İstatistik, Journal of the Turkish Statistical Association
1998, Vol. 1, No. 1, pp. 5 - 11

THE MANY FACETS OF STATISTICAL SCIENCE

Edmond Malinvaud
INSEE-CREST, France

Abstract

There are three main functions of statisticians; observing complex phenomena, providing simple representations of the facts, and inducing conclusions of some generality from data. Working in different fields, statisticians share a common culture and a common set of ethical principles, along these functions, which is based on a common discipline that orders objectivity, completeness and confidentiality.

Key Words: Statistical science, official statistics

1. Introduction.

In the modern world one finds statisticians in many sectors, in many industries, in many activities and with many different functions. One finds some of them in research in the natural sciences; they then are planning experiments, or searching for appropriate stochastic models that will fit the observed results; or still analyzing the implications of the models. One finds statisticians in research in the social sciences; they then are planning surveys, or analyzing large data sets, or still forecasting evolutions that will shape future social life. One finds statisticians in manufacturing industries; they then are in charge of quality control or human resource management. One finds statisticians in services; they are then involved in market research or in building softwares for processing large data sets. Last but not the least, far from that, one finds quite a few statisticians also in official statistical agencies; then they plan censuses, sample surveys and retrieval of information from administrative records; they organize the processing of large masses of data; they produce and disseminate economic indices, national accounts, social indicators and the like.

Although spread in many places, working in many different environments and having many different responsibilities, all statisticians contribute to a common science, and also rely on this science. All statisticians share a common culture based on this science, on the approaches that it implies and the norms of conduct that it recommends.

Knowledge of the statistical science and familiarity with the statistical culture are almost everywhere underdeveloped. People are not used the think that many features in nature and in human societies are quite variable, from one unit to another and through time for the same unit. People are not used to face the consequences of this variability, in particular the fact that it always introduces uncertainty, with which we have to cope. I may say in particular that the statistical culture is underdeveloped in my country, France, notwithstanding the efforts spent for it during the past half a century at least.

In this address I cannot deal with the full range of subjects that this account of the present situation suggests as interesting. The focus will be on three main general tasks that have to be performed, that contribute to the progress of statistical science and whose methodology is given by statistical science. These tasks are:

- observe complex phenomena for which variability cannot be ignored;
- provide simplified, synthetic and schematic but objective representations of the facts;
- induce from the data conclusions of some generality and then use appropriate techniques developed by mathematical statistics.

In the conclusion I shall say a little more about the implications for the statistical culture.

2. Observe Complex Phenomena

Observation is the first stage of knowledge. Collection of data on many units is required when the facts to be known vary from one unit to another one. Then statisticians are called on.

It would be too long to discuss here the various contexts in which such a collection of data occurs. It is appropriate to limit attention to collection by official statistical agencies.

When I think of it in retrospect, after same 45 years of personal involvement with it, the first comment coming to my mind is to stress the tremendous increase in the demand for statistical data, an increase that far exceeds what I could have anticipated, an increase that could not have been met without the technical progress in data processing and, to a more limited extent, the technical progress also in data gathering. This explosion of demand for statistical data has two sources.

In the first place, the needs for knowledge in a new field of investigation are commonly underestimated. One first believes that few concepts suffice to describe the world, that each concept has a simple definition and that data are easily accessible for measuring the concept. But one always realizes that the first measurements that are provided fulfill only a part of what is required and fulfill it imperfectly. They are subject to errors that one would want to control by approaching measurement in some other way; they apply definitions that are too loose for the level of accuracy required in some cases; they catch only one aspect of a phenomenon and one is interested in other aspects, and so on.

In the second place, in our modern societies the world is being made increasingly complex, particularly in the socio-economic field where new forms of life or of organization appear, in which new products and new markets are created, in which new regulations or new types of transfer are decided. Statisticians are requested to keep up with these developments.

Faced with such a rapid increase in demand, statisticians are often found to be too slow in answering. One does not realize that, in order to achieve the required level of accuracy, statistical operations have to be well accepted by those that contribute by their answers, well planned, well controlled. Moreover a single measurement is hardly ever enough, since one wants to compare. Hence, one has to view the setting up of a new statistical operation as an investment.

Speaking of the techniques used in statistical data collection would be long if I had to be fairly complete. Here, I shall mention only two quite different types of techniques: probability sampling and processing of administrative data. In both types reliance on the technique is imposed for efficiency. i.e. for having some needed information at the lowest possible cost.

If we observe a sample rather than the full universe it is in order to save on costs wherever this solution is possible without fully jeopardizing accuracy of the results. But in order to be sure to achieve accuracy, one needs the sample to be representative enough of the whole universe. Selecting the units to be observed at random. within the confines of a well patterned sampling frame, is the only way that guarantees representativeness of the sample.

One easily understands that this technique of random selection and of definition of sampling frames relies on a precise methodology. Thinking about the number of questions it raises and the diversity of the situations in which it is applied, one should not be surprised to learn that this methodology is the subject of a whole branch of statistical science.

If we now more and more produce statistics from data found in the files of various administrative offices, it is because direct collection of data from people or firms is costly, both for those collecting the data and for those providing them. Once the collection was made by an administration for its own needs, statisticians save much on costs if they also can use these data. The solution requires not only that statisticians have access to the files, but also that the data correspond to what statisticians are requested to measure. This last condition is hardly ever perfectly satisfied.

As with probability sampling, one easily imagines that a methodology exists for this statistical processing of administrative data. I do not insist on it. I just want to add a comment of another nature, namely that, for the solution to work, a rigorous organization is necessary, not only in the statistical office in charge of processing the data, but also in the administration collecting the data and using them for its own purpose. Efficiency is the joint business of both institutions.

3. Provide Simple Representations

Statisticians have to inform about facts and phenomena. They collect large quantities of data. Directly discharging these data would not much help the users, who could not easily make sense of them. It belongs to statisticians to extract the useful information.

Of course, what is useful may depend on the user. Statisticians have the duty to also serve someone who needs the raw data because he wants to apply to them a particular treatment for his own purpose. But there is enough in common in the needs of most users for general treatments made by statisticians to have proved indispensable.

The purpose then is to provide simple and synthetic representations of the facts. For serving this purpose methods were found and developed. Improvements are often brought to previously existing methods. New methods are also introduced. There is again a methodology, which is another part of statistical science. We may as well speak of several methodologies, because several kinds of operations have been found necessary in order to provide appropriate representations. I shall briefly consider three such parts, which I may call descriptive statistics, taxonomy and theory of indices and aggregates.

When data are available for a number of units, individuals or firms for instance, it is often the case that one is not interested in the identities of particular units and that one considers them as interchangeable from the point of view of the facts or phenomena to be known. One is then only interested in what is called the statistical distribution of the data.

If only one numerical character of the units had been observed, such as the age of the individuals, the statistical distribution would give the proportion of the units for which the value of this character does not exceed any given quantity one may want to consider. Fróm this distribution one can compute the average value of the character or indicators of the dispersion of the values of the character for the various units, such as the so-called standard deviation of the character.

Usually statistical distributions are much more complex object because many characters have been observed on each unit and some of these characters are not numerical but qualitative, such as the sex of the individual or the kind of job he or she is holding. Techniques exist in order to deal with these complex distributions, so as to extract from the data significant quantities or indicators, giving some useful information in the same way as is done by average values or standard deviations. These techniques belong to the broad field often called "data analysis", a field that has benefited, over the last century and since the last world war, of the work of many mathematicians.

Although not interested in the identification of the units, one often wants to be able to classify them among relevant groups and to look at the statistical distribution within each group. In some cases the grouping is obvious, such as when one wants to distinguish by sex. In other cases it is not immediately given and one would like to know how to proceed. For instance the kind of job held by an individual is defined by many features: the activity of the establishment where he works, the qualification of the job, perhaps also the type of employment contract and of work schedule. Moreover for each feature, such as the activity or the establishment, there are many possible modalities. It would serve no purpose to define groups at a greatest level of detail, because there would be too many groups.

The problem of knowing how to define useful classifications is the object of taxonomy and has foremost importance for statisticians. There are two quite different approaches to it, serving different kinds of needs. According to the first approach one considers the whole available data set and one defines groups by clustering units, according to proximities in the values taken by the various characters for the different units; the object then is to have, within each group, units that, as far as possible, are similar with respect to these values.

So, data analysis is used for defining classifications.

Instead of looking at the data to be summarized, the other approach considers the needs of the users. The principle is to define classifications by trying to best fulfill these needs and in order to do so to ask users. It is commonly the case that different users have different needs and wishes; they request different and conflicting classifications. So, a kind of compromise has to be found. This is the approach commonly used in official statistics, an approach applying procedures that have been learned from experience.

The definition, production and dissemination of aggregate indices and indicators, as well as of national economic accounts are now notorious statistical operations. They have, involved in the past and are still involving now for their progress, three kinds of specialists: (i) statisticians looking for the desirable formal properties of the aggregates and trying to derive formulas that would have these properties, as well as possible; (ii) subject matter specialists, mostly economists but also sociologists,studying what meaningful definitions ought to be and what would be most appropriate for research on economic and social phenomena; (iii) official statisticians concerned by the efficiency and accuracy of the methods to be applied in the regular and timely production of aggregates.

4. Induce From the Data

Knowing a phenomenon is not only to have well observed past events and present structures; it is also to be able to explain at least some aspects of these events and structures; it is to be able to make some production about the effects of occurring changes in the environment or of deliberate actions. Knowing a phenomenon then is to have an understanding of how it works in general and not only to have seen its manifestation in the particular circumstances that held when data were collected.

This general knowledge is built progressively by inference from statistical observation. An important part of statistical science is quite naturally concerned by how conclusions with a general meaning can be drawn from data. This part is currently called mathematical statistics, although mathematics are also used in descriptive statistics, as we saw.

Mathematical statistics is directly facing a fundamental difficulty, which has been widely discussed in the philosophy of sciences under the name "the problem of induction": How can one logically draw a general conclusion from particular observations? Entering a discussion on this logical issue would, of course, take us too far afield.

Let me simply say that mathematical statistics admits the existence of a model ruling the phenomenon under study. The model, which of course is taken as applying to the data, was not fully known to the statistician before he had looked at the data to be analyzed; the values of some parameters was unknown or incompletely known; but he knew already enough about the phenomenon to be able to write down the form of the model.

The model is not deterministic; it is not meant to provide a complete explanation of the data or the phenomenon; it recognizes the presence of some unexplained errors. The errors are taken as random and the model as probabilistic.

The object of statistical inference then is to derive from the data conclusions about the unknown parameters of the model, for instance to estimate them. The conclusions or estimation are inevitably subject to some margins of error, which are random in the same way as the error of the model.

Mathematical statistics has to find formulas for drawing inference from the data, for instance formulas for estimating the parameters of the model. It has to study the margins of error and to characterize them. It has to compare the respective performances of alternative formulas. This has to be done for all kinds of models and for all types of inferences that have to be considered in the various fields where statistics are used.

This explains why mathematical statistics grew so much as to become a field of its own, with its scientific journals, with its scientific associations, with its departments in many well known universities. Drawing on mathematics and probability theory, research is very active in this field and is attracting excellent people. It is stimulated by many applied inference problems: in economics, in psychology, in biology, in medicine, in agricultural and environmental sciences, and so on.

5. The Statistical Culture

Considering how different are the functions of statisticians, it may be surprising to learn that statisticians nevertheless share common culture. As I said at the beginning, this culture is based both on a common discipline, which is suited for all these functions, and on a common set of ethical principles. Let us look briefly at these two aspects of statistical culture.

The scientific discipline concerns how to deal with facts and phenomena that are subject to the influence of uncontrolled and variable factors. In order to grasp these facts and phenomena, in order to overcome the variability to which they are subject, one needs to have observations collected on many units. With few data the main tendencies would not appear; with large sets of data they are revealed to those who correctly look at the evidence. All statisticians have to be competent in dealing with large sets of data; they must know how to collect them in such a way that they are informative with respect to the underlying facts and phenomena; they must know how to process them so as to extract this relevant information.

Such a competence requires to think in terms of probabilities. Thinking in terms of purely deterministic causation, when facts and phenomena are subject to many uncontrolled and variable factors, would be much too complex, hence inefficient. A probabilistic approach is unavoidable. It is sometimes implicit, but a little reflection shows that randomness is always admitted; a little reflection also shows that making explicit reference to the assumed randomness helps to work correctly.

The ethical principles embodied in the statistical culture contain those that are common to all scientists. We may summarize this first group of principles under the label of objectivity: statisticians must aim at true statements and strive to prevent their subjective evaluation to interfere.

The ethical principles of statistics also impose two, sometimes conflicting, requirements. Statisticians must inform as completely as then can. But they must also protect the confidentiality of individual data, i.e. they must not disclose information on individual units. In most cases the statistical information requested by the users does not create any problem with respect to confidentiality because it concerns large populations. Problems may, however, arise with quite specific data on small geographical areas or with data concerning activity in industries where only one or two large firms operate. As usual with ethical problems, the profession then has to find appropriate rules of conduct and to make them known.

I am conscious that this tour around the many facets of statistical science has been fast. But I am sure that my fellow statisticians are ready to present our discipline at more leisure to those who want to know it more deeply.

ÖZET

İstatistikçilerin üç önemli işlevi vardır; karmaşık olayları gözlemek , gerçeklerin basit temsillerini vermek ve verilerden bazı genelleyici sonuçlara ulaşmak. Çeşitli alanlarda çalışan istatistikçiler bu işlevlerinde ortak bir kültürü ve ortak bir ahlaki ilkeler kümesini paylaşırlar ki bunlar objektifliği, bütünlüğü ve gizliliği emreden ortak bir disipline dayalıdır.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 1, pp. 13 - 19

ON DUALITY BETWEEN CONTROL AND ESTIMATION PROBLEMS

Agamirza Bashirov, Huseyin Etikan and Nidai Semi
Department of Mathematics
Eastern Mediterranean University
Gazimagosa - North Cyprus

Abstract

In this paper, the dual control problems for the linear estimation problems with noises having pointwise and distributed shifts are derived.

Key Words: Estimation, filtering, smoothing, prediction, duality, control.

1. Introduction

Estimation theory is a widely used concept in engineering studies such as space engineering, electronics, geophysics, etc. The estimation problems consist in estimating an unobservable signal process x at instant t on the basis of observation data z on time interval $[0, \tau]$. Depending on the relation of t and τ the three kinds of the estimation problems are considered: (a) filtering, if $t = \tau$, (b) smoothing, if $t < \tau$ and (c) prediction, if $t > \tau$.

The underlying ideas of the linear estimation were defined by Kalman and Bucy [1,2]. In particular, in Kalman [1], the duality between the linear filtering and linear regulator problems is obtained. This result determines the general approach for synthesizing the optimal estimators through the optimal controls in the dual linear regulator problems. This approach is used in Bashirov and Mishne [3,4] for synthesizing the optimal filters in the linear filtering problems when the noises on the signal and the observations have pointwise and distributed shifts.

Application of the approach, based on duality, in studying the estimation problems requires the construction of the dual optimal control problem. The aim in this paper is the construction of the dual optimal control problems to the linear smoothing and prediction problems, with noises having pointwise and distributed shifts, and proving the duality theorem.

2. Notations

In this paper, the following notations are used:

$(\Omega, \mathcal{F}, \mathbf{P})$	:	complete probability space
H, X, Z	:	real seperable Hilbert spaces
$\mathcal{L}\ (H,Z)$	:	space of linear bounded operators on H to Z
$\mathcal{L}(X) = \mathcal{L}\ (X\ ,\ X)$		
$< \cdot, \cdot >$	:	inner product
$\|\| \cdot \|\|$	:	norm
$B_2(a, b; \mathcal{L}\ (H, X))$	:	class of strongly measurable $\mathcal{L}\ (H, X)$-valued functions F on $[a,b]$ with $\int_a^b \|\| F_t \|\|^2\, dt < \infty$
$B_\infty\ (a, b; \mathcal{L}\ (H, X))$	:	class of strongly measurable $\mathcal{L}\ (H, X)$-valued functions F on $[a,b]$ with essup$\|\| F_t \|\| < \infty$
$L_2(a, b; Z)$	:	space of (equivalence classes of) Lebesque measurable and square integrable functions on $[a, b]$ to Z
A^*	:	adjoint to operator A
$\mathbf{E}$	:	expectation
cov(x,y)	:	covariance operator of random variables x and y
cov $x=$ cov(x, x)		
$\chi_{[a,b]}(s)$	:	characteristic function of the set $[a, b]$

3. Setting of Linear Estimation Problem for Shifted White Noises

Let (x_t, z_t) be a partially observed linear stochastic system

$$\dot{x}_t = Ax_t + \Phi w_t, \qquad t > 0, \qquad x_0 \text{ is given,} \tag{1}$$

$$\dot{z}_t = Cx_t + \Psi w_{t+\epsilon}, \qquad t > 0, \qquad z_0 = 0, \tag{2}$$

where x_t and z_t are the signal and observation processes, respectively, at time t, A is the infinitesimal generator of the strongly continuous semigroup $\mathcal{U}_t$, $t \geq 0$, $C \in \mathcal{L}(X, Z)$, $\Phi \in \mathcal{L}(H, X)$, $\Psi \in \mathcal{L}(H, Z)$, w is H-valued white noise process with $\mathbf{E}w_t = 0$, cov$(w_t, w_s) = \overline{W}\delta(t-s)$ in which δ is the Dirac's delta function, x_0 is a random variable with $\mathbf{E}x_0 = 0$ and cov$x_0 = P_0$. We suppose that x_0 is independent of w and $\epsilon > 0$. The system (1) - (2) can be written in the following integral form as well:

$$x_t = \mathcal{U}_t x_0 + \int_0^t \mathcal{U}_{t-s}\Phi w_s ds, \qquad t \geq 0, \tag{3}$$

$$z_t = \int_0^t Cx_s ds + \int_0^t \Psi w_{s+\epsilon} ds, \qquad t \geq 0. \tag{4}$$

Let $\tau > 0$ and $t > 0$ be two time moments. The best linear estimation $\hat{x}_t^\tau$ of x_t on the basis of observations z_s on $0 \leq s \leq \tau$ is a random variable in the form

$$\hat{x}_t^\tau = \int_0^\tau K_s \dot{z}_s ds; \quad K \in B_2(0,\tau;\mathcal{L}(Z,X)), \tag{5}$$

that minimizes the error

$$\mathbf{E} \parallel x_t - \int_0^\tau G_s \dot{z}_s ds \parallel^2 \tag{6}$$

over all $G \in B_2(0,\tau;\mathcal{L}\,(Z,X))$.

