with the help of the urn model. Denote by Ω_n the set of the results ω of such series of trainings; it is to see

that the number of the elements of the set Ω_n equals to $n!$: $\Omega_n = \{\omega_r = r = 1, 2, ..., n!\}$, and, moreover, $p\{\omega = \omega_r\} = \frac{1}{n!}$. The following statements are correct [1]:

Theorem 1. *Let be $F^*(u_1, u_2, ..., u_n)$ a joint distribution function of the sample G^*; then the joint distribution function $F^*_{\gamma^{(n)}}(u_1, u_2, ..., u_n)$ of the induced sample $\gamma^{(n)}$ obtained with the help of the urns model has the following form:*

$$F^*_{\gamma^{(n)}}(u_1, u_2, ..., u_n) = \frac{1}{n!} \sum_{(i_1,...,i_n)\in I_n} F^*(u_{i_1}, u_{i_2}, ..., u_{i_n})$$

where I_n is the group of permutations of the numbers $(1, 2, ..., n)$.

Corollary 1. *If G^* be the sample from the general population G with the distribution $F(u)$ then every component $\gamma_i, i = 1, 2, ..., n$ of the random vector $\gamma^{(n)}$ has distribution $F(u)$.*

Corollary 2. *If the elements of the sample G^* from the general population G be independent random variables then the distribution function $F^*_{\gamma^{(n)}}(u_1, u_2, ..., u_n)$ of the induced sample $\gamma^{(n)}$ obtained with the help of the urn model coincide with the distribution function $F^*(u_1, u_2, ..., u_n)$:*

$$F^*_{\gamma^{(n)}}(u_1, u_2, ..., u_n) = F^*(u_1, u_2, ..., u_n) = F(u_1)...F(u_n) .$$

Theorem 2. *Let G^* be a sample from the general population G with the distribution $F(u)$ where the sample values from G^* have the normal join distribution. The sample values from G^* are independent if and only if the elements $\gamma_1, \gamma_2, ..., \gamma_n$ of the induced sample $\gamma^{(n)}$ obtained with the help of the urn model are independent random variables.*

The second method is based on the concept of the variational series and randomized procedure. Consider the variational series $x^*_{(1)} \le x^*_{(2)} \le ... \le x^*_{(n)}$ constructed on the unordered set of the sample G^* obtained with the help of the simple sampling (it means that the random variables $x^*_1, x^*_2, ..., x^*_n$ are independent and have the same marginal distribution). Define m- dimensional random variable $\xi^{(h)} = (\xi_1, \xi_2, ..., \xi_m)$, $2 \le m \le n$ with the help of the following additional random experiment (randomized procedure): we take by the simple sampling without replacement the sample $g_r = (i_1, i_2, ..., i_m)$ from the set of the numbers $(1, 2, ..., n)$; where $r = 1, 2, ..., A^m_n$, and $A^m_n = \frac{n!}{(n-m)!}$. Put

$$\xi_1 = x^*_{(1)}, \xi_2 = x^*_{(2)}, ..., \xi_m = x^*_{(m)}, \ \xi^{(m)} = (\xi_1, \xi_2, ..., \xi_m) .$$

Theorem 3. *Let the distribution function $F(u)$ of the general population G has the probability density $f(u)$; then the components $\xi_1, \xi_2, ..., \xi_m$ of the multidimensional random variable $\xi^{(m)}$ are independent and identical distributed random variables, whose distribution functions $F_{\xi_i}(u)$ coincide with $F(u)$, $j = 1, 2, ..., m$.*

Definition 1. *Let G^* be an unordered set of the sample obtained from the general population G by the simple sampling. The multidimensional random variables $\gamma^{(n)}$ and $\xi^{(m)}$, mentioned above in the Theorems 1 and 3, are called renewed samples with the sizes n and m.*

The above mentioned conception of the general population points out that the sample $x = (x_1, x_2, ..., x_k), k < n$ obtained from the general population G with the help of the method of sampling S does not always consist of an independent random values, even if we use the simple sampling. This contradicts to the conventional well-known concepts of the classic mathematical statistics [2] which state that by using the simple sampling from the general population it is possible to obtain the sample values $x = (x_1, x_2, ..., x_k), k < n$, where x_i are only independent variable with the same distribution " [2]. Thus, the existing dependence in the sample G^* cannot be excluded with the help of any methods of sampling S.

In the definition of the general population G and the method of sampling S we do not impose any conditions on the multidimensional distribution function $F(u_1, u_2, ..., u_n$ (for example, the condition of oherence of the distributions and so on). However, hereinafter we shall study two methods of sampling:

1) the simple sampling when the random variables $x_1, x_2, ..., x_n$ are independent and identically distributed, so that $F(u_1, u_2, ..., u_n) = F(u_1)...F(u_n)$, where $F(u)$ is the marginal distribution;

2) random sampling where the random variables are independent but they can have different marginal distributions, so that $F_k(u_k)$, $k = 1, 2, ..., n$; $F(u_1, u_2, ..., u_n)$ $=F_1(u_1)...F_n(u_n)$.

On the basis of the above mentioned results about renewed sample we can guarantee the existence of the simple sampling or the random sampling in the case when the primary G^* consists of the independent sample (hypothesis H). Hereinafter we shall suppose that the hypothesis H is correct.

In applications of mathematical statistics the concepts of homogeneous and heterogeneous general populations are very important. *Homogeneous general population* consists of the similar objects such that the values $x_1, x_2, ..., x_n$ of the quantitative index x have the identical marginal distributions $F(u) : F_k(u) = F(u), k = 1, 2, ..., n$. However, this condition is not sufficient for the homogeneity of the general population. According to empirical statistical concepts it is conventional point of view that the homogeneous general population is characterized by the unimodal distribution function, that is why we shall assume that the general population is homogeneous if the marginal distribution $F(u)$ is unimodal. This definition is not logically faultless

because there counterexamples of the mixture of two normal distributions whose probability density is unimodal. Nevertheless, such normal distributions are sufficiently close each to other and it may be considered as practically identical. Thus, according to our concept the terms "unimodal" and "homogeneous" are the synonyms.

The cell population is the particular case of an abstract general population, which is investigated in contemporary mathematical statistics, and it is the arbitrary set G of the values of the quantitative parameters of some objects with specific distribution $F(u)$, which, probably, is unknown. As a rule, the distribution of the heterogeneous general population G is multimodal, but the sample $x = (x_1, x_2, ..., x_n)$ from the heterogeneous general population G may be in the simplest case an combination of two samples $x^{(1)} = (x_1, x_2, ..., x_k) \in G_1$ and $x^{(2)} = (x_{k+1}, x_{k+2}, ..., x_n) \in G_2$ from the different general populations G_1 and G_2 (we shall call such samples *combined* ones) or the mixture of two samples from these populations G_1 and G_2 (we shall call such samples mixed ones). The combined samples arise when an individuals, which form the sample, have the different localization (i.e. they are placed in the different regions of the space), while an individuals, which form the mixed sample may be placed in a random order.

Consider the problem of the recognition of the general population heterogeneity. It is obvious, that to recognize the combined sample is significantly more simply then mixed one. Really, if we assume that the parts $x^{(1)} = (x_1, x_2, ..., x_k) \in G_1$ and $x^{(2)} = (x_{k+1}, x_{k+2}, ..., x_n) \in G_2$ and of the sample $x = (x_1, x_2, ..., x_n) \in G$ have the equal size, then we can determine the difference between $x^{(1)}$ and $x^{(2)}$ computing the proximity measure between them. If this distinguish is significant then the sample $x = x^{(1)} \cup x^{(2)}$ is combined, otherwise it is should been verified whether this sample is mixed one.

Let us give the description of the computation of the proximity measure between the samples $x = (x_1, x_2, ..., x_n) \in G$ and $x' = \left(x'_1, x'_2, ..., x'_m\right) \in G'$, obtained by simple sampling from the general populations G and G' having unknown distribution functions $F_G(u)$ and $F_{G'}(u)$, respectively. Consider the most general situation when the distribution functions $F_G(u)$ and $F_{G'}(u)$ can have atoms. Explain the concept more detail. As well-known, any distribution function $F(u)$ is monotone decreasing continuous from the left function which is defined on the whole number line and $0 \leq F(u) \leq 1$. Among the points of the domain of the function $F(u)$, i.e. the real line, there exist the points of continuity u_0, then

$$F(u_0 - 0) = \lim_{u \to u_0, u < u_0} F(u) = F(u_0 + 0) = \lim_{u \to u_0, u > u_0} F(u).$$

In addition, on the real line there exist the points of discontinuity u_1, where

$$F(u_0 - 0) < F(u_0 + 0).$$

The points of discontinuity are called *atoms*. For these points the probability of the event "the sample value x exactly equals to u_1" is not zero while for the points

of continuity this probability equals to zero. On the basis of the sample such atoms can be determined if the sample is sufficiently large, of course. In this case atoms are such sample values which occur in sample more then one time.

Let be $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}, x'_{(1)} \leq x'_{(2)} \leq ... \leq x'_{(m)}$ variational series constructed on the samples $x = (x_1, x_2, ..., x_n) \in G$ and $x' = \left(x'_1, x'_2, ..., x'_m\right) \in G'$, respectively. If the order statistics $x_{(k)}$ occur in the variational series more then one time, then $x_{(k)}$ is the atom of the distribution $F_G(u)$. Suppose, that there are no atoms in the half-open interval $\left[x_{(i)}, x_{(j)}\right)$, then on the basis of the results of the paper [3] ,

$$p(A_{ij}) = p\left(\widetilde{x} \in \left(x_{(i)}, x_{(j)}\right)\right) = p_{ij} = \frac{j-i}{n+1} \tag{1}$$

where x is the next sample value from the general population G which does not depend on the sample $x = (x_1, x_2, ..., x_n) \in G$. In the case when the half-open interval $\left[x_{(i)}, x_{(j)}\right)$ contains the atoms, we can represent it in the form of the sum of the adjacent component half-open intervals:

$$\left[x_{(i)}, x_{(j)}\right) = \left[x_{(i)}, x_{(i+1)}\right) \cup \left[x_{(i+1)}, x_{(i+2)}\right) \cup ... \cup \left[x_{(j-1)}, x_{(j)}\right).$$

Suppose that the left end point of some half-open interval $\left[x_{(k)}, x_{(k+1)}\right)$ is an atom. Denote by n_k the number of the repetitions of x_k in the sample $x = (x_1, x_2, ..., x_n)$; put $\gamma_k = \frac{n_k}{n}$. It is readily seen that on the basis of the law of large numbers for sufficiently large n we have:

$$\gamma_k \approx F(x_k + 0) - F(x_k - 0).$$

In this case we must correct the formulae (1) in the following way:

$$p(A_{kk+1}) = p\left(\widetilde{x} \in \left[x_{(k)}, x_{(k+1)}\right)\right) = p_{kk+1} = p\left(\widetilde{x} \in \left\{x_{(k)}\right\} \cup \left(x_{(k)}, x_{(k+1)}\right)\right) =$$
$$p\left(\widetilde{x} = x_{(k)}\right) + p\left(\widetilde{x} \in \left(x_{(k)}, x_{(k+1)}\right)\right) \approx \gamma_k + \frac{1}{n+1}.$$

Taking into account this correction, we have:

$$p(A_{ij}) = p\left(\widetilde{x} \in \left[x_{(i)}, x_{(j)}\right)\right) = p_{ij} =$$
$$p\left(\widetilde{x} \in \left[x_{(i)}, x_{(i+1)}\right) \cup \left[x_{(i+1)}, x_{(i+2)}\right) \cup ... \cup \left[x_{(j-1)}, x_{(j)}\right)\right) =$$
$$p\left(\widetilde{x} \in \left[x_{(i)}, x_{(i+1)}\right)\right) + p\left(\widetilde{x} \in \left[x_{(i+1)}, x_{(i+2)}\right)\right) + ... + p\left(\widetilde{x} \in \left[x_{(j-1)}, x_{(j)}\right)\right),$$

so that

$$p(A_{ij}) = p_{ij} = \gamma_i + \gamma_{i+1} + ... + \gamma_{j-1} + \frac{j-i}{n+1}. \tag{2}$$

Note, hat the formula (2) is correct irrespectively of, whether the half-open interval $\left[x_{(i)}, x_{(j)}\right)$ contains the atoms or does not. In the case when there are no atoms in $\left[x_{(i)}, x_{(j)}\right)$ then $\gamma_i + \gamma_{i+1} + ... + \gamma_{j-1} = 0$, and formula (2) transforms into the formula (1).

If we have the sample $x' = \left(x'_1, x'_2, ..., x'_m\right)$, we can determine the frequency h_{ij} of the random event A_{ij} and the confidential limits $p_{ij}^{(1)}, p_{ij}^{(2)}$ for the probability p_{ij}, corresponding to the given significance level β, such that $1 - \beta = p(B)$, where $B = \left\{p_{ij} \in \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right)\right\}$. These limits can be calculated on the formulae[4]:

$$p_{ij}^{(1)} = \frac{h_{ij} + 0.5g^2 - g\sqrt{h_{ij}(1 - h_{ij})\, m + 0.25g^2}}{m + g^2},$$
$$p_{ij}^{(2)} = \frac{h_{ij} + 0.5g^2 + g\sqrt{h_{ij}(1 - h_{ij})\, m + 0.25g^2}}{m + g^2} \tag{3}$$

where g satisfies condition $\varphi(g) = 1 - \frac{1}{2}\beta, \varphi(u)$ is the density function of the normal normed distribution (if m is small, then we use the "3σ" -rules with $g = 3$).

Denote by N all confidence intervals $I_{ij} = \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right), N = \frac{n(n-1)}{2}$ and by L the number of such I_{ij}, which contain the probability p_{ij}. Put $h = \rho\left(F^*, F^{*'}\right) = \rho\left(x, x'\right) = \frac{L}{N}$. As far as h is a frequency of the random event $B = \{p_{ij} \in I_{ij}\}$ having the probability $p(B) = 1 - \beta$, then setting $h_{ij} = h$, $m = N$ and $g = 3$ in the formulae (3) we obtain the confidence interval $I = \left(p^{(1)}, p^{(2)}\right)$ for the probability $p(B)$, which has the confidence level 0.95. The test of hypothesis H with the significance level, which is approximately equal to 0.05, may be formulated in the following way: if the confidence interval $I = \left(p^{(1)}, p^{(2)}\right)$ contains the probability $p(B) = 1 - \beta$ then the hypothesis H is accepted, otherwise it is rejected. Statistics h is the proximity measure $\rho\left(x, x'\right)$ between samples x and x'.

The problem of mixed sample recognition is more difficult because the mixture of two unimodal distribution may be again an unimodal distribution, when there exist some relations between their mathematical expectations and variances. However, this situation is not typical and, as a rule, after mixing of two unimodal distributions we have bimodal one.

We propose new method to solve the problem of recognition of mixed sample. This method is based on the investigation of the graphics of the variational series of the sample. Let $x = (x_1, x_2, ..., x_n)$ be a sample from the general population G and $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ be its variational series. Consider on the plane the points $\left(i, x_{(i)}\right), i = 1, 2, ..., n$, draw the smooth curve through these points (for example, joining the adjacent points by the linear segments). This curve is called the Quetelet curve (fig. 1a - 6a). It is well-known that the Quetlet curve consists of the three

parts. Its initial part is obtained from the first minimal order statistics, then it have the middle part, which is formed by the points $x_{(i)}$ of the samples from the bulk of the general population G, and its final part formed by the maximal order statistics. In many cases the initial and final parts of the Quetelet curve are the curves formed by the outlyings stipulated by artifacts. That is why the middle part of the Quetelet curve is the most reliable one. If the general population is homogeneous then the middle part of the Quetelet curve is well approximated by the linear segment. Hence, when the middle part of the Quetelet curve is approximated by two segments generating broken line consisting of two links (the first case) or by two disconnected segments (the second case), we deal with the heterogeneous general population. In any case, the right end point of the first segment we shall call *singular point* of the Quetelet curve; in the first case, it is *the point of the break*, and in the second case it is *the point of discontinuity* of the Quetelet curve.

Justification of the proposed test for detection of the heterogeneity is base on the following reasons. If the general population is homogeneous then its density of probability is increasing at first, and the, after mode, it begin to decrease. Hence the distribution function of such general population is convex before mode and concave one after mode, and in the mode neighborhood (point of inflection) this curve is well approximated by the linear segment, so that the bulk of the general population is placed between abscissas of the end points of this segment. As far as the Quetelet curve approximates the inverse function to the distribution function, the graphics of this curve is the reflection of the graphics of the distribution function relatively to the bisectrix of the first and the third quadrant. Hence, in this case the middle part of the Quetelet curve is well approximated by the linear segment. If the density of probability of the general population G is the bimodal curve then the graphics of its distribution function consists of two similar parts: for one part we have the curve changing the convexity on a concavity in the first mode neighborhood and the second one which change the concavity on a convexity in the second mode neighborhood. That is why, the graphics if the distribution function is well approximated by two linear segments. Hence, the Quetelet curve is also well approximated by the linear segments (see, for example, fig. 3a). Note, however, that under calculations it as more preferable to deal with the Quetelet curve, because it is more slanting one in comparison with the distribution function.

The algorithm for the detection and computation of the coordinates of the singular points of the Quetelet curve is the following. At first, on the sample $x = (x_1, x_2, ..., x_n) \in G$ we construct the variational series $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ and remove 10% of its extreme order statistics (5% of the first minimal order statistics and 5% of the maximal ones). Then, using the rest order statistics we construct the curve formed by the increments of the ordinates of the vertecies of the adjacent links of the Quetelet curve (i, y_i), where $y_i = x_{(i+1)} - x_{(i)}$, $i = l, l+1, ..., n-l-1$ connecting the adjacent vertecies $\left(i, y_{(i)}\right)$, $\left(i+1, y_{(i+1)}\right)$ by the linear segments. The values $y_{(i)}$ are called *spacings*, so we call the curve constructed in such way as *the spacing curve*

(fig. 1b-6b). Then, we divide the set of the abscissas and ordinates of the vertecies of the spacing curve on the three equal (or almost equal) parts : U_1, U_2, U_3 and Y_1, Y_2, Y_3 :

$$U_1 = \{1, 2, ..., k\}, U_2 = \{k+1, k+2, ..., m\}, U_3 = \{m+1, m+2, ..., n-l-1\},$$
$$Y_1 = \{y_1, y_2, ..., y_k\}, Y_2 = \{y_{k+1}, y_{k+2}, ..., y_m\}, Y_3 = \{y_{m+1}, y_{m+2}, ..., y_{n-l-1}\},$$

where $k = \left[\frac{n-2l-1}{3}\right]$, $m = n-l-k-2$.

Put $y_{\min}^{(i)} = \min Y_i, y_{\max}^{(i)} = \max Y_i, i = 1, 2, 3$. Then the criteria for the recognition of the heterogeneity of the general population we call formulate in the following way:

1) If $y_{\min}^{(1)} \leq y_{\min}^{(3)}$ and $y_{\max}^{(1)} \geq y_{\max}^{(3)}$, then beginning from the right end point of the graphics of the spacing curve we verify the inequality $y_{\min}^{(3)} \leq y_{n-l-1-j} \leq y_{\max}^{(3)}$; $j = 1, 2, ..., n-l-2$; the first point with the order number n_0 outlying from the interval limits is the singular point and, respectively, the bound between two subpopulations in the general population. In this case the set $W_1 = \{y_i\}$, $i = 1, 2, ..., n_0$ belongs to the first subpopulation and the set s the singular point and, respectively, the bound between two subpopulations in the general population. In this case the set $W_1 = \{y_i\}$, $i = n_0, n_0+1, ..., n-l-1$ belongs to the second one.

2) Otherwise, we begin to sort out the points of the spacing curve from the left end point verifying the inequality $y_{\min}^{(1)} \leq y_j \leq y_{\max}^{(1)}$; $j = 1, 2, ..., n$; similarly the previous case the first point with the order number n_0 outlying from the limits of the interval is the singular point and the set $W_1 = \{y_i\}$, $i = 1, 2, ..., n_0$ belongs to the first subpopulation and the set $W_1 = \{y_i\}$, $i = n_0, n_0+1, ..., n-l-1$ belongs to the second one.

