GENERATION OF PROBABILITY MEASURES WITH THE GIVEN SPECIFICATION OF THE SMALLEST BANS

Alexander A. Grusho, Nick A. Grusho and Elena E. Timonina
Institute of Informatics Problems of RAS
Vavilova 44-2,
119333, Moscow, Russia
Email: grusho@yandex.ru

KEYWORDS

Consistent sequences of criteria, bans of probability measures in the discrete spaces, generation of probability measures

ABSTRACT

A ban means a sequence which has zero probability in a finite space. Generation of probability models is carried out, as a rule, from simpler models by introduction of additional restrictions. However fulfilment of required properties for stochastic processes requires the proof in case of introduction of additional restrictions. In particular, the proof is required that restrictions on admissibility of trajectories don't destroy the property of being a random process that is to satisfy to the Kolmogorov's theorem.

The paper deals with conditions under which introduction of restrictions on trajectories of random sequences according to the given specification of the smallest bans again generates random process.

When the probability measure Q is generated by restrictions defined by bans we consider testing of sequence of hypotheses $H_{0,n} : Q_n$ against $H_{1,n} : P_n$, where Q possesses the specification of the smallest bans, P is a uniform measure and Q_n, P_n are projections of measures Q and P. The existence of consistency of sequence test defined by bans is investigated.

INTRODUCTION

There is a problem of anomalies search in behavior of researched processes in monitoring and control systems and in case of simulation of complex systems. One of the main tools in the solution of such problems is mathematical statistics. However in case of a nonzero error of decision-making based on observation, anomaly detection process generates a large number of false alarms (Axelson, 1999) which complicate or do impossible the analysis of the reasons of anomalies. For random processes with discrete time and finite set of states we found a case when probability of false alarm is equal zero when hypotheses are tested. But at the same time the probability of the correct decision on existence of anomalies tends to 1. This approach is based on bans of probability measures (Grusho and Timonina, 2011; Grusho et al., 2013a,b).

In the previous studies (Grusho and Timonina, 2011; Grusho et al., 2013a,b) we introduced a definition of a ban for a probability measure on a finite space. A ban means a sequence which has zero probability in a finite space. We have shown that the notion of bans is convenient because it allows to determine the critical sets of statistical tests in the simplest way for calculation (Grusho and Timonina, 2011; Grusho et al., 2013a,b).

The analysis of anomalies in observed processes requires statistical modelling with use of an assessment of results by statistical techniques. However the application of statistical techniques is correct only in the conditions of correctly constructed probability models. In particular, if during experiment a remoted future behavior of the system is extrapolated, the probability model shall be represented by stochastic process at least (Prokhorov and Rozanov, 1993). In other words, extrapolation of any limited sections of a trajectory shall be consistent with probability distributions of process continuation. Such conditions are determined by Kolmogorov's theorem about the unique continuation of a probability measure due to the consistent finite-dimensional probability distributions. Therefore in case of study of the aberrant behavior of process the model shall meet conditions of stochastic process and restrictions on admissible trajectories of such process. Example of necessary extrapolation of a statistical output is property of consistency. At testing statistical hypotheses the property of consistency of the decision guarantees the reliability of outputs about model in case of increase in a segment of observation.

Generation of probability models is carried out, as a rule, from simpler models by introduction of additional restrictions. However fulfilment of required properties for stochastic processes requires the proof in case of introduction of additional restrictions. In particular, the proof is required that restrictions on admissibility of trajectories doesn't destroy the property of being a random process that is to satisfy to the Kolmogorov's theorem.

In this paper conditions under which introduction of restrictions on trajectories of random sequences according to the given specification of the smallest bans again generates random process are researched.

The article is structured as follows. Section 2 introduces definitions and previous results. Section 3 defines conditions for the case when probability measure is generated according to given specification of smallest bans. In Section 4 we give an example of usage of proved conditions. In Section 5 we shortly analyze applications to consistent sequences of tests.