Lemma 1 : *Under the above conditions $\hat{x}_t^\tau$, defined in (5), is a best linear estimation for the system (3) and (4), if and only if K satisfies the following Wiener-Hopf equation*

$$K_s V + \int_0^\tau K_r C \Lambda_{r,s} C^* dr + \int_0^{\max(0,s-\epsilon)} K_r N^* \mathcal{U}_{s-r-\epsilon}^* C^* dr \; +$$

$$\int_{\min(\tau,s+\epsilon)}^\tau K_r C \mathcal{U}_{r-s-\epsilon} N dr = \Lambda_{t,s} C^* + \chi_{[0,t-\epsilon]}(s) \mathcal{U}_{t-s-\epsilon} N, \;\; 0 \leq s \leq \tau, \tag{7}$$

where

$$\Lambda_{r,s} = \mathcal{U}_r P_0 \mathcal{U}_s^* + \int_0^{\min(r,s)} \mathcal{U}_{r-\sigma} W \mathcal{U}_{s-\sigma}^* d\sigma, \qquad r \geq 0, \;\; s \geq 0, \tag{8}$$

and

$$W = \Phi \overline{W} \Phi^*, \;\; V = \Psi \overline{W} \Psi^*, \;\; N = \Phi \overline{W} \Psi^*. \tag{9}$$

Proof. By the orthogonal projection lemma (see Curtain [5]) (5) is the best linear estimation of x_t on the base z_s, $\;0 \leq s \leq \tau$, if and only if

$$cov \left(x_t - \hat{x}_t^\tau, \int_0^\tau G_s \dot{z}_s ds \right) = 0, \qquad \text{for all } G \in B_2(0,\tau;\mathcal{L}(Z,X)).$$

Evaluating the above equality and using the arbitrariness of G one can obtain the equation (7) with (8) and (9) for K and vice versa.

Note that the equation (7) is not constructive in synthesizing of K. Nevertheless, we will use (7) in proving of the duality theorem.

4. Dual Control Problem in the Case of Shifted White Noises

Consider the control problem

$$\xi_s = -\left\{ \begin{array}{ll} \mathcal{U}^*_{s-\max(0,\tau-t)}f, & s \geq \tau - t \\ 0, & s < \tau - t \end{array} \right\} + \tag{10}$$

$$\int_{\max(0,t-\tau)}^{\max(s,t-\tau)} \mathcal{U}^*_{s-r} C^* \eta_r dr, \quad 0 \leq s \leq \max(\tau, t),$$

and

$$J(\eta) = \left\langle \xi_{\max(\tau,t)}, P_0 \xi_{\max(\tau,t)} \right\rangle + \int_0^{\max(\tau,t)} \langle \xi_s, W\xi_s \rangle \, ds + \tag{11}$$

$$\int_{\max(0,t-\tau)}^{\max(\tau,t)} \langle \eta_s, V\eta_s \rangle \, ds + 2\int_{\max(\varepsilon,t-\tau)}^{\max(\varepsilon,\tau,t)} \langle \eta_s, N^*\xi_{s-\varepsilon} \rangle \, ds,$$

where ξ is a state process, $J(\eta)$ is the functional to be minimized, η is a control from the set of admissible controls $L_2(\max(0, t-\tau), \max(\tau, t); Z)$, f is an arbitrary fixed vector in X, N, V, W are defined in (9), $\mathcal{U}$, P_0 and ε are as defined in Section 3.

Note that in the control problem (10) - (11), the cases $t - \tau = 0$, $t - \tau < 0$ and $t - \tau > 0$ are available and in fact (10) - (11) is the combination of these three cases.

Lemma 2 : *Under the above conditions η is an optimal control in the control problem* (10) - (11) *if and only if it satisfies the following equation*

$$V\eta_s + \int_{\max(0,t-\tau)}^{\max(\tau,t)} C\Lambda^*_{\max(\tau,t)-r,\max(\tau,t)-s} C^*\eta_r dr \; +$$

$$\int_{\max(0,t-\tau)}^{\max(0,s-\varepsilon,t-\tau)} N^*\mathcal{U}^*_{s-r-\varepsilon} C^*\eta_r dr + \int_{s+\varepsilon}^{\max(s+\varepsilon,\tau,t)} C\mathcal{U}_{r-s-\varepsilon} N\eta_r dr = \tag{12}$$

$$C\Lambda^*_{t,\max(\tau,t)-s} f + \chi_{[\max(\varepsilon,\varepsilon+\tau-t),\infty]}(s) N^*\mathcal{U}^*_{s-\varepsilon+\min(0,t-\tau)} f,$$

where $\max(0, t-\tau) \leq s \leq \max(\tau, t)$ and $\Lambda_{r,s}$ *is defined in* (8).

Proof. If η is the optimal control then $\mathcal{J}(\eta + \lambda\Delta\eta) - \mathcal{J}(\eta) \geq 0$ for all real numbers λ and admissible controls $\Delta\eta$. Evaluating the above inequality and using the arbitrariness of η and $\Delta\eta$ one can obtain the equation (12) for η and vice versa. Note that the proof of Lemma 2 in general case has difficulties. To overcome this, it is convenient to prove it for $t-\tau<0$ and $t-\tau>0$ cases separately. The case $t-\tau=0$ needs no separate consideration since it can seen in any of the above cases.

Theorem 1 : *Under the above conditions* (5) *is the best linear estimation for the system* (3) - (4) *if and only if* $\eta_s = K^*_{\max(\tau,t)-s}f$, $\max(0, t-\tau) \le s \le \max(\tau, t)$ *is an optimal control in the control problem* (10) - (11).

Proof. Let η be optimal in the control problem (10) - (11). Then, by Lemma 2, it satisfies the equation (12). Substituting $\eta_s = K^*_{\max(\tau,t)-s}f$ in (12), using arbitrariness of f and taking adjoint in both sides of (12) one can obtain that K satisfies (7). So, by Lemma 1, the best linear estimation for the system (3) -(4) has the form (5). Conversely, if (5) is the best linear estimation for the system (3) - (4), then by Lemma 1 K in (5) satisfies (7). Taking adjoint in both sides of (7), one can obtain that $\eta_s = K^*_{\max(\tau,t)-s}f$ satisfies (12) for all $f \in X$. So, by Lemma 2 the control $\eta_s = K^*_{\max(\tau,t)-s}f$ is optimal in the control problem (10) - (11).

Theorem 1 states the duality between the estimation problem for the system (3) - (4) and the control problem (10 - (11). By this theorem synthesizing of K in (5) is equivalent to finding optimal control as in Theorem 1.

5. Setting of Linear Estimation Problem for Shifted White and Wide-Band Noises

In this and next section, the results of the previous sections will be modified to the linear estimation problem (13) - (14) defined below.

Consider the partially obsered linear stochastic system

$$x_t = \mathcal{U}_t x_0 + \int_0^t \int_{\max(0,s-\varepsilon)}^s \mathcal{U}_{t-s}\Phi_{\theta-s}w_\theta d\theta ds, \quad t \ge 0, \tag{13}$$

$$z_t = \int_0^t Cx_s ds + \int_0^t \Psi w_s ds, \quad t \ge 0, \tag{14}$$

where the conditions of Section 3 hold and $\Phi \in B_\infty(-\epsilon, 0; \mathcal{L}(H, X))$. One can verify that the noise

$$\varphi_t = \int_{\max(0,t-\epsilon)}^t \Phi_{\theta-t}w_\theta d\theta \tag{15}$$

of the signal system (13) is obtained by distributed shift of the observation noise and satisfies

$$\operatorname{cov}(\varphi_t, \varphi_s)\begin{cases} = 0, & |t-s| \ge \varepsilon \\ \ne 0, & |t-s| < \varepsilon. \end{cases} \tag{16}$$

So φ is a wide-band noise and one can specify the system (13) - (14) as a system with shifted white and wide-band noises. Note that in the case of system (3) - (4) the pointwise shift of the white noises was used.

The following result can be proved similar to the proof of Lemma 1.

Lemma 3 : *$\hat{x}_t^\tau$, defined in* (5), *is a best linear estimation for the system* (13) - (14) *if and only if K satisfies the following Wiener-Hopf equation*

$$K_sV + \int_0^\tau K_r C\Lambda_{r,s}C^*dr + \int_0^s \int_{\max(-\epsilon,r-s)}^0 K_r N_\sigma^* \mathcal{U}_{s-r+\sigma}^* C^* d\sigma dr +$$

$$\int_s^\tau \int_{\max(-\epsilon,s-r)}^0 K_r C\mathcal{U}_{r-s+\sigma} N_\sigma d\sigma dr = \tag{17}$$

$$\Lambda_{t,s}C^* + \int_{\max(-\epsilon,s-t)}^{\max(0,s-t)} \mathcal{U}_{t-s+\theta} N_\theta d\theta, \quad 0 \le s \le \tau,$$

where

$$\Lambda_{r,s} = \mathcal{U}_r P_0 \mathcal{U}_s^* + \int_0^{\min(r,s)} \int_{\max(-\epsilon,\theta-r,)}^0 \int_{\max(-\epsilon,\theta-s)}^0 \mathcal{U}_{r-\theta+\sigma} W_{\sigma,\alpha} \mathcal{U}_{s-\theta+\alpha}^* d\alpha d\sigma d\theta, \tag{18}$$

$$W_{\sigma,\alpha} = \Phi_\sigma \overline{W} \Phi_\alpha^*, \quad N_\theta = \Phi_\theta \overline{W} \Psi^*, \quad V = \Psi \overline{W} \Psi^*. \tag{19}$$

6. Dual Control Problem in the Case of Shifted White and Wide-Band Noises

Consider the state process ξ, defined in (10), and the functional, where η is a control from the set of admissible controls $L_2(\max(0,t-\tau), \max(\tau,t); Z)$ and W,N,V are as defined in (19).

The following result can be proved similar to the proof of Lemma 2.

Lemma 4 : *Under the above conditions η is an optimal control in control problem* (10) *and* (20) *if and only if it satisfies the following equation*

$$V\eta_s + \int_{\max(0,t-\tau)}^{\max(\tau,t)} C\Lambda^*_{\max(\tau,t)-r,\max(\tau,t)-s} C^* \eta_r dr +$$

$$\int_{\max(0,t-\tau)}^{s} \int_{\max(-\epsilon,r-s)}^0 N_\sigma^* \mathcal{U}_{s-r+\sigma}^* C^* \eta_r d\sigma dr +$$

$$\int_s^{\max(\tau,t)} \int_{\max(-\epsilon,s-\tau)}^{0} C\mathcal{U}_{r-s+\sigma} N_\sigma \eta_r \, d\sigma \, dr = \tag{20}$$

$$C\Lambda^*_{t,\max(\tau,t)-s} f \; + \int_{\max(-\epsilon,-s,\tau-t-s)}^{\max(0,\tau-t-s)} N^*_\theta \mathcal{U}^*_{s+\theta-\max(0,\tau-t)} f \, d\theta,$$

where $\max(0, t-\tau) \le s \le \max(\tau, t)$ *and* $\Lambda_{r,s}$ *is defined in* (18).

Theorem 2 : *Under the above conditions* (5) *is the best linear estimation for the system* (13) - (14) *if and only if* $\eta_s = K^*_{\max(\tau,t)-s} f$, $\max(0, t-\tau) \le s \le \max(\tau, t)$, *is an optimal control in the control problem* (10) *and* (20).

References

[1] Kalman, R.E. (1960) *A New Approach to Linear Filtering And Prediction Problems.* J. Basic Engineering, 82, 35-45

[2] Kalman R.E. and Bucy R.S. (1961) *New Results in Linear Filtering And Prediction Theory.* J. Basic Engineering, 83, 65-108

[3] Bashirov, A.E. (1988) *On Linear Filtering Under Dependent Wide-Band Noise.* Stochastics, 23, 413-437.

[4] Bashirov, A.E. and Mishne L.R. (1991) *On Linear Filtering Under Dependent White Noises.* Stochastics and Stochastic Reports, 35, 1-23.

[5] Curtain, R.F. (1970) *Estimation Theory for Abstract Evalution Equations Excited by General White Noise Processes.* SIAM J. Contr. and Optim., 14, 1124-1150.

ÖZET

Bu çalışmada, kestirim problemlerinden olan düzleme ve öngörü problemleri ele alınmış ve bu problemlerin noktasal ve dağıtılmış kaydırma içeren gürültüler için çifteş denetim problemleri türetilmiştir. Kısmen gözlemlenebilen doğrusal türel sistemi için dikey izdüşüm ön kuramı ve artırım yöntemi kullanılmış ve en iyi doğrusal kestirim bulunmuştur.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 1, pp. 21 - 27

THE UNIFORMLY BETTER UNBIASED ESTIMATOR FOR THE SLOPE PARAMETER IN SIMPLE LINEAR REGRESSION WITH ONE- FOLD NESTED ERROR

Bilgehan Güven
Middle East Technical University
Department of Statitiscs
06531, Ankara Turkey

Abstract

Simple linear regression with one fold nested error is defined as one way covariance classification in which the treatment effects are random. As in one way covariance classification, there are two independent unbiased estimators of the slope parameter. The condition for obtaining the more efficient unbiased estimator of the slope parameter and its numerical verification are presented.

Key Words: Recovery Block Information, Uniformly Minimum Variance Unbiased Estimator, More Efficient Estimator.

1. Introduction

Fuller and Battese (1973) introduced the following model

$$Y_{jk} = \mu + \beta x_{jk} + a_j + e_{jk} \quad j = 1, 2 \ldots, J \quad k = 1, 2, \ldots, K$$

where Y_{jk} and x_{jk} are dependent and independent variables respectively, μ and β are the unknown parameters. The terms a_j and e_{jk} are independent normal variables with zero expectations and variances σ_a^2 and σ_e^2 respectively. This model is called simple linear regression with one fold nested error.

The model can be used for an experiment on J randomly selected units with K subunits taken on each selected unit. The term a_j represents an error component associated with units and the term e_{jk} represents an error component associated with the subunit. A classical example of the model is two stage sampling.

Fuller and Battese described the method of generalized least squares estimation for the unknown parameters of the model. Tong and Cornelius (1989, 1991) studied both the estimation and the hypothesis testing of the slope parameter β.

As it is explained in Section 2, the uniformly minimum variance unbiased (UMVU) estimator for the slope parameter β does not exist when the variance components σ_a^2, σ_e^2 are unknown. However there are two different unbiased estimators for β. The main purpose of this article is to find the conditions that an unbiased estimator of β obtained by combining two unbiased estimators of β has smaller variance than any of two unbiased estimators.

2. Uniformly Minimum Variance Unbiased Estimator

Simple linear regression with one fold nested error can be considered as one-way analysis of covariance model in which block effects are random. Then we have two independent unbiased estimators of β given by

$$\hat{\beta}_a = \frac{SS_a(Yx)}{SS_a(xx)} \quad \hat{\beta}_e = \frac{SS_e(Yx)}{SS_e(xx)} \tag{1}$$

where

$$SS_a(xY) = K\sum_{j=1}^{J}(Y_{j.} - Y_{..})(x_{j.} - x_{..})$$

$$SS_e(xY) = \sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{jk} - Y_{j.})(x_{jk} - x_{j.})$$

$x_{j.} = \sum_{k=1}^{K} x_{jk}/K$, $x_{..} = \sum_{j=1}^{J}\sum_{k=1}^{K} x_{jk}/JK$. $SS_a(xx)$, $SS_e(xx)$, $Y_{j.}$, and $Y_{..}$ are defined similarly, $\hat{\beta}_a \sim N(\beta, (\sigma_e^2 + K\sigma_a^2)/SS_a(xx))$ and
$\hat{\beta}_e \sim N(\beta, \sigma_e^2/SS_e(xx))$.

The minimal sufficient statistic for the four unknown parameters based on JK observations is

$$(SS_a(YY), SS_e(YY), SS_a(xY), SS_e(xY), Y_{..}) \tag{2}$$

and says that the dimension of the natural parameter space is five. However, the dimension of the parameter space is four. It follows that the minimal sufficient statistic in (2) is not complete and consequently, the Rao-Blackwell theorem can not be used for obtaining the UMVU estimator of β. Therefore, the theorem 1.1 in the page 77 of Lehman (1983) is used

The class of unbiased estimators of β is:

$$\mathcal{U} = \{\hat{\beta}_U = w\hat{\beta}_a + (1 - w)\hat{\beta}_e \quad w \in [0,1]\}.$$

Then $\hat{\beta}_U$ is the UMVU estimator of β if

$$E[\delta\hat{\beta}_U] = w\frac{\sigma_e^2 + K\sigma_a^2}{SS_a(xx)} - (1 - w)\frac{\sigma_e^2}{SS_e(xx)} = 0$$

where $\delta = \hat{\beta}_a - \hat{\beta}_e$ is the unbiased estimator of 0. It implies that the weight function w for the UMVU estimator of β must be:

$$w = \frac{\sigma_e^2/SS_e(xx)}{\sigma_e^2/SS_e(xx) + (\sigma_e^2 + K\sigma_a^2)/SS_a(xx)}. \tag{3}$$

An UMVU estimator is unique when sufficient statistics is either complete or not complete (see Barankin , 1949). So it is impossible to obtain the UMVU estimator of β without knowing the variance components.

3. Uniformly Better Unbiased Estimator

When the variance components are unknown, the problem turns to be how to combine two independent estimators $\hat{\beta}_a$ and $\hat{\beta}_e$ to get a more efficient estimator of β. An unbiased estimator more efficient than both $\hat{\beta}_a$ and $\hat{\beta}_e$ is called a uniformly better unbiased estimator of β. It is well known that for any constant $c \in [0,1]$, the variance of $c\hat{\beta}_a + (1-c)\hat{\beta}_e$ is greater than either $Var(\hat{\beta}_a)$ or $Var(\hat{\beta}_e)$ for some values of $Var(\hat{\beta}_a)$ and $Var(\hat{\beta}_e)$. Therefore, in order to get a more efficient estimator of β, we first find an estimator $\hat{w}$ of (1) as a random weight function.