3) The variant when the singular point does not exist is possible also, so that $y_{\min}^{(1)} = y_{\min}^{(3)}$ and $y_{\max}^{(1)} = y_{\max}^{(3)}$. In this case the sample is homogeneous.

Note, that above mentioned method can be applied for detection of the combined samples also.

On the fig. 1-6 the Quetelet curves and spacing curves constructed on the variational series of the samples formed by the areas of the nuclei of the cells of the Guerin's carcinoma of rates under investigation of the generations of the radioiresistance cells are shown. Using of the propose method permitted to detect that the general population of the control (uniradiated)cells is homogeneous and the general populations of the cells in the consequent generations are heterogeneous (combined and mixed).

The results of the investigations of the heterogeneity obtained by the computation of the proximity measure between two halves $X_1 = \{x_1, x_2, ..., x_k\}$ and $X_2 = \{x_k, x_{k+1}, ..., x_n\}$ of the sample $X = \{x_1, x_2, ..., x_n\}$, where $k = \left[\frac{n}{2}\right]$ is the entire part

Generation	$\rho(X_1, X_2)$	LCL	UCL	$\mu(X_1, X_2)$	LCL	UCL
Control	1.000	0.944	1.000	0.000	0.000	0.056
3	0.974	0.950	0.987	0.026	0.013	0.050
5	0.249	0.190	0.318	0.751	0.682	0.810
6	0.998	0.991	0.999	0.002	0.001	0.009
7	0.493	0.419	0.566	0.507	0.434	0.581
12	0.441	0.409	0.475	0.559	0.525	0.591

Table 1: Indecies of the sample heterogeneity.

Generation	NC	SP	NC1	Part1	Min1	Max1	NC2	Part2	Min2	Max2
Control	45	No	45	1.00	55	85	-	-	-	-
3	89	83	83	0.93	30	125	6	0.07	125	225
5	89	44	44	0.49	30	80	45	0.51	100	225
6	135	118	118	0.87	65	125	17	0.13	125	160
7	89	69	69	0.78	30	60	20	0.22	60	90
12	135	110	110	0.82	25	125	25	0.18	125	200

Table 2: Estimations of the population heterogeneity by spacing curve method

of the number $\frac{n}{2}$, are represented in the Table 1. This proximity measure we denote by $\rho(X_1, X_2)$ and introduce the value $\mu(X_1, X_2) = 1 - \rho(X_1, X_2)$. The small values of $\mu(X_1, X_2)$ mean that the sample is combined.

Here LCL and UCL are the lower and upper confidence limits of the proximity measure corresponding to the 5% significance level.

It is readily seen, that the samples of the cell from 5th, 7th and 12th generations are combined one.

Then we investigated the type of the samples by the method of the spacing curves.We computed the singular points of the Quetelet curve (fir.1-6), part of the subpopulations (modal classes) which form the generation and the limits of the variation of the area of the nuclei of the cells from these subpopulations (Table 2).

Notes. In the Table 2 NC is the number of cells in the sample, SP is singular point of the Quetelet curve, NC1 is the number of cells in the first subpopulation, Part1and Part2 are the part of the first and second subpopulation in the sample, Min1 and Max2 are minimal and maximal order statistics of the first and second subpopulation, respetively. We removed 5% of the first minimal order statistics and 5% of the last maximal statistics of the samples . In addition, we do not consider the sample consisting from the cells which number is less then 5% as a separate subpopulation.

Acknolegements. Authors are very grateful to the senior researcher of the Ukrainian Institute of Oncology and Radiology V.A.Zinchenko for kindly submitted experimental data.

References

[1] Cramer H. Mathematical Methods of Statistics. - Stockholm, 1946.

[2] Petunin Yu.I., Semeiko N.G. Random point processes with independent marking // Reports of the Academy of the USSR. - 1986. - V.288. - 4. - P.823-827 (In Russian).

[3] Van der Waerden B.L. Mathematische Statistic. - Springer-Verlag, Berlin, 1957; English. transl. of 2nd (1965) ed. Springer-Verlag, Berlin and New York, 1969.

[4] Madreimov I., Petunin Yu.I. Characterisation of the uniform distribution with the help of the order statistics // Theory of probability and mathematical statistics. - 1982. - V.27. - P. 96-102 (In Russian).

ÖZET

Homojen ve homojen olmayan kitleler kavramları verilmiştir. Homojenliği analiz etmek için Quetlet eğrilerine ve komşu sıra istatistiklerinin farklarına dayalı testler tasarlanmıştır. Sonuçların onkolojide bir uygulaması verilmiştir.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 31 - 42

ASYMPTOTIC BEHAVIOUR OF SOLUTIONS OF STOCHASTIC EQUATIONS AND APPLICATIONS IN STATISTICS

Vladimir V. Anisimov
Bilkent University
Dept. of Industrial Engineering
Bilkent, 06533, Ankara, Turkey

Hussein Salem Kaibah
Kiev University
Dept. of Applied Statistics
Kiev 17, 252017, Ukraine

Abstract

Asymptotic properties of solutions of general stochastic equations (consistency and the behaviour of deviations) are studied. Applications to the analysis of moments method estimators constructed by observations on the trajectories of stochastic systems in stationary and transient conditions are considered.

Key Words: stochastic equation, solution, convergence in probability, weak convergence, moments method.

1. Introduction.

In different models that appear in numerical mathematics, stochastic optimization problems, statistical parameter estimation we come across the necessity to study the behaviour of solutions of stochastic equations.

Let us consider the following examples.

Example 1. Suppose that we would like to find a solution of a deterministic equation

$$f(\theta) = 0, \tag{1}$$

where $f(\theta)$ is some continuous function, $\theta \in \Theta \subset \mathbb{R}^r$, and Θ is some bounded region. But according to the real scheme of calculations we measure the function $f(\theta)$ with random errors in the form:

$$r_k(\theta) = f(\theta) + \xi_k(\theta),\ 1 \le k \le n,$$

where $\{\xi_k(\theta), \theta \in \Theta\}, k \ge 1$ are jointly independent families of random functions (fields) such that $\mathbf{E}\xi_k(\theta) = 0$.

In this case it is reasonable to approximate the function $f(\theta)$ by the averaging

$$f_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} r_k(\theta). \tag{2}$$

Therefore a natural question arises: in what sense and under which conditions a solution of a stochastic equation

$$f_n(\theta) = 0 \tag{3}$$

approximates a solution of the equation (1) as $n \to \infty$.

Example 2. In different statistical schemes, estimators can be represented as solutions of stochastic equations where a corresponding random function is some additive type functional on a trajectory of the observed system.

Consider the following quite general scheme of observations. Let $S(t), t \geq 0$ be the trajectory of some (random or non-random) system with values in $\mathbb{R}^r$ and $t_1 < t_2 < \ldots$ be the times of observations on the interval $[0, T]$. Let also independent parametric families of random variables $\{\gamma_k(\alpha), \alpha \in \mathbb{R}^r\}, k \geq 0$ with values in $\mathbb{R}^r$ and independent on $S(\cdot)$ be given. Suppose for simplicity that distributions of random variables $\gamma_k(\alpha)$ do not depend on index k. Consider various statistical methods.

1. Maximum likelihood method. Consider the case of complete observations. Let densities of random variables $\{\gamma_k(\alpha), \alpha \in \mathbb{R}^r\}$ exist and belong to the parametric family of densities $\{p(z, \theta, \alpha), z \in \mathbb{R}^r, \theta \in \Theta, \alpha \in \mathbb{R}^r\}$ where Θ is some bounded closed region in $\mathbb{R}^r$. We observe variables $s_k = S(t_k)$ and $y_k = \gamma_k(s_k), k \leq \nu(T)$, where $\nu(T)$ is the total number of observations on the interval $[0, T]$.

Then a logarithmic maximum likelihood function $L(\theta, T)$ is represented as an additive random functional on the trajectory of $S(\cdot)$ of the form:

$$L(\theta, T) = T^{-1}\sum_{k=1}^{\nu(T)} \ln p(y_k, \theta, s_k). \tag{4}$$

If derivatives exist then the maximum likelihood estimator is a solution of the vector equation

$$T^{-1}\sum_{k=1}^{\nu(T)} \psi(y_k, \theta, s_k) = 0, \tag{5}$$

where $\psi(y, \theta, s) = \frac{\partial}{\partial\theta}\ln(y, \theta, s)$ (in a vector case the expression $\frac{\partial}{\partial\theta}\ln(y, \theta, s)$ means a vector of partial derivatives).

2. Moments method. Suppose that the 1st moment's functions of variables $\{\gamma_k(\alpha), \alpha \in \mathbb{R}^r\}$ exist and belong to a parametric family of functions $\{g(\theta, \alpha), \theta \in \Theta, \alpha \in \mathbb{R}^r\}$.

Then under the same scheme of observations the moments method estimator is a solution of the equation

$$T^{-1}\sum_{k=1}^{\nu(T)} g(\theta, s_k) = T^{-1}\sum_{k=1}^{\nu(T)} y_k. \tag{6}$$

3. Least squares method. Suppose that the function $f(\theta, \alpha)$ is given and we observe variables s_k and $z_k = f(\theta_0, s_k) + \varepsilon_k$, $0 \le k \le \nu(T)$, where θ_0 is the unknown parameter, and variables ε_k represent a random noise, $E[\varepsilon_k / s_k] = 0$, $E[\varepsilon_k / s_k]^2 < \infty$. Denote

$$F(\theta, T) = T^{-1} \sum_{k=1}^{\nu(T)} (z_k - f(\theta, s_k))^2 \tag{7}$$

Then if derivatives exist a least squares method estimator is a solution of the equation

$$T^{-1} \sum_{k=1}^{\nu(T)} \frac{\partial}{\partial \theta} f(\theta, s_k) \, (z_k - f(\theta, s_k)) = 0. \tag{8}$$

These examples show that the analysis of various type statistical models leads to the necessity to investigate the following classes of problems:

1. Studying the asymptotic properties of solutions of general stochastic equations.
2. The description of classes of stochastic systems for which it is possible to study an asymptotic behaviour of additive type functionals.

In this paper, a new approach in statistical parameter estimation for observations on trajectories of stochastic systems is suggested. At first an estimator is represented as a solution (set of solutions) of some stochastic equation with function which is some additive functional on the trajectory of a system and then, using solutions of both above mentioned problems, an asymptotic behaviour of the estimator itself is studied.

In the paper it is shown how to realize this approach for moments method estimators constructed on the trajectory of a system satisfying some conditions of averaging (or ergodicity).

The paper consists of two main parts. The 1st one deals with studying of asymptotic properties of solutions of stochastic equations. Theorems of the convergence to the limit point (consistency) and the convergence of the normed deviation to some random value which is in general a solution of some limiting stochastic equation are investigated. We mention that a general theorem about the asymptotic behaviour of solutions in Banach space was given by Van der Vaart [7] for the case when a limiting equation has a linear form. In our paper a finite-dimensional case is considered but the limiting equation has a general form (in particular non-linear).

In the 2d part applications to the analysis of moments method estimators constructed by the observations on trajectories of stochastic systems in transient and stationary conditions are studied.

2. Analysis of solutions of stochastic equations

2.1. Asymptotic behaviour of solutions for general stochastic equations

Here we prove some general results devoted to the asymptotic behaviour of solutions of stochastic equations.

At first let us give some necessary notations. For any function $g(\theta)$ and any compact set $K \subset \Theta$ denote by

$$\Delta_U(c, g(\cdot), K) = \sup\{|g(q_1) - g(q_2)| : |q_1 - q_2| \leq c, q_1, q_2 \in K\}$$

the modulus of continuity in a uniform metric for the function $g(\theta)$ on the set K.

Definition 1. The sequence of random functions $f_n(\theta)$ uniformly converges (U-converges) to the function (deterministic or random) $f_0(\theta)$ on the set K if for any $k > 0$ and for any $\theta_1 \in K, \ldots \theta_k \in K$ a multidimensional distribution of a vector $(f_n(\theta_1), \ldots, f_n(\theta_k))$ weakly converges to the distribution of a vector $(f_0(\theta_1), \ldots, f_0(\theta_k))$, and for any $\varepsilon > 0$

$$\lim_{c \to +0} \limsup_{n \to \infty} P\{\Delta_U(c, f_n(\cdot), K) > \varepsilon\} = 0.$$

This means that the sequence of measures generated by the sequence of random functions $f_n(\cdot)$ in Skorokhod space D_K weakly converges to the measure generated by $f_0(\cdot)$.

Definition 2. Let G_n is some sequence of random sets in Θ. We say that G_n converges in probability to g_0 ($G_n \xrightarrow{\mathrm{P}} g_0$), where g_0 is a deterministic or a random point in Θ, if

$$\rho(g_0, G_n) \xrightarrow{\mathrm{P}} 0,$$

where $\rho(a, A) = \sup_{x \in A} |a - x|$.

Definition 3. We say that the function $g(\theta), \theta \in \Theta$ satisfies the condition of separateness **S** if there exists such $\delta > 0$ that for any $y \in \mathbb{R}^r$, $|y| < \delta$ the equation

$$g(\theta) = y$$

has a unique solution and the solution θ_0 of the equation $g(\theta_0) = 0$ is the inner point of the region Θ (if the function $g(\theta)$ is random, it means that the condition of separateness is satisfied with probability one).

Now let $f_n(\theta), t \geq 0, \theta \in \Theta, n > 0$ be a sequence of continuous random functions with values in $\mathbb{R}^r$, where Θ is some bounded region in $\mathbb{R}^r$. Let us consider a stochastic equation

$$f_n(\theta) = 0, \tag{9}$$

and denote the set of all possible solutions by $\{\theta_n\}$.

Theorem 1. 1). Suppose that the sequence of functions $f_n(\theta)$ U-converges in each set $K \subset \Theta$ to the function $f_0(\theta)$ (random or non-random) which satisfies the condition of separateness **S**, and the point θ_0 is the solution of a limiting equation:

$$f_0(\theta_0) = 0. \tag{10}$$

Then with probability which tends to one the solution of the equation (9) exists and the sequence of sets $\{\theta_n\}$ converges in probability to θ_0.

2). Suppose further that θ_0 is a non-random point and there exists $\beta > 0$ and a non-random sequence $v_n \to \infty$ such that for any $L > 0$ the sequence of random functions $v_n^\beta f_n(\theta_0 + v_n^{-1}u)$ U-converges in the region $\{|u| \le L\}$ to some (random) function $\eta_0(u)$, which satisfies the condition **S** and the point κ_0 is the solution of the equation $\eta_0(\kappa_0) = 0$.

Then there exists a solution $\widehat{\theta_n}$ of the equation (9) such that the sequence $v_n(\widehat{\theta_n} - \theta_0)$ weakly converges to the (random) variable κ_0.

Consequence 1. If in conditions of the 2nd part of Theorem 1 the function $\eta_0(u)$ can be represented in the form $\xi_0 + G_0 u$, where ξ_0 and G_0 are vector and matrix-valued (possibly dependent) random variables, and the matrix G_0 is not degenerated with probability one, then there exists a solution $\widehat{\theta_n}$ such that the sequence $v_n(\widehat{\theta_n} - \theta_0)$ weakly converges to the random variable $-G_0^{-1}\xi_0$.

Remark 1 If the solution of limiting equation (10) is not unique and the limiting function $f_0(\theta)$ is deterministic then

$$P \lim_{n\to\infty} \{\theta_n\} \subset \{\theta_0\}$$

where $\{\theta_0\}$ is the set of all possible solutions of the equation (10) and symbol P lim denotes the convergence in probability.

Remark 2. An asymptotic behaviour of solutions of stochastic equations with some applications in statistics is considered in [2]. A generalization of these results will appear in [5].

Proof. The proof of Theorem 1 uses the method of a common probabilistic space developed by Skorokhod [6]. According to this method we can construct our random functions $f_n(\theta,\omega), n \ge 0$ on the same probabilistic space Ω in such a way that $f_n(\theta,\omega)$ uniformly converges for all $\omega \in \Omega_0$ to the function $f_0(\theta,\omega)$ in the region Θ where $\mathbf{P}(\Omega_0) = 1$. Let us fix some $\omega \in \Omega_0$. Denote by $\theta_n(\omega)$ one of solutions of the equation

$$f_n(\theta,\omega) = 0, \tag{11}$$

if the set $\{\theta_n\}$ is not empty. As the region Θ is bounded, we can choose a subsequence $\theta_{n_k}(\omega)$ such that there exists a limit

$$\lim_{k\to\infty} \theta_{n_k}(\omega) = \theta'(\omega).$$

Using the uniform convergence of $f_n(\theta,\omega)$ we obtain that

$$f_n(\theta_{n_k}(\omega),\omega) \to f_o(\theta'(\omega),\omega) = 0.$$

It means that all partial limits of the sequence $\theta_n(\omega)$ belong to the set $\{\theta_0(\omega)\}$ and proves the Remark 1.

Further if the condition **S** of separateness takes place then a solution of the equation (11) exists at large n according to the continuity of $f_n(\theta,\omega)$. Now if $\theta_n(\omega)$ is one of solutions of the equation (11) and it does not tends to $\theta_0(\omega)$, then there exists a subsequence n_k such

that $\theta_{n_k}(\omega) \to \theta'(\omega) \not= \theta_0(\omega)$. Using the uniform convergence of $f_n(\theta, \omega)$ we again obtain that $f_n(\theta_{n_k}, (\omega), \omega) \to f_0(\theta'(\omega)) = 0$. But this equation has a unique solution according to the condition **S**. This contradiction shows that $\theta'(\omega) = \theta_0(\omega)$ and $\theta_n(\omega) \to \theta_0(\omega)$ for all $\omega \in \Omega_0$ and finally proves the 1st part of Theorem 1.

Let us consider the behaviour of deviations. Again using the method of a common probabilistic space we can construct the sequence of functions $f_n(\theta_o + v_n^{-1}u)$ and random variables $\eta_0(u)$ on the same probabilistic space Ω in such a way that $v_n f_n(\theta_o + \alpha_n u, \omega) = \eta_0(u, \omega) + \beta_n(t, u, \omega)$, where for each $L > 0$ and almost for all $\omega \in \Omega$

$$\sup_{|u|<L} |\beta_n(u, \omega)| \to 0. \tag{12}$$

Consider the equation

$$\eta_0(u, \omega) = -\beta_n(u, \omega). \tag{13}$$

Due to the condition **S** and continuity of left and right parts in (13) as $\sup_{|u|<L} |\beta_n(u, \omega)| \leq \delta$, then at least one solution of the equation (13) exists. Denote it by $\widehat{u}_n(\omega)$. As the function $\eta_0(u, \omega)$ has a reversed function $\eta_0^{-1}(u, \omega)$ (at least at small enough u) we can re-write at large n an equation (13) in the form:

$$\widehat{u}_n(\omega) = \eta_0^{-1}(-\beta_n(u, \omega), \omega). \tag{14}$$

According to the relation (12) right part in (14) tends to the value $\eta_0^{-1}(0, \omega) = \gamma_0(\omega)$. This relation proves the 2nd part of Theorem 1.

□

2.2. Asymptotic behaviour of solutions under stochastic errors

In this part we consider as one of possible directions of applications of the results of Theorem 1 the behavior of approximately calculated solutions of deterministic equations under stochastic errors at calculations.

Consider Example 1 from Introduction. Let us study the asymptotic behavior of solutions of the equation

$$f_n(\theta) = 0 \tag{15}$$

where the function $f_n(\theta)$ is defined in (2). As before, denote by $\{\theta_0\}$ the set of possible solutions of the equation (1) and by $\{\theta_n\}$ the set of possible solutions of the equation (15). Denote also

$$\zeta_n(\theta) = \frac{1}{n} \sum_{k=1}^{n} \xi_k(\theta).$$

Theorem 2. Let families of random functions $\{\xi_k(\theta), \theta \in \Theta\}, k \geq 1$ be independent (at different k), identically distributed and the following conditions are satisfied:

1. $\mathbf{E}\xi_1(\theta) = 0$, $\theta \in \Theta \subset \mathbb{R}^r$;
2. for any $\varepsilon > 0$ and any compact set $K \subset \Theta$

$$\lim_{c \to +0} \limsup_{n \to \infty} P\{\Delta_U(c, \zeta_n(\cdot), K) > \varepsilon\} = 0; \tag{16}$$

3. The function $f(\theta)$ satisfies the condition **S** with unique point θ_0.

Then at $n \to \infty$ with probability that tends to one the set $\{\theta_n\}$ is not empty and the sequence $\{\theta_n\}$ converges in probability to θ_0.