Proceedings 28th European Conference on Modelling and Simulation ©ECMS Flaminio Squazzoni, Fabio Baronio, Claudia Archetti, Marco Castellani (Editors)
ISBN: 978-0-9564944-8-1 / ISBN: 978-0-9564944-9-8 (CD)

MATHEMATICAL MODEL. BASIC DEFINITIONS AND PREVIOUS RESULTS

Let $X = \{x_1, ..., x_m\}$ be a finite set, X^n be a Cartesian product of X, X^∞ be a set of all sequences when i-th element belongs to X. Define $\mathcal{A}$ be a σ-algebra on X^∞, generated by cylindrical sets. $\mathcal{A}$ is also Borel σ-algebra in Tychonof product X^∞, where X has a discrete topology (Bourbaki, 1968; Prokhorov and Rozanov, 1993).

On $(X^\infty, \mathcal{A})$ a probability measure P is defined. Assume P_n is a projection of P on the first n coordinates of sequences from X^∞. It is clear that for every $B_n \subseteq X^n$

$$P_n(B_n) = P(B_n \times X^\infty).$$

Let D_n be a support of measure P_n:

$$D_n = \{\overline{x}_n \in X^n,\ P_n(\overline{x}_n) > 0\}.$$

Denote $\Delta_n = D_n \times X^\infty$. The sequence Δ_n, n=1,2,..., is nonincreasing and

$$\Delta_P = \lim_{n\to\infty} \Delta_n = \bigcap_{n=1}^{\infty} \Delta_n.$$

The set Δ_P is closed and it is a support of P.

If $\overline{x}_k \in X^k$, then $\widetilde{x}_{k-1}$ is obtained from $\overline{x}_k$ by dropping the last coordinate.

Definition 1. Ban in measure P_n is a vector $\overline{x}_k \in X^k$, $k \le n$, such that

$$P_n(\overline{x}_k \times X^{n-k}) = 0.$$

If

$$P_{k-1}(\widetilde{x}_{k-1}) > 0,$$

then $\overline{x}_k$ is the smallest ban.

If $\overline{x}_k$ is a ban in P_n then for every $k \le s \le n$ and for every $\overline{x}_k$ sequence starting with $\overline{x}_k$ we have

$$P_s(\overline{x}_s) = 0.$$

In fact, if $P_k(\overline{x}_k) = 0$ then

$$P(\overline{x}_k \times X^\infty) = 0,$$

and

$$P(\overline{x}_k \times X^{s-k} \times X^\infty) = 0.$$

It follows that

$$P_s(\overline{x}_s) = P(\overline{x}_s \times X^\infty) \le P(\overline{x}_k \times X^{s-k} \times X^\infty) = 0.$$

If there exists $\overline{x}_n \in X^n$ such that $P_n(\overline{x}_n) = 0$ then there exists the smallest ban.

Let for all n the support of measure P_n equals to X^n. Then the support of measure P equals to X^∞. Let's assume that the specification of the smallest bans is a set

$$\nu = \{\nu_n,\ n = 1, 2, ...\},$$

where ν_i be a number of the smallest bans of length n. The problem consists in using a measure P and the specification ν to construct a probability measure Q on space $(X^\infty, \mathcal{A})$ at which the set of the smallest bans possesses the specification ν. For creation Q at first we will construct the consistent system of probability measures $Q_n, n = 1, 2...$, which under the Kolmogorov's theorem will unambiguously determine measure Q. Let's denote $D_n, n = 1, 2, ..., D_n \subseteq X^n$, supports of measures Q_n, and through d_n powers of these supports. In paper (Grusho et al., 2013a) it is proved that numbers d_n, $n = 1, 2, ...$, are unambiguously connected to the specification of ν in the following ratios

$$\nu_1 m^{n-1} + \cdots + \nu_{n-1} m + \nu_n + d_n = m^n. \qquad (1)$$

for all $n = 1, 2, ...$ Thus, it is necessary to construct the consistent family of probability measures $\{Q_n\}$ which powers of supports are unambiguously determined by ratios (1).