The unbiased estimators $\hat{\beta}_a$ and $\hat{\beta}_e$ in (1) can be rewritten in the matrix forms as

$$\hat{\beta}_a = \frac{K\mathbf{X}_{J.}^T\mathbf{P}\mathbf{Y}_{J.}}{\mathbf{X}_{J.}\mathbf{P}\mathbf{X}_{J.}} \quad \hat{\beta}_e = \frac{\mathbf{X}^T\mathbf{Q}\mathbf{Y}}{\mathbf{X}^T\mathbf{Q}\mathbf{X}} \tag{4}$$

where the symmetric and idempotent matrices $\mathbf{P}$ and $\mathbf{Q}$ are:

$$\mathbf{P} = [\mathbf{I}_J - \frac{1}{J}\mathbf{1}_J\mathbf{1}_J^T] \quad \mathbf{Q} = [\mathbf{I}_J \otimes (\mathbf{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}_K^T)]$$

and $\mathbf{X}_{J.}^T = (x_{1.}, x_{2.}, \ldots x_{J.})$, $\mathbf{X}^T = (x_{11}, x_{12} \ldots x_{1K}, x_{21}, \ldots x_{J1}, x_{J2}, \ldots, x_{JK})$. $\mathbf{Y}_{J.}$ and $\mathbf{Y}$ are defined similarly. $\mathbf{1}_n$ is an $n \times 1$ vector of ones. The symbol $\otimes$ denotes the Kronecker matrix product.

To estimate w, we find the unbiased estimators for $\sigma_e^2 + K\sigma_a^2$ and σ_e^2. Consider the following quadratic forms,

$$K\sum_{j=1}^{J}(Y_{j.} - Y_{..} - \hat{\beta}_a(x_{j.} - x_{..}))^2 = K(\mathbf{Y}_{J.} - \hat{\beta}_a\mathbf{X}_{J.})^T\mathbf{P}(\mathbf{Y}_{J.} - \hat{\beta}_a\mathbf{X}_{J.}) \tag{5}$$

$$\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{jk} - Y_{j.} - \hat{\beta}_e(x_{jk} - x_{j.}))^2 = (\mathbf{Y} - \hat{\beta}_e\mathbf{X})^T\mathbf{Q}(\mathbf{Y} - \hat{\beta}_e\mathbf{X}). \tag{6}$$

Denote (5) by $S_a^2(\hat{\beta}_a)$ and (6) by $S_e^2(\hat{\beta}_e)$. Substituting the matrix form of $\hat{\beta}_a$ in (4) into (5), we get

$$S_a^2(\hat{\beta}_a) = K\mathbf{Y}_{J.}^T(\mathbf{P} - \frac{\mathbf{P}\mathbf{X}_{J.}\mathbf{X}_{J.}^T\mathbf{P}}{\mathbf{X}_{J.}^T\mathbf{P}\mathbf{X}_{J.}})\mathbf{Y}_{J.} \tag{7}$$

Let $\mathbf{u}_{J.}^T = (e_{1.} + a_1, e_{2.} + a_2, \ldots, e_{J.} + a_J)$ where $e_{j.} = \sum_{k=1}^{K} e_{jk}/K$. Then the $J \times 1$ vector $\mathbf{Y}_{J.}$ can be expressed as

$$\mathbf{Y}_{J.} = \mu\mathbf{1}_J + \beta\mathbf{X}_{J.} + \mathbf{u}_{J.} \tag{8}$$

where the random variable $K^{1/2}\mathbf{u}_{J.}$ is distributed as a multivariate normal distribution with zero mean vector and covariance matrix $(\sigma_e^2 + K\sigma_a^2)\mathbf{I}_J$. Using (8), (7) becomes

$$S_a^2(\hat{\beta}_a) = K\mathbf{u}_{J.}^T(\mathbf{P} - \frac{\mathbf{P}\mathbf{X}_{J.}\mathbf{X}_{J.}^T\mathbf{P}}{\mathbf{X}_{J.}^T\mathbf{P}\mathbf{X}_{J.}})\mathbf{u}_{J.} \tag{9}$$

since $\mathbf{P}\mathbf{1}_J = \mathbf{0}$ and $(\mathbf{P} - \frac{\mathbf{P}\mathbf{X}_{J.}\mathbf{X}_{J.}^T\mathbf{P}}{\mathbf{X}_{J.}^T\mathbf{P}\mathbf{X}_{J.}})\mathbf{X}_{J.} = \mathbf{0}$.

Similarly substituting the matrix form of $\hat{\beta}_e$ in (4) into (6), we get

$$S_e^2(\hat{\beta}_e) = \mathbf{Y}^T(\mathbf{Q} - \frac{\mathbf{QXX}^T\mathbf{Q}}{\mathbf{X}^T\mathbf{QX}})\mathbf{Y} \tag{10}$$

where the $JK \times 1$ vector $\mathbf{Y}$ can be expressed as

$$\mathbf{Y} = \mu \mathbf{1}_{JK} + \beta \mathbf{X} + (\mathbf{I}_J \otimes \mathbf{1}_K)\mathbf{a} + \mathbf{e}. \tag{11}$$

Using (11), (10) becomes

$$S_e^2(\hat{\beta}_e) = \mathbf{e}^T(\mathbf{Q} - \frac{\mathbf{QXX}^T\mathbf{Q}}{\mathbf{X}^T\mathbf{QX}})\mathbf{e} \tag{12}$$

since $\mathbf{Q1}_{JK} = \mathbf{0}$, $(\mathbf{I}_J \otimes \mathbf{1}_K)\mathbf{Q} = \mathbf{0}$ and $(\mathbf{Q} - \frac{\mathbf{QXX}^T\mathbf{Q}}{\mathbf{X}^T\mathbf{QX}})\mathbf{X} = \mathbf{0}$, and where the random vector $\mathbf{e}$ is distributed as a multivariate normal distribution with the zero mean vector and covariance matrix $\sigma_e^2 \mathbf{I}_{JK}$. Note that both $(\mathbf{P} - \frac{\mathbf{PX}_{J.}\mathbf{X}_{J.}^T\mathbf{P}}{\mathbf{X}_{J.}^T\mathbf{PX}_{J.}})$ and $(\mathbf{Q} - \frac{\mathbf{QXX}^T\mathbf{Q}}{\mathbf{X}^T\mathbf{QX}})$ are idempotent since the matrices $\mathbf{P}$ and $\mathbf{Q}$ are idempotent.

From (9) and (12), it can be proved that $S_a^2(\hat{\beta}_a)/(\sigma_e^2 + K\sigma_a^2)$ and $S_e^2(\hat{\beta}_e)/\sigma_e^2$ are distributed as chi square distributions with degrees of freedoms

$$tr(\mathbf{P} - \frac{\mathbf{PX}_{J.}\mathbf{X}_{J.}^T\mathbf{P}}{\mathbf{X}_{J.}^T\mathbf{PX}_{J.}}) = J - 2$$

and

$$tr(\mathbf{Q} - \frac{\mathbf{QXX}^T\mathbf{Q}}{\mathbf{X}^T\mathbf{QX}}) = J(K-1) - 1.$$

Then $S_a^2(\hat{\beta}_a)/(J-2)$ and $S_e^2(\hat{\beta}_e)/(J(K-1)-1)$ are the unbiased estimators of $\sigma_e^2 + K\sigma_a^2$ and σ_e^2 respectively. Consequently, the estimator $\hat{w}$ of (3) will be:

$$\hat{w} = \frac{[S_e^2(\hat{\beta}_e)/(J(K-1)-1)SS_e(xx)]}{[S_a^2(\hat{\beta}_a)/(J-2)SS_a(xx)] + [S_e^2(\hat{\beta}_e)/(J(K-1)-1)SS_e(xx)]}.$$

Let $\mathcal{P} = (\mathbf{P} - \frac{\mathbf{PX}_{J.}\mathbf{X}_{J.}^T\mathbf{P}}{\mathbf{X}_{J.}^T\mathbf{PX}_{J.}})$ and $\mathcal{Q} = (\mathbf{Q} - \frac{\mathbf{QXX}^T\mathbf{Q}}{\mathbf{X}^T\mathbf{QX}})$. Since the matrices $\mathcal{P}$ and $\mathcal{Q}$ are idempotent, the quadratic forms (9) and (12) can be rewritten as

$$S_a^2(\hat{\beta}_a) = K\mathbf{u}_{J.}^T\mathcal{P}\mathcal{P}\mathbf{u}_{J.} \quad S_e^2(\hat{\beta}_e) = \mathbf{e}^T\mathcal{Q}\mathcal{Q}\mathbf{e}.$$

Using the normal theory, it can be shown that the random variables $K^{1/2}\mathcal{P}\mathbf{u}_{J.}$, $\mathcal{Q}\mathbf{e}$, $\hat{\beta}_a$ and $\hat{\beta}_e$ are independent. It follows that $S_a^2(\hat{\beta}_a)$, $S_e^2(\hat{\beta}_e)$, $\hat{\beta}_a$ and $\hat{\beta}_e$ are independent.

The estimator $\hat{w}$ of (3) is of the form

$$\hat{w} = \frac{1}{1 + rF}$$

where

$$F = \frac{S_a^2(\hat{\beta}_a)/[(J-2)(\sigma_e^2 + K\sigma_a^2)]}{S_e^2(\hat{\beta}_e)/(J(K-1)-1)\sigma_e^2}$$

and

$$r = \frac{(\sigma_e^2 + K\sigma_a^2)/SS_a(xx)}{\sigma_e^2/SS_e(xx)} = \frac{Var(\hat{\beta}_a)}{Var(\hat{\beta}_e)}$$

where F is a F random variable with degrees of freedoms $J-2$ and $J(K-1)-1$.

Now we first create an unbiased estimator of β and then find the condition that this unbiased estimator is more efficient than both $\hat{\beta}_a$ and $\hat{\beta}_e$. Consider the estimator $\hat{\beta}_* = \hat{w}\hat{\beta}_a + (1-\hat{w})\hat{\beta}_e$. Then

$$E[\hat{\beta}_*] = E_{\hat{w}}[E[\hat{\beta}_*|\hat{w}]] = E_{\hat{w}}[E[\hat{w}\hat{\beta}_a|\hat{w}] + E[(1-\hat{w})\hat{\beta}_e|\hat{w}]]$$
$$= E_{\hat{w}}[\hat{w}E[\hat{\beta}_a] + (1-\hat{w})E[\hat{\beta}_e]] = \beta$$

since $\hat{w}$ is independent of $\hat{\beta}_a$ and $\hat{\beta}_e$.

We find the conditions on J and K for which $\hat{\beta}_*$ is more efficient than both $\hat{\beta}_a$ and $\hat{\beta}_e$. The variance of $\hat{\beta}_*$ is

$$Var(\hat{\beta}_*) = E[\hat{w}(\hat{\beta}_a - \beta) + (1-\hat{w})(\hat{\beta}_e - \beta)]^2$$
$$= Var(\hat{\beta}_a)E[\hat{w}^2] + Var(\hat{\beta}_e)E[(1-\hat{w})^2]$$
$$= Var(\hat{\beta}_a)E[\frac{1}{(1+rF)^2}] + Var(\hat{\beta}_e)E[\frac{r^2F^2}{(1+rF)^2}]$$

since the cross term $E[(\hat{\beta}_a - \beta)(\hat{\beta}_e - \beta)] = 0$. Two cases must be considered: $Var(\hat{\beta}_a) \le Var(\hat{\beta}_e)$ and $Var(\hat{\beta}_e) \le Var(\hat{\beta}_a)$.

If the first case holds, then $0 \le r \le 1$ and

$$Var(\hat{\beta}_*) = Var(\hat{\beta}_a)(E[\frac{1}{(1+rF)^2}] + \frac{1}{r}E[\frac{r^2F^2}{(1+r^2F^2)^2}])$$
$$= Var(\hat{\beta}_a)E[\frac{1+rF^2}{(1+rF)^2}].$$

If the second case holds, then $1 \le r < \infty$ and

$$Var(\hat{\beta}_*) = Var(\hat{\beta}_e)(rE[\frac{1}{(1+rF)^2}] + E[\frac{r^2F^2}{(1+rF)^2}]).$$

If we let $a = 1/r$ and $W = 1/F$, then $0 < a < 1$ and W is a F random variable with degrees of freedoms $J(K-1)-1$ and $J-2$, we get

$$Var(\hat{\beta}_*) = Var(\hat{\beta}_e)E[\frac{1+aW^2}{(1+aW)^2}].$$

The unbiased estimator $\hat{\beta}_*$ is more efficient if and only if

$$E[\frac{1+rF^2}{(1+rF)^2}] < 1 \quad \text{and} \quad E[\frac{1+aW^2}{(1+aW)^2}] < 1. \tag{13}$$

Using the method of Graybill (1959), we find which values of J and K allow the uniformly better estimator to exist. Let $m_1 = J(K-1)-1$ and $m_2 = (J-2)$ and let $E[\frac{1+rF^2}{(1+rF)^2}]$ be denoted by $f(r)$. In what follows, it is assumed that both m_1 and m_2 are greater than 4. We examine the derivative of $f(r)$ at the point $r = 0$.

$$f'(r) = \frac{d}{dr}E[\frac{1+rF^2}{(1+rF)^2}] = E[\frac{d}{dr}(\frac{1+rF^2}{(1+rF)^2})]$$

and so

$$f'(0) = E[F^2 - 2F] = \frac{m_1}{m_1-4}[\frac{m_1(m_2+2)}{m_2(m_1-4)} - 2]. \quad (14)$$

Observe that $f(0) = 1$. Consider $f'(0) < 0$ and then the slope of $f(r)$ is negative at $r = 0$. Therefore $f(r)$ must be less than 1 for some values of r in the neighborhood of 0. From (14), this happens if

$$\frac{m_1(m_2+2)}{m_2(m_1-4)} < 2.$$

Taking m_1 fixed, we find the values of m_2 such that $f'(0) < 0$. For the values of $m_1 = 10, 11, 12, 13, 14, 15, 16$, we get $m_2 \geq 10, 8, 7, 6, 5, 5, 2$

If we use the similar approach for $E[\frac{1+aW}{(1+aW)^2}]$, we get $m_1 \geq 10, 8, 7, 6, 5, 5, 2$ for the fixed values of $m_2 = 10, 11, 12, 13, 14, 15, 16$. We conclude that the uniformly better estimator exists when both m_1 and m_2 are equal or greater than 10. That implies, the uniformly better estimator for β exists when $J \geq 12$ and $K \geq 2$.

4. Numerical Verification

From (13), the uniformly better estimator of β exists if

$$E[\frac{1+rF^2}{(1+rF)^2}] = \frac{(m_2/m_1)^{m_1/2}}{B(m_1/2, m_2/2)} \quad (15)$$

$$\times \int_0^\infty \frac{(1+rv)^2 v^{m_2/2-1}}{(1+rv)^2(1+m_2v/m_1)^{(m_1+m_2)/2}} dv$$

is less than 1 where $m_1 = J(K-1) - 1$ and $m_2 = K - 2$. The integral is numerically evaluated for some values of J, K and the ratio r of the variances of $\hat{\beta}_a$ and $\hat{\beta}_e$. Using the subroutine **dqagi** in the CM computer library, the following table is tabulated

(J, K)	r=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(16, 2)	0.947	0.899	0.855	0.815	0.778	0.745	0.713	0.685	0.559
(16, 3)	0.940	0.887	0.839	0.796	0.757	0.721	0.689	0.659	0.631
(16, 4)	0.938	0.833	0.834	0.789	0.749	0.713	0.680	0.650	0.622
(16, 5)	0.936	0.881	0.831	0.786	0.745	0.710	0.675	0.645	0.616
(18, 2)	0.946	0.898	0.855	0.814	0.778	0.744	0.713	0.685	0.657
(18, 3)	0.940	0.886	0.840	0.796	0.756	0.720	0.688	0.657	0.630
(18, 4)	0.937	0.882	0.833	0.789	0.749	0.712	0.679	0.648	0.620
(18, 5)	0.936	0.880	0.830	0.785	0.744	0.708	0.674	0.643	0.615
(20, 2)	0.947	0.899	0.855	0.815	0.778	0.744	0.713	0.684	0.657
(20, 3)	0.940	0.887	0.838	0.795	0.755	0.720	0.687	0.656	0.628
(20, 4)	0.938	0.882	0.832	0.788	0.748	0.711	0.677	0.646	0.618
(20, 5)	0.936	0.880	0.830	0.784	0.743	0.706	0.672	0.641	0.613
(22, 2)	0.947	0.898	0.854	0.814	0.778	0.743	0.712	0.683	0.656
(22, 3)	0.940	0.886	0.838	0.794	0.755	0.719	0.686	0.655	0.627
(22, 4)	0.937	0.882	0.832	0.787	0.747	0.710	0.676	0.646	0.617
(22, 5)	0.936	0.879	0.829	0.784	0.743	0.705	0.671	0.641	0.612

Table 1: The Values of Integral in (15) for Some Degrees of Freedoms J and K and the Ratio r of the Variances $\hat{\beta}_a$ and $\hat{\beta}_e$

From the table, we conclude that the expectations in (13) can be considerd as a decreasing function of r and a. Also, when the variances of $\hat{\beta}_a$ and $\hat{\beta}_e$ are close to each other, it is possible to get unbiased estimator whose variances is half of the maximum of the variances of the unbiased estimators $\hat{\beta}_a$ and $\hat{\beta}_e$ of β.

References

[1] Barankin E. W. (1949). Locally Best Unbiased Estimates. Ann. Math. Stat., 20, pp.477-501

[2] Fuller F. A. and Battese G. E. (1973). Transformation for Estimation of Linear Models with Nested Error. JASA, 68, pp.626-632

[3] Graybill F. A. and Deal R. B. (1959). Combining Unbiased Estimators. Biometrics, 15, pp.543-550

[4] Lehmann E. L. (1983). Theory of Point Estimation, John Wiley & Sons Inc.