Proof. We represent the function $f_n(\theta)$ in the form $f_n(\theta) = f(\theta) + \zeta_n(\theta)$. ¿From the Law of Large Numbers it follows that at each $\theta \in \Theta$

$$P \lim_{n\to\infty} \zeta_n(\theta) = 0, \tag{17}$$

and the condition (16) implies that the sequence of functions $\zeta_n(\theta)$ U-converges to 0 in each compact set K, that is the sequence $f_n(\cdot)$ U-converges to $f(\cdot)$. Then our statement directly follows from Theorem 1.

□

Remark 3. The condition (16) is rather general and in some cases it is not so simple to check it. But using Chebyshev's inequality it can be shown that the following condition is sufficient for (16):

$$\lim_{c\to+0} \mathrm{E}\Delta_U(c, \xi_1(\cdot), K) = 0 \tag{18}$$

for any compact set $K \subset \Theta$.

Remark 4. Suppose that $\xi_1(\theta) = q(\theta, \xi_1)$ where $q(\theta, y)$ is uniformly bounded and continuous in θ at each fixed y function. Then the condition (18) is satisfied.

For this scheme the behavior of deviations $v_n(\theta_n - \theta_0)$ also can be studied.

3. Statistical parameter estimation

Now let us consider applications of Theorem 1 to problems of statistical parameter estimation. For visualization we consider a one-dimensional case ($r = 1$). Let $s_{nk}, k \geq 1$ be a trajectory of some (random or non-random) system with values in $\mathbb{R}$. Let also $\{\gamma_k(\alpha), \alpha \in \mathbb{R}\}, k \geq 0$ be parametric families of random variables with values in $\mathbb{R}$, which are jointly independent and independent on $s_{nk}, k \geq 1$. For simplicity we assume that distributions of random variables $\gamma_k(\alpha)$ do not depend on index k.

Suppose that we observe variables s_{nk} and $y_k = \gamma_k(s_{nk}), k \leq n$, where n is the number of observations.

3.1. Moments method (transient case)

At first we study the behaviour of moment's method estimators in non-stationary conditions. Suppose that first moment's functions of variables $\{\gamma_k(\alpha), \alpha \in \mathbb{R}\}$ exist and belong to the parametric family of functions $\{g(\theta, \alpha),\ \theta \in \Theta,\ \alpha \in \mathbb{R}\}$ and $\mathrm{E}\gamma_1(\alpha) = g(\theta_0, \alpha) = g(\alpha)$ where θ_0 is some inner point in the region Θ.

Then the moments method estimator is a solution of the equation

$$n^{-1} \sum_{k=1}^{n} g(\theta, s_{nk}) - n^{-1} \sum_{k=1}^{n} y_k = 0. \tag{19}$$

Denote as before by $\{\theta_n\}$ the set of possible solutions of the equation (19). We study an asymptotic behavior $\{\theta_n\}$ as $n \to \infty$.

Theorem 3. Suppose that the sequence s_{nk} satisfies the following averaging condition: there exists a deterministic continuous function $s(t)$ on the interval $[0,1]$ such that

$$P \lim_{n\to\infty} \max_{0\le k\le n} |s_{nk} - s(k/n)| = 0, \tag{20}$$

variables $\gamma_k(\alpha)$ satisfy the following condition: for any $L > 0$

$$\lim_{N\to\infty} \sup_{|\alpha|\le L} \mathbf{E}|\gamma_1(\alpha)|\chi\{|\gamma_1(\alpha)| > N\} = 0, \tag{21}$$

the function $g(\theta,\alpha)$ is continuous in both arguments (θ,α) and there exists $\delta > 0$ such that the equation

$$\int_0^1 g(\theta, s(u))\mathrm{d}u - \int_0^1 g(s(u))\mathrm{d}u = v \tag{22}$$

has a unique solution for any $|v| < \delta$.

Then with probability which tends to one a solution of the equation (19) exists and $\{\theta_n\} \to \theta_0$.

Proof. It can be easily seen that under conditions (20),(21) the second term at the left part of (19) converges in probability to the value $-\int_0^1 g(s(u))\mathrm{d}u$. The first term for any $L > 0$ uniformly in $|\theta| \le L$ converges to the function $\int_0^1 g(\theta, s(u))\mathrm{d}u$. And finally our statement follows from Theorem 1.

□

Let us consider now the behaviour of deviations.

Theorem 4. Suppose that in addition to Theorem 3 the following conditions hold:

1. there exists bounded integrable functions $b((\pm 1,\alpha)$ such that for some $\beta > 0$ as $h \to +0$

$$h^{-\beta}(g(\theta_0 + he,\alpha) - g(\theta_0,\alpha)) \to \mathrm{sign}(e)\, b(e,\alpha),\ \alpha \in \mathbb{R}, \tag{23}$$

where $e = \pm 1$ and $b(+1,\alpha)b(-1,\alpha) \ge 0$;

2. there exists γ, $1 < \gamma \le 2$, such that at each α

$$\mathbf{E}\exp\{i\lambda(\gamma_1(\alpha) - g(\alpha))\} = 1 - |\lambda|^\gamma a(\lambda,\alpha) + o(|\lambda|^\gamma,\alpha), \tag{24}$$

where $a(\lambda,\alpha) = c(\alpha)(1 - i\beta(\alpha)\mathrm{sign}(\lambda)\tan(\pi\gamma/2))$, $c(\alpha) > 0, \beta(\alpha) \in [-1,1]$,
and for any $L > 0$

$$\lim_{\lambda\to 0} \sup_{|\alpha|<L} |\lambda|^{-\gamma} o(|\lambda|^\gamma,\alpha) \to 0;$$

Then there exists a solution $\widehat{\theta}_n$ of the equation (19) such that

$$n^{\frac{\gamma-1}{\beta\gamma}}(\widehat{\theta}_n - \theta_0) \stackrel{\mathrm{w}}{\Rightarrow} \gamma_0, \tag{25}$$

where symbol $\stackrel{w}{\Rightarrow}$ denotes a weak convergence (in distribution),

$$\gamma_0 = |\frac{\zeta}{\widehat{b}(\delta(\zeta))}|^{1/\beta}\,\delta(\zeta), \tag{26}$$

$$\delta(\zeta) = \mathrm{sign}(-\frac{\zeta}{\widehat{b}(+1)}),\ \widehat{b}(e) = \int_0^1 b(e, s(v))\mathrm{d}v,\ e = \pm 1, \tag{27}$$

and the value ζ has a stable distribution with characteristic function

$$\mathbf{E}\exp\{i\lambda\zeta\} = \exp\{-|\lambda|^\gamma \int_0^1 a(\lambda, s(v))\mathrm{d}v\}. \tag{28}$$

Proof. Denote by $f_n(\theta)$ the left part of (19). Put $v_n = n^{\frac{\gamma-1}{\beta\gamma}}$. Then we can write a representation

$$v_n^\beta f_n(\theta_0 + v_n^{-1}u) = n^{-1}\sum_{k=1}^n v_n^\beta(g(\theta_0 + v_n^{-1}u, s_{nk}) - g(\theta_0, s_{nk}))- \tag{29}$$

1 $\mathrm{n}^{-1/\gamma}\sum_{k=1}^n(\gamma_k(s_{nk}) - g(\theta_0, s_{nk}))$.1

It is not hard to prove using conditions (20),(24) that the second term at the right part of (29) weakly converges to the variable ζ (see (28)).

Condition (23) implies that

$$v_n^\beta(g(\theta_0 + v_n^{-1}u, \alpha) - g(\theta_0, \alpha)) \approx \mathrm{sign}(u)\,|u|^\beta b(\mathrm{sign}(u), \alpha).$$

According to it the first term at the right part of (29) can be represented in the form

$$n^{-1}\sum_{k=1}^n \mathrm{sign}(u)\,|u|^\beta b(\mathrm{sign}(u), s_{nk}) + o(1),$$

and this term U-converges in any bounded region $\{|u| \le L\}$ to the value $\mathrm{sign}(u)\,|u|^\beta\widehat{b}(\mathrm{sign}(u))$. Finally we get a limiting equation in the form

$$\mathrm{sign}(u)\,|u|^\beta\widehat{b}(\mathrm{sign}(u)) + \zeta = 0,$$

and a solution of it exists and can be written in the form (26).

□

Consequence 2. Suppose that conditions of Theorem 4 hold and there exists a continuous in both arguments derivative $R(\theta, \alpha) = \frac{\partial}{\partial\theta}g(\theta, \alpha)$ and a continuous variance $\sigma^2(\alpha) = \mathbf{E}(\gamma_1(\alpha) - g(\alpha))^2$. Denote

$$\widehat{R}(\theta_0) = \int_0^1 R(\theta_0, s(v))\mathrm{d}v,\ \widehat{\sigma}^2 = \int_0^1 \sigma^2(s(v))\mathrm{d}v. \tag{30}$$

Suppose that $\widehat{R}(\theta_0) > 0$ and variables $\gamma_k(\alpha)$ satisfy Lindeberg condition: for any $L > 0$

$$\lim_{N\to\infty}\sup_{|\alpha|\le L}\mathbf{E}\gamma_1(\alpha)^2\chi\{|\gamma_1(\alpha)| > N\} = 0. \tag{31}$$

Then there exists a solution $\widehat{\theta}_n$ of the equation (19) such that the sequence $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ weakly converges to a gaussian distribution with mean 0 and variance $\widehat{R}^{-2}\widehat{\sigma}^2$.

Proof. We put $v_n = \sqrt{n}$, $\beta = 1$. Then it can be easily seen using conditions (20),(31) that the second term in the right part of (29) weakly converges to the gaussian distribution with mean 0 and variance $\widehat{\sigma}^2$. The first term can be represented in the form

$$n^{-1}\sum_{k=1}^{n} R(\theta_0 + n^{-1/2}q_{nk}u, s_{nk})\ u$$

where $|q_{nk}| \le 1, k \ge 0$, and this term U-converges in any bounded region $\{|u| \le L\}$ to the value $u\int_0^1 R(\theta_0, s(v))dv = u\widehat{R}(\theta_0)$. It means that the limiting equation can be written in the form:

$$u\widehat{R}(\theta_0) + \widehat{\sigma}\mathcal{N}(0,1) = 0$$

where $\mathcal{N}(0,1)$ denotes a standard gaussian random variable, which proves our statement. □

Example 3. Suppose that

$$g(\theta,\alpha) = \begin{cases} g(\alpha) - b(-1,\alpha)|\theta - \theta_0|^\beta & \text{as } \theta \le \theta_0, \\ g(\alpha) + b(1,\alpha)|\theta - \theta_0|^\beta & \text{as } \theta > \theta_0, \end{cases}$$

where $\beta > 0$ and $b(e,\alpha), g(\alpha)$ are some continuous functions, $b(\pm 1,\alpha) \ge 0$.

Suppose also that for any α $\mathbf{Var}\gamma_1(\alpha) = \sigma^2(\alpha) < C < \infty$. We introduce values $\widehat{b}(e)$ and $\widehat{\sigma}^2$ according to notations (27),(30).

The if condition (20) is true, the equation (22) has a form:

$$\mathrm{sign}(\theta - \theta_0)\widehat{b}(\mathrm{sign}(\theta - \theta_0))|\theta - \theta_0|^\beta = v,$$

and obviously has a unique solution if $\widehat{b}(e) > 0, e = \pm 1$.

Further the relation (23) has the same form, in (24) $\gamma = 2$, $a(\lambda,\alpha) = \sigma^2(\alpha)/2$, the variable ζ can be written in the form $\zeta = \widehat{\sigma}\mathcal{N}(0,1)$ and the value γ_0 in (25) can be represented in the form

$$\gamma_0 = \left|\frac{\widehat{\sigma}\mathcal{N}(0,1)}{\widehat{b}(\mathrm{sign}(\mathcal{N}(0,1))}\right|^{1/\beta}\mathrm{sign}(\mathcal{N}(0,1)).$$

If functions $g(\theta,\alpha)$ are differentiable the value γ_0 has a gaussian distribution.

Remark 5. We mention that the condition (20) is satisfied for wide classes of stochastic systems that have a recurrent character of developing (for instance special classes of Markov systems) and it is mostly oriented on non-stationary (transient) conditions. An average principle for rather general stochastic recurrent sequences in transient conditions is given in Anisimov (1991,1995). That gives the possibility of applications to statistical parameter estimation for these classes of random sequences.

2.2. Moments method (stationary case)

Analogous results can be obtained when the sequence s_{nk} is in some sense ergodic. Suppose that the following condition holds:

there exists a probability measure $\pi(A)$, $A \in B_{\mathbb{R}}$ such that for any bounded measurable function $\varphi(\alpha)$, $\alpha \in \mathbb{R}$

$$P \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \varphi(s_{nk}) = \int_{\mathbb{R}} \varphi(\alpha)\pi(\mathrm{d}\alpha) \tag{32}$$

(for instance s_{nk} can be a Markov ergodic sequence).

Denote $\widehat{g}(\theta) = \int_{\mathbb{R}} g(\theta,\alpha)\pi(\mathrm{d}\alpha)$.

Theorem 5. Suppose that conditions (21),(32) are satisfied, the function $g(\theta,\alpha)$ is continuous in both arguments, $\sup_{\theta,\alpha} |g(\theta,\alpha)| < C$, and there exists $\delta > 0$ such that the equation

$$\widehat{g}(\theta) - \widehat{g}(\theta_0) = v$$

has a unique solution for any $|v| < \delta$.

Then with probability that tends to 1 a solution of the equation (19) exists and $\{\theta_n\} \xrightarrow{\mathrm{P}} \theta_0$.

Proof. Under our conditions the second term at the left part of (19) converges in probability to the value $\widehat{g}(\theta_0)$. According to Theorem 1 we need to prove that the first term for any $L > 0$ uniformly converges in $|\theta| \le L$ to the function $\widehat{g}(\theta)$. Denote

$$\delta(c,L,N) = \sup\{\ |g(\theta_1,\alpha) - g(\theta_2,\alpha)| : |\theta_1 - \theta_2| < c,\ \theta_1 \vee \theta_2 \le L,\ |\alpha| \le N\ \}.$$

According to continuity of $g(\theta,\alpha)$ $\delta(c,L,N) \to 0$ at any fixed L, N and $c \to +0$. Then

$$P \limsup_{n\to\infty}\ n^{-1} \sup_{|\theta_1-\theta_2|<c,\theta_1\vee\theta_2\le L} |\ \sum_{k=1}^{n} (g(\theta_1,s_{nk}) - g(\theta_2,s_{nk})\ | \le$$

$\mathrm{Plim\,sup}_{n\to\infty} \left(\ \delta(c,L,N)\ n^{-1}\ \sum_{k=1}^{n}\ \chi\{|s_{nk}| \le N\} + 2C\ n^{-1}\ \sum_{k=1}^{n} \chi\{|s_{nk}| > N\ \}\ \right) \le$
$\delta(c,L,N) + 2C\pi(z:\ |z| \ge N)$ (33)
and the right part of (33) tends to 0 as $c \to +0$ and then $N \to \infty$. Finally our statement follows from Theorem 1.

□

Let us consider now the behaviour of deviations.

Theorem 6. Suppose that conditions of Theorem 5 hold, the condition 1 of Theorem 4 takes place, the condition 2 of Theorem 4 holds where the function $a(\lambda,\alpha)$ is bounded at each fixed λ.

Then the relation (25) takes place, γ_0 is given in (26) where for our case

$$\widehat{b}(e) = \int_{\mathbb{R}} b(e,\alpha)\pi(\mathrm{d}\alpha), \tag{34}$$

and vector ζ has a stable distribution with characteristic function

$$\mathrm{E}\exp\{i\lambda\zeta)\} = \exp\{-|\lambda|^{\gamma}\widehat{a}(\lambda)\}, \tag{35}$$

where $\widehat{a}(\lambda) = \int_{\mathbf{R}} a(\lambda,\alpha)\pi(\mathrm{d}\alpha)$.

Proof. The proof follows the same steps as in Theorem 4. Put $v_n = n^{\frac{\gamma-1}{\beta\gamma}}$. Then under our conditions it can be proved that the second term at the right part of (29) weakly converges to the variable $\widehat{\zeta}$ (see (35)).

The first term U-converges in any bounded region $\{|u| \le L\}$ to the value $\mathrm{sign}(u)\,|u|^{\beta}\widehat{b}(\mathrm{sign}(u))$ where the function $\widehat{b}(e)$ is given in (34). According to Theorem 1 this implies our statement.

□

Analogous results can be given for maximum likelihood and least squares methods estimators.

We mention that asymptotic properties of parameter estimators for switched by some ergodic sequence Poisson type processes and asymptotic properties of maximum likelihood estimators constructed by observations on trajectories of recurrent processes of semi-Markov type on the base of the same technique (analysis of maximum likelihood equation) are studied in Anisimov (1991) and Anisimov and Orazklychev (1993).

References

[1] Anisimov, V.V. (1985) On estimation the parameters of switched Poisson processes. *Theory Probability & Math. Statist.*, 31, 1–11.

[2] Anisimov, V.V. (1985) Asymptotic behaviour of solutions of some classes of stochastic equations and applications to statistical problems. *Numerical and Applied Math.*, 57, 103–112 (Russian).

[3] Anisimov, V.V. (1991) Averaging principle for switching recurrent sequences. *Theory Probability & Math. Statist.*, 45, 1–8.

[4] Anisimov, V.V. and Orazklychev, A.O. (1993) Asymptotic parameter estimation of recurrent processes of semi-Markov type. *Theory Probability & Math. Statist.*, 49, 1–13.

[5] Anisimov, V.V. and Pflug, G., Z-theorems: Limits of stochastic equations, (forthcoming).

[6] Skorokhod, A.V. (1956) Limit theorems for random processes, *Theory Probability & Applications*, 1, 3, 289–319.

[7] Van der Vaart, A.W. (1995) Efficiency of infinite dimensional M-estimators. *Statistica Neerlandica*, 49, 1, 9–30.

ÖZET

Bu çalışmada genel rassal denklem çözümlerının asimptotik özellikleri (tutarlılığı ve sapmaların davranışı) incelendi. Bunlardan elde edilen sonuçlar, durağan ve geçici durumlarda rassal sistemlerin yörüngelerindeki gözlemlerle elde edilen momentler yöntemi kestiricilerinin analizine uygulandı.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 43 - 46

LIMIT BEHAVIOR OF THE TIMES OF ONE-SIDED SUCCESSIVE APPROXIMATIONS

A. Stepanov
Math. Department
Kaliningrad State Technical University,
Kaliningrad, Russia

Abstract

Let $X_1, X_2, ...$be independent random variables with a common continuous distribution function $F(x)$. Let a is an arbitrary point in the support of the probability measure. The attention has been focused on random indices of those X's which tend to a from the left. We study the limiting behavior of such random indices.

Keywords: Limit theorems, times and values of one-sided successive approximation, records.

1. Introduction

Let $X_1, X_2, ...$ be independent random variables (r.v.'s) with a common continuous distribution function $F(x)$. Let $I = (\gamma, \beta)$, $-\infty \leq \gamma < \beta \leq \infty$, be a support of the probability measure. Consider the following sequences of r.v.'s- the sequence of values of one-sided successive approximations (o.-s.s.a's) $V^a_{(n)}$ and the sequence of times of the o.-s.s.a's $T^a(n)$. Let

$$T^a(0) = 0, \quad X_0 = V^a_{(0)} = \gamma,$$
$$T^a(n+1) = \min\left\{j : j > T^a(n), \ X_{T^a(n)} < X_j < a\right\},$$
$$V^a_{(n)} = X_{T^a(n)} \ \ (n = 0, 1, ...),$$

where $a \in I$. If $a = \beta$ then the values of o.-s.s.a's $V^a_{(n)}$ coincide with the well-known record values $X_{(n)}$ and the times of o.-s.s.a's $T^a(n)$ turn out to be the record times $L(n)$. Many papers have been devoted to the records (see, for example, [2]).