GENERATION OF PROBABILITY MEASURES WITH THE GIVEN SPECIFICATION OF THE SMALLEST BANS

Let $D_n, n = 1, 2, ..., D_n \subseteq X^n$, be some family of the sets fitting (1), $\overline{x}_n$ be a arbitrary element of X^n. For all $n, n = 1, 2, ...$, we will define functions

$$g_{n+1}:\ D_{n+1} \to X^n$$

as follows. For all $\overline{x}_{n+1} \in D_{n+1}$, $\overline{x}_{n+1} = \overline{x}_n x$,

$$g_{n+1}(\overline{x}_{n+1}) = \overline{x}_n.$$

In addition to (1) on sets $\{D_n,\ D_n \subseteq X^n,\ n = 1, 2...,\}$ we will superimpose the following two restrictions connected to functions g_n. For all $n, n = 1, 2, ...$, and all $\overline{x}_{n+1} \in D_{n+1}$

$$g_{n+1}(\overline{x}_{n+1}) \in D_n, \qquad (2)$$

$$g_{n+1}:\ D_{n+1} \longrightarrow^{on} D_n. \qquad (3)$$

Functions h_n are determined by analogy to functions g_n for sequence of sets X^n so that for all n, $n = 1, 2, ...$, and all $\overline{x}_n x \in X^{n+1}$

$$h_{n+1}(\overline{x}_n x) = \overline{x}_n.$$

As measures P_n have supports X^n, it is obvious that conditions (1), (2) and (3) are executed for sequence $\{X^n, h_n\}$.

Let's take arbitrary sequence of surjective functions

$$f_n:\ X^n \to D_n,\ n = 1, 2, ...$$

Each such function generates on X^n a probability measure Q_n with the support D_n.

Then for all n functions g_{n+1} and measures Q_{n+1} generate on X^n probability measures Q'_n with supports D_n (because functions f_{n+1} and g_{n+1} are surjective).

Theorem 1. *Let arbitrary family of probability measures $\{Q_n\}$ with supports $\{D_n\}$ and family of functions $\{g_n\}$, which satisfy the conditions (2) and (3), be set. Family of probability measures $\{Q_n\}$ is the consistent if and only if for all n equalities $Q_n = Q'_n$ are executed.*

Proof. We will prove the sufficiency. For consistency of measures it is enough that for all $\overline{x}_n \in X^n$

$$Q_n(\overline{x}_n) = Q_{n+1}(\overline{x}_n,\, X).$$

From the finiteness of probability schemes

$$Q_{n+1}(\overline{x}_n,\, X) = \sum_{x \in X} Q_{n+1}(\overline{x}_n x).$$

By definition

$$Q'_n(\overline{x}_n) = Q_{n+1}(g^{-1}_{n+1}(\overline{x}_n)) = \sum_{x \in D_{n+1}} Q_{n+1}(\overline{x}_n x) =$$

$$= \sum_{x \in X} Q_{n+1}(\overline{x}_n x) = Q_{n+1}(\overline{x}_n,\, X).$$

Under the theorem condition

$$Q'_n(\overline{x}_n) = Q_n(\overline{x}_n).$$

for all $\overline{x}_n \in X^n$. From this it follows that for all $\overline{x}_n \in X^n$

$$Q_n(\overline{x}_n) = Q_{n+1}(\overline{x}_n,\, X).$$

Sufficiency is proved.

We will prove necessity. If $\{Q_n\}$ is the consistent family of probability measures, then for all $\overline{x}_n \in X^n$

$$Q_{n+1}(\overline{x}_n,\, X) = Q_n(\overline{x}_n).$$

Besides, for all $\overline{x}_n \in X^n$

$$Q'_n(\overline{x}_n) = Q_{n+1}(g^{-1}_{n+1}(\overline{x}_n)) = Q_{n+1}(\overline{x}_n,\, X) = Q_n(\overline{x}_n).$$

The theorem is proved.

Family of functions $\{f_n\}$ and a probability measure P generate family of probability measures $\{Q_n\}$ with supports $\{D_n\}$. Let functions $\{g_n\}$ satisfy conditions (2) and (3). Then fairly following proposition.

Corollary 1. Family of functions $\{f_n\}$ and probability measure P generate the only probability measure Q if and only if when for all $n = 1, 2, ...,$ the equality $Q_n = Q'_n$ is executed.