[5] Tong L. I. and Cornelius P. L. (1989). Studies on the Estimation of the Slope Parameter in Simple Linear Regression Model With One-Fold Nested Error Structure. Commun. Statis. -Simula, 18,pp.201-225.

[6] Tong L. I. and Cornelius P. L. (1991). Studies on the Hypothesis Testing of the Slope Parameter in the Simple Linear Regression Model With One-Fold Nested Error Structure. Commun. Statis. -Theory Meth, 20, pp.2023-2043.

ÖZET

Birli katlanmış hata terimli basit regresyon modelinde eğim parametresi için iki farklı yansız tahmin edici vardır. Varyanslar bilinmediği zaman tahmin problemi bu iki tahmin edicinin hangi ağırlık katsayılarıyla birleştirilerek daha küçük varyanslı yansız bir tahmine dönüştürülmesi problemidir. Bu çalışmada, rasgele ağırlık katsayıları elde edilerek eğim parametresinin daha etkin yansız tahmin edicisi bulunmuş ve sayısal bir örnekle kanıtlanmıştır.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 1, pp. 29 - 41

SOME LIMIT THEOREMS FOR DECOMPOSABLE BRANCHING PROCESSES WITH DECREASING IMMIGRATION

I. Rahimov and A. Teshabaev
Tashkent Mathematical Institute, Uzbekistan

Abstract

Decomposable branching processes with decreasing immigration are considered. Some limit theorems for these processes are provided.

Key words: Decomposable branching processes, single type branching processes, limit theorems.

1. Introduction

We consider a non homogeneous, in general, branching process with particles of three types T_1, T_2 and T_3. Assume that within the interval of time $(t, t+\Delta t)$, $\Delta t \to 0$, a particle of type T_i is transmuted into a collection of particles w=(w_1, w_2, w_3) of types T_1, T_2, T_3 with probabilities $\delta_i^w + p_i^w(t)\,\Delta t + o(\Delta t)$, where $\delta_k^w = 1$ for $w_k = 1$, $w_i = 0$, $i \neq k$ and $\delta_k^w = 0$ otherwise. We shall assume that particles of the type T_1 are final, that is under any change a particle of the type T_1 yields exactly one particle of type T_1 and a certain collection of particles of types T_2 and T_3 which cannot revert back into T_1 (Sevast'yanov 1971).

In addition we assume that particles of type T_2 may transmute into particles of types T_2 and T_3, and particles of type T_3 only into those of type T_3 and probabilities of these transmutions are independent of time. This means that our process has only transitions of the form $T_1 \to T_2 \to T_3$.

We denote by $\mu_{kj}^{\tau}(t)$ the number of particles of type T_j obtained from one particle of type T_k within the interval of time (τ, t) and $\mu_{kj}(t) = \mu_{kj}^{0}(t)$. The process $\mu(t) = (\mu_{12}(t), \mu_{13}(t))$ can be interpreted as two type decomposable branching process with time-dependent immigration.

Single-type branching processes and multi-type indecomposable branching processes with time-dependent immigration were considered by Foster and Williamson (1971), Badalbaev and Rahimov (1978,1982,1985), Rahimov (1986). The survey by Vatutin and Zubkov (1985) and papers by Vatutin and Sagitov (1988,1989) contain quite a complete list of references concerning decomposable processes.

Some limit theorems for the process $\mu(t)$ will be proved in this paper, when reproduction processes are critical and intensities of the number of "immigrants" are decreasing.

Introduce the following generating functions

$$f_k(t,X) = \sum_w P_k^w(t)\, X_1^{w_1} X_2^{w_2} X_3^{w_3}\ ,\ X = (X_1, X_2, X_3)$$

with $f_k(t,X) = 0$ for $X_1 = X_2 = X_3 = 1$ and for any $t \in [0,\infty)$. In our assumptions these generating functions can be represented in a form

$$f_1(t,X) = X_1 g(t, X_2, X_3)\,, f(t,X_2) = f_2(X_2,X_3)\,, f_3(t,X) = f_3(X_3)\,.$$

Introduce the generating functions

$$F^\tau(t,X) = \sum_w P\{\mu_{k1}^\tau(t) = w_1, \mu_{k2}^\tau(t) = w_{2,}\mu_{k3}^\tau(t) = w_{3,}\}\, X_1^{w_1} X_2^{w_2} X_3^{w3}$$

and put $H(t,X_2,X_3) = F_1^0(t,1,X_2,X_3)$. By the same arguments as Sevast'yanov (1971) it can be obtained that

$$H(t,X_2,X_3) = \exp\left\{ \int_0^t g(u, F_2(t-u,X)\, F_3(t-u,X))\, du \right\}. \tag{1}$$

We shall assume that, at the points $X_2 = X_3 = 1$, the functions $f_2(X)$ and $f_3(X)$ have all derivatives up to order three and denote

$$\frac{\partial f_k}{\partial x_j} |_{X=1} = a_{kj}\ ,\ \frac{\partial^2 f_k}{\partial X_j^2} = b_k\ ,\ k,j = 2,3.$$

We also put for k,j =2,3

$$\frac{\partial g(t,X_2,X_3)}{\partial X_k} |_{X=1} = \alpha_k(t)\ ,\ \frac{\partial^2 g(t,X_2,X_3)}{\partial X_k \partial X_j} = \beta_{kj}(t)\ ,$$

and assume that

$$\sup_t \alpha_k(t) < \infty\ ,\ \sup_t \beta_{kj}(t) < \infty. \tag{2}$$

2. Asymptotics of the Probability $\mathbf{P}(\mu(\mathbf{t}) \neq \mathbf{0})$.

First we shall investigate the probability $P(\mu(t) \neq 0)$, 0=$(0,0)$. It is known from Savin and Chistyakov (1962) that

$$P\{\mu_{22}(t) + \mu_{23}(t) > 0\} \sim \frac{\sigma_2}{\sqrt{t}}\ ,\ t \to \infty \tag{3}$$

where $\sigma_2 > 0$. Using same arguments of Polin (1976) we obtain that

$$\sigma_2 = 2\sqrt{\frac{a_{23}}{b_2 b_3}}$$

in our case.

We assume that $\alpha_k(t)$, $k = 2,3$ vary regularly as $t \to \infty$ and

$$\beta_{22}(t)\ln t \to 0 \ , \ \beta_{ij}(t) \to 0 \ , \ i+j \neq 4. \tag{4}$$

Theorem 1. If $a_{22} = a_{33} = 0$, $b_2 > 0$, $b_3 > 0$, condition (4) is satisfied and as $t \to \infty$

$$\alpha_2(t)\sqrt{t} \to c_2 \ , \ \alpha_3(t)\ln t \to c_3 \ , \ c_i \in [0,\infty) \tag{5}$$

then

$$P = \lim_{t \longrightarrow \infty} P(\mu(t) \neq 0) = 1 - \exp\left\{-2\pi c_2\sqrt{\frac{a_{23}}{b_2 b_3}} - \frac{2c_3}{b_3}\right\} .$$

It follows from Theorem 1 that $P = 0$, if $\max(c_2, c_3) = 0$ and it can be shown that $P = 1$, if $\max(c_2, c_3) = \infty$.

Proof. Consider relation (1) with $X_2 = X_3 = 0$. Expanding the function $g(u, F_2, F_3)$ in a neighborhood of the point $F_2 = F_3 = 1$, we get

$$\begin{aligned} \ln P\{\mu(t) = 0\} = -\sum_{i=2}^{3}\int_0^t \alpha_i(u)(1 - F_i(t-u,0))\,du + \\ \tfrac{1}{2}\sum_{i,j=2}^{3}\int_0^t \hat{\beta}_{i,j}(u,t)(1 - F_i(t-u,0))(1 - F_j(t-u,0))\,du \end{aligned} \tag{6}$$

where $0 \leq \hat{\beta}_{i,j}(u,t) \leq \beta_{i,j}(u)$ for any $t \in [0,\infty)$.

Let $c_i > 0$. Using (3) and condition (5), we obtain that as $t \longrightarrow \infty$

$$\int_0^{t-\ln t} \alpha_2(u)(1 - F_2(t-u,0,0))\,du \sim 2c_2\sigma_2 \arcsin\sqrt{1 - \tfrac{\ln t}{t}} \ ,$$

$$\int_{t-\ln t}^{t} \alpha_2(u)(1 - F_2(t-u,0,0))\,du \to 0 \ ,$$

and , consequently,

$$\lim_{t\to\infty}\int_0^t \alpha_2(u)(1 - F_2(t-u,0,0))\,du = c_2\sigma_2\pi. \tag{7}$$

Since the $\mu_{33}(t)$ is a simple one dimensional critical branching process,

$$1 - F_3(t,0) \sim \frac{2}{b_3 t} \ , \ t \to \infty \ . \tag{8}$$

Using (8) and condition $\alpha_3(u) \sim c_3/\ln t$ it can be shown that the second summand of the first sum in (6) has limit $-2c_3/b_3$.

It remains to find the limit of the second sum in (6). First we use estimations of the functions $\hat{\beta}_{i,j}(u,t) \leq \beta_{ij}(u)$. Using (3) in the summand with $i = j = 2$, we obtain that it is less than

$$const\int_0^{t-t_0} \beta_{22}(u)\,\frac{du}{t-u}$$

for some $t_0 > 0$. But under condition $\beta_{22}(t)\ln t = o(1)$, this integral is an infinitesimal. Relating to other summands, if we use (3) and (8), we obtain for them estimations of the following form :

$$const \int_0^{t-t_0} \beta_{ij}(u) \frac{du}{(t-u)^{1+\alpha}} \ , \ i+j \neq 4 \ ,$$

where $t_0 > 0$ and α is equal to either 1/2 or 1. If $\beta_{ij}(t) \to 0$, $t \to \infty$, $i+j \neq 4$, then these integrals are also infinitesimal. Hence the limit of the second sum in (6) is equal to zero.

The theorem is proved for $c_i > 0$. In the case when $c_i = 0$ it can be proved by similar arguments.

Now we consider the case

$$\alpha_2(t)\sqrt{t} \sim l_2(t) \ , \ \alpha_3(t)\ln t \sim l_3(t) \tag{9}$$

where $l_i(t) \to 0$ and are slowly varying functions as $t \to \infty$. It follows from Theorem 1 that $P = \lim P(\mu(t) \neq 0) = 1$ in this case.

We also need the following condition

$$\beta_{22}(t)\ln t = o(L(t)) \ , \ \beta_{ij}(t) = o(L(t)) \ , \ i+j \neq 4 \ , \ t \longrightarrow \infty \tag{10}$$

where $L(t) = l_2(t) + l_3(t)$.

Theorem 2. If $a_{22} = a_{33} = 0$, $b_2, b_3 > 0$, conditions (9) and (10) are satisfied, then

$$P(\mu(t) \neq 0) \sim 2\pi\sqrt{\frac{a_{23}}{b_2 b_3}} l_2(t) + \frac{2l_3(t)}{b_3} \ , \ t \longrightarrow \infty.$$

Proof. Since P =0, we have

$$P(\mu(t) \neq 0) = 1 - e^{\ln P\{\mu(t)\neq 0\}} \sim \ln P\{\mu(t) \neq 0\}$$

Consider relation (6). Since $l_i(t)$ are slowly varying functions, there are the functions $\lambda_{l_i}(t) \longrightarrow \infty$ as $t \longrightarrow \infty$ such that for any function
$\lambda(t)$, $1 \leq \lambda(t) \leq \lambda_{l_i}(t)$

$$\lim_{t \longrightarrow \infty} \frac{l_i(t/\lambda(t))}{l_i(t)} = 1 \tag{11}$$

(See Badalbaev and Rahimov (1978)).
Using the same arguments, as in the proof of Theorem 1, we obtain that as $t \longrightarrow \infty$

$$\int_{t/\lambda_{l_2}(t)}^{t} \alpha_2(u)(1 - F_2(t-u,0,0))\,du \sim 2\sigma_2 l_2(t) \arcsin\left(\frac{x}{\sqrt{t}}\right) \Big|_{\sqrt{t/\lambda_{l_2}(t)}}^{\sqrt{t}} \ ,$$

Let $\delta > 0$ such that $1/2+\delta < 1$. Then it is clear that

$$\sup_{0 \leq u \leq t/\lambda_{l_2}(t)} \left\{u^\delta l_2(u)\right\} = O\left(t^\delta l_2(t)\right) \ , \ t \longrightarrow \infty.$$

Therefore using (3), we have

$$\int_{\epsilon}^{t/\lambda_{l_2}(t)} \alpha_2(u)(1-F_2(t-u,0,0))\,du = O\left(t^{\delta} l_2(t) \int_{\epsilon}^{t/\lambda_{l_2}(t)} \frac{du}{u^{1/2+\delta}\sqrt{t-u}}\right)$$

for some $\epsilon > 0$. It is easy to see that

$$\int_{\epsilon}^{t/\lambda_{l_2}(t)} \frac{du}{u^{1/2+\delta}\sqrt{t-u}} = o\left(t^{-\delta}\right),$$

and consequently,

$$\int_0^t \alpha_2(u)(1-F_2(t-u,0,0))\,du \sim \pi\sigma_2 l_2(t). \tag{12}$$

Using (8), by the same arguments by Rahimov (1986) it can be shown that

$$\int_0^t \alpha_3(u)(1-F_3(t-u,0))\,du \sim \frac{2l_3(t)}{b_3} \tag{13}$$

It remains to estimate the second sum in (6). It is clear that we have the same estimations as in proof of Theorem 1 for summands of this sum. But under conditions (10) these estimations are $o(L(t))$. Theorem is proved.

In the case $a_{22} = 0$ the process $\mu_{12}(t)$ is a one-dimensional critical branching process with time dependent immigration in which intensity of the number of immigrants is $\alpha_2(t)$. Asymptotical behavior of the probability $P\{\mu_{12}(t) > 0\}$ was considered, for example, by Badalbaev and Rahimov (1978). It is interesting to compare asymptotics of probabilities of "non-extinction" for processes $\mu(t)$, $\mu_{12}(t)$, $\mu_{13}(t)$.

Theorem 3. If $a_{22} = a_{33} = 0$, $b_2, b_3 > 0$ and conditions (4) and (5) are satisfied, then

$$\lim_{t\to\infty} P\{\mu_{13}(t) > 0\} = 1 - \exp\left\{-\frac{2\pi c_2}{b_2}\sqrt{a_{23}} - \frac{2c_3}{b_3}\right\}.$$

Now we present an analogy of Theorem 2. We need the following conditions.

$$\alpha_2(t) \sim l_2(t)/\sqrt{t}\ ,\ \alpha_3(t)\ln t = o(l_2(t)) \tag{14}$$

where $l_2(t) \to 0$ and slowly varies as $t \longrightarrow \infty$;

$$\beta_{22}(t)\ln t = o(l_2(t))\ ,\ \beta_{ij}(t) = o(l_2(t))\ ,\ i+j \neq 4\ . \tag{15}$$

Theorem 4. If $a_{22} = a_{33} = 0$, $b_2, b_3 > 0$, conditions (14) and (15) are satisfied , then

$$P\{\mu_{13}(t) > 0\} \sim \frac{2\pi l_2(t)}{b_2}\sqrt{a_{23}}\ .$$

Proofs of this theorems are similar to proofs of Theorems 1 and 2

3. Limit Theorems

We shall start from the case

$$\alpha_2(t) \sim l_2(t)/\sqrt{t}\ ,\ t \to \infty \tag{16}$$

where $l_2(t) \to \infty$ and varies slowly as $t \to \infty$. In this case

$$P = \lim_{t\to\infty} P\{\mu(t) \neq 0\} = 1$$

as it was noted already.

Theorem 5. If $a_{22} = a_{33} = 0$, $a_{23}, b_2, b_3 > 0$, conditions (2), (4) and (16) are satisfied , then

$$\frac{2\mu_{13}(t)}{b_3 a_{23} t l_2^2(t)} \xrightarrow{D} \xi\ ,$$

as $t \longrightarrow \infty$, where ξ has a stable distribution with exponent 1/2 and density function

$$P(X,b) = \frac{2}{b_2 X\sqrt{\pi X}} \exp\left\{-4/b_2^2 X\right\}\ ,\ \mathrm{X} > 0\ .$$

Consider the case, when $l_2(t) = const.$ in (16).

Theorem 6. If $a_{22} = a_{33} = 0$, $a_{23}, b_2, b_3 > 0$, conditions (2) and (4) are satisfied and

$$\alpha_2(t)\sqrt{t} \to c_2\ ,\ \mathrm{t}^{-1}\int_0^t \alpha_3(u)\,du \to c_3\ ,\ c_i \in [0,\infty) \tag{17}$$

then $2\mu_{13}(t)/b_3 t \to \xi$, where

$$\Psi(\lambda) \stackrel{def}{=} Ee^{-\lambda\xi} = (1+\lambda)^{-2c_3/b_3} \exp\left\{-\frac{4c_2\sqrt{a_{23}}}{b_2} \arcsin\sqrt{\frac{\lambda}{1+\lambda}}\right\}\ .$$

Now we shall describe some properties of the limit random variable. It can be verified easily, that

$$\lim_{\lambda\longrightarrow 0} \Psi'(\lambda) = -\infty\ .$$

Consequently $E\xi = \infty$.

It is obvious that ξ has the gamma distribution, if $c_2 = 0$, $c_3 > 0$.

Let now $c_3 = 0$, $c_2 > 0$. In this case

$$\lim_{\lambda\longrightarrow\infty} \Psi(\lambda) = \exp\left\{-\frac{2\pi c_2\sqrt{a_{23}}}{b_2}\right\} \stackrel{def}{=} \Delta$$

From well-known relation between distribution function and Laplace transform (Feller 1968, p.418) it follows that the limit distribution has an atom of the mass Δ at the point zero. This is connected with the fact that under conditions of Theorem 3

$$\lim_{t\to\infty} P\{\mu_{13}(t) = 0\} = \Delta$$

It must be noted that if $c_i > 0$, i=2,3 the limit distribution has no atom, that is the "presence" of immigration of T_3 type "splits" the atom Δ .