Some results for $V^a_{(n)}$ and $T^a(n)$ have been obtained in [3] and [4]. It was shown that the behavior of the r.v.'s $V^a_{(n)}$ taken from a sequence of independent X's with a common continuous distribution function $F(x)$ is the same as the behavior of the

record values $X_{(n)}$ taken from a sequence of independent $\widetilde{X}$'s with a common continuous distribution function $F(x)/F(a)$. As a result all facts known for the record values could be reproduced for $V^a_{(n)}$ for free. Thus, in particular,

$$P\left\{V^a_{(n)} < x\right\} = \frac{1}{(n-1)!} \int_0^{-\log(1-F(x)/F(a))} v^{n-1}e^{-v}dv.$$

On the contrary the behavior of $T^a(n)$ and $L(n)$ is far from similarity. One of the basic facts known for record times is that the distribution of $L(n)$ does not depend on the shape of the initial continuous distribution function $F(x)$. In contrast with this we have for the times of the o.-s.s.a's [3]

$$P\{T^a(1) = k\} = (1-F(a))^{k-1}F(a),$$
$$P\{T^a(1) = k_1, ..., T^a(n-1) = k_{n-1}, T^a(n) \geq k_n\} =$$

$$= \sum_{i=1}^{n} \frac{(1-F(a))^{k_i-1}}{(k_1-k_i)...(k_{i-1}-k_i)(k_{i+1}-k_i)...(k_n-k_i)}.$$

In this paper we study the limit behavior of $T^a(n)$ $(n \to \infty)$ under some conditions.

2. Results

Theorem 1. Let $x(n) = \frac{n-kF(a)}{\sqrt{kF(a)(1-F(a))}}$ and $x(n) \to \infty$, $\frac{x(n)}{\sqrt{k}} \to 0$ as $k > n$ and $n \to \infty$. Then

$$P\{T^a(n) = k\} \sim \frac{\log^{n-2}n}{\sqrt{2\pi}k(n-1)!x(n)}.e^{-\frac{x^2}{2}}$$

as $n \to \infty$.

Let us now we have a sequence of points a_n instead of one single point. Let $a_n \to \beta$ in such manner that

$$\lim_{n\to\infty} n(1-F(a_n)) = \lambda \tag{1}$$

$(\lambda > 0)$. Consider the Poisson's scheme of independent trials with a common continuous distribution function $F(x)$. The first series runs till $V^{a_1}_{(1)}$ appears, the second one runs till $V^{a_2}_{(2)}$ appears and e.t.c.

Theorem 2. For r.v. $\Delta(n) = T^{a_n}(n) - L(n)$ the following result is valid

$$P\{\Delta(n) = k\} \to \frac{\lambda^k}{k!}e^{-\lambda}$$

where $k = 0, 1, ...$ and $n \to \infty$.

3. Proofs

Proof of the Theorem 1. M. Westcott [5] has discovered that

$$P\{L(n)=m\} \sim \frac{\log^{n-2} m}{(n-2)!m^2} \tag{2}$$

($m \geq n$, $n \to \infty$). It was shown in [3] that

$$P\{T^a(n)=k \mid L(n)=m\} = C_{k-1}^{k-m} F^m(a)(1-F(a))^{k-m}. \tag{3}$$

Taking into consideration (2),(3) and the formula of total probability we get

$$P\{T^a(n)=k\} = \frac{1}{k(n-2)!} \sum_{m=n}^{k} C_k^m F^m(a)(1-F(a))^{k-m} . \frac{\log^{n-2} m}{m}. \tag{4}$$

Let $x(m) = \frac{m-kF(a)}{\sqrt{kF(a)(1-F(a))}}$. Using the techniques of the De Moivre - Laplace theorem (see [1]) we can approximate our sum by an integral. Under the conditions

$$x(m) \to \infty, \quad \frac{x^3(m)}{\sqrt{k}} \to 0$$

the equality (4) can be rewritten as the expression

$$P\{T^a(n)=k\} \sim$$

$$\sim \frac{1}{\sqrt{2\pi}k(n-2)!} \int\limits_{\frac{n-kF(a)}{\sqrt{kF(a)(1-F(a))}}}^{\infty} e^{-\frac{x^2}{2}} \frac{\log^{n-2}(kF(a)+x\sqrt{kF(a)(1-F(a))}}{kF(a)+x\sqrt{kF(a)(1-F(a))}} dx$$

Using the substitution

$$y = kF(a) + x\sqrt{kF(a)(1-F(a))}$$

we get

$$P\{T^a(n)=k\} \sim$$

$$\sim \frac{1}{\sqrt{2\pi}k(n-2)!\sqrt{kF(a)(1-F(a))}} \int_n^{\infty} \exp\left(-\frac{(y-kF(a))^2}{2kF(a)(1-F(a))}\right) \frac{\log^{n-2} y}{y} dy.$$

If we are interested in funding the basic term of the expression we can reduce it to

$$P\{T^a(n)=k\} \sim \frac{\log^{n-2} n}{\sqrt{2\pi}k(n-1)!x(n)} e^{-\frac{x^2}{2}}.$$

Proof of the Theorem 2. It is true that

$$P\{\Delta(n)=k\}=\sum_{m=n}^{\infty} C_{m+k-1}^{k} F^{m}(a_m)(1-F(a_m))^{k} P\{L(n)=m\}$$

we focus our attention on the term

$$A_m=(m+k-1)...(m+1).m.(1-F(a_m))^{k}.$$

According to (1)

$$A_m \to \lambda^{k} \quad (m\to\infty) \quad \text{and} \quad F^{m}(a_m)\to e^{-\lambda}.$$

It follows that

$$P\{\Delta(n)=k\}\to\sum_{m=n}^{\infty}\frac{\lambda^{k}e^{-\lambda}}{k!}P\{L(n)=m\}=\frac{\lambda^{k}e^{-\lambda}}{k!}.$$

References

[1] Feller, W. (1967) An introduction to probability theory and its applications. Vol 1. New York: John Wiley & Sons, Inc.,London: Chapman & Hall, Limited, 498 pp.

[2] Nevzorov, V. B. (1987) Records. *Theory Probab. Appl.*, V. 32, p.201-208.

[3] Stepanov, A.V. (1996) Extreme order statistics under changes of ordering relation. *Theory Probab. Appl.*, V. 41, p.896-900.

[4]Stepanov, A.V. (1998) Limit laws for times of one-sided sequential approximations. In Abs. of communications of Int. conf. "Asymptotic methods in probab. and math stat." dedicated to the 50-th Ann.of the chair of probab. and stat. In St. Petersburg university.

[5] Westcott M. (1987) A note on record times. *J. Appl. Probab.* V. 14. p.637-639.

ÖZET

$X_1, X_2, ...$ birbirinden bağımsız ve aynı sürekli $F(x)$ dağılım fonksiyonuna sahip rasgele değişkenler ve a, olasılık ölçüsünün tanım uzayında keyfi bir nokta olsun. X'lerin rasgele indislerinin a'ya soldan yaklaştığı durum dikkate alınarak, bunların limitsel davranışları araştırılmıştır.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 47 - 53

THE DISCRETE FOURIER TRANSFORM APPROXIMATION FOR PERIODICALLY CORRELATED TIME SERIES

Yilmaz Akdi
Ankara University, Department of Statistics
06100 Tandogan-Ankara
e-mail : akdi@science.ankara.edu.tr

Abstract

The discrete Fourier transform is used in many statistical inference problems. An approximation of the dicrete Fourier transform becomes very important especially in approximating the spectral density function and deriving some distributional properties. In this study, it is shown that a similar approximation derived for stationary time series is also valid for periodically correlated series.

Key Words: Discrete Fourier Transform, Spectral Density Function

1. Introduction

There are occasions when a pair of real-valued time series can most generally be regarded as a single complex-valued series. In general we are faced with strictly real-valued data. These can always be regarded as complex numbers with zero imaginary parts, although this may seem an unnecessary. technicality. However, certain algebratic simplifications that arise make the required stretch of the imagination interestingly appealing.

Many problems of statistical inference for time series are based on the frequency-domain properties of the series. Spectral analysis for time series and in particular the estimation of the spectral density function, depends heavily on the asymptotic distribution as $n \to \infty$ of the periodogram ordinates which is defined in terms of the discrete Fourier transform of the series $\{X_0, X_1, ..., X_{n-1}\}$. Given a set of observations $\{X_0, X_1, ..., X_{n-1}\}$ the discrete Fourier transform of the series is defined by

$$d_X(f) = \frac{1}{n} \sum_{t=0}^{n-1} X_t e^{-2\pi i f t} \tag{1}$$

where $f = 0, 1/n, 2/n, ..., (n-1)/n$ (known as fundamental frequencies or Fourier frequencies and) $i = \sqrt{-1}$.

In this paper, we study the asymptotic behavior of $d_X(f)$ for periodically correlated time series with a known period τ. A stochastic process $\{X_t : t \in T\}$ defined on a probability space $(\Omega, \Im, P)$ and having finite second moments for all t is called second order stationary (sometimes covariance stationary, or simply stationary), if the following conditions hold:

1. $E(X_t) = \mu$ for all t (that is, the mean is constant in time)
2. $\gamma(h) = Cov(X_t, X_{t+\tau})$ for all t

and the process is called periodically correlated (PC) with period τ, if

1a. $E(X_t) = E(X_{t+\tau})$ for all t

2a. $Cov(X_t, X_s) = Cov(X_{t+\tau}, X_{s+\tau})$ for all t and s.

Without any loss of generality, we can assume $E(X_t) = 0$. Then (2a) reduces to $E(X_t X_s^*) = E(X_{t+\tau} X_{s+\tau}^*)$. Here X^* denotes the complex conjugate of X. In the above definition, if T is taken to be integers, then the process is called periodically correlated random sequence, and if T is taken to be all the real numbers, then $\{X_t : t \in T\}$ is called continuous time periodically correlated stochastic process. The periodically correlated processes are also known as periodically non stationary processes, cyclostationary, or sometimes periodically stationary time series. Some of the properties of the PC processes have been discussed extensively by many author. Gladyshev (1961) gives the necessary and sufficient conditions for a function to be a correlation function of some periodically correlated series and provides a representation of a PC process in terms of a stationary time series. This is a very useful representation between the PC processes and stationary time series and this will be the key result in our study. Hurd (1989) gives a representation of a continuous time strongly harmonizable PC process and their covariances. Troutman (1979) considers a representation of a PC process as an infinite linear combination of independent, periodically distributed random variables. Jones and Brelsford (1967) give a method of prediction of time series with periodic structure. Bloomfield, Hurd and Lund (1992) look at the periodic correlation in the stratospheric ozone data. Brockwell and Davis (1987) give a good summary of the discrete Fourier transform of a time series in the context of Hilbert spaces and present some of its consistency properties. For a stationary time series a powerful identity is used in many statistical inference problems. Now, we will consider a stationary time series and review some of the properties of the identity mentioned above.

For an illustration assume that the set of random variables $\{X_0, X_1, ...X_{n-1}\}$ come from a stationary first order auto regressive time series $X_t = \varphi X_{t-1} + \epsilon_t$ with $|\varphi| < 1$ and that ϵ_t's are independent and identically distributed random variables with mean 0 and variance σ^2. When we write the discrete Fourier transform for this series

$$d_X(f) = \frac{1}{n}\sum_{t=0}^{n-1} X_t e^{-2\pi i f t} = \frac{1}{n}\sum_{t=0}^{n-1}(\varphi X_{t-1} + \epsilon_t)e^{-2\pi i f t}$$

$$= \varphi \sum_{t=0}^{n-1} X_{t-1} e^{-2\pi i f t} + d_\epsilon(f) = \varphi e^{-2\pi i f} d_X(f) + d_\epsilon(f) - \varphi X_{n-1} e^{-2\pi i f n}$$

This implies that when f is taken to be $0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}$ (known as Fourier frequencies or harmonic frequencies) and using the Euler's identity ($e^{ix} = \cos(x) + i\sin(x)$) we have

$$(1 - \varphi e^{-2\pi i f}) d_X(f) = d_\epsilon(f) - \varphi \frac{1}{n} X_{n-1}$$

and

$$\sqrt{n} d_X(f) = \sqrt{n} \frac{1}{1 - \varphi e^{-2\pi i f}} d_\epsilon(f) - \sqrt{n} \frac{1}{n} \frac{1}{1 - \varphi e^{-2\pi i f}} X_{n-1}$$

This can be written as

$$\sqrt{n} d_X(f) = \sqrt{n} a(f) d_\epsilon(f) + R(n, f)$$

where $a(f) = \frac{1}{1-\varphi e^{-2\pi i f}}$ and $R(n, f) = -\frac{1}{\sqrt{n}} \frac{1}{1-\varphi e^{-2\pi i f}} X_{n-1}$. Using the stationarity of X_t, we can show that $R(n, f) \xrightarrow{P} 0$ as $n \to \infty$ and $\sqrt{n} a(f) d_\epsilon(f) = O_P(1)$.

It can easily be shown that for any stationary process we have the following identity holds:

$$\sqrt{n} d_X(f) = \sqrt{n} a(f) d_\epsilon(f) + R(n, f) \qquad (2)$$

where $R(n, f) \xrightarrow{P} 0$ as $n \to \infty$. The periodogram ordinates of the series is defined in terms of the discrete Fourier transform as $I(f) = n|d_X(f)|^2$ and the spectral density function is the expected value of the periodogram ordinates. When ϵ_t's is a sequence of independent and identically distributed random variables with mean zero and variance σ^2, we have $E(n|d_\epsilon(f)|^2) = \sigma^2$ and the spectral density function of the series $\{X_t : t = 0, 1, 2, ..., n-1\}$ can be approximated by $S(f) = \sigma^2 |a(f)|^2$. The spectral density function of stationary first order auto regressive time series can be seen to be $S(f) = \frac{\sigma^2}{1+\varphi^2+2\varphi\cos(2\pi f)}$. Thus the discrete Fourier transform of the series $\{X_t : t = 0, 1, 2, ..., n-1\}$ can be written as $d_X(f) \cong \sqrt{S(f)} d_\epsilon(f)$ where $d_\epsilon(f)$ is the discrete Fourier transform of the white noise sequence and the periodogram ordinate of the series is approximated as $I_X(f) \cong S_X(f) I_\epsilon(f)$. Here $S(f)$ is the spectral density function of the series $\{X_t : t = 0, 1, 2, ..., n-1\}$ which is sometimes written as $S(f) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i f h}$ where $\gamma(h) = Cov(X_t, X_{t+h})$ is the autocovariance function of the series. Our purpose is to derive an identity similar to (2) for periodically correlated time series.

2. Periodically Correlated Series

As already mentioned, there is a useful representation between the periodically correlated series and stationary series. Gladyshev (1961) gives such a representation as

$$X_t = \sum_{k=0}^{\tau-1} z_t^k e^{2\pi ikt/\tau} \tag{3}$$

where z_t^k is the kth component of a τ-dimensional stationary (in the wide sense) vector process $Z_t = (z_t^0, z_t^1, ..., z_t^{\tau-1})'$. Obviously, if a vector process is known to be stationary, then each component is marginally stationary. Therefore, for each k, $(k = 0, 1, 2, ..., \tau - 1)$ z_t^k is a univariate stationary time series and we can apply (2) to z_t^k

$$\sqrt{n} d_{z^k}(f) = \sqrt{n} a_k(f) d_{\epsilon^k}(f) + R^k(n, f) \tag{4}$$

where $R^k(f, n) \xrightarrow{P} 0$ as $n \to \infty$. Using the equations (1) and (3), we obtain

$$d_X(f) = \frac{1}{n} \sum_{t=0}^{n-1} X_t e^{-2\pi ift} = \frac{1}{n} \sum_{t=0}^{n-1} \sum_{k=0}^{\tau-1} z_t^k e^{2\pi ifkt/\tau} e^{-2\pi ift}$$

$$d_X(f) = \sum_{k=0}^{\tau-1} \frac{1}{n} \sum_{t=0}^{n-1} z_t^k e^{-2\pi it(f-k/\tau)} = \sum_{k=0}^{\tau-1} d_{z^k}(f - k/\tau) \tag{5}$$

for each k $(k = 0, 1, 2, ..., \tau - 1)$, $d_{z^k}(f - k/\tau)$ satisfies the identity (2), and thus from (5) we get

$$\sqrt{n} d_X(f) = \sqrt{n} \sum_{k=0}^{\tau-1} d_{z^k}(f - k/\tau)$$

$$= \sqrt{n} \sum_{k=0}^{\tau-1} a(f - k/\tau) d_{\epsilon^k}(f - k/\tau) + \sum_{k=0}^{\tau-1} R^k(n, f - k/\tau) \tag{6}$$

In the final equation $R^k(n, f - k/\tau) \xrightarrow{P} 0$ as $n \to \infty$ without depending on k. Since the summand ranges from 0 to $\tau - 1$, the summation will go to zero in probability because the period τ does not depend on n. Thus we can give the following theorem which give the identity we seek:

Theorem: If the time series $\{X_t : t = 0, 1, 2, ..., n - 1\}$ is periodically correlated time series with a known period τ, then

$$\sqrt{n} d_X(f) = \sqrt{n} \sum_{k=0}^{\tau-1} a(f - k/\tau) d_{\epsilon^k}(f - k/\tau) + \sum_{k=0}^{\tau-1} R^k(f - k/\tau) \tag{7}$$

where ϵ^k is a white noise sequence corresponding to the kth component of the τ-dimensional stationary vector time series, and $\sum_{k=0}^{\tau-1} R^k(f - k/\tau) \xrightarrow{P} 0$ as $n \to \infty$

3. Examples

1. Assume that a set of random variables $\{Y_0, Y_1, ..., Y_{n-1}\}$ form a first order moving average time series $Y_t = \theta\epsilon_{t-1} + \epsilon_t$, where ϵ_t is a sequence of independent and normally distributed random variables with mean zero and variance σ^2. The autocovariance function of this moving-average series is

$$\gamma_Y(h) = \begin{cases} \sigma^2(1+\theta)^2 & h = 0 \\ \sigma^2\theta & h = \pm 1 \\ 0 & \text{elsewhere} \end{cases}$$

The spectral density function can be calculated as

$$S_Y(f) = \sum_{h=-\infty}^{\infty} \gamma(h)e^{-2\pi ifh} = \sigma^2(\theta e^{2\pi if} + 1 + \theta^2 + \theta e^{-2\pi if})$$

$$= \sigma^2\left(1 + \theta^2 + 2\theta\frac{e^{2\pi if} + e^{-2\pi if}}{2}\right) = \sigma^2(1 + \theta^2 + 2\theta\cos(2\pi f))$$

We can find the spectral density function by using the discrete Fourier transform of the series. The discrete Fourier transform is given by

$$d_Y(f) = \frac{1}{n}\sum_{t=0}^{n-1} Y_t e^{-2\pi ift} = \frac{1}{n}\sum_{t=0}^{n-1}(\epsilon_t + \theta\epsilon_{t-1})e^{-2\pi ift} = d_Y(f) + \frac{\theta}{n}\sum_{t=0}^{n-1}\epsilon_{t-1}e^{-2\pi ift}$$

$$= d_\epsilon(f) + \theta e^{-2\pi if}\frac{1}{n}\sum_{t=0}^{n-1}\epsilon_t e^{-2\pi ift} - \frac{\theta}{n}\epsilon_{n-1} = (1 + \theta e^{-2\pi if})d_\epsilon(f) + R(n,f)$$

$$= \theta(f)d_\epsilon(f) + R(n,f)$$

where $R(n,f) \xrightarrow{P} 0$ as $n \to \infty$ and $\theta(f) = 1 + \theta e^{-2\pi if}$. It is easy to see that $\sqrt{n}d_\epsilon(f) = O_P(1)$. Defining the periodogram ordinate in terms of the discrete Fourier transform $I_Y(f) = n|d_Y(f)|^2$ and using the fact that $E(I_\epsilon(f)) = \sigma^2$ we can write the spectral density function of the series as follows:

$S(f) = \sigma^2|\theta(f)|^2 = \sigma^2(1 + \theta e^{2\pi if})(1 + \theta e^{-2\pi if}) = \sigma^2(1 + \theta^2 + 2\theta\cos(2\pi f))$ which is the same as before.