Theorem 2. *For consistency of a set of probability measures $\{Q_n\}$ generated by functions $\{f_n\}$ and projections of a measure P on X^∞ it is enough that for all n the following diagrams are commutative*

$$\begin{array}{ccc} X^n & \longleftarrow^{h_{n+1}} & X^{n+1} \\ f_n \downarrow & & \downarrow f_{n+1} \\ D_n & \longleftarrow^{g_{n+1}} & D_{n+1} \end{array} \qquad (4)$$

where $\{g_n\}$ satisfy (2) and (3)

Proof. Each function f_n and measure P_n on X^n generate on D_n a probability distribution Q_n. Owing to consistency of projections of a measure P each function h_{n+1} generates a measure P_n from a measure P_{n+1}. Therefore it is possible to consider that the measure Q_n is generated from a measure P_{n+1} by means of composition of functions $(f_n \star f_{n+1})$.

In turn function f_{n+1} and the measure P_{n+1} generate distribution of probabilities Q_{n+1} on D_{n+1}. This measure and function g_{n+1} generate a measure Q'_n on D_n. That is the measure Q'_n on D_n is generated from measure P_{n+1} by means of a function composition $(g_{n+1} \star f_{n+1})$. Under the condition (4) functions $(f_n \star h_{n+1})$ and $(g_{n+1} \star f_{n+1})$ are equal. Therefore, these functions and measure P_{n+1} generate on D_n the same measure. That is $Q_n = Q'_n$. From here and theorems 1 consistency of family of probability measures $\{Q_n\}$ follows. The theorem is proved.

EXAMPLE OF GENERATING OF PROBABILITY MEASURES WITH THE GIVEN SPECIFICATION OF THE SMALLEST BAN

Let's consider $\nu = \{\nu_n = 1, n = 1, 2, ...\}$, $X = \{0, 1, ..., m-1\}$, $m > 2$, and P be a uniform measure on X^∞. We will give an example of creation of measure Q with the specification of the smallest bans ν.

All vectors $\overline{x}_n \in X^n$ can be considered as numbers in m-dimensional numeration system. Then in each set $B_n \subseteq X^n$ there is the smallest vector $\overline{x}_n$ considered as a number. We will build a required measure inductively. In D_1 the smallest ban we will equal to 0. Function f_1 maps X on $X \setminus \{0\}$. We will assume that D_n and f_n are defined. We will define D_{n+1}. Let $\overline{x}^0_n$ be the smallest number in D_n. In a set $D_n \times X$ we will define the smallest ban equals to $(\overline{x}^0_n 0)$. Let's suppose

$$D_{n+1} = (D_n \times X) \setminus \{(\overline{x}^0_n 0)\}.$$

Let's construct a surjective function

$$f_{n+1} :\ X^{n+1} \longrightarrow^{on} D_{n+1}.$$

For all $\overline{x}_n x \in X^{n+1}$ except that in which $f_n(\overline{x}_n) = \overline{x}^0_n$ and $x = 0$, we will suppose that

$$f_{n+1}(\overline{x}_n x) = (f_n(\overline{x}_n), x) \in D_n \times X.$$

Let's denote $\overline{y}^{(i)}_n$, $i = 1, ..., k$, all members of set $f^{-1}_n(\overline{x}^0_n, 0)$. We will define

$$f_{n+1}(\overline{y}^{(i)}_n, 0) = (\overline{x}^0_n, 1) \in D_n \times X.$$

We will notice that $(\overline{x}^0_n, 1)$ is the smallest number in D_{n+1}. By definition f_n is mapping X^n on D_n. Therefore f_{n+1} maps X^{n+1} on $D_{n+1} = (D_n \times X) \setminus \{(\overline{x}^0_n, 0)\}$. Under construction D_{n+1} from D_n, for D_{n+1} there is only one smallest ban $(\overline{x}^0_n, 0)$, that is $\nu_{n+1} = 1$.