In order to find a nonsingular limit distribution for $\mu_{13}(t)$ under conditions of Theorems 2 and 4, we shall consider conditional distributions of the process.First we use the event $\{\mu_{13}(t) > 0\}$ as a condition.

Theorem 7. If $a_{22} = a_{33} = 0$, $b_2, b_3 > 0$, and conditions (14) and (15) are satisfied, then

$$\lim_{t\to\infty} P\left\{\frac{2\mu_{13}(t)}{b_3 t} < X \mid \mu_{13}(t) > 0\right\} = F(X) ,$$

$$\int_0^\infty e^{-\lambda X} dF(X) = 1 - \frac{2}{\pi} \arcsin\sqrt{\frac{\lambda}{1+\lambda}} .$$

We find another limit distribution, if we use the event $\{\mu(t) \neq 0\}$ as a condition.

Theorem 8. If $a_{22} = a_{33} = 0$, $b_2, b_3 > 0$, and conditions (9) and (10) are satisfied and

$$\lim_{t\to\infty} \frac{l_3(t)}{l_2(t)} = \theta \in [0,\infty) ,$$

then

$$\lim_{t\to\infty} P\left\{\frac{2\mu_{13}(t)}{b_3 t} < X \mid \mu(t) \neq 0\right\} = T(x) ,$$

where

$$\int_0^\infty e^{-\lambda X} dT(X) = 1 - \frac{\frac{4\sqrt{a_{23}}}{b_2} \arcsin\sqrt{\frac{\lambda}{1+\lambda}}}{\pi\sqrt{\frac{a_{23}}{b_2 b_3} + \frac{2\theta}{b_3}}}.$$

4. Proofs of Limit Theorems

Proof of Theorem 5. The density function $P(X, b_2)$ corresponds to the Laplace transform $\exp\left\{-4\sqrt{\lambda}/b_2\right\}$, Feller (1968, Chapter XIII). Hence it suffices to show that

$$H(t, 1, X_3(t)) \to \exp\left\{-4\sqrt{\lambda}/b_2\right\}$$

as $t \to \infty$, where $X_3(t) = \exp\left[-2\lambda/b_3 a_{23} t l^2(t)\right]$, $\lambda > 0$.

Consider relation (1) with $X_2 = 1$, $X_3 = X_3(t)$. Expanding the function $g(u, F_2, F_3)$ in a neighborhood of the point $F_2 = F_3 = 1$, we get

$$\ln H(t, 1, X_3) = -\sum_{i=2}^{3} \int_0^t \alpha_i(u)(1 - F_i(t-u, X))\,du + \epsilon(t,\lambda) \qquad (18)$$

where $X = (X_2, X_3)$ and $\epsilon(t,\lambda)$ is less than the second sum in relation (6). It follows from proof of Theorem 1 that under our conditions $\epsilon(t,\lambda) \to 0$ as $t \to \infty$, $\lambda > 0$.

It remains to find the limit of the first sum in (18). It is known, Chistyakov (1970), that the function $R_2 = 1 - F_2$ can be represented in the form

$$R_2(t,1,X_3) = Z_\gamma(t) - \frac{2\theta(t)}{2}\left[\sqrt{\frac{\gamma}{a_{23}}} - \int_0^t \theta(u)\,du\right]^{-1}, \tag{19}$$

for all $t \geq 0$, $\gamma = 2/b_3(1 - X_3) \geq \gamma_0 > 0$, where

$$\theta(t) = \exp\left\{-4\sqrt{a_{23}\gamma}\left(\sqrt{1+\frac{t}{\gamma}} - 1\right) + O\left(\ln\left(1+\frac{t}{\gamma}\right)\right)\right\}$$

and $Z_\gamma(t)$ satisfies the inequality

$$\left|Z_\gamma(t) - \frac{2}{b_2}\left[\frac{a_{23}}{t+\gamma}\right]^{1/2}\right| \leq \frac{const.}{t+\gamma} \tag{20}$$

for all $t \geq 0$, $\gamma \geq \gamma_0 > 0$, $Z_\gamma(0) = 2\left[a_{23}/b_2\gamma\right]^{1/2}$.

First we shall consider the following integral :

$$I_1 = \int_0^t \frac{2\alpha_2(u)}{b_2}\left[\frac{a_{23}}{t-u+\gamma}\right]^{1/2} du = \int_0^t \varphi_t(u)\,du .$$

Since $l(t)$ is a slowly varying function, there is the function $\lambda_1(t) \to \infty$ as $t \to \infty$ such that for any function $\lambda(t)$, $1 \leq \lambda(t) \leq \lambda_1(t)$

$$\lim_{t\to\infty} l(t/\lambda(t))/l(t) = 1 \tag{21}$$

Then

$$\int_{t/\lambda_1}^t \varphi_t(u)\,du \sim \frac{\sqrt{4a_{23}}}{b_2} l(t) \int_{t/\lambda_1}^t \frac{du}{\sqrt{u(t-u+\gamma)}},$$

as $t \to \infty$and since $\gamma \sim a_{23}tl^2(t)/\lambda, t \to \infty$. Last integral is equal to

$$2\arcsin\left[\frac{X}{t+\gamma}\right]^{1/2} \Big|_{t/\lambda}^{t} \sim 2\left[\lambda/a_{23}l^2(t)\right]^{1/2}$$

From the following estimation

$$\int_0^{t/\lambda_1} \varphi_t(u)\,du \leq const. \int_0^{t/\lambda_1} \alpha_2(u)\,t^{-1/2}du ,$$

we have that

$$\lim_{t\to\infty} I_1 = 4\lambda^{1/2}/b_2. \tag{22}$$

It can be shown by similar arguments that

$$I_2 = \int_0^t \frac{\alpha_2(u)\,du}{t-u+\gamma} \to 0\ ,\ t \to \infty. \tag{23}$$

Now we shall consider the integral

$$I_3 = \int_0^t \alpha_2(u)\,\frac{2\theta(t-u)}{b_2}\left(\left(\frac{\gamma}{a_{23}}\right)^{1/2} - \int_0^{t-u}\theta(x)\,dx\right)^{-1} du$$

We use the following estimations of the function $\theta(u)$, $0 \le u \le t$, Chistyakov (1970) :

$$\theta(u) = \exp\left\{-2\left(\frac{a_{23}}{\gamma}\right)^{1/2} u\right\}\left[1 + O\left(\frac{u}{\gamma} + \frac{u^2}{\gamma}\right)\right]\ , \tag{24}$$

$$\int_0^t \theta(u)\,du = \frac{1}{2}\sqrt{\frac{\gamma}{a_{23}}}\left(1 - \exp\left\{-2\left(\frac{a_{23}}{\gamma}\right)^{1/2} t\right\}\right) + O(1) \tag{25}$$

as $t \to \infty$, $\gamma \sim tl^2(t)\,a_{23}/\lambda$. By the same arguments as in estimation of the integral I_1, using (24) and (25) we find that

$$\lim_{t\to\infty}\sup I_3 \le \lim_{t\to\infty}\sup\left\{const.\frac{l(t)}{\sqrt{t}}\int_0^t \Psi_t(u)\,du\right\}$$

where

$$\Psi_t(u) = \frac{1}{\sqrt{u}}\exp\left\{-2\left(\frac{a_{23}}{\gamma}\right)^{1/2}(t-u)\right\}\ ,\ u > 0\ .$$

Let $L(t) \to \infty$, $L(t) = o(t)$, a function such that $\alpha_2(t)\,L(t) \to \infty$. Using

$$\int_0^{t-t/L(t)} \Psi_t(u)\,du \le const.\exp\left\{-2\sqrt{a_{23}}/\alpha_2(t)\,L(t)\right\}\sqrt{t}\ ,$$

$$\int_0^{t-t/L(t)} \Psi_t(u)\,du \le const.\sqrt{t}/L(t)\ ,$$

we obtain that

$$\lim_{t\to\infty} I_3 = 0. \tag{26}$$

Since

$$\left|\int_0^t \alpha_2(u)\,R_2(t-u,1,X_3)\,du - I_1\right| \le I_2 + I_3\ ,$$

it follows from relations (22), (23) and (26) that the summand with i=2 of the first sum in (18) has the limit $-4\sqrt{\lambda}/b_2$.

The summand with i =3 in this sum is an infinitesimal because of the following inequality

$$\int_0^t \alpha_3(u)\, R_3(t-u, X_3(t))\, du \le \sup_u \alpha_3(u)\, t\, (1 - X_3(t))$$

Theorem is proved.

Proof of Theorem 6. Consider the relation (18) with $X_2 = 1$, $X_3 = X_3(t) = \exp\left\{-\frac{2\lambda}{b_3 t}\right\}$. Under condition (4) $\epsilon(t,\lambda) \to 0$ as $t \to \infty$ for any $\lambda > 0$.

First we consider the following integral from the proof of Theorem 5 :

$$I_1 = \int_0^t \frac{2\alpha_2(u)}{b_2} \left[\frac{a_{23}}{t-u+\gamma}\right]^{1/2} du$$

where

$$\gamma = \frac{2}{b_3(1-X_3)} \sim \frac{t}{\lambda} \text{ in this case.}$$

Using the same arguments as in the proof of Theorem 5, obtain that

$$\lim_{t\to\infty} I_1 = \frac{4c_2\sqrt{a_{23}}}{b_2} \arcsin\sqrt{\frac{\lambda}{1+\lambda}}\ ,\ \lambda > 0.$$

Next, it is easy to see that

$$I_2 = \int_0^t \alpha_2(u)\frac{du}{t-u+\lambda} \le const.\frac{1}{t}\int_0^t \alpha_2(u)\, du\ ,$$

that is $I_2 \to 0$ as $t \to \infty$. Using estimations (24) and (25), we obtain that $I_3 \to 0$ as $t \to \infty$, where I_3 is the same as in proof Theorem 5.It follows from (19) that

$$\left|\int_0^t \alpha_2(u)\, R_2(t-u, 1, X_3)\, du - I_1\right| \le I_2 + I_3\ ,$$

and, consequently ,

$$\lim_{t\to\infty}\int_0^t \alpha_2(u)\, R_2(t-u,1,X_3)\, du = \frac{4c_2\sqrt{a_{23}}}{b_2}\arcsin\sqrt{\frac{\lambda}{1+\lambda}}. \tag{27}$$

Consider the summand with i=3 in (18). It is easy to see that

$$\begin{aligned} I_4 = \int_0^t \alpha_3(u)\, R_3(t-u, X_3)\, du = \int_0^t c_3 R_3(t-u, X_3)\, du + \\ + \int_0^t R_3(t-u, X_3)\, d(u\epsilon(u)) = R_1 + R_2\ , \end{aligned} \tag{28}$$

where

$$\epsilon(u) = \frac{1}{u}\int\limits_0^u \alpha_3(X)\,dX - c_3.$$

It is known from Sevast'yanov (1971) that for critical processes the generating function $F_3(t, X_3) \to 1, t \to \infty$ and $F_3(t+\tau, X_3) \geq F_3(t, x_3)$. If integrate by parts, we have

$$R_3 = u\epsilon(u)\,R_3(t-u, X_3)\,|_0^t - \int\limits_0^t \epsilon(u)uR_3(t-du, X_3)\ . \qquad (29)$$

Since $\epsilon(t) \to 0$ by condition (17) for any $\epsilon > 0$, there is $T > 0$, that $|\epsilon(u)| \leq \epsilon$ for $u > t$. Therefore

$$\int\limits_0^t \epsilon(u)\,uR_3(t-du, X_3) \leq \int\limits_0^T \epsilon(u)\,uR_3(t-du, X_3) + \epsilon\int\limits_T^t uR_3(t-du, X_3)$$

But last integral is less than

$$t\,(R_3(0, X_3) - R_3(t-T, X_3)) = O(1)$$

as $t \to \infty$ for fixed T.

First integral is an infinitesimal as $t \to \infty$ for any fixed T. Consequently

$$\lim_{t\to\infty} R_2 = 0\ , \qquad (30)$$

By similar arguments it can be shown that

$$\lim_{t\to\infty} R_1 = \frac{2c_3}{b_3}\ln(1+\lambda)\ . \qquad (31)$$

Theorem is proved.

Proof of Theorem 7. Consider the following relation :

$$E\left[\exp\left\{-\frac{2\lambda\mu_{13}(t)}{b_3 t} \mid \mu_{13}(t) > 0\right\}\right] = 1 - \frac{1 - H(t, 1, X_3)}{P\{\mu_{13}(t) > 0\}} \qquad (32)$$

where $X_3 = X_3(t) = \exp\left\{-\frac{2\lambda}{b_3 t}\right\}$.

Since

$$1 \geq H(t, 1, X_3) \geq H(t, 1, 0)$$

and $H(t, 1, 0) \to 1$ as $t \to \infty$ under the conditions of Theorem 7,

$$1 - H = 1 - e^{\ln H} \sim -\ln H\ ,\ t \to \infty\ ,$$

therefore we need to consider relation (18) with $X_3 = X_3(t)$. It follows from the condition (15) that $\epsilon(t, \lambda) \to 0$ as $t \to \infty$ for any $\lambda > 0$.

By the same arguments as in the proof of Theorem 5 we obtain that

$$I_1 = \int_0^t \frac{2\alpha_2(u)}{b_2}\left[\frac{a_{23}}{t-u+\gamma}\right]^{1/2} du \sim \frac{4\sqrt{a_{23}}}{b_2} l_2(t) \arcsin\sqrt{\frac{\lambda}{1+\lambda}} \tag{33}$$

Using relations (19) and (20) we find from (33) that

$$\int_0^t \alpha_2(u) R_2(t-u,1,X_3)\, du \sim \frac{4\sqrt{a_{23}}}{b_2} l_2(t) \arcsin\sqrt{\frac{\lambda}{1+\lambda}}. \tag{34}$$

Since $F_3(t, X_3)$ is the generating function of the critical branching process, using the following inequality

$$\int_0^t \alpha_3(u) R_3(t-u, X_3)\, du \le \int_0^t \alpha_3(u) R_3(t-u, 0)\, du$$

we have that the summand with $i = 3$ in (18) is $o(l_2(t))$ as $t \to \infty$.

Hence

$$1 - H(t, 1, X_3) \sim \frac{4\sqrt{a_{23}}}{b_2} l_2(t) \arcsin\sqrt{\frac{\lambda}{1+\lambda}}$$

as $t \to \infty$. We have that proof of Theorem 7 from relations (32), (35) and from Theorem 4.

Theorem 7 is proved.

Proof of Theorem 8. Proof of this theorem is similar to the proof of Theorem 7. Namely, consider relation (32) with $P\{\mu(t) \neq 0\}$ instead of $P\{\mu_{13}(t) > 0\}$. Under conditions of Theorem 8, by similar arguments as in the proof of Theorem 7 we obtain that

$$1 - H(t, 1, X_3) \sim \frac{4\sqrt{a_{23}}}{b_2} l_2(t) \arcsin\sqrt{\frac{\lambda}{1+\lambda}} + \frac{2c_3}{b_3}\ln(1+\lambda) \tag{35}$$

Using asymptotic of $P\{\mu(t) \neq 0\}$ from Theorem 2 it is not difficult to find that

$$\lim_{t\to\infty} E\left[\exp\left\{-\frac{2\lambda\mu_{13}(t)}{b_3 t}\right\} \mid \mu(t) \neq 0\right] = 1 - \frac{\frac{4\sqrt{a_{23}}}{b_2}\arcsin\sqrt{\frac{\lambda}{1+\lambda}}}{\pi\sqrt{\frac{a_{23}}{b_2 b_3}} + \frac{2\theta}{b_3}}.$$

Theorem is proved.

References

[1] Badalbaev I.S.,Rahimov I.(1978) Critical Branching Processes with Immigration Decreasing Intensity . *Theory of Probab. Appl.*, 23,No2,p.259-268.

[2] Badalbaev I.S.(1982) Limit Theorems for Critical Multitype Branching Processes with Immigration Decreasing Intensity. In the book *"Limit Theorems for Random Processes."* Tashkent Fan, p.41-54.

[3] Badalbaev I.S., Rahimov I. (1985) New Limit Theorems for Multitype Branching Processes with Immigration, *Informations of ASc. of Uzbek SSR*, No2, p.17-22.

[4] Chistyakov, V.P. (1970) Some Limit Theorems for Branching Processes with Particles of Final Type. *Theory Probab. Appl.*,15, No 3,p.515-521.

[5] Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol.2, Wiley, New York.

[6] Foster J. H., Williamson J. A. (1971) Limit Theorems for the Galton-Watson Process with Time-Dependent Immigration, *Z. Wahrschein, und Verw. Geb.*, V.20, No3,p.227-235.

[7] Polin A. K. (1976) Limit Theorems for Decomposable Critical Branching Processes. *Matematicheski Sbornik*, V.100, No 3(7),p.420-435.

[8] Rahimov I. (1986) Critical Branching Processes with Infinite Variance and Decreasing Immigration. *Theory Probab. Appl.* Vol.31 No 1,p.88-100.

[9] Savin, A.A.,Chistyakov, V.P. (1962) Some Theorems for Branching Processes with Several Types of Particles. *Theory Probab. Appl.*,7, No 1,p.93-100.

[10] Sevast'yanov, B.A. (1971) *Branching Processes*, Nauka,Moscow.

[11] Vatutin,V.A.,Zubkov, A.M. (1985) Branching Processes, 1, In :*Progress in Science and Technology :Probability Theory. Math. Statist. Theor. Cybernetics*, 23,3-67.

[12] Vatutin,V.A.,Sagitov, S.M. (1988,1989) Decomposable Critical Branching Bellman-Harris Process with Particles of Two Different Types 1,2. *Theory Probab. Appl.*, V.33, No3, 495-507, V.34,No.2, 251-262.