2. Now we consider the first order periodically correlated moving average series with period 2. The series $Y_t = \epsilon_t + \theta\epsilon_{t-1}$ satisfies the conditions that the series to be periodically correlated, since mean is constant and

$$Cov(Y_{t+\tau}, Y_{s+\tau}) = Cov(\epsilon_{t+\tau} + \theta\epsilon_{t+\tau-1}, \epsilon_{s+\tau} + \theta\epsilon_{s+\tau-1})$$

$$= Cov(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_s + \theta\epsilon_{s-1}) = Cov(Y_t, Y_s)$$

From (7), we can write

$$\sqrt{n}d_Y(f) = \sqrt{n}\sum_{k=0}^{1}\theta(f-k/2)d_{\epsilon^k}(f-k/2) + \sum_{k=0}^{1}R^k(f-k/2)$$

and the periodogram ordinate is

$$I_Y(f) = n|d_Y(f)|^2 \doteq n|\sum_{k=0}^{1}\theta(f-k/2)d_{\epsilon^k}(f-k/2)|^2 \tag{8}$$

If ϵ_t^0 and ϵ_t^1 independently distributed random variables then the identity (8) implies that the spectral density function can be written as $S_Y(f) = 2\sigma^2(1+\theta^2)$. Now, assume that the vectors

$$\epsilon_t = \begin{pmatrix}\epsilon_t^0\\ \epsilon_t^1\end{pmatrix} \text{are } i.i.d \text{ with } \left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{bmatrix}\sigma^2 & \rho\\ \rho & \sigma^2\end{bmatrix}\right)$$

under the circumstance the spectral density function of the series can be calculated as follows:

$$S_Y(f) = E\{n|d_Y(f)|^2\} \doteq nE\left\{|\sum_{k=0}^{1}\theta(f-k/2)d_{\epsilon^k}(f-k/2)|^2\right\}$$

$$= nE\{(\theta(f)d_{\epsilon^0}(f) + \theta(f-1/2)d_{\epsilon^1}(f-1/2))^*(\theta(f)d_{\epsilon^0}(f) + \theta(f-1/2)d_{\epsilon^1}(f-1/2))\}$$

$$= |\theta(f)|^2E\{n|d_{\epsilon^0}(f)|^2\} + |\theta(f-1/2)|^2E\{n|d_{\epsilon^1}(f-1/2)|^2\}$$

$$+n\theta^*(f)\theta(f-1/2)E\{d_{\epsilon^0}^*(f)d_{\epsilon^1}(f-1/2)\} + n\theta^*(f-1/2)\theta(f)E\{d_{\epsilon^1}^*(f-1/2)d_{\epsilon^0}(f)\}$$

From the Euler's formula $a(f-1/2) = 1-\theta e^{-2\pi if}$ and $d_{\epsilon^1}(f-1/2) = \frac{1}{n}\sum_{t=0}^{n-1}(-1)^t\epsilon_t^1 e^{-2\pi ift}$

Since

$$E\{d_{\epsilon^0}^*(f)d_{\epsilon^1}(f-1/2)\} = E\left\{\frac{1}{n}\sum_{t=0}^{n-1}\epsilon_t^0 e^{2\pi ift}\frac{1}{n}\sum_{s=0}^{n-1}\epsilon_t^1 e^{-2\pi is(f-1/2)}\right\}$$

$$= \frac{1}{n^2}E\left\{\sum_{t=0}^{n-1}\epsilon_t^0 e^{2\pi ift}\sum_{s=0}^{n-1}\epsilon_t^1(-1)^s e^{-2\pi isf}\right\} = \frac{1}{n^2}\sum_{t=0}^{n-1}\sum_{s=0}^{n-1}(-1)^s e^{2\pi if(t-s)}E\{\epsilon_t^0\epsilon_s^1\}$$

$$= \frac{\rho}{n^2}\left(\sum_{t=0}^{n-1}e^{2\pi ift}\right)\left(\sum_{s=0}^{n-1}(-1)^s e^{-2\pi ifs}\right) = 0, \text{ because } \sum_{t=0}^{n-1}e^{2\pi ift} = 0 \text{ for } f = k/n.$$

Then the spectral density function becomes

$$S(f) = |\theta(f)|^2 E\{n|d_{e^0}(f)|^2\} + |\theta(f-1/2)|^2 E\{n|d_{e^1}(f-1/2)|^2\} = 2\sigma^2(1+\theta^2)$$

Since

$$|\ a(f-1/2)|^2 = |1-\theta e^{-2\pi if}|^2 = (1-\theta e^{2\pi if})(1-\theta e^{-2\pi if}) = 1+\theta^2-2\theta\cos(2\pi f)$$

and

$$|a(f)|^2 = |1+\theta e^{-2\pi if}|^2 = (1+\theta e^{2\pi if})(1+\theta e^{-2\pi if}) = 1+\theta^2+2\theta\cos(2\pi f).$$

Hence, the spectral density function is

$$S_Y(f) = 2\sigma^2(1+\theta^2).$$

Conclusion

In this paper, the discrete Fourier transform approximation is applied to the periodically correlated time series. For an illustration, periodically correlated first order moving-average series is discussed.

Acknowledgement

I greatfully acknowledge the invaluable comments from Prof.Dr. Yalcin Tuncer and Prof.Dr. Ismihan Bairamov.

References

[1] Bloomfield, P. (1976) Fourier Analysis of Time Series: An Introduction, John Wiley & Sons, New York

[2] Bloomfield, P. Hurd, H.L. and Lund, R.B. (1992) Periodic Correlation in Stratospheric Ozone Data, Mimeo Series 2076, University of North Carolina

[3] Brockwell, P.C. and Davis, R.A. (1987) Time Series: Theory and Methods, Springer-Verlag, New York

[4] Gladyshev, E.G. (1961), Periodically Correlated Random Sequences, Soviet Math 2, 285-288

[5] Hurd, H.L. (1989), Representation of Strongly Harmonizable Periodically Correlated Processes and Their Covariances, Journal of Multivariate Analysis, 29, 53-67

[6] Jones, R.H. and Brelsford, W.M. (1967), Time Series with Periodic Structure, Biometrika, 54, 403-408

[7] Troutman, B.M. (1979), Some Results in Periodic Autoregressions, Biometrika, 66, 219-228

ÖZET

Kesikli Fourier dönüşümü birçok istatistiki sonuç çıkarım problemlerinde kullanılmaktadır. Kesikli Fourier dönüşümünün bir yaklaşımı özellikle spektral yoğunluk fonksiyonunun tahmininde ve bazı dağılım özelliklerinin elde edilmesinde önemli olmaktadır. Bu çalışmada, durağan zaman serileri için var olan bir yaklaşımın benzerinin periyodik zaman serileri için de geçerli olduğu gösterilmiştir.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 55 - 63

THE INFLUENCE OF MOMENTS IN BOOTSTRAP APPROXIMATION

İhsan Karabulut
Department of Statistics, Ankara University
06100, Tandoğan, Ankara, TURKEY

Abstract

The influence of moments on the rate of convergence of bootstrap distribution function, $F_n^*(x)$, of standardized arithmetic mean to its true distribution function $F_n(x)$ is studied in the case of i.i.d random variables $X_1, X_2, ..., X_n$ with $E(X) = 0$ and $E(X^2) < \infty$. It is verified that the existence of moment $E|X|^{2+\delta}$ is sufficient to have the rate of convergence to be $o(n^{-\delta/2})$ almost surely (a.s.) for $0 < \delta < 1$.

Key Words: Edgeword expansion, central limit theorem, characteristic function.

1. Introduction

Let, $X_1, X_2, X_3, ..., X_n$ be a sequence of independent and identically distributed(i.i.d.) random variables(r.v.'s) with a symmetric distribution function(d.f) $F(x)$ and $EX_j = 0$, $EX_j^2 = \sigma^2 < \infty$. Define, $T_n = \sum_{j=1}^n X_j$, $\bar{X}_n = n^{-1}T_n$, $S_n^2 = n^{-1}\sum_{j=1}^n (X_j - \bar{X}_n)^2$. The distribution function of the standardized arithmetic mean will be denoted by $F_n(x) = P(T_n/\sqrt{n}\sigma \leq x)$. The specified r.v. of interest for the bootstrap is $T_n/\sqrt{n}\sigma$. Let $\hat{F}_n(x)$ denote empirical distribution function of $X_1, X_2, X_3, ..., X_n$, the distribution that puts mass $1/n$ at each point. The bootstrap is the name of a variety of resampling methods, namely, simple random sampling with replacement from the original sample. It is to approximate the d.f. of $T_n/\sqrt{n}\sigma$ by $T_n^*/\sqrt{n}S_n$ under $\hat{F}_n(x)$ where $T_n^* = n^{-1}\sum_{j=1}^n X_j^*$. Here, $X_1^*, X_2^*, X_3^*, ..., X_n^*$ is a random sample of size n from $\hat{F}_n(x)$. We use notation $F_n^*(x) = P(T_n^*/\sqrt{n}S_n \leq x)$ for the bootstrap approximation of $F_n(x)$. The main concern of this paper is to find the conditions by means of the moments to have

$$\sum_{n=1}^{\infty} n^{-1+\delta/2} \sup_x |F_n(x) - F_n^*(x)| < \infty \ a.s.$$

where $0 < \delta < 1$. This requires finding some sort of lower bound for the rate of convergence of the bootstrap approximation $F_n^*(x)$ to the true distribution $F_n(x)$. The studies of this kind date back to 1960's and its development can be found in Hall(1982). The rationale of considering such a summation for the rate of convergence of $\sup_x |F_n(x) - F_n^*(x)|$ can

be explained by the words of Baum and Katz(1965): One way of measuring the rate of convergence of non-negative and bounded sequence $\{c_n\}$ is to determine $r \geq -1$ the series $\sum_{n=1}^{\infty} n^r c_n$ converges if there is any. The idea of connecting the rate of convergence to the moments however, mainly due to Ibragimov(1966).

2. Background and Result

There are several results on the issue which are related with the result of this paper. Freidman, Katz and Koopmans(1966), Ibragimov(1966), Heyde(1967) gave connected results on the rate of convergence of $F_n(x)$ to the normal distribution function $\Phi(x)$.The following result belongs to Heyde(1967):

Theorem 2.1. *Let* $X_i, i = 1, 2, 3, ...$ *be a sequence of i.i.d. random variables with* $EX_i^2 = \sigma^2 < \infty$, $EX_i = 0$, *Then*

$$\sum_{n=1}^{\infty} n^{-1+\frac{\delta}{2}} \sup_x \left|F_n(x) - \Phi(x)\right| < \infty, 0 \leq \delta < 1$$

if and only if $E|X_i|^{2+\delta} < \infty$, $0 < \delta < 1$, $EX_i^2 ln(1 + |X_i|) < \infty$, $\delta = 0$.

This result has been extended for the difference between $F_n(x)$ and a portion of its (k+1) term Edgeworth expansion, $G_{kn}(x)$, by Galtsyan(1971) and Heyde and Leslie(1972) independently of the other. This will be given below for the sake of completeness and to give some idea about the possible extensions:

Theorem 2.2. *In order that*

$$\sum_{n=1}^{\infty} n^{-1+(k+\delta)/2} \sup_x \left|F_n(x) - G_{kn}\right| < \infty$$

where k *is a non-negative integer and* $0 < \delta < 1$, *it is necessary and for* $k=0$ *or for distributions satisfying Cramer's condition also sufficient that* $E|X_j|^{k+\delta+2)} < \infty$.

The last theorem was established for the case $\delta = 0$ in both Galtsyan(1971), Heyde and Leslie(1972)

An another exploitation of the forementioned results is to set up them as theorems to find out the rate of convergence to zero of $\sup_x \left|F_n^*(x) - F_n(x)\right|$; the rate of convergence of naive bootstrap approximation $F_n^*(x)$ to the d.f. of the standardized arithmetic mean $F_n(x)$. One result has been obtained by Hall(1988) (Theorem 3.1.i) in this direction. It is excerpted from Hall(1988):

Theorem 2.3. *Let* $X_1, X_2, X_3, ..., X_n$ *be a sequence of i.i.d. r.v.'s with d.f.* $F(x)$ *and* $EX_i = 0$, $EX_i^2 = 1$, *and define the bootstrapped* T_n *as* $T_n^* = n^{-1/2} \sum_{i=1}^{n} X_i^*$ *and its d.f.* $F_n^*(x)$. *Then,*

$$\sup_{-\infty<x<\infty} \left|F_n^*(x) - \{\Phi(x) - n^{-1/2}\frac{\beta}{6}(x^2 - 1)\}\right| = o(n^{-1/2}) \ a.s.$$

as $n \to \infty$ *if and only if* $E|X_i|^3 < \infty$ *and* $EX_i^3 = \beta$.

What Hall is utilized as a technique to derive this result is called leading term approach and it is a bootstrap counterpart of Theorem 2.2. above for $\delta = 0$ and $k = 1$. Next, we will state the result of this work.

Theorem 2.4. *Let $X_1, X_2, X_3, ..., X_n$ be a sequence of i.i.d. r.v.'s with a symmetric d.f. $F(x)$, $EX_j = 0$, $EX_j^2 = \sigma^2 < \infty$ and $\hat{F}_n(x)$ denote symmetrized empirical distribution function. Then,*

$$\sum_{n=1}^{\infty} n^{-1+\frac{\delta}{2}} \sup_x \left|F_n^*(x) - F_n(x)\right| < \infty \ a.s. \tag{2.1}$$

if $E\left|X_j\right|^{2+\delta} < \infty$, where $0 < \delta < 1$.

Remark 1. The symmetry assumption in Theorem 2.4. on the underlying distribution is not indispensable. It can be eliminated as it is done in Heyde(1971).

Remark 2. No further justification is needed for the existence of the second moment; because, it has been shown by Athreya(1987) that naive bootstrap fails unless $EX^2 < \infty$ in approximating the d.f. of the appropriately normalized $\bar{X}_n$.

The empirical d.f. to resample is symmetrized in order to comply with the underlying d.f. $F_n(x)$. The symmetrization can be achieved by following either Efron(1979)'s or Babu and Singh(1984)'s suggestions. According to the first suggestion the both positive and negative values of the differences $\left|X_j(w) - \bar{X}_n\right|$ are taken into the consideration in constructing symmetrical empirical d.f. $\hat{F}_n(x)$. In this case the symmetrical empirical d.f. is formed as the following:

$$\hat{F}_n(x) = \frac{1}{2n} \sum_{j=1}^{2n} I(X_j(w) - \bar{X}_n \leq x).$$

Babu and Singh's method(1984) corrects the skewness as follows:

$$\hat{F}_n(x) = \frac{1}{2}[\hat{F}_n(x) + \hat{F}_n(2x - \bar{X}_n)].$$

The first approach will be preferred in the proof of Theorem 2.4. The result of this paper is a verification the bootstrap approximation $F_n^*(x)$ of $F_n(x)$ is as close as the approximation supplied by the normal theory by means of sufficient moment condition $E\left|X_j\right|^{2+\delta} < \infty$ for $0 < \delta < 1$.

3. Proof

It will be convenient to give the main ingredients of the proof before outlining it. Three lemmas will be stated below.

Lemma 3.1. *Let $Q(t/\sqrt{n}\sigma)$ and $Q^*(t/\sqrt{n}s_n)$ be the characteristic functions of the r.v.'s $T_n/\sqrt{n}\sigma$ and its bootstrap counterpart respectively. If $E|X_j| < \infty$ then*

$$\sup\{\left|Q^*(t/\sqrt{n}s_n) - Q(t/\sqrt{n}\sigma)\right| : |t| < \sqrt{n}\sigma M\} = o(1)$$

a.s. as $n \to \infty$ for a real number $M > 0$.

The proof of the lemma is made by combining the discretization method given by Babu and Singh(1984) and Lemma 4.2 of Lahiri(1989). This lemma is used for the proof of Lemma 3.2 in the sequel , and given as:

Lemma 3.2. *Let $X_1, X_2, ..., X_n$, be a sequence of i.i.d. random variables with a symmetric d.f. $F(x)$ and $EX_j = 0$, $E|X_j|^r < \infty$ for a real number $r \geq 2$. Let κ_s^* denote sth order cumulants of the X_j^* with the symmetrized empirical d.f. $F_n^*(x)$. The characteristic function of r.v. $X_j^* - \bar{X}_n$ can be written as*

$$Q^*(t) = e^{\sum_{s=2}^{r} \frac{(it)^s}{s!}\kappa_s^* + |t|^r \gamma^*(t)} \ a.s.$$

as $n \to \infty$ Here, $\gamma^*(t) = o(1)$ a.s. as $t \to 0$, $n \to \infty$ and there exist $\epsilon > 0$ small enough for which $|\gamma^*(t)| \neq 0$ a.s. $n \to \infty$. Under the same assumptions, the characteristic function can be represented as

$$Q^*(t) = \sum_{s=1}^{r} \frac{(it)^s}{s!}\mu_s^* + |t|^r \beta^*(t)$$

a.s. as $n \to \infty$ where μ_s^* is the s th order expected value of the r.v. $X_j^* - \bar{X}_n$ and similarly $|\beta^*(t)| = o(1)$ a.s. as $t \to 0$ and $n \to \infty$.

The detailed proof of the lemma can be found in Karabulut(1995). It is a modified form of Theorem 1.6.1. of Ibragimov(1966) or Heyde and Leslie(1972).

Lemma 3.3. *If $S_n^2 \to \sigma^2$ and $0 < u \leq \sqrt{n}S_n$ then*

$$\left| e^{-\frac{1}{2}u^2 n S_n^2} - e^{-\frac{1}{2}u^2 n \sigma^2} \right| = o(1) \ a.s.$$

It is a simple result of Proposition 1.2.16. of Rao(1987)

Now we are ready to give the outline of the proof of Theorem 2.4. of this paper. It will follow the line that of Heyde (1967). That method is based on the development made by Ibragimov(1967). First, note that

$$\sup_x \left| F_n(x) - F_n^*(x) \right| \leq \sup_x \left| F_n(x) - \Phi(x) \right| + \sup_x \left| F_n^*(x) - \Phi(x) \right|.$$

The result related for the first term on the right hand side provided by Heyde(1967), so it is enough to consider the results related with the second term on the right hand side. The outline of the proof begins with the verification of the following lemma.

Lemma 3.4. *Let $X_1, X_2, ..., X_n$ be a sequence of r.v.'s as in Theorem 2.4. and $\gamma^*(t)$ as defined in Lemma 3.2 The series in (2.1) converges if and only if*

$$\int_0^A \frac{|\gamma^*(u)|}{u^{1+\delta}} du < \infty \tag{3.2}$$

a.s. as $n \to \infty$ for a real $A > 0$ and $0 < \delta < 1$.

Using this lemma, it will be shown that (3.2) is equivalent to the existence of the moment condition of Theorem 2.4..

