CRITERIA ON STATISTICALLY DEFINED BANS

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

KEYWORDS

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

ABSTRACT

The method of statistical determination of smallest bans of probability measures on discrete spaces is offered. Consistency of the constructed estimates is shown. The application of the received estimates for testing of statistical hypotheses in the discrete spaces is constructed. It is shown that estimates of bans can generate consistent sequences of criteria in some sense.

INTRODUCTION

Finite probability spaces play an essential role in cases of different simulation problems of network and computer security, cryptography, etc. Information flows analysis is less effective if there is positive probability of false alarms. Besides complexity of statistical analysis sometimes prevents from big data controls in realtime. Bans of finite probability measures can help to solve these problems. We consider the problem of testing of sequence of simple hypotheses $H_{0,n}$ against complex alternatives $H_{1,n}$ in finite spaces. Let $X = \{x_1, ..., x_m\}$ be a finite set, X^n be the Cartesian product of a set X, X^∞ be a set of the infinite sequences with elements from X, $\mathcal{A}$ be the minimum σ-algebra generated by all cylindrical sets (Lehmann, 1997; Neveu, 1964; Bourbaki, 1968). If on X the discrete topology is considered, Tychonof product X^∞ is compact topological space with countable basis (Bourbaki, 1968; Prokhorov and Rozanov, 1993). In this case Borel σ-algebra $\mathcal{B}$ is also σ-algebra $\mathcal{A}$.

Let P_0 - a probability measure on $(X^\infty, \mathcal{A})$. For each n $P_{0,n}$ is a projection of P_0 to the first n coordinates of sequences from X^∞.

Earlier (Grusho and Timonina, 2011a,b) we defined concept of a ban of a finite probability measure. Bans are understood as initial sections of sequences from X^∞ which probabilities are equal to zero. In statistical problems there exist the criteria which critical sets completely are defined by bans. In (Grusho and Timonina, 2011a,b) necessary and sufficient conditions of existence of the consistent sequences of criteria (CSC) which critical sets are completely defined by bans are proved.

The statistical criterions defined by bans, possess the property that in case of a null hypothesis $P_{0,n}(S_n) = 0$, where S_n, $n = 1, 2, ...$, - critical sets of criteria. In some cases of probability measures it is possible to define bans of these measures quite easily.

Let Λ_n be a finite set of bans of measure $P_{0,n}$. Let's assume that we can independently produce vectors from X^n with probability measures $P_{0,n}$ for any finite n. Then it is possible to construct a consistent estimation of a set Λ_n.

In section 2 estimates of sets Λ_n, $n = 1, 2, ...$, are constructed, and their consistency are proved. In section 3 we use bans for testing of hypotheses $H_{0,n}$ against $H_{1,n}$ taking into account possibility of inexact determination of sets Λ_n. In conclusion the received results are discussed.

ESTIMATES OF SETS OF SMALLEST BANS

Let on $(X^\infty, \mathcal{A})$ the probability measure P_0 is defined. It is obvious that for any $B_n \in X^n$

$$P_{0,n}(B_n) = P_0(B_n \times X^\infty). \tag{1}$$

Let $D_{0,n}$ be the support of measure $P_{0,n}$:

$$D_{0,n} = \{\overrightarrow{x}_n \in X^n, \, P_{0,n}(\overrightarrow{x}) > 0\}. \tag{2}$$

Denote $\Delta_{0,n} = D_{0,n} \times X^n$. Sequence $\Delta_{0,n}$, $n = 1, 2, ...$, is nonincreasiing and

$$\Delta_0 = \lim_{n\to\infty} \Delta_{0,n} = \bigcap_{n=1}^{\infty} \Delta_{0,n}. \tag{3}$$

The set Δ_0 is closed and is the support of the measure P_0.

We also consider a set of probability measures $\{P_\theta, \, \theta \in \Theta\}$ on $(X^\infty, \mathcal{A})$ for which $P_{\theta,n}$, $D_{\theta,n}$, $\Delta_{\theta,n}$, Δ_θ are defined.

If $\overline{\omega}_k \in X^k$, then $\widetilde{\omega}_{k-1}$ is received from $\overline{\omega}_k$ discarding of the last coordinate.

Definition 1. In a measure $P_{0,n}$ the vector $\overline{\omega}_k \in X^k$, $k \le n$, such that

$$P_{0,n}(\overline{\omega}_k \times X^{n-k}) = 0 \tag{4}$$

is called the ban.