ÖZET

Tek ve çok tipli dallanma süreçleri için zamana bağımlı göç etkisi altında limitte davranış özellikleri pek çok araştırmacı tarafından incelenmiştir. Bu çalışmada, azalan göç içeren ayrıştırılabilen dallanma süreçleri ele alınmış ve bu tür olasılık süreçleri için bazı limit teoremleri tanıtlarıyla birlikte sunulmuştur.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 1, pp. 51 - 61

ON THE DISTANCE SEQUENCE OF DISTRIBUTION FUNCTIONS AND GOODNESS OF FIT TESTS

Ismihan G. Bairamov and Omer L. Gebizlioglu
Ankara University, Faculty of Science, Department of Statistics
06100, Tandogan, Ankara, Turkey

Abstract

This study attempts to show that it is possible to get a class of criteria with good properties to test several hypotheses by using a sequence of functionals , that we call "distance sequence of distribution functions", which characterize the proximity of two distribution functions. Choosing an optimal criterion among the said class of criteria is shown. Distance sequence of distribution functions are based on confidence intervals that are built with the help of order statistics . Asymptotic properties are proved for statistics which are obtained by using distance sequence of distribution functions . It is presented that useful criteria, derived from these statistics, in goodness of fit problem is achievable.

Key Words: Distance measure, goodness of fit test, consistent criteria.

1. Introduction

Let $X_1, X_2, ..., X_n$ be a sample from a population with unknown distribution function (d.f.). Consider the hypothesis $H_0 = \{X_1, X_2, ..., X_n$ has d.f. $P\}$ and the alternative hypothesis $H_1 = \{X_1, X_2, ..., X_n$ has d.f. $Q \in \Im, Q \neq P\}$, where $\Im$ is some class of continuous d.f.. On the basis of constructing statistical criteria to check simple hypothesis H_0 against the composite hypothesis H_1, there usually lies the consideration of how far the empirical distribution of $X_1, X_2, ..., X_n$ is from the distribution in some sense of distance $d(P,Q)$. This distance functional has the desirable properties $d(P,Q) \geq 0$ and $d(P,Q) = 0$, if and only if $P = Q$; and the property of continuousness in neighborhood of points so that small deviations do not result in large deviations in distance. There exist many classical statistical goodness of fit tests and their modifications based on distance measures, such as Kolmogorov-Smirnov, Wilcoxon, Cramer-Von-Mises and Moran tests, among others. In this paper we consider a different approach for the construction of goodness of fit tests based on the, so called, distance sequence for distribution functions (d.s.d.f.). Distance between distribution functions is an important subject of study in statistical inference. For instance, classification of a data item to one of several populations is a problem that requires some distance measures.(See , Hajek and Sidak (1965), and Kendall and Stuart (1979)).

Let$\{W_a^m\}_m$ be the class of criteria for testing the hypothesis H_0 against the alternative H_1 with

$$\lim_{n\to\infty} P\{X \in W_a^m / H_0\} = 1 - \alpha, m = 0, 1, 2, ...$$

Definition 1. If for $\forall\, Q(u) \in \Im$, there exists a number $m_0 = m_0(Q)$, such that

$$P\{(X_1, X_2, ..., X_n) \in W_\alpha^{m_0} \mid Q(u)\} = P_{W_\alpha^{m_0}}(Q) \to 1, \text{ as } n \to \infty$$

then the class of criteria W_α^m, $m = 0, 1, 2, ...$, is called consistent for checking hypothesis H_0 against the composite hypothesis H_1.

Suppose that $X_1, X_2, ..., X_m, X_{m+1}$ are continuous i.i.d. random variables with d.f. $Q(u)$ and $Y_1, Y_2, ..., Y_m$ are continuous i.i.d. random variables with d.f. $P(u)$; then

$$P\left\{X_{m+1} \in (X_{(1)}, X_{(m)})\right\} = \frac{m-1}{m+1},$$

where $X_{(1)} = \min(X_1, X_2, ..., X_m)$, $X_{(m)} = \max(X_1, X_2, ..., X_m)$. That is; $(X_{(1)}, X_{(m)})$ is an invariant confidence interval containing the main distributed mass (see Bairamov, Petunin,1990). Also consider the extreme order statistics $Y_{(1)}, Y_{(m)}$ obtained from random sample $Y_1, Y_2, ..., Y_m$. Then one can write

$$P\left\{X_{m+1} \in (Y_{(1)}, Y_{(m)})\right\} = 1 - \int_{-\infty}^{\infty} \left[(P(u))^m + (1 - P(u))^m\right] dQ(u).$$

Now the absolute value of difference of these probabilities is obtained as the expression

$$\left|P\left\{X_{m+1} \in (X_{(1)}, X_{(m)})\right\} - P\left\{X_{m+1} \in (Y_{(1)}, Y_{(m)})\right\}\right| = \left|\int_{-\infty}^{\infty} \left[(P(u))^m + (1 - P(u))^m\right] dQ(u) - \tfrac{2}{m+1}\right|.$$

Below, there is a theorem which states that under the given features we receive d.s.d.f.'s. Let $P^{-1}(u) = \inf\{x : P(x) \geq u\}$ be the inverse function of P.

Let $\Im_c$ be the class of continuous d.f.'s and $\Im_a$ be the class of all continuous d.f.'s that are symmetric about number a. $\Im_a \subset \Im_c$ .Consider

$$d_m^{(1)}(P, Q) = \left|\int_{-\infty}^{\infty} (P(u))^m dQ(u) - \frac{1}{m+1}\right| , m = 0, 1, 2, ... \qquad (1)$$

$$d_m^{(2)}(P, Q) = \left|\int_{-\infty}^{\infty} (1 - P(u))^m dQ(u) - \frac{1}{m+1}\right| , m = 0, 1, 2, ... \qquad (2)$$

$$d_m^{(3)}(P, Q) = \left|\int_{-\infty}^{\infty} \left[(P(u))^m + (1 - P(u))^m\right] dQ(u) - \frac{2}{m+1}\right| , m = 0, 1, 2, ... \qquad (3)$$

Theorem 1. The sequences $\left\{d_m^{(i)}(P, Q)\right\}_{m=0}^{\infty}$, $i = 1, 2, 3$ defined in (1), (2),(3) satisfy the following conditions:

1) $d_m^{(i)}(P, Q) \geq 0$, $m = 0, 1, 2, ...$; $i = 1, 2, 3$.

2) $d_m^{(i)}(P, Q) = 0$ if and only if $P = Q$, $P, Q \in \Im_c$ $i = 1, 2$; $d_m^{(3)}(P, Q) = 0$ if and only if $P = Q$, $P, Q \in \Im_a, a \in R$.

(Note that we will call a sequence of functionals of d.f.'s $\{d_m(P,Q)\}_{m=0}^{\infty}$, $P, Q \in \Im$ with properties 1) and 2) a distance sequence for distribution functions for class $\Im$ (d.s.d.f.)).

Proof. It is clear that $d_m^{(i)}(P,Q) \geq 0$ and $d_m^{(i)}(P,Q) = 0$ if $P = Q$, $i = 1,2,3$. Note that $d_m^{(i)}(P,Q), i = 1,2,3$ may be written as

$$d_m^{(1)}(P,Q) = \left| m \int_0^1 (Q(P^{-1}(u)) - u)u^{m-1}du \right|, \tag{4}$$

$$d_m^{(2)}(P,Q) = \left| m \int_0^1 (Q(P^{-1}(u)) - u)(1-u)^{m-1}du \right|, \tag{5}$$

$$d_m^{(3)}(P,Q) = \left| m \int_0^1 (Q(P^{-1}(u)) - u)(u^{m-1} - (1-u)^{m-1})du \right|, \ m = 0,1,2,\dots. \tag{6}$$

Consider (4). According to Stone-Weierstarass theorem (see Rudin,1964); the set of functions $1, u, u^2, u^3, \dots$ is closed in the space of all continuous functions $C_{[0,1]}$ on $[0,1]$. If $d_m^{(1)}(P,Q) = \left| m \int_0^1 (Q(P^{-1}(u)) - u)u^{m-1}du \right| = 0, m = 0,1,2,\dots$, then

$$\int_0^1 (Q(P^{-1}(u)) - u)u^m du = 0, \quad m = 0,1,2,\dots.$$

Hence $Q(P^{-1}(u)) - u = 0, u \in [0,1]$ and $Q(u) = P(u), u \in R$.

Consider (6). It is clear that without loss of generality we may take $a = 0$. In fact, considering the random variable $X^* = X - a$ with d.f. $Q_1(x) = Q(x+a)$ and $Y^* = Y - a$ with d.f. $P_1(x) = P(x+a)$ one can see that $Q_1(x)$ and $P_1(x)$ are symmetric about zero and

$$P\left\{Y_{(1)} < X_{n+1} < Y_{(n)}\right\} = P\left\{Y_{(1)}^* < X_{n+1}^* < Y_{(n)}^*\right\}.$$

Let P and Q be symmetric about zero. Now suppose $d_m^{(3)}(P,Q) = 0, m = 0,1,2,..$ Then from

$$\left| m \int_0^1 (Q(P^{-1}(u)) - u)(u^{m-1} - (1-u)^{m-1})du \right| = 0, m = 0,1,2,\dots \quad ,$$

one can obtain

$$\int_0^1 (Q(P^{-1}(u)) - u)u^{m-1}du = \int_0^1 (Q(P^{-1}(1-u)) - (1-u))u^{m-1}du, m = 0,1,2,\dots \tag{7}$$

Denote $\varphi(t) = Q(P^{-1}(t)) - t$. Then (7) may be written as

$$\int_0^1 (\varphi(t) - \varphi(1-t))t^m dt = 0, m = 0,1,2,\dots \tag{8}$$

From (8) we have $\varphi(t) = \varphi(1-t), t \in [0,1]; Q(P^{-1}(t)) - t = Q(P^{-1}(1-t)) - (1-t)$, $t \in [0,1]$. Next using $1 - P(t) = P(-t)$ and $1 - Q(t) = Q(-t)$ we have $Q(t) - P(t) = Q(P^{-1}(P(-t))) - (1 - P(t)) = Q(-t) - (1 - P(t)) = 1 - Q(t) - 1 + P(t) = P(t) - Q(t)$ and $Q(t) = P(t), t \in R$.

Remark (Counter example). Denote by $U_{a,b}(x)$ the d.f. of uniform distribution on $[0,1]$. It is easy to see that if we take $P(x) = U_{0,2}(x)$ and $Q(x) = U_{0,1}(x)$ then $d_m^{(3)}(P,Q) = 0, m = 0,1,2,\ldots$. That is $d_m^{(3)}(P,Q) = 0, m = 0,1,2,\ldots$. is not a d.s.d.f. for class $\Im_c$.

2. Asymptotic Normality of Statistics Obtained by d.s.d.f.

Let $X_1, X_2, \ldots, X_n$ be a random sample from general set with d.f. F_0. Let F_n^* be the empirical d.f. of the sample $X_1, X_2, \ldots, X_n$. Here we consider some examples of statistics obtained by the d.s.d.f.. Consider for example

$$d_m^{(3)}(F_0, F_n^*) = \left| \int_{-\infty}^{\infty} (F_0(u))^m dF_n^*(u) - \tfrac{1}{m+1} \right| = \left| \tfrac{1}{n} \textstyle\sum_{i=1}^n (F_0(X_i))^m - \tfrac{1}{m+1} \right|$$
$$\stackrel{d}{=} \left| \tfrac{1}{n} \textstyle\sum_{i=1}^n U_i{}^m - \tfrac{1}{m+1} \right|$$

where $U_1, U_2, \ldots, U_n$ are i.i.d. r.v.'s with d.f. $U_{0,1}(x)$. It is clear that $EU_i{}^m = \frac{1}{m+1}$, $Var(U_i{}^m) = \frac{1}{2m+1} - \frac{1}{(m+1)^2} \equiv \sigma_m^{(1)}$. Using the law of large numbers, central limit theorem and Glivenko-Cantelli Theorem (one can see Fergusson 1996, Gaensler, Stute 1987, Borovkov, 1984 among others) we can give the following theorems:

Theorem 2. Let Q be some continuous d.f. Then

$$d_m^{(i)}(Q, F_n^*) \to d_m^{(i)}(Q, F_0), \ a.s.$$

for every integer m, as $n \to \infty$.

Denote

$$I_1(X_1, X_2, \ldots, X_n) = d_m^{(1)}(F_0, F_n^*) = \left| \frac{1}{n} \sum_{i=1}^n (F_0(X_i))^m - \frac{1}{m+1} \right|,$$

$$I_2(X_1, X_2, \ldots, X_n) = d_m^{(2)}(F_0, F_n^*) = \left| \frac{1}{n} \sum_{i=1}^n (1 - F_0(X_i))^m - \frac{1}{m+1} \right|,$$

$$I_3(X_1, X_2, \ldots, X_n) = d_m^{(3)}(F_0, F_n^*) = \left| \frac{1}{n} \sum_{i=1}^n [(F_0(X_i))^m + (1 - (F_0(X_i))^m] - \frac{2}{m+1} \right|.$$

It is easy to verify that

$$\sigma_m^{(3)} \equiv Var\,[(F_0(X_i))^m + (1 - (F_0(X_i))^m] = Var(U_i{}^m + (1 - U_i)^m)$$
$$= \tfrac{2}{2m+1} + \tfrac{2(m!)}{(m+1)(m+2)\ldots(2m+1)} - (\tfrac{2}{m+1})^2.$$

Theorem 3. For every finite integer m, it is true that

$$\lim_{n\to\infty}\sup_x\left|P\left\{\frac{\sqrt{n}}{\sigma_m^{(i)}}I_i(X_1,X_2,...,X_n)\le x\right\}-\frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt\right|=0,\ i=1,2,3.$$

2.1. Separability of Distributions

Now we investigate the behavior of $d_m(P,Q)$ in (3) for large m.

$$d_m(P,Q)=\left|\int_{-\infty}^{\infty}[(P(u))^m]\,dQ(u)-\tfrac{2}{m+1}\right|=\left|\int_{\{u:P(u)=0\}}dQ(u)+\int_{\{u:P(u)=1\}}dQ(u)+\int_{\{u:0<P(u)<1\}}[(P(u))^m+(1-P(u))^m]\,dQ(u)-\tfrac{2}{m+1}\right| \tag{9}$$

Letting to limit in (9) ,as $m\to\infty$, we obtain

$$\lim_{m\to\infty}d_m(P,Q)=\left|\int_{\{u:P(u)=0\}}dQ(u)+\int_{\{u:P(u)=1\}}dQ(u)\right|\equiv d(P,Q).$$

Definition 2. Let $F(u)$ and $Q(u)$ be distribution functions corresponding to random variables X and Y. Denote $A_0\equiv\{u:F(u)=0\}$, $A_1\equiv\{u:F(u)=1\}$, $Q(A_0)=P\{Y\in A_0\}=\int_{A_0}dQ(u)$, $Q(A_1)=P\{Y\in A_1\}=\int_{A_1}dQ(u)$. If $Q(A_0)+Q(A_1)=1$, then we say that F and Q are separated.

It is easy to see that if the supports of distributions F and Q are noncrossing, then F and Q are separated.

Theorem 4. Let $X_1,X_2,...,X_n$ be i.i.d. random variables with distribution function $F(u)$ and $Y_1,Y_2,...,Y_n,Y_{n+1}$ be i.i.d. random variables with distribution function $Q(u)$. If

$$P\left\{(X_{(1)},X_{(n)})\cap(Y_{(1)},Y_{(n)})=\emptyset\right\}=1,\quad n=2,3,...$$

then distributions F and Q are separated.

Proof. One can write

$$P\left\{Y_{n+1}\in(X_{(1)},X_{(n)})\cup(Y_{(1)},Y_{(n)})\right\}=P\left\{Y_{n+1}\in(X_{(1)},X_{(n)})\right\}+P\left\{Y_{n+1}\in(Y_{(1)},Y_{(n)})\right\}. \tag{10}$$

Using (10), we have

$$1\ge P\left\{Y_{n+1}\in(X_{(1)},X_{(n)})\cup(Y_{(1)},Y_{(n)})\right\}$$
$$=\tfrac{n-1}{n+1}+\left[1-\int_{-\infty}^{\infty}[(F(u))^n+(1-F(u))^n]\,dQ(u)\right]=\tfrac{n-1}{n+1}+$$

$$+\left\{1-\left[\int_{A_0} dQ(u)+\int_{A_1} dQ(u)+\int_{0<F(u)<1}[(F(u))^n+(1-F(u))^n]\,dQ(u),\right]\right\} \tag{11}$$

Letting to limit as $n\to\infty$ in (11) we obtain $\int_{A_0} dQ(u)+\int_{A_1} dQ(u)\geq 1$. Hence

$$\int_{A_0} dQ(u)+\int_{A_1} dQ(u)=1.$$

Theorem is proved.

It should be noted that $d_m(P,Q)$, for large values m , approaches to 1 if the overlap of supports of d.f.'s P and Q becomes sufficiently small.

3. Class of Criteria Defined by d.s.d.f.'s

Suppose that $d_m(P,Q)$ has the from $d_m(P,Q)=|G_m(Q)-G_m\,(P)|$ where

$$G_m(Q)=h(\int g_m\,(x))dQ\,(x) \text{ where } g_m\,(x)\,,\ m=0,1,2,...$$

are some sequences of Borel functions, and $h\,(x)$ is a continuous function.

Theorem 5. Let $\{d_m(P,Q)\}_{m=0}^{\infty}$ be the d.s.d.f defined as in (1), (2) and (3). Then, class of criteria $W_\alpha^m=\left\{\frac{\sqrt{n}}{\sigma_m}d_m(P,P_n^*)>x_\alpha\right\}$, $m=0,1,2,...$ is consistent for checking H_0 against H_1.