First assume that the series in (2.1.) is convergent, i.e.

$$\sum_{n=1}^{\infty} n^{-1+\delta/2} \sup_x \left| F_n^*(x) - \Phi(x) \right| < \infty \ a.s. \tag{3.3}$$

Now, it will be shown that (3.3) and

$$\sum_{n=1}^{\infty} n^{-1+\delta/2} \int_0^1 \left|Q_n^*\left(\frac{t}{\sqrt{n}S_n}\right) - e^{-t^2/2}\right| e^{-t^2/2} dt < \infty \tag{3.4}$$

a.s. are equivalent. By employing Parseval relation(see Feller(1971)) to (3.4) one can have

$$\begin{aligned}
&\sum_{n=1}^{\infty} n^{-1+\delta/2} \left| \int_{-\infty}^{\infty} \left(Q_n^*\left(\frac{t}{\sqrt{n}S_n}\right) - e^{-t^2/2}\right) e^{-t^2/2} dt \right| \\
&\qquad = \sum_{n=1}^{\infty} n^{-1+\delta/2} \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (F_n^*(x) - \Phi(x)) x e^{-x^2/2} dx \right| \\
&\qquad \leq \sum_{n=1}^{\infty} n^{-1+\delta/2} \sup_x \left|F_n^*(x) - \Phi(x)\right|
\end{aligned}$$

When Mill ratio is used the equivalence of (3.3) and (3.4) follows. Because of the symmetry assumption in Lemma 3.2, the sign of the integrand in (3.4) does not change in intervals $(0, \epsilon)$ or $(-\epsilon, 0)$, and thus in the interval $0 < t/\sqrt{n}S_n < \epsilon$. If some manipulations is made on the integrand of (3.4) to be able to use the inequality

$$\left|1 - e^x\right| \leq \left|x\right| e^{|x|}$$

and using the fact that we can find a constant (possibly random) $c > 0$ such that

$$\max_{0 \leq t/\sqrt{n}S_n \leq c} \left|\gamma^*\left(\frac{t}{\sqrt{n}S_n}\right)\right| \leq 1/2 \ a.s. \tag{3.5}$$

as $n \to \infty$. Using (3.5) and the transformation $u = t/\sqrt{n}S_n$, (3.4) becomes

$$\sum_{n=1}^{\infty} n^{\frac{1}{2}(1+\delta)} \int_0^{1/\sqrt{n}S_n} S_n^3 \frac{u^2}{2} |\gamma^*(u)| du < \infty \ a.s. \tag{3.6}$$

Since the summand in (3.6) is monotonically decreasing, by Lemma 3.2 $|\gamma^*(u)| \to 0$ a.s. as $u \to 0$ in addition to decreasing upper limit of the integral. The summation in (3.6) can be changed to integration by the integral test theorem as in Apostol (1974). To do so n is replaced by $x > 1$. Let [x] denote integer valued function and remember that $[x] \leq x < 1 + [x]$. Also it is possible to find a constant $c > 0$ such that $(n+1)^{\frac{1}{2}(1+\delta)} < cn^{\frac{1}{2}(1+\delta)}$

Therefore, by making use of these results it is seen that

$$\begin{aligned}
&\int_0^X x^{\frac{1}{2}(1+\delta)} \Big(\int_0^{1/\sqrt{x}S_n} u^2 |\gamma^*(u)| du \Big) dx \\
&\qquad \leq \sum_{n=1}^{[X]} \int_n^{n+1} x^{\frac{1}{2}(1+\delta)} \Big(\int_0^{1/\sqrt{x}S_n} u^2 |\gamma^*(u)| du \Big) dx \\
&\qquad < \sum_{n=1}^{[X]} n^{\frac{1}{2}(1+\delta)} \int_0^{1/\sqrt{n}S_n} u^2 |\gamma^*(u)| du
\end{aligned}$$

and

$$\int_1^{\infty} x^{\frac{1}{2}(1+\delta)}\Big(\int_0^{1/\sqrt{x}S_n} u^2|\gamma^*(u)|du\Big)dx < \infty \ a.s.$$

When w tends to infinity we should have

$$\int_w^{2w} x^{\frac{1}{2}(1+\delta)}\Big(\int_0^{1/\sqrt{x}S_n} u^2|\gamma^*(u)|du\Big)dx \to 0 \ a.s. \tag{3.7}$$

The integration in the parenthesis is monotonically decreasing. Therefore, the transformation $v = \frac{1}{\sqrt{2w}S_n}$ after x is replaced by 2w in the upper limit of the integral gives that

$$\frac{1}{v^{(3+\delta)}}\int_0^v u^2|\gamma^*(u)|du \to 0$$

as $v \to 0$. If the partial integration is performed to (3.7) after making transformation $v = \frac{1}{\sqrt{x}S_n}$ and taking $A = S_n^{-1}$ at the end we obtain (3.2).

For the sufficiency we will begin with the use of Berry- Esseén Lemma 16.3.2 of Feller(1971). To do so first, take the integral limit as $T = B\sqrt{n}S_n$ with the constant $B > 0$ and make use of Lemma 3.2 again to choose a constant c, (possibly random) satisfying $0 < c < B$. Hence,

$$\Big|Q_n^*(t) - e^{-t^2/2}\Big| \le \frac{t^2}{2}|\gamma^*(\frac{t}{\sqrt{n}S_n})| \ a.s. \tag{3.8}$$

The second term on the right hand side of Berry - Esseén inequality can be neglected because of $o(n^{-1/2})$ a.s.. Replacing (3.8) in (2.1) by using the symmetry property of (3.8) and omitting all the constants we find the inequality

$$\sum_{n=1}^{\infty} n^{-1+\frac{\delta}{2}}\int_{-T}^{T}\Big|\frac{Q_n^*(t) - e^{-t^2/2}}{t}\Big|dt \le \sum_{n=1}^{\infty} n^{-1+\frac{\delta}{2}}\int_0^{c\sqrt{n}S_n} t|\gamma^*(\frac{t}{\sqrt{n}S_n})|dt.$$

By changing variable, $u = \frac{t}{\sqrt{n}S_n}$ and changing the order of summation and integration, the positiveness of all terms permits this, the right hand side of the last inequality becomes

$$\sum_{n=1}^{\infty} n^{-1+\frac{\delta}{2}}\int_{-T}^{T}\Big|\frac{Q_n^*(t) - e^{-t^2/2}}{t}\Big|dt \le S_n^2\int_0^c u|\gamma^*(u)|\Big[\sum_{n=1}^{\infty} n^{\frac{\delta}{2}}e^{-\frac{1}{4}u^2nS_n^2}\Big]du$$

An Abelian theorem in Apostol(1974) will be utilized with the random terms to simplify the right hand side of the last inequality above. Now, Lemma 3.3. is invoked to make sure that the random sum is finite. Combining all the forementioned results together with the fact that $1 - e^{-x} > x - x^2/2$ and $1/4S_n^2u^2 < 1$ for n large enough we can find a positive constant c such that

$$\sum_{n=1}^{\infty} n^{-1+\frac{\delta}{2}}\int_{-T}^{T}\Big|\frac{Q_n^*(t) - e^{-t^2/2}}{t}\Big|dt \le c\int_0^c \frac{|\gamma^*(u)|}{u^{(1+\delta)}(1 - \frac{1}{8}S_n^2u^2)^{(1+\frac{1}{2})}}du$$

The result follows after taking partial integration of the right hand side of the inequality and using (3.2).

The next part of the proof will be completed by showing that (3.2) is equivalent to the existence of the moment with the stated order in Theorem 2.4..

By replacing $(t^2/2)|\gamma^*(t)|$ in (3.2) with its equivalence in Lemma 3.2 and remembering that $\hat{F}_n(x)$ is symmetric, the following calculation is possible. Assume that $\gamma^*(t) < 0$ without losing any generality and make use of that $(\cos x - 1 + x^2/2) \geq 0$ (see Ibragimov(1967) or Hall(1982)) to reach the following

$$\begin{aligned}\frac{S_n^2}{2}\int_0^c \frac{|\gamma^*(t)|}{t^{(1+\delta)}}dt &= \int_0^c \frac{\int_{-\infty}^{\infty}(e^{\frac{it(X_j-\bar{X}_n)}{S_n}} - 1 + \frac{t^2}{2}\frac{(X_j-\bar{X}_n)^2}{S_n^2})d\hat{F}_n(x)}{t^{(3+\delta)}}dt \\ &= \int_{-\infty}^{\infty}\int_0^c \frac{(\cos\frac{t(X_j-\bar{X}_n)}{S_n} - 1 + \frac{t^2}{2}\frac{(X_j-\bar{X}_n)^2}{S_n^2})}{t^{(3+\delta)}}dtd\hat{F}_n(x) \\ &< \infty\end{aligned}$$

a.s. as $n \to \infty$. After applying partial integration two times to the integral on the right hand side

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\Big[&(\frac{-\cos\frac{c(X_j-\bar{X}_n)}{S_n} - 1 + \frac{c^2}{2}\frac{(X_j-\bar{X}_n)^2}{S_n^2}}{(2+\delta)c^{(2+\delta)}}) \\ &+ (\frac{\frac{-(X_j-\bar{X}_n)}{S_n}\sin c\frac{(X_j-\bar{X}_n)}{S_n} + c\frac{(X_j-\bar{X}_n)^2}{S_n^2}}{(2+\delta)c^{(2+\delta)}}) \\ &+ (\int_0^c \frac{(X_j-\bar{X}_n)^2(1-\cos\frac{t(X_j-\bar{X}_n)}{S_n})}{(2+\delta)c^{(2+\delta)}}dt)\Big]\end{aligned}$$

is obtained. The verification of the existence and finiteness of integral placed on the most right will give the necessary and sufficient condition which we seek. By making transformation $v = t(X_j - \bar{X}_n)/S_n$ we get

$$\frac{1}{n}\sum_{j=1}^{n}|X_j-\bar{X}_n|^{(2+\delta)}\int_0^{c|X_j-\bar{X}_n|/S_n}\frac{1-\cos v}{v^{(1+\delta)}}dv$$

The integral is finite and it can be neglected. Hence at the end, the integral in (3.2) is finite a.s. if the summation is finite a.s.. It is to be finite we should have

$$\frac{1}{n}\sum_{j=1}^{n}\left|X_j-\bar{X}_n\right|^{2+\delta} \to E\left|X_j\right|^{(2+\delta)} < \infty$$

by the strong law of large numbers. Thus, it is true that $E|X_j|^{(2+\delta)} < \infty$ is sufficient to have (2.1).

Acknowledgment

This is a part of my Ph.D. thesis. Thank to Professor Soner Gönen for his encouragement.

References

[1] Apostol, T. M. (1974) Mathematical Analysis,2nd ed., Addison-Wesley, Reading.

[2] Athreya, K. B. (1987) Bootstrap of the mean in the infinite variance case, The Annals of Statistics, 15(2), 724-731.

[3] Babu, G. J. and Singh, K. (1984) On one term Edgeworth correction by Efron's Bootstrap, Sankhyā, 46, Series A, Pt.2, 219-232.

[4] Baum, L. E. and Katz, M. L. (1965) Convergence rates in the law of large numbers, Trans. Am. Math. Soc., 120, 108-123.

[5] Efron, B. (1979) Bootstrap methods: Another look at the jackknife, The Annals of Statistics, 17(1), 1-26.

[6] Feller, W. (1971) An Introduction to Probability Theory and Its Applications, Volume 2, 2nd ed., John Wiley & Sons, New York.

[7] Friedman, N., Katz, M. L., Koopmans, L. H. (1966) Convergence rates for the central limit theorem, Proc. Nat. Acad. Sci. U.S.A., 56, 1042-1065.

[8] Galtsyan, F. N. (1971) On asymptotic expansions in the central limit theorem, Theory of Probability and Its Applications, 16(3), 528-533.

[9] Hall, P. (1982) Rates of Convergence in the Central Limit Theorem, Pitman Books Limited, London.

[10] Hall, P. (1988) Rate of convergence in bootstrap approximation, The Annals of Probability, 16(4), 1665-1684.

[11] Heyde, C. C. (1967) On the influence of moments on the rate of convergence to the normal distribution, Z. Wahrscheinlichkeitstheorie Verw. Geb., 8, 12-18.

[12] Heyde, C. C. and Leslie, J. R. (1972) On the influence of moments on approximations by portion of a Chebyshev series in central limit convergence, Z. Wahrscheinlichkeitstheorie Verw. Geb., 21, 255-268.

[13] Ibragimov, I. A. (1966) On the accuracy of Gaussian approximation to the distribution functions of sums of independent variables, Theory of Probability and Its Applications, 11(4), 539-570.

[14] Karabulut, İ. (1995) Bootstrap Yaklaşımında Beklenen Değerlerin Yakınsama Hızına Etkisi. Unpublished Ph.D. thesis , Gazi Üniversitesi, Ankara Türkiye.

[15] Lahiri, S. N. (1989) Bootstrap approximations and Edgeworth expansion for the distributions of M-estimators of a linear regression parameter, Technical Report, Department of Statistics, Iowa State University.

[16] Rao, B.L.S.P. (1987) Asymptotic Theory of Statistical Inference, John Wiley & Sons, New York.

Özet

Birbirinden bağımsız ve aynı dağılımlı ve $EX = 0$, $EX^2 < \infty$ olan $X_1, X_2, ..., X_n$ rassal değişkenlerinin aritmetik ortalamalarına ait bootstrap dağılım fonksiyonu $F_n^*(x)$'in yine bu rassal değişkenin dağılımı $F_n(x)$'e yakınsamasında beklenen değerlerin etkisi araştırıldı. Yakınsama hızının hemen hemen heryerde $o(n^{-\delta/2})$ olabilmesi için $0 < \delta < 1$ olmak üzere $E|X|^{2+\delta} < \infty$ olmasının yeterli olduğu gösterildi.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 65 - 75

PREDICTION OF HIGH WATER LEVELS IN THE BALTIC

A. Svenson and J. Holst
Department of Mathematical Statistics
Lund Institute of Technology, Box 118, SE-22100, Lund Sweden
Phone:+46 46 2224679 (ASv); +46 46 2229538 (JH)
E-mail: andress@maths.lth.se (ASv); jahn@maths.lth.se (JH)

Abstract

This paper presents a level crossing predictor for Gaussian ARMAX processes, which is optimal in the sense that it minimizes the number of false alarms for a given probability of detecting the level-crossings. It is applied to real data for predicting and warning for high water levels at the Danish coast in the Baltic Sea. The optimal alarm system is shown to work better than a simpler and more conventional alarm system. A method to optimally predict the crossings also when the external signals are not known is presented. In this particular case most of the variability of the predictions are due to system noise, so the performance of the system with predicted external signals are almost identical to the performance when the external signals are known. A smaller simulation study shows that the water level process is hard to predict and that the choice of model can be rather important.

Key Words: level crossings, flooding alarm, catastrophe prediction, optimal alarm, ARMAX process.

1. Introduction

A flooding incident can be disastrous, especially if people are not warned. Hence, in many situations it is important to be able to give an alarm some time before the incident occurs. It is also important to give as few false alarms as possible, but still find a sufficient number of the flooding incidents.

In a more general setting, the problem is to predict level crossings, catastrophes, of a stochastic process a sufficient time in advance. This catastrophe prediction problem was treated by de Maré [2] and Lindgren [3], and a definition of the optimal catastrophe predictor was given as the predictor that gives a minimum number of false alarms for a given detection probability. This idea was further treated in Svensson, Holst, Lindquist & Lindgren [7], and leads to an explicit catastrophe predictor for Gaussian ARMA processes with constant catastrophe level. Since the construction of the optimal catastrophe predictor requires quite a large amount of calculations, two suboptimal predictors were also introduced. In Svensson & Holst [6] the technique was extended to cover both ARMAX and SETARMAX processes with a deterministic but changing catastrophe level. This made it possible to use the optimal catastrophe predictor on real data, describing water levels in the Baltic Sea, presented in this paper. Modelling of the water levels in the Baltic Sea is treated in Berntsen [1], Nielsen [4] and Spliid & Nielsen [5]. A complication with ARMAX processes is that the external signals might not be known in advance, which means that they have to be predicted too. An idea how this can be treated in the same framework as above is also included in this paper and applied to the data sets used.

2. The data set

The data sets used in this paper are from 1978, 1979 and 1980. They consist of the following measurements.

Location	Water level	Head wind	Side wind	Air pressure	Temp.
Korsør	X				
Rødbyhavn	X				
Gedser	X				
Visby	X				
Kadetrenden/ Maribo (78)		X	X	X	X
Møn-Sydøst lightship (79,80)		X	X	X	X
Møn lighthouse:		X	X	X	X
Christiansø lighthouse	X	X	X	X	X
Hammer Odde lighthouse:		X	X	X	X

Only three of these signals are used in the final model describing the water level at Rødbyhavn. They are the water level at Rødbyhavn, the head wind at Christiansø lighthouse and the air pressure at Kadetrenden/Maribo (78) or Møn-Sydøst lightship (79,80). The original data sets contained measurements every hour, but since the process is oversampled, only one sample per 3 hours was used for modelling the water level. Before they have been used for modelling, the mean value using data from all three years has been subtracted. However, the catastrophe levels used later are related to the original data. In Figure 1 the water level at Rødbyhavn is shown for the data set from 1978. The complete data sets with a short description can be found at the address: http://www.maths.lth.se/matstat/staff/anderss/data/data.html.

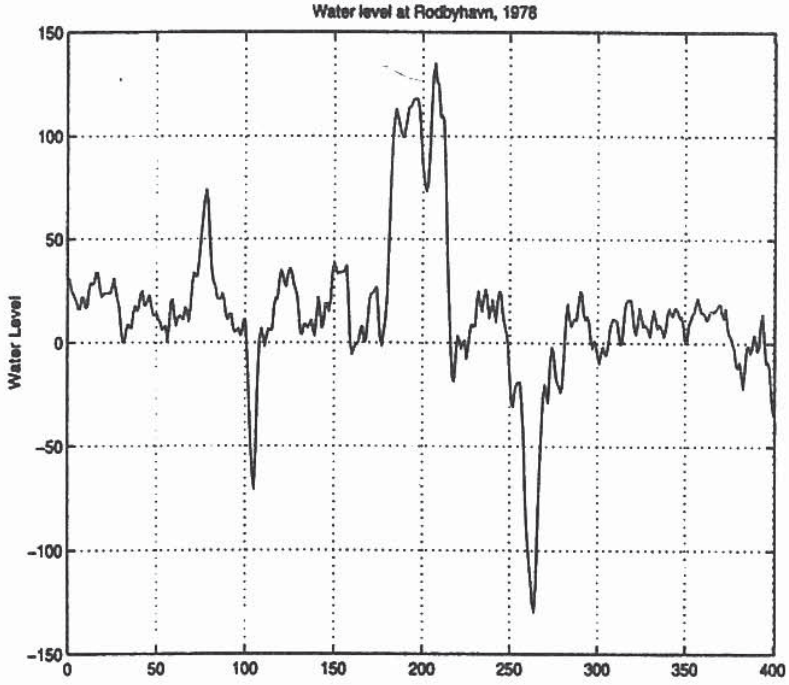


Figure 1: The water level at Rødbyhavn during the period that is covered by the data set from 1978.

3. The models

The water level at Rødbyhavn has been modelled as ARMAX and SETARX processes, denoted X_t, with two external signals, denoted $u_{1,t}$ and $u_{2,t}$. The structure of an ARMAX(p,q,r_1,r_2) process is

$$X_t + a_1 X_{t-1} + \ldots + a_p X_{t-p} =$$
$$b_{1,0}u_{1,t} + \ldots + b_{1,r_1}u_{1,t-r_1} + b_{2,0}u_{2,t} + \ldots + b_{2,r_2}u_{2,t-r_2} + c_0 e_t + \ldots + c_q e_{t-q},$$

or shorter

$$A(z^{-1})X_t = B_1(z^{-1})u_{1,t} + B_2(z^{-1})u_{2,t} + C(z^{-1})e_t,$$

where $\{e_t\}_{t=-\infty}^{\infty}$ is white noise and e_t is uncorrelated with X_s, $u_{1,s}$ and $u_{2,s}$ for $s < t$. It is furthermore assumed that $e_t \in N(0,1)$.

After trying a number of different models three were chosen and estimated on the data from 1978, and optimal alarm systems were calculated. The models are ARMAX(2,1,1,1), ARMAX(4,2,1,1) and SETARX(2;2,2;1,1). The noise is assumed to be independent and Gaussian with variance 1.

The ARMAX(2,1,1,1)-model is

$$\begin{aligned} A(z^{-1}) &= 1.0000 - 1.2794z^{-1} + 0.3786z^{-2} \\ C(z^{-1}) &= 5.5678 + 3.1836z^{-1} \\ B_1(z^{-1}) &= -0.0072z^{-1} \\ B_2(z^{-1}) &= 0.0232z^{-1} \end{aligned}$$

The empirical density functions for the one and two-step prediction error for the data set from 1978 are shown in Figure 2, together with the normal density function, and normal probability plots. It can be seen that the residuals have slightly heavier tails than in the normal distribution. However, in spite of these deviations the normal distribution has been used for modelling and calculation of the alarm systems. It seems to work rather well.