We will prove that diagrams (4) are commutative. Under construction f_n and h_{n+1}

$$f_n(\overline{y}^{(i)}_n) = \overline{x}^0_n.$$

Therefore for all $x \in X$

$$(f_n * h_{n+1})(\overline{y}^{(i)}_n, x) = \overline{x}^0_n.$$

In case of $\overline{x}_n \neq \overline{y}^{(i)}_n, i = 1, ..., k,$

$$(f_n * h_{n+1})(\overline{x}_n, x) = f_n(\overline{x}_n).$$

Further

$$f_{n+1}(\overline{y}_n^{(i)},0) = (\overline{x}_n^0, 1) \in D_{n+1},\ i = 1, ..., k,$$

$$(f_n * g_{n+1})(\overline{y}_n^{(i)},0) = f_n(\overline{y}_n^{(i)}) = \overline{x}_n^0.$$

For elements $(\overline{y}_n^{(i)}, x)$, $x \neq 0$, $i = 1, ..., k$,

$$f_{n+1}(\overline{y}_n^{(i)},x) = (f_n(\overline{y}_n^{(i)}, x) = (\overline{x}_n^0, x) \in D_{n+1}.$$

Therefore in case of $x \neq 0$

$$(f_{n+1} * g_{n+1})(\overline{y}_n^{(i)},x) = \overline{x}_n^o.$$

In case of $\overline{x}_n \neq \overline{y}_n^{(i)}, i = 1, ..., k$, by determination f_{n+1} we have that for all $x \in X$

$$f_{n+1}(\overline{x}_n, x) = (f_n(\overline{x}_n), x) \in D_{n+1}.$$

Then under construction

$$(g_{n+1} * f_{n+1}) = (f_n * h_{n+1}).$$

Commutativity of diagrams (4) is proved.

From here it follows the existence of a measure Q on $(X^\infty, \mathcal{A})$ with the specification of the smallest bans $\nu = \{\nu_i = 1,\ i = 1, 2, ...\}$.

APPLICATION TO THE CONSISTENCY ANALYSIS

From ratios (1) we receive ratios (5)

$$d_{n+1} - md_n + \nu_{n+1} = 0,\ n = 1, 2, ... \tag{5}$$

Let P be a uniform measure on $(X^\infty, \mathcal{A})$. Then the relation $\frac{d_n}{m^n}$ is probability of the set D_n in a measure P_n. For the specification $\nu = \{\nu_i = 1,\ i = 1, 2, ...\}$, from (5) we receive the following ratios

$$\frac{d_n}{m^n} = 1 - \frac{1}{m-1} + \frac{1}{(m-1)m^n}.$$

In case of $n \to \infty$ a limit of this probability is equal to probability P of support Δ_Q of measure Q

$$P(\Delta_Q) = 1 - \frac{1}{m-1} > 0.$$

From necessary and sufficient conditions (Grusho and Timonina, 2011) of existence of consistent sequences of tests, determined by bans, for testing hypotheses $H_{0,\,n} : Q_n$ against $H_{1,\,n} : P_n$ it follows that such sequences of tests don't exist.

The vector $\overline{x}_k \in X^k$ is called the minimum ban if for all vectors $\overline{x}_n \in X^n$, $P_n(\overline{x}_n) > 0$, the vector $(\overline{x}_n, \overline{x}_k) \in X^{n+k}$ is the smallest ban.

Let P be a uniform measure which generates measure Q. We will notice that if in a measure Q all bans are defined by nonempty finite set of the minimum bans, there is the consistent sequence of tests determined by bans. It follows from the fact that in measure P_n the probability of absence of the fixed vector of length k tends to 0 in case of $n \to \infty$. It means that $P(\Delta_Q) = 0$ and it is possible to apply the theorem in (Grusho and Timonina, 2011).