Proof. By Theorem 3 we have

$$\lim_{n\to\infty} P\left\{\frac{\sqrt{n}}{\sigma_m}d_m(P,P_n^*)>x_\alpha/H_0\right\}=1-\alpha=1-\Phi(x_\alpha)$$

where P_n^* is the empirical d.f. of $X_1,X_2,...,X_n$. Consider

$$P\left\{\frac{\sqrt{n}}{\sigma_m}d_m(P,P_n^*)>x_\alpha\mid Q\,(x)\right\}=P_Q\left\{\frac{\sqrt{n}}{\sigma_m}\mid G_m(P_n^*)-G_m\,(P)\mid>x_\alpha\right\} \tag{12}$$

where

$$G_m(Q)=\left|\int[(P\,(x))^m+(1-P\,(x))^m]\,dQ\,(x)-\frac{2}{m+1}\right|,$$

$$G_m(P_n^*)=\left|\int[(P\,(x))^m+(1-P\,(x))^m]\,dP_n^*\,(x)-\frac{2}{m+1}\right|$$

and $G_m\,(P)=0$. Letting to limit in (12)

$$\lim_{n\to\infty} P_Q\left\{|G_m(P_n^*)-G_m\,(P)|>\frac{x_\alpha}{\sqrt{n}}\sigma_m\right\}=$$

$$= \lim_{n\to\infty} P_Q \left\{ |G_m(P_n^*) - G_m(P)| > C_\alpha^m(n) \right\}, C_\alpha^m(n) \to 0, n \to \infty \quad (13)$$

since $G_m(P_n^*) \to G_m(Q)$, almost sure, and since $P \neq Q$ there exist an m_0, such that

$$|G_{m_0}(Q) - G_{m_0}(P)| = d_{m_0}(P,Q) > 0.$$

Hence from (13) we obtain

$$\lim_{n\to\infty} P_Q \left\{ |G_{m_0}(P_n^*) - G_{m_0}(P)| > 0 \right\} = 1$$

Thus, the class of criteria

$$W_\alpha^m = \left\{ (X_1, X_2, ..., X_n) : \frac{\sqrt{n}}{\sigma_m} \left| \sum_{i=1}^{n} [(P(X_i))^m + (1 - P(X_i))^m] - \frac{2}{m+1} \right| > x_\alpha \right\}$$

is consistent for testing H_0 against composite H_1.

Analogously, we may define the class of criteria by $d_m^{(1)}(P,Q)$ and $d_m^{(2)}(P,Q)$. It is clear that the statistics following from $d_m^{(1)}(P,Q)$ and $d_m^{(2)}(P,Q)$ has similar properties.

Let $\mathbf{P} = \{P\}$ be some class of d.f.'s.

$$H_0 = \{X_1, X_2, ..., X_n \text{ has d.f. } P_0\} \text{ and } H_1 = \{X_1, X_2, ..., X_n \text{ has d.f. } P, P \in \mathbf{P} - \{P_0\}\}.$$

Let $\{d_m(P,Q)\}_{m=0}^{\infty}$ be a d.s.d.f.'s . Then there exist $m_0 = m_0(Q)$, such that for $\forall Q \in \mathbf{P}$ and $Q \neq P_0$, $d_{m_0}(P_0, Q) > 0$.

Definition 3. We say that d.s.d.f. has a property (A) for P_0 and class $\mathbf{P} = \{P\}$ if there exist an m_0 such that , for $\forall Q \in \mathbf{P}$ and $Q \neq P_0$ $d_{m_0}(P_0, Q) > 0$ (m_0 independent from Q).

Denote $M_0 = \{m_0 : d_{m_0}(P_0, Q) > 0, Q \neq P_0, Q \in \mathbf{P}\}$. Let

$$W_\alpha^m = \{(X_1, X_2, ..., X_n) : d_m(P_0, P_n^*) > C_\alpha\}$$

be class of criteria with asymptotic level α and depend on parameter m for testing hypothesis H_0 against alternative H_1, and $\{d_m(P,Q)\}_{m=0}^{\infty}$ has the property (A) for P_0 and $\mathbf{P} = \{P\}$. Then for any $m_0 \in M_0$ (if M_0 contains more than one element) the criteria $W_\alpha^{m_0}$ is consistent for testing H_0 against H_1. Therefore, with property (A), we have the class of consistent criteria for testing H_0 against H_1.

Definition 4. Criteria $W_\alpha^{m^*}$ is called the best criteria in class W_α^m if m_0^* is selected according to

$$\sup_{m_0 \in M_0} \inf_{Q \in P} d_{m_0}(P_0, Q) = d_{m_0^*}(P_0, Q_*).$$

The following examples illustrate some applications of the ideas that we presented above.

Example 1. Consider the d.s.d.f. expressed by (1) . Let $P_0(x) = 1 - e^{-x}, x \geq 0; \mathbf{P} = \left\{P(x) = 1 - e^{-\theta x}, x \geq 0, \theta = 1, 2, 3, 4, ...\right\}$, $Q(x) \in \mathbf{P}$ and $Q \neq P_0$

$$d_m^{(1)}(P, Q) = \left| \int\limits_{-\infty}^{\infty} (P_0(u))^m dQ(u) - \tfrac{1}{m+1} \right| =$$
$$= \left| \theta \int\limits_0^{\infty} (1 - e^{-x})^m e^{-\theta x} dx - \tfrac{1}{m+1} \right|.$$

It is clear that

$$\theta \int_0^{\infty} (1 - e^{-x})^m e^{-\theta x} dx = \theta \int_0^1 y^m (1 - y)^{\theta - 1} dy = \theta B(m + 1, \theta) = \frac{m!\theta!}{(m + \theta)!}$$

We obtain

$$d_{m_0}^{(1)}(P, Q) = \left| \frac{m_0!\theta!}{(m_0 + \theta)!} - \frac{1}{m_0 + 1} \right| > 0 \text{ for } m_0 = 1, 2, ...$$

So $d_{m_0}^{(1)}(P, Q)$ has the property (A) for class $\mathbf{P}$, and P_0 . Here $M_0 = \{m_0 : m_0 = 1, 2, 3, ...\}$. It is easy to show that

$$\inf_{Q \in P} d_{m_0}^{(1)}(P_0, Q) = d_{m_0}^{(1)}(P_0, Q_2)$$

where $Q_k = 1 - e^{-kx}, x \geq 0, k = 2, 3, 4, ...$and

$$\sup_{m_0 \in M_0} \inf_{Q \in P} d_{m_0}^{(1)}(P_0, Q) = d_1^{(1)}(P_0, Q_1) = d_2^{(1)}(P_0, Q_2) = \frac{1}{6}.$$

This concludes that the best test is reached by choosing $m_0^* = 1$ or 2.

Example 2. Consider the d.s.d.f.'s given by (1) . Let $P_0(x) = U_{0,1}(x)$

$$\mathbf{P} = \left\{Q_b(x) : Q_b(x) = U_{(2,b)}(x), \quad 2 < b \leq 3\right\}. \text{ For } Q \in \mathbf{P}, \text{ we have}$$

$$d_m(P, Q) = \left| \int\limits_2^b dQ(x) - \tfrac{2}{m+1} \right| = \left| 1 - \tfrac{2}{m+1} \right|, 2 < b \leq 3$$

So , $\inf\limits_b d_m(P_0, Q_b) = \left| 1 - \frac{2}{m+1} \right|$.

Therefore the best test is received for a very large value of m_0^*.

5. Close Alternatives

Now we investigate the behavior of criteria defined by (3) for any fixed m for close alternatives, which converge to each other at the rate of $\frac{1}{\sqrt{n}}$. Consider $F(x) = F_0(x) + \frac{1}{\sqrt{n}} P(x)$, where $P(x)$ is some continuous function with appropriate properties , such that $F(x)$ is d.f..

Let $H_0 = \{X_1, X_2, ..., X_n$ has d.f $F_0(x)\}$ and $H_1 = \{X_1, X_2, ..., X_n$ has d.f $F(x)\}$. Now we intend to claim and prove that using a statistic $I_3(X_1, X_2, ..., X_n)$ in testing H_0 against H_1 under criteria W_α^m, close alternatives are distinctively identified.

Theorem 6. It is true that

$$\lim_{n\to\infty} P\{X \in W_\alpha^m / H_1\} = \lim_{n\to\infty} P_F\{I_3(X_1, X_2, ..., X_n) > x_\alpha\} = P\{|\eta + P_m| > x_\alpha\},$$ where

$$P_m = \frac{m}{\sigma_m} \int_{-\infty}^{\infty} \left((1 - F(x))^{m-1} - F^{m-1}(x)\right) P(x)\, dF(x)$$ and

$P\{\eta \le x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \quad (\sigma_m \equiv \sigma_m^{(3)}).$

Proof. Consider

$$P\{X \in W_\alpha^m / H_1\} = P\left\{\tfrac{\sqrt{n}}{\sigma_m} I_3(X_1, X_2, ..., X_n) > x_\alpha / H_1\right\} =$$

$$= P\left\{\frac{\sqrt{n}}{\sigma_m}\left|\sum_{i=1}^{n} [(F_0(X_i))^m + (1 - F_0(X_i))^m] - \frac{2}{m+1}\right| > x_\alpha / H_1\right\} =$$

$$= P_F\left\{\frac{\sqrt{n}}{\sigma_m}\left|\sum_{i=1}^{n} [(F_0(X_i))^m + (1 - F_0(X_i))^m] - \frac{2}{m+1}\right| > x_\alpha\right\} =$$

$$= P_F\left\{\frac{\sqrt{n}}{\sigma_m}\left|\int \left[\left(F(u) - \frac{1}{\sqrt{n}}P(u)\right)^m + \left(1 - F(u) + \frac{1}{\sqrt{n}}P(u)\right)^m\right] dF_n^*(u) - \frac{2}{m+1}\right| > x_\alpha\right\} =$$

$$\begin{aligned} &= P_F\Big\{\Big|\tfrac{\sqrt{n}}{\sigma_m}\left(\int [(F(u))^m + (1 - F(u))^m]\, dF_n^*(u) - \tfrac{2}{m+1}\right) \\ &+ \tfrac{1}{\sigma_m}\int \left[\left(1 - F(u) + \tfrac{1}{\sqrt{n}}P(u)\right)^m - (1 - F(u))^m\right] \sqrt{n}\, dF_n^*(u) \\ &+ \tfrac{1}{\sigma_m}\int \left[\left(F(u) - \tfrac{1}{\sqrt{n}}P(u)\right)^m - (F(u))^m\right] \sqrt{n}\, dF_n^*(u) > x_\alpha\Big|\Big\} \end{aligned} \tag{14}$$

where $F_n^*(u)$ is the empirical d.f of sample $X_1, X_2, ..., X_n$. Denote,

$$\varphi_n^{(m)}(x) = \left[\left(F(u) - \frac{1}{\sqrt{n}}P(u)\right)^m - (F(u))^m\right]\sqrt{n} + \left[\left(1 - F(u) + \frac{1}{\sqrt{n}}P(u)\right)^m - (1 - F(u))^m\right]\sqrt{n}$$

and $\lim_{n\to\infty} \varphi_n^{(m)}(x) = \varphi^{(m)}(x)$. (14) may be written as

$$\begin{aligned} &P\{X \in W_\alpha^m / H_1\} = \\ &= \left\{\tfrac{\sqrt{n}}{\sigma_m}\left(\int [(F(u))^m + (1 - F(u))^m]\, dF_n^*(u) - \tfrac{2}{m+1}\right) + \tfrac{1}{\sigma_m}\int \varphi_n^{(m)}(u)\, dF_n^*(u) \right. \\ &\left. > x_\alpha \right\} \end{aligned} \tag{15}$$

Considering the absolute value of differences, we get

$$\begin{aligned} &\left|\int \varphi_n^{(m)}(x)\, dF_n^*(x) - \int \varphi^{(m)}(x)\, dF(x)\right| \\ &= \left|\tfrac{1}{n}\sum_{i=1}^{n} \varphi_n^{(m)}(X_i) + \int(\varphi_n^{(m)}(x) - \varphi^{(m)}(x)\, dF(x) - \int \varphi_n^{(m)}(x)\, dF(x)\right| \\ &= \left|\frac{1}{n}\sum_{i=1}^{n} \varphi_n^{(m)}(X_i) - \frac{1}{n}\sum_{i=1}^{n} E\varphi_n^{(m)}(X_i) + \int \left(\varphi_n^{(m)}(x) - \varphi^{(m)}(x)\right) dF(x)\right|. \end{aligned} \tag{16}$$

Lemma. 1. Let $H(x), x \in R$, be a real valued differentiable function; $F(x)$ and $P(x)$ some continuous functions; and b_n a sequence of real numbers which converge to zero, as $n \to \infty$. Then

$$\lim_{n\to\infty} \frac{1}{b_n}\left(H\left(F(x)+b_nP(x)\right)-H\left(F(x)\right)\right)=H'\left(F(x)\right)P(x)$$

Proof. Consider $h(y)=\begin{cases}\frac{1}{y}\left(H\left(F(x)+y\right)\right)-H\left(F(x)\right), y\neq 0\\ H'\left(F(x)\right), \ y=0\end{cases}$

It is clear that $h(y)$ is continuous at point $y=0$ and $b_nP(x)\to 0$. We have $\frac{1}{b_n}(H(F(x)+b_nP(x)))$ $-H(F(x))=h(b_n(P(x)))$. Letting to limit we obtain

$$\lim_{n\to\infty}\frac{1}{b_n}\left(H\left(F(x)+b_nP(x)\right)-H\left(F(x)\right)\right)=\lim_{n\to\infty} h\left(b_n\left(P(x)\right)P(x)\right)h(0)P(x)=$$

$=H'(F(x))P(x)$. This concludes the proof of Lemma 1.

By Lemma 1 for $H(x)=x^m$, we have

$$\varphi^{(m)}(x)=-m\left(F(x)\right)^{m-1}P(x)+m\left(1-F(x)\right)^{m-1}P(x).$$

Consider $\int\varphi_n^{(m)}(x)\,dF_n^*(x)$. By Glivenko-Canteli theorem; $F_n^*(x)\to F(x)$, almost sure (See Gaensler, Stute). It is clear that $\int\varphi_n^{(m)}(x)\,dF_n^*(x)=\frac{1}{n}\sum_{i=1}^{n}\varphi_n^{(m)}(X_i)$ and $\int\varphi_n^{(m)}(x)\,dF(x)$ $=E\varphi_n^{(m)}(X_1)$. Here $\varphi_n^{(m)}(X_1),\varphi_n^{(m)}(X_2),...,\varphi_n^{(m)}(X_n)$ are i.i.d.random variables. It is clear that

$$\int\left(\varphi_n^{(m)}(x)-\varphi^{(m)}(x)\right)dF(x)\to 0 \text{ as } n\to\infty, m=0,1,2,.... \tag{17}$$

And now we consider the differences $\frac{1}{n}\sum_{i=1}^{n}\varphi_n^{(m)}(X_i)-\frac{1}{n}\sum_{i=1}^{n}E\varphi_n^{(m)}(X_i)$ in (16)

Lemma 2. It is true that

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_n^{(m)}(X_i)-\frac{1}{n}\sum_{i=1}^{n}E\varphi_n^{(m)}(X_i)\xrightarrow{P}0, \text{ as } n\to\infty.$$

Proof.

Denote $\sum_{i=1}^{n}\varphi_n^{(m)}(X_i)=y_n$, $n=1,2,\dots$. Obviously, $Ey_n=n\int\varphi_n^{(m)}(u)\,dF(u)$.$n=$ $1,2,\dots$ .We have $E\left(\sum_{i=1}^{n}\varphi_n^{(m)}(X_i)\right)^2=Ey_n^2=$

$$\sum_{i=1}^{n}E\left(\varphi_n^{(m)}(X_i)\right)^2+2\sum_{1\le k\le l\le n}E(\varphi_n^{(m)}(X_k)\varphi_n^{(m)}(X_l))=\sum_{i=1}^{n}E\left(\varphi_n^{(m)}(X_i)\right)^2$$
$$+2\left(n(n-1)/2\right)\left(E(\varphi_n^{(m)}(X_1))\right)^2=nE\left(\varphi_n^{(m)}(X_1)\right)^2+n(n-1)\left[E\left(\varphi_n^{(m)}(X_1)\right)\right]^2$$

Hence $var(y_n)=Ey_n^2-(Ey_n)^2=n\int\left(\varphi_n^{(m)}(x)\right)^2dF(x)-\left(n\int\varphi_n^{(m)}(x)\,dF(x)\right)^2+$

$+n(n-1)\left(E\left(\varphi_n^{(m)}(X_1)\right)\right)^2 = n\int\left(\varphi_n^{(m)}(x)\right)^2 dF(x) - n\left(\int\varphi_n^{(m)}(x)\,dF(x)\right)^2$

Denote $\frac{1}{n}y_n = \zeta_n$. $E\zeta_n = \frac{1}{n}Ey_n = \int\varphi_n^{(m)}(x)\,dF(x)$ and

$var(\zeta_n) = \frac{1}{n^2}var(y_n) = \frac{1}{n}\int\left(\varphi_n^{(m)}(x)\right)^2 dF(x) - \frac{1}{n}\left(\int\left(\varphi_n^{(m)}(x)\right)dF(x)\right)^2$

By Chebyshev inequality, for $\forall\epsilon > 0$

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\varphi_n^{(m)}(X_i) - \frac{1}{n}\sum_{i=1}^{n}E\varphi_n^{(m)}(X_i)\right| > \epsilon\right\} \le \frac{var(\zeta_n)}{\epsilon^3}$$
$$= \frac{1}{\epsilon^3}\left[\frac{1}{n}\int\left(\varphi_n^{(m)}(x)\right)^2 dF(x) - \frac{1}{n}\left(\int\varphi_n^{(m)}(x)\,dF(x)\right)^2\right]$$

So, it can be easily shown that, $\frac{1}{n}\int(\varphi_n^{(m)}(x))^2 dF(x) - \frac{1}{n}\left(\int\varphi_n^{(m)}(x)\,dF(x)\right)^2 \to 0$ as $n \to \infty$.

Thus

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_n^{(m)}(X_i) - \frac{1}{n}\sum_{i=1}^{n}E\varphi_n^{(m)}(X_i) \xrightarrow{P} 0 \qquad (18)$$

This concludes the proof of Lemma 2.

Using Lemma 1 and Lemma 2, if we insert (16), (17), (18) in (15) and let to limit as $n \to \infty$, we obtain ;

$$\lim_{n\to\infty} P\{X \in W_\alpha^m/H_1\} = \lim_{n\to\infty} P_F\{I_{F_0}(X_1, X_2, ..., X_n) > x_\alpha\} = P\{|\eta + P_m| > x_\alpha\},$$

where

$$P_m = \frac{m}{\sigma_m}\int_{-\infty}^{\infty}\left((1-F(x))^{m-1} - F^{m-1}(x)\right)P(x)\,dx\ ,\ P\{\eta \le x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt.$$

This proves the theorem.