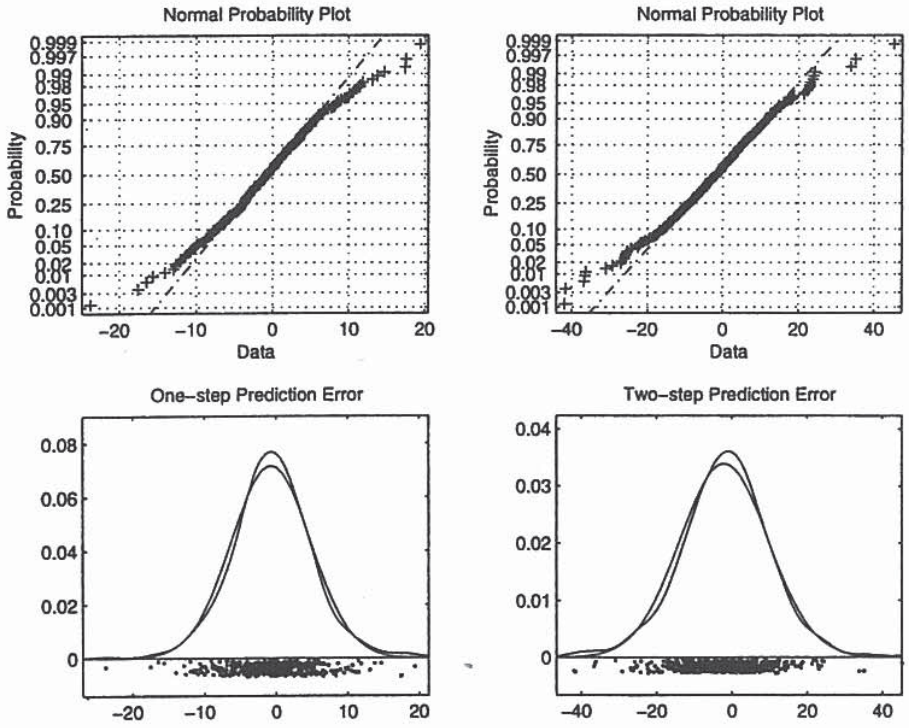


Figure 2: The one and two-step prediction errors for data78, when modelled as ARMAX(2,1,1,1).

The ARMAX(4,2,1,1)-model is

$$\begin{aligned} A(z^{-1}) &= 1.0000 - 1.7227z^{-1} + 1.7602z^{-2} - 1.5950z^{-3} + 0.6652z^{-4} \\ C(z^{-1}) &= 5.1942 + 0.4259z^{-1} + 3.7263z^{-2} \\ B_1(z^{-1}) &= -0.0074z^{-1} \\ B_2(z^{-1}) &= 0.0225z^{-1}. \end{aligned}$$

The SETARX-model is composed of two ARX-models where

$$\begin{aligned} A(z^{-1}) &= 1.0000 - 1.5280z^{-1} + 0.6137z^{-2} \\ C(z^{-1}) &= 6.0661 \\ B_1(z^{-1}) &= -0.0049z^{-1} \\ B_2(z^{-1}) &= 0.0155z^{-1} \end{aligned}$$

is used when the process value $X_{t-2} < 30$ and

$$\begin{aligned} A(z^{-1}) &= 1.0000 - 1.4226z^{-1} + 0.7434z^{-2} \\ C(z^{-1}) &= 6.0661 \\ B_1(z^{-1}) &= 0.0202z^{-1} \\ B_2(z^{-1}) &= 0.1312z^{-1} \end{aligned}$$

when the process value $X_{t-2} \geq 30$.

In cases when models for the external signals are needed these signals have been modelled as AR processes. The model for the head wind at Christiansø lighthouse is an AR(3) process with the parameters

$$\begin{aligned} A_1(z^{-1}) &= 1.0000 - 1.7042z^{-1} + 0.5411z^{-2} + 0.1713z^{-3} \\ C_1(z^{-1}) &= 7.7460 \end{aligned}$$

and the model for the air pressure at Kadetrenden/Maribo (78) or Møn-Sydøst lightship (79,80) is an AR(1) process with the parameters

$$\begin{aligned} A_2(z^{-1}) &= 1.0000 - 0.9657z^{-1} \\ C_2(z^{-1}) &= 27.8675. \end{aligned}$$

It could be considered using one model for the air pressure at Kadetrenden/Maribo (78) and another model for Møn-Sydøst lightship (79,80), but since the locations are rather close to each other, the same model has been used. This also requires fewer calculations.

4. The optimal alarm system

The optimal alarm systems used in this paper are optimal in the sense that they minimize the probability of false alarms for a given probability of detecting the catastrophes. Optimality is reached by the alarm system defined through the likelihood ratio,

$$\frac{dP_{Y(t-k)}(y|C_t^*)}{dP_{Y(t-k)}(y|C_t)} \leq \text{constant},$$

where $Y(t)$ denotes the available information at time t, C_t is the event that a catastrophe occurs at time t and C_t^* is the complementary event that no catastrophe occurs at time t. This condition can be simplified, so that the alarm system can be based on only the predictor $(\hat{x}_{t-1}, \hat{x}_t)$ of the process X_t at times $t-1$ and t, instead of all the available information $Y(t)$. The result is

$$P(C_t|\hat{x}_{t-1}, \hat{x}_t) > P_b,$$

which was shown in Svensson et al. [7], to be the optimal alarm system for ARMA processes with a constant catastrophe level. It is then possible to calculate the alarm region in advance, which makes the alarm system rather fast. A typical alarm region in the $(\hat{x}_{t-1}, \hat{x}_t)$-plane is shown in Figure 3. The model is the ARMAX(2,1,1,1) described above, with the influence of the external signals subtracted. The predictor is using 6 old process values and the prediction horizon is 2. This idea was further

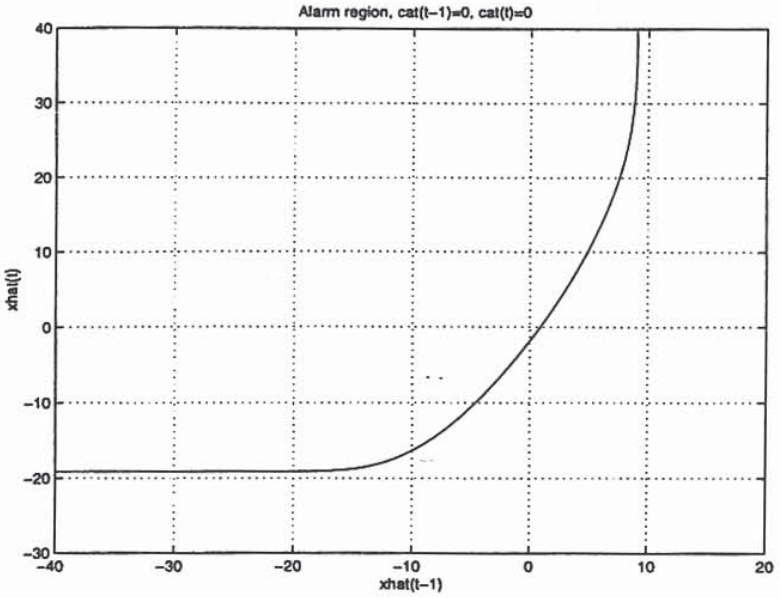


Figure 3: A typical optimal alarm region for an ARMA-process with catastrophe level 0 at times $t-1$ and t (the region above the solid line). The model is the ARMAX(2,1,1,1) described above, with the influence of the external signals subtracted. The predictor is using 6 old process values and the prediction horizon is 2.

developed in Svensson & Holst [6], to cover ARMAX and SETARX processes when the external signals are known and the process is stationary.

5. Alarm system using predicted external signals

Since predictions of the process values are needed in the level crossing predictor, also predictions of the external signals are needed when the external signals are not known in advance. In case of known external signals, the effect can be included in the catastrophe level, giving a catastrophe level that changes through time, see Svensson & Holst [6]. However it is not that simple in case of stochastic external signals. A few assumptions on the signals have to be added in order to get an explicit level crossing predictor.

If we assume that the external signals and the process are stationary Gaussian processes, the covariance of the process value predictor $Cov(\hat{x}_{t-1}, \hat{x}_t)$ will include both the effects of the process noise and the external signals.

Suppose the process can be written,

$$\begin{aligned} A(z^{-1})X_t &= B(z^{-1})u_t + C(z^{-1})e_t \\ A_1(z^{-1})u_t &= C_1(z^{-1})w_t. \end{aligned}$$

Due to linearity, X_t can be decomposed into one part, $X_{u,t}$, describing the influence of the external signals and one part, $X_{e,t}$, describing the influence of the system noise.

$$X_t = X_{u,t} + X_{e,t}$$

$$\begin{aligned} A(z^{-1})X_{e,t} &= C(z^{-1})e_t \\ A(z^{-1})A_1(z^{-1})X_{u,t} &= B(z^{-1})C_1(z^{-1})w_t. \end{aligned}$$

The same deductions can be done for the predictions, leading to

$$\hat{X}_t = \hat{X}_{u,t} + \hat{X}_{e,t}.$$

If the noise processes e_t and w_t are assumed to be independent, the covariance of the predictions $\hat{x}_{t-1}, \hat{x}_t$ is

$$Cov(\hat{x}_{t-1}, \hat{x}_t) = Cov(\hat{x}_{u,t-1}, \hat{x}_{u,t}) + Cov(\hat{x}_{e,t-1}, \hat{x}_{e,t}).$$

This means that if $X_{u,t}$ can be optimally predicted, the technique presented in Svensson & Holst [6], can still be used and thus the resulting critical levels for the stochastic part of the process will be

$$[L_{cat}(t-1|t-k), L_{cat}(t|t-k)] = [L(t-1), L(t)] - [\hat{x}_u(t-1|t-k), \hat{x}_u(t|t-k)].$$

The catastrophe level $L(t)$ for the original process, is assumed to be deterministic and known, and need not be predicted. The part of the process that is due to the external signals, influences the mean value of the process and will thus enter as an addition to the catastrophe level. Predictions for times $t-1$ and t are needed and the information is available up until $t-k$.

One model, the ARMAX(2,1,1,1)-model with the external signals modelled as AR(3) and AR(1) as above, have been tested and the results are shown and compared to the other alarm systems in Table 1, Table 2 and Table 3. The alarm system works well for the data set that was used for estimating the model, but poorer for the other two data sets. The reason for this could be that the fixed models for the process and the external signals are not totally correct. This is similar to the alarm system where the external signals are not predicted, which is expected since almost all the variability is due to process noise.

In order to check how much the departures from normality and model type influence the performance, a smaller simulation study based on the ARMAX(2,1,1,1) model above with the external signals simulated as AR(3) and AR(1), was also performed. It shows that the process is very hard to predict, and will give a large amount of false alarms if a high detection probability is desired. An alarm is denoted false if it does not predict the catastrope exactly in time. The influence of the inputs are rather easy to predict when the prediction horizons are short, leading to almost the same alarm system as for known inputs. The variability of the predictions of process values is almost entirely due to the influence of the system noise, e_t. The results from the simulation are shown in Table 1. When the wrong model is used the detection probability can become a lot lower than calculated. This is obvious, especially for the SETARX model. The performance would have been better if the models had been estimated on the simulated data and not on the water data. Worth noting is that the maximal detection probability for the naive-naive alarm system is 0.29, so it is not comparable to the other alarm systems.

Alarm system	Alarms		Catastrophes	
	False	Total	Detected	Total
best naive-naive	628 (0.92)	685	36 (0.29)	195
ARMAX2111pred (70%)	1676 (0.93)	1808	132 (0.68)	195
ARMAX2111 (70%)	1676 (0.93)	1808	132 (0.68)	195
ARMAX2111 (90%)	3218 (0.95)	3400	182 (0.93)	195
ARMAX4211 (90%)	2860 (0.95)	3019	159 (0.82)	195
SETARX (90%)	4119 (0.98)	4218	99 (0.51)	195

Table 1: Results for the naive-naive alarm system, compared to the optimal alarm system using three different process models, and one with predicted external signals. Simulated data from the ARMAX(2,1,1,1) process has been used in all the simulations. The alarm systems have been constructed to give the detection probabilities in paranthesis if the data comes from the models above.

6. Results

The optimal alarm systems for the different models were compared to some simpler alarm systems. The simplest alarm system, called the naive-naive alarm system, gives an alarm when the process value k steps before a possible catastrophe crosses a certain level. This alarm system did work, but not as well as the optimal alarm systems. The most important disadvantage is that the naive-naive alarm system will have a maximum detection probability, that cannot be exceeded and is rather low.

Another simple alarm system models the process and gives alarm when the predicted process values crosses a level, that was determined from the data sets. This alarm system did not work, hence it has not been included in the tables below.

The optimal alarm system has a nonlinear alarm region, that changes depending on the catastrophe level and the process values. This makes the optimal alarm system rather complex. In many cases when the performance is important this is the alarm system that ought to be used. In other cases it might be good to compare a proposed simpler alarm system to the optimal in order to check how close to the optimal the simpler alarm system is.

The parameters describing the process have all been estimated on the data set from 1978, and then tested on all three data sets. The alarm level for the naive-naive alarm system has been optimized over the three data sets together. As can be seen in Table 2 the naive-naive alarm system has a rather low maximal detection probability, and thus is not possible to use if a high detection probability is required. The performance of the optimal alarm systems for these three data sets does not differ very much from each other and they have almost the same number of false alarms. The detection probabilities used are shown in parenthesis. They were in most cases set to 90%. ARMAX2111pred is the alarm system where also the external signals are predicted.

The optimal alarm systems with the highest detection probabilities have quite a few false alarms according to the strictest definition, where an alarm is considered false if it does not predict the catastrophe exactly right in time, but it could be questioned if all of these should be considered false. In Figure 4, it can be seen that a few of the so called false alarms are early alarms, or alarms given when the levels are still critical. In case of early alarms, at least for one or two steps early which means 3-6 hours early, the additional cost should not be too large. Also, the confidence in the alarm system will not be damaged too much. In the case of alarms when still over the critical level, it means that it will take a little longer to get back to normal state from the emergency state, caused by the process being alarmed. The cost should be small compared to the cost of the catastrophe. If these ideas, i.e. one and two steps early alarms are counted as correct alarms and alarms given when in catastrophe state are not counted at all are taken into account, Table 2 will turn into Table 3.

The alarm level for the naive-naive alarm system is optimized over all three sets. It only reaches a total detection probability of approximately 40 %, which is far below the detection probabilities reached by the different optimal alarms. However, a higher detection probability will inevitably lead to more false alarms, and that is a trade-off that has to be made in each individual case.

In Figure 5 close-ups at some different times are shown to give an explanation for the rather high rate of false alarms. 95% one-dimensional confidence intervals based on the one and two step predictions are also shown. The process is rather hard to predict which leads to wide confidence intervals and a high number of false alarms if a high detection probability is wanted.

7. Conclusions

This paper has presented an optimal alarm for processes described by linear or piecewise linear processes applied to prediction of high water levels in the Baltic. The optimal alarm technique gives as few false alarms as possible for a given probability of detecting the catastrophes.

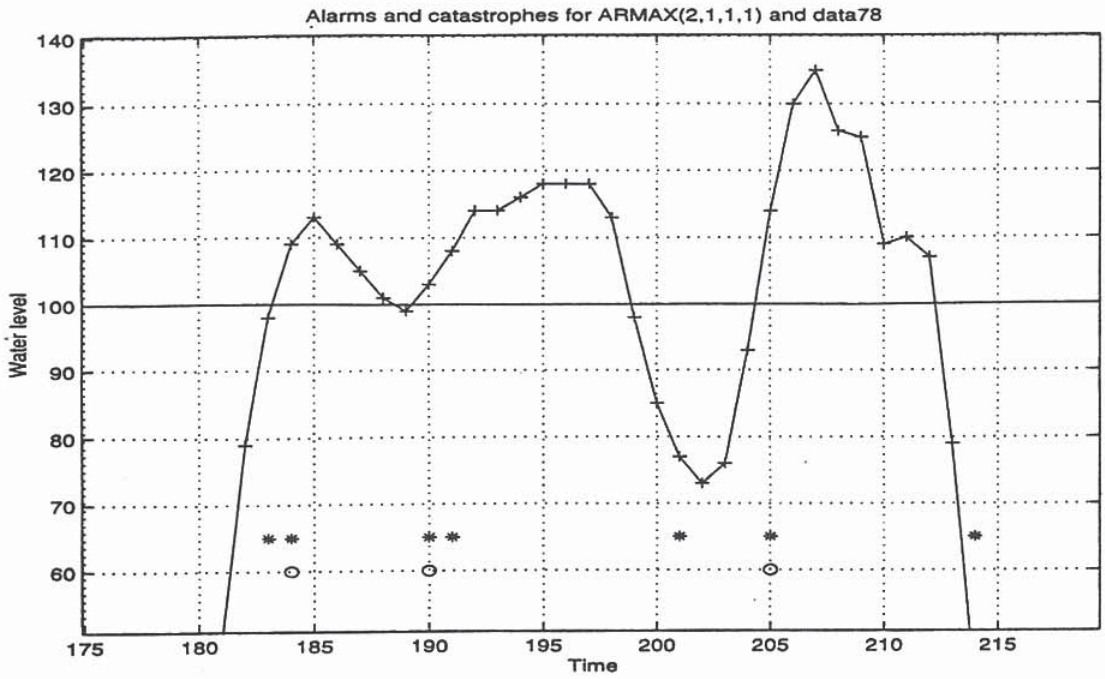


Figure 4: Part of the data set from 1978. The detection probability was set to 90 % and the prediction horizon was set to two steps. The circles denote the actual catastrophes i.e. the crossings of level 100, and the stars show the predicted catastrophe times.

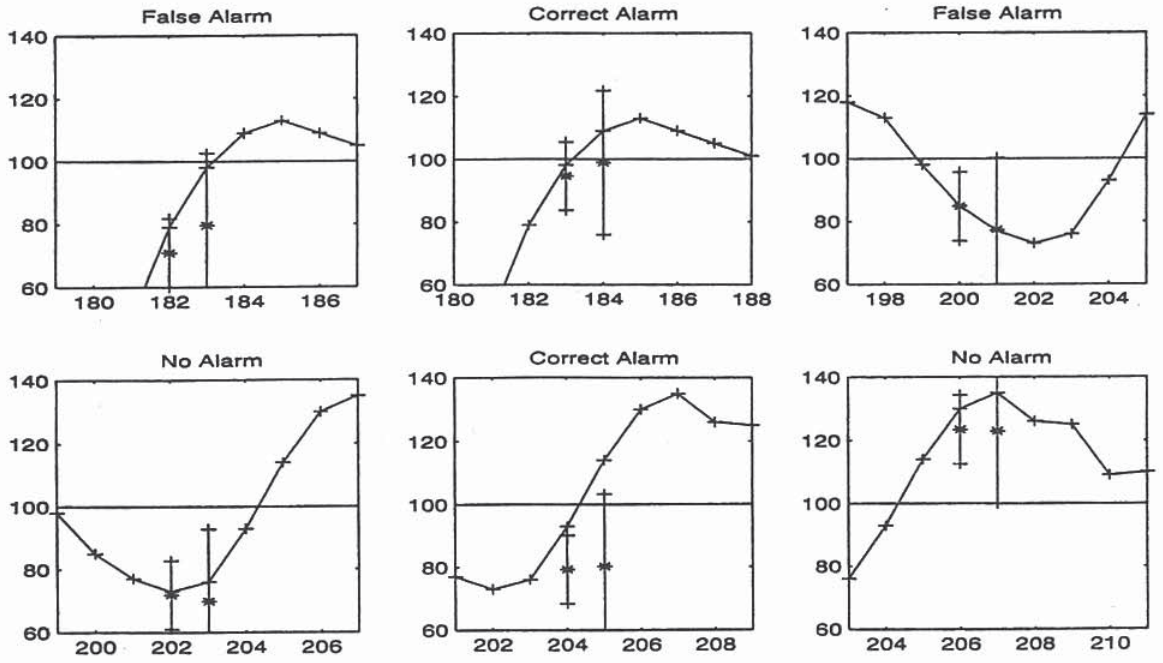


Figure 5: Close-ups for some of the times in Figure 4. The one-dimensional 95% prediction intervalls are plotted. The predictions are plotted as stars.