From this it follows that having divided both parts of ratio (1) on m^n, there exists a limit in case of $n \to \infty$. We will receive the following equality

$$\frac{\nu_1}{m} + \frac{\nu_2}{m^2} + \cdots + \frac{\nu_n}{m^n} \cdots = 1.$$

Thus, the set of numbers $\{\frac{\nu_n}{m^n},\ n = 1, 2, ...\}$, form a distribution of probabilities on a set of natural numbers. In particular,

$$\frac{\nu_n}{m^n} \to 0,\ n \to \infty.$$

Ratios (5) allow receiving sufficient conditions when the specification ν of the smallest bans guarantees existence of consistent sequence of tests, determined by bans, in testing of sequence of hypotheses $H_{0,\,n} : Q_n$ against $H_{1,\,n} : P_n$, where Q possesses the specification ν, and P is a uniform measure. From (5) equality follows

$$\frac{d_{n+1}}{m^{n+1}} = \frac{d_n}{m^n} - \frac{\nu_{n+1}}{m^{n+1}}.$$

We will suppose

$$\nu_{n+1} = \varepsilon_n m d_m,\ 0 < \varepsilon_n < 1.$$

Then

$$\frac{d_{n+1}}{m^{n+1}} = \frac{d_n}{m^n}(1 - \varepsilon_n). \tag{6}$$

From (6) it follows

$$\frac{d_{n+1}}{m^{n+1}} = \frac{d_1}{m} \exp\{\sum_{i=1}^{n} \ln(1 - \varepsilon_n)\}. \tag{7}$$

In case of $n \to \infty$ the right part of (7) tends to 0 for ε_n, satisfying to inequality

$$\frac{1}{n^\alpha} < \varepsilon_n < 1,\ 0 < \alpha \leq 1.$$

In these cases there exists (Grusho and Timonina, 2011; Grusho et al., 2013a) a consistent sequence of tests determined by bans.

CONCLUSION

In the paper the conditions of correct construction of probability models with the given specification of the smallest bans are received. Correct construction of stochastic models allows to use well developed tools of the theory of random sequences and processes in the analysis of statistical data.

We will mark that in the case of $d_n \equiv 1$, there is no open set in the support of measure Q. In the example of section 4 in case of $\nu = \{\nu_i = 1,\ i = 1, 2, ..., \}$ in Δ_Q such open set exists. All open set in a uniform measure has probability bigger than 0.

Authors couldn't prove or refute a hypothesis that $P(\Delta_Q) > 0$ if and only if in case when in Δ_Q there is an open set. Correctness of this proposition would help to simplify the proof of existence of consistent sequence of tests, determined by bans, for the alternatives, which dominate a uniform measure.

Acknowledgements
This work was supported by the Russian Foundation for Basic Research (grant 13-01-00480).

REFERENCES

Axelson, S. 1999. "The Base-Rate Fallacy and its Implications for the Difficulty of Intrusion Detection". In *Proc. of the 6th Conference on Computer and Communications Security.*

Bourbaki, N. 1968. *Topologie G'en'erale. Russian translation.* Science, Moscow.

Grusho, A., N. Grusho and E. Timonina. 2013. "Consistent sequences of tests defined by bans". *Springer Proceedings in Mathematics and Statistics, Optimization Theory, Decision Making, and Operation Research Applications*, 281-291.

Grusho, A., N. Grusho and E. Timonina. 2013. "Criteria on statistically defined bans". In *Proceedings of 27th European Conference on Modelling and Simulation* (May 27-30, 2013, Alesund, Norway). Digitaldruck Pirrot GmbHP, Dudweiler, Germany, 610-614.

Grusho, A. and E. Timonina. 2011. "Prohibitions in discrete probabilistic statistical problems". *Discrete Mathematics and Applications* 21, No.3, 275-281.

Prokhorov, U.V.; and U.A. Rozanov. 1993. *Theory of probabilities.* Science, Moscow.

AUTHOR BIOGRAPHIES

ALEXANDER A. GRUSHO was born 28.08.1946, Doctor of sciences (Math), Professor of Moscow State University, leading scientist of Institute of Informatics Problems of Russian Academy of Sciences. His email is `grusho@yandex.ru`.

NICK A. GRUSHO was born 15.06.1982, PhD (Math), senior scientist of Institute of Informatics Problems of Russian Academy of Sciences. His email is `info@itake.ru`.

ELENA E. TIMONINA was born 27.02.1952, Doctor of sciences (Tech), Professor, leading scientist of Institute of Informatics Problems of Russian Academy of Sciences. Her email is `eltimon@yandex.ru`.