References

[1] Borovkov A.A. (1984) *Mathematical Statistics.* Moskow, Nauka. (In Russian)

[2] Bairamov I.G. and Petunin Yu.I. (1990) Structure of Invariant Confidence Intervals Containing the Main Distributed Mass. *Theory of Probability and Its Applications*, 35,15-26.

[3] Gaensler P. and Stute W. (1987) *Seminar on Empirical Processes.* BirkhauserVerlag Basel. Boston.

[4] Ferguson T.S. (1996) *A Course in Large Sample Theory.* Chapman & Hall.

[5] Kendall M. and Stuart A. (1979) *The Advanced Theory of Statistics*, Volume 2,Macmillan.

[6] Hajek J. and Sidak Z. (1967) *Theory of Rank Tests.* Academic Press.

[7] Lehman W. (1959) *Testing Statistical Hypothesis.* Wiley.

[8] Rudin W. (1953) *Principles of Mathematical Analysis.* Wiley.

ÖZET

Bu çalışmada, dağılım fonksiyonlarının uzaklık fonksiyonelleri kullanılarak çeşitli hipotezlerin testleri için iyi özelliklere sahip olan bir ölçüt sınıfının elde edilebileceği gösterilmiştir. Dağılım fonksiyonlarının uzaklık dizisi, sıra istatistikleri yardımıyla elde edilen güven aralıklarına dayandırılmış, bunları kullanarak çıkarılan istatistiklerin asimptotik özellikleri sunulmuştur.

İstatistik, Journal of the Turkish Statistical Association
1998, Vol.1, No. 1, pp. 43 - 50

LOT QUALITY DISTRIBUTION FOR AUTOCORRELATED DATA

Donald S. Holmes
Stochos Inc.
14 N.College Street
Schenectady, N.Y. 12305
U.S.A.

A.Erhan Mergen
Rochester Institute of Technology
College of business, Decision Sciences
107 Lomb Memorial Drive
Rochester, N.Y. 14623-5608
U.S.A

Abstract

The number of nonconforming items contained in a manufacturing lot determines "the lot quality". The probability distribution of the lot quality is referred to as the "lot quality distribution" (LQD). In an autocorrelated (dependent) process it is assumed that the quality of the item produced at any time interval t depends upon the quality of the item(s) produced at previous time interval(s). In this study we assume that lots are formed from items sequentially produced by an autocorrelated process, and use first order Markov Chain with two states ("conforming," "nonconforming") to describe such a process. The model, in the context of manufacturing, assumes that the quality of the (k+1)th item is positively correlated with the quality of the (k)th item. The result of this study is a probability function, that describes the lot quality distribution for a first order autocorrelated process, which can be an integral part of Sequential process control (SPC).

Key Words: Lot quality distribution, Markov chain, dependent process.

1. Introduction

Most of the techniques in the statistical quality control area consider the continuous manufacturing process as an independent process. However, use of advanced manufacturing technology, e.g., use of automated process control (APC), etc., may cause autocorrelation in the process which then the independence assumption would be violated (Montgomery and Mastrangelo (1991)). The failure to take autocorrelation into account may lead to erronous conclusion about the process status (Holmes and Gordon (1992), Dodson (1995), and Runger and Willemain (1995)). If the manufacturing process is an independent process, the quality of an item produced at a given time is independent of the quality of the item produced at the previous time, and the lot quality distribution would be the Binomial distribution. However, not all the empirical lot quality distributions seem to be of the Binomial type (Wetherill (1977)). This paper describes some of the empirical lot quality distributions by assuming existence of an autocorrelated process (i.e., dependent process) which is modeled as a first order Markov Chain. Knowing the true nature of the process will lead to better use of acceptance sampling procedures.

The existence of autocorrelation, which is mainly due to the increasing use of high technology in processes, is an emerging problem for many SPC practioners. In this paper we discuss the probability function for the number of nonconforming items in lots (i.e., lot quality distribution) of items produced sequentially by a dependent process. This probability function has been used in several different applications. Mergen and Holmes (1986), for example, used this function to approximate the lot quality distribution of a subassembly of an aircraft jet engine, and the result was very close to the empirical distribution. Deligonul and Mergen (1987) used this probability function to show the dependence bias in p-charts when the production process is not independent and proposed a way to correct this bias. Mergen (1981), and Holmes and Mergen (1988), again by employing this probability function, derived a new measure called expected average outgoing quality (EAOQ) as an alternative to average outgoing quality limit (AOQL) to evaluate the performance of acceptance sampling plans in terms of outgoing quality. The advantage of the EAOQ over AOQL is that the former incorporates the lot quality distribution into the process of determining the outgoing quality and as a result the status of the process becomes an integral part of the selection of the sampling plan. Using the EAOQ approach, the sampling plan decision is made based on the expected value rather than the maximum value of the outgoing quality level. This leads to smaller sample sizes which in turn lower inspection costs.

2.The Model

When the manufacturing process is described as a first order Markov Chain, the process can be represented in matrix form as follows:

		(k+1)th item	
		g	b
kth item	g	x	$1-x$
	b	y	$1-y$

Table 1. First Order Markov Chain Matrix.

Where,

g = conforming,

b = nonconforming,

x =probability that the (k+1)th item is conforming given that the (k)th item was conforming,

$1-x$ = probability that the (k+1)th item is nonconforming given that the (k)th item was conforming,

y = probability that the (k+1)th item is conforming given that the (k)th item was nonconforming,

$1-y$ = probability that the (k+1)th item is nonconforming given that the (k)th item was nonconforming.

If the process described by this matrix continues for a sufficient period of time, the steady state probability of nonconformance, Π, (i.e., the fraction nonconforming) can be shown to be:

$$\Pi = \frac{1-x}{1-x+y} \tag{1}$$

This fraction defective will be used as a starting point for developing the distribution of lot qualities. Thus when the process starts to form a new lot, the first item will be either a conforming one or a nonconforming one, with probability $1-\Pi$ and Π respectively. Now, let (d,g) and (d,b) be two states with (d) number of nonconforming items given that the last item lotted was conforming and nonconforming, respectively; and also let $p(d:g,n)$ and $p(d:b,n)$ denote the probabilities of (d) number of nonconforming items in a lot size of (n) if the last (nth) item produced was conforming and nonconforming, respectively. For example, if the lot size (n) is determined as two, then the probability that there will be no nonconforming item after the second item is lotted would be

$$p(0:2) = p(0:g,2)p(0:g,2) = (1-\Pi)x$$

The probability that there will be one conforming item in the lot is
$p(1:2) = p(0:g,2)p(1:b,2) + p(1:b,2)p(1:g,2) = (1-\Pi)(1-x) + \Pi y$
The probability that there will be two nonconforming items in the lot is
$p(2:2) = p(1:b,2)\,p(2:b,2) = \Pi\,(1-y)$
The lot quality distribution for lot size two will be
$p(0:2) = (1-\Pi)\,x$
$p(1:2) = (1-\Pi)\,(1-x) + \Pi y$
$p(2:2) = \Pi\,(1-y)$
$\overline{\sum = 1.0}$

To generalize this for any lot size n, the following Markov matrix may be used to describe the process.

The transition matrix in Table 2 will be referred to as Matrix T. For example, the state description $(2,b)$ means that in the lotting process there are presently two nonconforming items and the last item generated is nonconforming. Thus the state $(0,b)$ is a nonexistent state since one cannot have both no nonconforming items in a lot and the last item in the lot be nonconforming. After the first item is produced and lotted, the initial state vector would be as follows:

$$\begin{array}{lcccccc} & (0,g) & (1,b) & (1,g) & (2,b) & \dots & (n,g) \\ V_1 = (& 1-\Pi, & \Pi, & 0, & 0, & \dots & 0) \end{array}$$

The state vector after the second, third, ..., nth item is produced and lotted would be respectively,
$V_2 = V_1.T, V_3 = V_2.T, \dots\dots V_n = V_{n-1}.T$.

k k+1	0,g	1,b	1,g	2,b	2,g	3,b	3,g	4,b	...	n,g
0,g	x	1-x	0	0	0	0	0	0	...	0
1,b	0	0	y	1-y	0	0	0	0	...	0
1,g	0	0	x	1-x	0	0	0	0	...	0
2,b	0	0	0	0	y	1-y	0	0	...	0
2,g	0	0	0	0	x	1-x	0	0	...	0
3,b	0	0	0	0	0	0	y	1-y	...	0
3,g	0	0	0	0	0	0	x	1-x	...	0
4,b	0	0	0	0	0	0	0	0	...	0
...	.	.	.	.	.	.	.	.	...	.
...	.	.	.	.	.	.	.	.	...	.
n,g	0	0	0	0	0	0	0	0	...	0

Table 2: Transition matrix

An appropriate combination of the elements of the n state vector will give the probabilities of the number of nonconforming items in a lot of size n (i.e., the lot quality distribution). To obtain the general solution for the lot quality distribution, the following difference equations can be written from Matrix T:

$$p(d,n) = p(d:b,n) + p(d:g,n) \tag{2}$$

where

$$p(d:b,n) = (1-x)\,p(d-1:g,n-1) + (1-y)\,p(d-1:b,n-1), \tag{3}$$

$$p(d:g,n) = xp(d:g,n-1) + yp(d:b,n-1) \tag{4}$$

for $n \geq d \geq 1$ and

$$p(0:b,n) = 0, \tag{5}$$

$$p(0:g,n) = xp(0:g,n-1) \tag{6}$$

These equations can be solved recursively for given x, y and n values. Details of the solution procedure are found in Mergen (1981). The results of the solution gives the following probability function (i.e., LQD):

$$p(d:n) = \sum_{i=1}^{\min(n-d,d-1)} \left[\binom{n-d-1}{n-d-i} x^{n-d-i} (1-x)^{i-1} y^{i} (1-y)^{d-i-1} [a_i] \right] + x^{n-2d-1} (1-x)^{d-1} y^{d} [c] \tag{7}$$

for $d = 1, 2, ..., n-1,$

where

$$\binom{r}{q} = 0 \text{ if } r \leq 0, q < 0, \tag{8}$$

$$a_i = \frac{\binom{d-1}{d-i-1}(1-x)^2 + 2\binom{d-1}{d-i}(1-x)(1-y) + \binom{d-1}{d-i-1}(1-y)^2}{1-x+y}, \tag{9}$$

$$c = \frac{2\binom{n-d-1}{n-2d} x(1-x) + \binom{n-d-1}{n-2d}(d-1)x(1-y) + \binom{n-d-1}{n-2d-1} y(1-x)}{1-x+y}, \tag{10}$$

$$p(0:n) = (1-\Pi)\, x^{n-1} = \frac{yx^{n-1}}{1-x+y},$$

$$p(n,n) = \Pi\,(1-x)^{n-1} = \frac{(1-x)(1-y)^{n-1}}{1-x+y}.$$

As a point of interest Mergen (1981) showed that the above lot quality distribution for dependent processes reduces to Binomial distribution (as it should) for the independent process situation where transition probabilities are equal (i.e., $x = y$).

For the case where x and y are not equal, the lot quality distribution given in equation (7) does not reduce to any known distribution (because of the assumption of dependence). Thus an exact convergence for the case where x is not equal to y could not be found, even though some approximate results were obtained through Beta and Compound Poisson distribution (Mergen (1981)). It could very well be the case that this equation might not converge to any known distribution. It is believed that it could be the subject for another research paper.

The Beta distribution could be a good approximation for the lot quality distribution of the dependent process for large lot size n (Mergen (1981)). The reason behind this stems from the fact that as the lot size increases, the effect of dependence disappears, and in turn, Beta approximation improves. However, when the lot size is small, large peaks in the tails are observed. Naturally, these can not be handled by the Beta approximation, because Beta distribution describes an independent process.

While the virtue of Beta distribution is that it is mathematically simple, the virtue of the first order Markov Chain model is that it is based on a plausible process description (Mergen and Holmes (1986)). In other words, Beta distribution assumes independence in the process. However, by assuming an independent production process, there is no way of getting some of the lot quality distributions which are encountered in practice. Thus, first order Markov Chain model is a more realistic approach, because it describes the process conditions first and then derives the corresponding lot quality distribution for it. Also to our knowledge, there is no underlying process model which gives rise to Beta distribution.

3. Lot Quality Distribution For Acceptance Sampling

The following example demonstrates how the LQD discussed in this paper is applied to acceptance sampling in terms of determining the outgoing quality of the lots passed through inspection. The proposed measure, as mentioned above, is called the expected value of average outgoing quality (EAOQ) (Holmes and Mergen (1988)). EAOQ would be a better measure than the average outgoing quality limit (AOQL), because EAOQ describes the typical average outgoing quality, whereas AOQL represents the worst average outgoing quality. The way to calculate EAOQ is to use the LQD in determining the average outgoing quality, namely:

$$\text{EAOQ} = \sum \theta p_a(\theta)\, p(\theta) \tag{11}$$

where θ = fraction defective in the lot, i.e., d/n

$p(\theta)$ = the probability distribution of θ (i.e.,LQD)

$p_a(\theta)$ = probability of accepting a lot which has θ fraction defective for a given acceptance sampling plan. The table below (Table 3) displays EAOQ and AOQL values for various acceptance sampling plans. The transition probabilities of x and y are also listed; these are used to derive the LQD by using equations (7)-(12).

$x = 0.95$ $y = 0.15$ Acceptance number used in the sampling plans=0

LOT SIZE	SAMPLE SIZE	EAOQ	AOQL
40	4	0.0515	0.0719
40	5	0.0392	0.0670
40	6	0.0307	0.0566
40	7	0.0246	0.0490
40	8	0.0201	0.0430

50	5	0.0426	0.0670
50	6	0.0334	0.0567
50	7	0.0267	0.0500
50	8	0.0217	0.0432
50	9	0.0180	0.0388
50	10	0.0150	0.0349

$x = 0.95$ $y = 0.15$ Acceptance number used in the sampling plans=1

LOT SIZE	SAMPLE SIZE	EAOQ	AOQL
40	4	0.1251	0.1974
40	5	0.1004	0.1596
40	6	0.0817	0.1337
40	7	0.0674	0.1153
40	8	0.0563	0.1013
50	5	0.1086	0.1595
50	6	0.0888	0.1339
50	7	0.0734	0.1153
50	8	0.0613	0.1014
50	9	0.0518	0.0904
50	10	0.0442	0.0814

Table 3. EAOQ and AOQL Values for Various Accettance Sampling Plans.

As one expects, the EAOQ values are smaller than the AOQL values. This implies that a desired average outgoing quality can be maintained by using smaller size samples if we integrate the knowledge of the process (i.e., LQD) into the evaluation of the performance of the sampling plan. This, in turn, leads to lower inspection costs.

4. Conclusion

The increasing appearance of autocorrelation (i.e., dependence) in manufacturing processes necessitates a different approach to model the lot quality distribution. In this paper the derivation and various applications of a probability function for the number of nonconforming items in lots of items produced sequentially by a dependent process is discussed. This function, based on a model of a production process, has the ability to fit a number of empirical lot quality distributions (LQD). The LQD's in this paper reduce to the Binomial for an independent process. Thus the probability model discussed in this paper may be considered as a generalization of the Binomial. As is discussed above, this LQD (or LQD's like this one) can be made an integral part of the acceptance sampling procedure to maintain the desired protection level with minimum sample sizes. This lot quality distribution can also be used to correct the potential bias on the p-control chart limits in the case of presence of autocorrelation as shown by Deligonul and Mergen (1987). The presence of positive autocorrelation in practice leads to the width of the area between control limits of a p-chart generally being underestimated. In such cases, the control limits erroneously give more frequent out-of-control signals (i.e., large Type I (alpha) error) than so warranted under its appropriate statistical model. Grant and Leavenworth (1980) note that this has adverse effects on the implementation of quality control procedure since it causes operating personnel to discredit the use of control charts. The lot quality distribution discussed in this paper is used to fix this problem by integrating the variance of this distribution in calculating the control limits of the $p-$chart (Deligonul and Mergen(1987)).

References

[1] Deligonul, Z.S. and Mergen, A.E. (1987) Dependence Bias in Conventional p-Charts and Its Correction with an Approximate Lot Quality Distribution. Journal of Applied Statistics, 14, 1, 75-81.

[2] Dodson, B. (1995) Control Charting Dependent Data: A Case Study. Quality Engineering, 7, 4, 757-768.

[3] Grant, E.L. and Leavenworth, R.S. (1980) Statistical Quality Control, McGraw-Hill, New York, p.249.

[4] Holmes, D.S. and Gordon, B. (1992) Automation can Make Your SPC Charts Send Wrong Messages. Quality, June, 20-21.

[5] Holmes, D.S. and Mergen, A.E. (1988) Selecting Acceptance Sampling Plans by Expected Average Outgoing Quality. North East Decision Sciences Institute Proceedings, 177-179.

[6] Mergen, A.E. (1981) The Lot Quality Distribution for a Dependent Production Process and its Impact on Quality Assurance Plans, Ph.D. Thesis, University Microfilms International, Michigan, U.S.A.

[7] Mergen, A.E. and Holmes, D.S. (1986) Lot Quality Distribution for Dependent Processes. ASQC Annual Quality Congress Transactions, 264-268.

[8] Montgomery, D.C. and Mastrangelo, C.M. (1991) Some Statistical Control Methods for Autocorrelated Data. Journal of Quality Technology, 23, 3, 179-193.

[9] Runger, G.C. and Willemain, T.R. (1995) Model-Based and Model-Free Control of Autocorrelated Processes. Journal of Quality Technology, 27, 4, 283-292.

[10] Wetherill,G.B. (1977) Sampling Inspection and Quality Control, Halsted Press, 2nd Ed., p.17.

ÖZET

Bir üretilmiş mallar öbeğinde yer alan aranan niteliklere uygun olmayan mallar o öbeğin kalitesini belirler. Öbek kalitesi için olasılık dağılımı öbek kalite dağılımı (LQD) adını alır. Teknolojik gelişmelere bağlı olarak, üretilen malların kalitesi zaman içinde değişiklik gösterir. Bu, bir üretim sürecinin otokorrelasyon özelliğinin dikkate alınmasını gerektirir. Bu çalışmada bir üretim sürecinin Markov Zinciri özelliğine sahip olduğu varsayılarak Bernoulli dağılımının genelleştirilmiş hali olan bir LQD elde edilmiştir.