PREDICTION OF HIGH WATER LEVELS

Data are collected in the southern part of the Baltic and high water levels in Rødbyhavn in Denmark are to be predicted.

The models that are used to describe the water levels all contain external variables, with future values that are unknown at prediction time. This means that also these external signals have to be predicted, which influences alarm levels and probabilities for detection and for alarm. Three different models for the water level have been considered.

The optimal alarm systems presented in the paper work well, and have the ability to reach any specified detection probability. The more conventional alarm algorithm that the optimal alarm is compared to, i.e. the alarm is sounded when the process reaches a certain level, has a maximal detection probability which in these cases is rather low. This means that if a high detection probability is required, the optimal alarm system *has to* be used. A drawback with a high detection probability is that the number of false alarms also becomes rather large, even though the optimal alarm systems give a minimum of false alarms. In particular the SETARX model for the water level shows this balance, it has a fast response and detect almost all catastrophes on all datasets, but at the expence of giving a high amount of false alarms, in particular on a dataset (from 1980) to which the model was not adapted.

A possibility to lower the number of false alarms is to find a better model, e.g. by using more external information for the predictions or by taking the timevariations of the water level process into account. Furthermore, in the flooding data case the prediction errors are not exactly normally distributed, which introduces further approximations in the calculations.

References

[1] P. Berntsen. Prediction of water levels in the Baltic (in Swedish). Master's thesis, Department of Mathematical Statistics, Lund Institute of Technology, 1987. CODEN: LUTFD2/(TFMS–5008)/1–59/(1987).

[2] J. de Maré. Optimal prediction of catastrophes with applications to Gaussian processes. *Ann. Probability*, 8:841–850, 1980.

[3] G. Lindgren. Model processes in non-linear prediction, with application to detection and alarm. *Ann. Probability*, 8:775–792, 1980.

[4] B. K. Nielsen. Flooding alarm at Lolland-Falster (in Danish). Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, 1983. No. 2/83.

[5] H. Spliid and B. K. Nielsen. Prediction of high tides in the Baltic Sea. Technical Report 13/84, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, 1984.

[6] A. Svensson and J. Holst. Optimal prediction of level-crossings in piecewise linear Gaussian processes with changing catastrophe level. Technical Report 1997:14, Department of Mathematical Statistics, Lund Institute of Technology, 1997. ISRN: LUTFD2/TFMS–3138–SE.

[7] A. Svensson, J. Holst, R. Lindquist, and G. Lindgren. Optimal prediction of catastrophes in autoregressive moving-average processes. *J. Time Series Analysis*, 17:511–531, 1996.

Alarm system	Alarms		Catastrophes	
	False	Total	Detected	Total
Data78 and cat=70				
naive-naive (44)	2 (1.00)	2	0 (0.00)	2
ARMAX2111pred (70%)	3 (0.75)	4	1 (0.50)	2
ARMAX2111 (70%)	3 (0.75)	4	1 (0.50)	2
ARMAX2111 (90%)	5 (0.83)	6	1 (0.50)	2
ARMAX4211 (90%)	4 (0.80)	5	1 (0.50)	2
SETARX (90%)	9 (0.90)	10	1 (0.50)	2
Data79 and cat=70				
naive-naive (44)	6 (0.86)	7	1 (0.50)	2
ARMAX2111pred (70%)	6 (0.86)	7	1 (0.50)	2
ARMAX2111 (70%)	7 (0.88)	8	1 (0.50)	2
ARMAX2111 (90%)	15 (0.88)	17	2 (1.00)	2
ARMAX4211 (90%)	14 (0.88)	16	2 (1.00)	2
SETARX (90%)	28 (0.93)	30	2 (1.00)	2
Data80 and cat=70				
naive-naive (44)	13 (0.80)	16	3 (0.50)	6
ARMAX2111pred (70%)	26 (0.84)	31	5 (0.83)	6
ARMAX2111 (70%)	26 (0.84)	31	5 (0.83)	6
ARMAX2111 (90%)	64 (0.93)	69	5 (0.83)	6
ARMAX4211 (90%)	48 (0.91)	53	5 (0.83)	6
SETARX (90%)	71 (0.93)	76	5 (0.83)	6
Data78 and cat=100				
naive-naive (70)	1 (0.50)	2	1 (0.33)	3
ARMAX2111pred (70%)	0 (0.00)	1	1 (0.33)	3
ARMAX2111 (70%)	0 (0.00)	1	1 (0.33)	3
ARMAX2111 (90%)	5 (0.62)	8	3 (1.00)	3
ARMAX4211 (90%)	5 (0.71)	7	2 (0.67)	3
SETARX (90%)	4 (0.57)	7	3 (1.00)	3
Data79 and cat=100				
naive-naive (70)	1 (0.50)	2	1 (1.00)	1
ARMAX2111pred (70%)	2 (0.66)	3	1 (1.00)	1
ARMAX2111 (70%)	2 (0.66)	3	1 (1.00)	1
ARMAX2111 (90%)	5 (0.83)	6	1 (1.00)	1
ARMAX4211 (90%)	4 (0.80)	5	1 (1.00)	1
SETARX (90%)	8 (0.89)	9	1 (1.00)	1
Data80 and cat=100				
naive-naive (70)	6 (1.00)	6	0 (0.00)	1
ARMAX2111pred (70%)	9 (1.00)	9	0 (0.00)	1
ARMAX2111 (70%)	9 (1.00)	9	0 (0.00)	1
ARMAX2111 (90%)	17 (1.00)	17	0 (0.00)	1
ARMAX4211 (90%)	15 (1.00)	15	0 (0.00)	1
SETARX (90%)	17 (0.94)	18	1 (1.00)	1

Table 2: Results for the naive-naive alarm system, compared to the optimal alarm system using three different process models, and one with predicted external signals. The stricter definition of correct alarm is used.

Alarm system	Alarms		Catastrophes	
	False	Total	Detected	Total
Data78 and cat=70				
naive-naive (44)	0 (0.00)	1	1 (0.50)	2
ARMAX2111pred (70%)	1 (0.33)	3	1 (0.50)	2
ARMAX2111 (70%)	1 (0.33)	3	1 (0.50)	2
ARMAX2111 (90%)	1 (0.25)	4	1 (0.50)	2
ARMAX4211 (90%)	2 (0.50)	4	1 (0.50)	2
SETARX (90%)	5 (0.62)	8	1 (0.50)	2
Data79 and cat=70				
naive-naive (44)	5 (0.83)	6	1 (0.50)	2
ARMAX2111pred (70%)	6 (0.86)	7	1 (0.50)	2
ARMAX2111 (70%)	7 (0.88)	8	1 (0.50)	2
ARMAX2111 (90%)	13 (0.87)	15	2 (1.00)	2
ARMAX4211 (90%)	13 (0.87)	15	2 (1.00)	2
SETARX (90%)	25 (0.86)	29	2 (1.00)	2
Data80 and cat=70				
naive-naive (44)	11 (0.79)	14	3 (0.50)	6
ARMAX2111pred (70%)	20 (0.71)	28	5 (0.83)	6
ARMAX2111 (70%)	20 (0.71)	28	5 (0.83)	6
ARMAX2111 (90%)	52 (0.85)	61	5 (0.83)	6
ARMAX4211 (90%)	38 (0.81)	47	5 (0.83)	6
SETARX (90%)	60 (0.85)	71	6 (1.00)	6
Data78 and cat=100				
naive-naive (70)	1 (0.50)	2	1 (0.33)	3
ARMAX2111pred (70%)	0 (0.00)	1	1 (0.33)	3
ARMAX2111 (70%)	0 (0.00)	1	1 (0.33)	3
ARMAX2111 (90%)	2 (0.33)	6	3 (1.00)	3
ARMAX4211 (90%)	2 (0.50)	4	2 (0.67)	3
SETARX (90%)	0 (0.00)	6	3 (1.00)	3
Data79 and cat=100				
naive-naive (70)	1 (0.50)	2	1 (1.00)	1
ARMAX2111pred (70%)	1 (0.50)	2	1 (1.00)	1
ARMAX2111 (70%)	1 (0.50)	2	1 (1.00)	1
ARMAX2111 (90%)	4 (0.80)	5	1 (1.00)	1
ARMAX4211 (90%)	4 (0.80)	5	1 (1.00)	1
SETARX (90%)	5 (0.62)	8	1 (1.00)	1
Data80 and cat=100				
naive-naive (70)	5 (1.00)	5	0 (0.00)	1
ARMAX2111pred (70%)	8 (1.00)	8	0 (0.00)	1
ARMAX2111 (70%)	8 (1.00)	8	0 (0.00)	1
ARMAX2111 (90%)	16 (1.00)	16	0 (0.00)	1
ARMAX4211 (90%)	14 (1.00)	14	0 (0.00)	1
SETARX (90%)	16 (0.94)	17	1 (1.00)	1

Table 3: Results for the naive-naive alarm system, compared to the optimal alarm system using three different process models, and one with predicted external signals. The alternative definition of correct alarm is used.

ÖZET

Bu çalışmada Gaussian ARMAX süreçleri için yanlış alarmların sayılarının minimize edilmesi anlamında optimal kestiriciler incelenmiştir. Sonuçlar Baltık Denizi'ndeki su seviyelerinin kestirimleri için uygulanmış ve burada verilen optimal alarm sisteminin daha basit sistemlere göre daha iyi sonuç verdiği gözlenmiştir.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 3, pp. 77 - 81

A USEFUL IDENTITY IN COMBINATORIAL ANALYSIS

Fahrettin ARSLAN
State Institute of Statistics, Necatibey Cad. No:114
06100 , Ankara-TURKEY

Abstract

This paper presents a new identity related to permutations. The identity is considered to be a useful tool in combinatorial analysis and hence in the algebra of some probability distributions.

Key Words: Legendere polynomial, combinatorial analysis

Introduction

A new identity is obtained on the basis of results of Henkel and Lacki (1986). This new identity seems to be useful in computation of sums of factorials. These sums are obviously well-known and the way of calculation of these sums are obtained in terms of Bernoulli polynomial coefficients and Bessel differential equations (c.f. , e.g. Prudnikov, Brickov, Marichev (1981),p.596) and hypergeometric families (c.f. Damjanovic (1986)). However, the new identity we developed has some advantage over the earlier-mentioned approaches in that calculations are carried out in an easier and less time consuming manner.

Given the positive integers n and k, it is well known that one can get the following sums which we come across in probability distributions such as Binomial, Poisson and many other discrete probability distributions.

$$\sum_{k=1}^{n} k^1, \sum_{k=1}^{n} k^2, \sum_{k=1}^{n} k^3, ..., \sum_{k=1}^{n} k^m \text{ where } m \text{ is a positive integer } m \geq 1.$$

However, the derivation of these sums are cumbersome. There are closed forms for calculations of the sums $\sum_{k=1}^{n} k^1, \sum_{k=1}^{n} k^2, \sum_{k=1}^{n} k^3$, but for large n, even the computers have difficulties in calculating the sums $\sum_{k=1}^{n} k^m$ for large n and m i.e, $m > 3$ and for large n. However, using the results of the theorem given below, any sum in the form of $\sum_{k=1}^{n} k^m$ can be easily calculated for any n and m. An identity is obtained below, by means of which these sums can easily be derived by utilizing the equation $z^n - 1 = 0$. Through the factorization techniques, $z^n - 1$, can be written as

$$z^n - 1 = (z-1)(1 + z + z^2 + ... + z^{n-1}). \tag{1}$$

Similarly, given the positive integers m, k and complex number $w = e^{2\pi ik/n}$, (1) and the Euler's identity ($e^{ix} = \cos(x) + i\sin(x)$) yield the following result

$$n = 1 + \sum_{k=1}^{n-1} 2(1 - w^k)^{-1}, \tag{2}$$

which can be obtained through routine calculation. The identity to be discussed in the following section can easily be applied to this problem, so we may get the result quickly.

2. The Identity

As discussed above, there is a useful identity which is often used in many mathematical and statistical problems. Hence, we state the following theorem

Theorem. For a complex number z and positive integers n and m, the following identity holds:

$$\prod_{j=0}^{m}(n-j)z^{n-(m+1)} = (m+1)\sum_{k=0}^{n-1}\prod_{j=0}^{m-1}(k-j)z^{k-m} + (z-1)\sum_{k=0}^{n-1}\prod_{j=0}^{m}(k-j)z^{k-(m+1)}$$

where m and n are positive integers. From this equations we can find

1. $\sum_{k=1}^{n} k^1, \sum_{k=1}^{n} k^2, \sum_{k=1}^{n} k^3, ..., \sum_{k=1}^{n} k^m$, without doing excessive computation.
2. $n = 1 + \sum_{k=1}^{n-1} 2(1 - w^k)^{-1}$ where n and k are positive integers and $w = \exp(2\pi ik/n)$.

Proof. Consider the equation $z^n-1=0$ or $z^n-1=(z-1)(1+z+z^2+...+z^{n-1})$. We may write this equation in a compact form as

$$z^n-1=(z-1)\sum_{k=0}^{n-1}z^k=\sum_{k=0}^{n-1}z^{k-1}-\sum_{k=0}^{n-1}z^k \tag{3}$$

Now taking the first derivative of (3) with respect to z, we have

$$nz^{n-1}=1\sum_{k=0}^{n-1}z^k+(z-1)\sum_{k=0}^{n-1}kz^{k-1}=\sum_{k=0}^{n-1}(k+1)z^k-\sum_{k=0}^{n-1}kz^{k-1} \tag{4}$$

Taking the first derivative of (4) with respect to z, we obtained

$$n(n-1)z^{n-2}=\sum_{k=0}^{n-1}kz^{k-1}+\sum_{k=0}^{n-1}kz^{k-1}+(z-1)\sum_{k=0}^{n-1}k(k-1)z^{k-2},$$

which, after some simplifications, boils down to

$$n(n-1)z^{n-2}=2\sum_{k=0}^{n-1}kz^{k-1}+(z-1)\sum_{k=0}^{n-1}k(k-1)z^{k-2}. \tag{5}$$

Again taking the first derivative of (5) with respect to z, one can obtain

$$n(n-1)(n-2)z^{n-3}=2\sum_{k=0}^{n-1}k(k-1)z^{k-2}+\sum_{k=0}^{n-1}k(k-1)z^{k-2}+\sum_{k=0}^{n-1}k(k-1)(k-2)z^{k-3}$$

$$=\sum_{k=0}^{n-1}k(k+1)(k-1)z^{k-2}-\sum_{k=0}^{n-1}k(k-1)(k-2)z^{k-3}$$

$$n(n-1)(n-2)z^{n-3}=3\sum_{k=0}^{n-1}k(k-1)z^{k-2}+(z-1)\sum_{k=0}^{n-1}k(k-1)(k-2)z^{k-3} \tag{6}$$

and taking the first derivative of (6) with respect to z, we then have

$$\begin{aligned}n(n-1)(n-2)(n-3)z^{n-4}&=4\sum_{k=0}^{n-1}k(k-1)(k-2)z^{k-3}\\ +(z-1)\sum_{k=0}^{n-1}k(k-1)(k-2)(k-3)z^{k-4}&-\sum_{k=0}^{n-1}k(k+1)(k-1)z^{k-3}\\ -\sum_{k=0}^{n-1}k(k-1)(k-2)(k-3)z^{k-4}&\end{aligned} \tag{7}$$

Finally if we take the repeated derivative of (7) with respect to z we can get the desired result

$$\prod_{j=0}^{m}(n-j)z^{n-(m+1)} = (m+1)\sum_{k=0}^{n-1}\prod_{j=0}^{m-1}(k-j)z^{k-m}+(z-1)\sum_{k=0}^{n-1}\prod_{j=0}^{m-1}(k-j)z^{k-(m+1)} \quad (8)$$

If we set $z = 1$, the equation (8) becomes;

$$\prod_{j=0}^{m}(n-j) = (m+1)\sum_{k=0}^{n-1}\prod_{j=0}^{m-1}(k-j)$$

Now, let $m = n-1$ then

$$\prod_{j=0}^{m}(n-j) = n\sum_{k=0}^{n-1}\prod_{j=0}^{m-2}(k-j) \text{ and } \prod_{j=0}^{n-1}(k-j) = n!$$

Therefore,

$$n! = n\sum_{k=0}^{n-1}(k(k-1)(k-2)...(k-(n-2)))$$

$$(n-1)! = \sum_{k=0}^{n-1}\prod_{j=0}^{m-2}(k-j)$$

And hence, $\prod_{j=0}^{m}(n-j) = (m+1)\sum_{k=0}^{n-1}\prod_{j=0}^{m-1}(k-j)$.

Example. For any polynomial $f(k)$, $f(k) = k^m+a_1k^{m-1}+a_2k^{m-2}+...+a_mk$; where $a_1, a_2, ..., a_m$ are integers, the sum $\sum_{k=0}^{n-1} f(k)$ can easily be calculated by using the theorem without doing excessive calculation through the following formula which is a trivial result of the theorem

$$\prod_{j=0}^{m-1}(n-j) = P(n,m+1) = (m+1)\sum_{k=0}^{n-1}\prod_{j=0}^{m-1}(k-j)$$

where

$$P(n,m+1) = \frac{n!}{(n-(m+1))!}.$$

For example if $m = 3$, we may desire to find the value of

$$(m+1)\sum_{k=0}^{n-1}\prod_{j=0}^{m-1}(k-j).$$

Now by putting $m = 3$ in the equation, we have

$$P(n, 3+1) = 4 \sum_{k=0}^{n-1} \prod_{j=0}^{2} (k-j),$$

and, after some simplifications,

$$\frac{1}{4} P(n,4) = \sum_{k=0}^{n-1} k(k-1)(k-2) = \sum_{k=0}^{n-1} k^3 - \sum_{k=0}^{n-1} 3k^3 + \sum_{k=0}^{n-1} 2k \tag{9}$$

$$= \frac{n^2(n-1)^2}{4} - \frac{3n(n-1)(2n-1)}{6} + \frac{2n(n-1)}{2} = \frac{n(n-1)}{4}(n^2 - 5n + 6)$$

As it can be noted, the RHS of the equation (9) is cumbersome to tackle but the LHS of the equation is less time consuming and much easier to compute. Finally if we set $m = n - 1$, then we can write a more compact form of (8)

$$P(n, m+1) = n! = n \sum_{k=0}^{n-1} \prod_{j=0}^{n-2} (k-j) \quad \text{or} \quad P(n,m) = n \sum_{k=0}^{n-1} \prod_{j=0}^{n-2} (k-j)$$

and $m - 1 \leq n$. Note that it will be an interesting task to seek for a relationship between this new identity and Legendere polynomial approximation to sampling distribution.

References

[1] Bhonsle, B.R. (1957), On a Series of Rainville Involving Legendere Polynomials, Proc. Amer. Math. Soc., 8, 10-14

[2] Damjanovic, S.LJ. (1986), Siam Review, Vol.28, Number 1.

[3] Earl D.R. (1960), Special Functions The Macmillian Company, New York

[4] Henkel, M. , Lacki, J. (1986), Siam Review, Vol.28, Number 1.

[5] Louis, L.P. , Louis, I.G. and Lasher, S. (1976), Elements of Complex Variables Holt, Richard and Winston, New York

[6] Lebedev, N.N. and Silverman, A.S. (1972), Special Functions and Their Applications, Dover Publications Inc. New York

[7] Slomson, A. (1990), An Introduction to Combinatorics, Chapman and Hall, London

[8] Prudnikov, A.P. , Brickov, Yu.A. , Marichev, O.I. (1981), Integrals and Series, Moskow (in Russian).

ÖZET

Bu çalışmada permutasyonlara ilişkin yeni bir özdeşlik sunulmuştur. Bu özdeşliğin kombinatorik hesaplamalarda ve dolayısıyla bazıolasılık hesaplamalarında yararlıolabileceği düşünümektedir.