If $P_{0,k-1}(\widetilde{\omega}_{k-1}) > 0$, then $\overline{\omega}_k$ is called the smallest ban.

If $\overline{\omega}_k$ is the ban of a measure $P_{0,n}$, then for any $k \le s \le n$ and for any vectors $\overline{\omega}_s$ beginning with $\overline{\omega}_k$, we have

$$P_{0,s}(\overline{\omega}_s) = 0. \tag{5}$$

Proceedings 27th European Conference on Modelling and Simulation ©ECMS Webjørn Rekdalsbakken, Robin T. Bye, Houxiang Zhang (Editors)
ISBN: 978-0-9564944-6-7 / ISBN: 978-0-9564944-7-4 (CD)

Further under Λ_n we will understand a set of the smallest bans of $P_{0,n}$.

All bans of a measure $P_{0,n}$ are defined by the smallest bans (Grusho and Timonina, 2011b), i.e. any ban contains at least one element from a set Λ_n. Assume that we don't know what elements enter into a set Λ_n and set capacity. Let's construct a consistent estimate of a set Λ_n in the specified assumptions.

Let $\overline{\omega}_n^{(1)}$, $\overline{\omega}_n^{(2)}$, ..., $\overline{\omega}_n^{(N)}$ be a sample from the distribution $P_{0,n}$. In the received sample we consider frequencies $\nu_n^{(i)}$, $i = 1, 2, ..., m^n$, of occurrence of all vectors. We select vectors from X^n which frequencies are equal to zero. The selected vectors define potential smallest bans of $P_{0,n}$ due to (5).

The constructed set of potential smallest bans Λ'_n as a result of specified algorithm can differ from the true set Λ_n. The set Λ'_n is an estimate of the set Λ_n. The following errors are possible. The s-chain which isn't the smallest ban, randomly never met in the viewed sample and belongs to a set Λ'_n. Then in case of changeover Λ_n by an estimation Λ'_n we receive excess smallest bans. Thus, Λ'_n contains Λ_n.

Let's consider probability of that the n-chain didn't belong to the sample randomly. Let's denote probabilities of appearance of all n-chains through $p_1^{(n)}$, $p_2^{(n)}, ..., p_{m^n}^{(n)}$. The probability that the frequency of the chain i is equal to 0, equals to $(1 - p_i^{(n)})^N$. Then mathematical expectation of a number of elements in Λ'_n equals to $\sum_{i=1}^{m^n}(1 - p_i^{(n)})^N$.

Let's denote

$$\epsilon = \min_{1 \le i \le m^n,\, p_i^{(n)} > 0} p_i^{(n)}. \tag{6}$$

Then mathematical expectation of power of the set Λ'_n equals to $|\Lambda_n| + E\xi$, where ξ - a random variable, equal to number of randomly not met n-chains. It is obvious that

$$E\xi = \sum_{i=1,\, p_i^{(n)} > 0}^{m^n} (1 - p_i^{(n)})^N. \tag{7}$$

For $E\xi$ the following estimation is fair

$$E\xi \le m^n(1 - \epsilon)^N. \tag{8}$$

$E\xi$ tends to 0. Let's denote through λ_N probability of that $|\Lambda'_n| > |\Lambda_n|$. Using Markov's inequality, we receive that in case of $N \longrightarrow \infty$ $\lambda_N \longrightarrow 0$.

CONSISTENT SEQUENCES OF CRITERIA WITH USE OF THE SMALLEST BANS

Let's create the criteria depending on bans. Earlier we assumed that there is a family of distributions $\{P_\theta,\, \theta \in \Theta\}$. There is (Grusho and Timonina, 2011a) necessary and sufficient condition of existence of CSC determined by bans. That is the condition $P_\theta(\Delta_0) = 0$ for $\forall\theta \in \Theta\}$.

Let the condition of existence of CSC determined by bans $\bigcup_{n=1}^{\infty} \Lambda_n$ be satisfied. Then the sequence of criteria with the critical sets consisting of all sequences, containing smallest bans, is consistent sequence of criteria for testing of hypotheses $H_{0,n} : P_{0,n}$, $n = 1, 2, ...$, against alternatives $H_{1,n}$.

In (Grusho and Timonina, 2011b) it was proved that the following equations between capacities of supports of $P_{0,n}$, $n = 1, 2, ...$, and numbers of the smallest bans of these measures take place

$$\nu_1 m^{n-1} + \cdots + \nu_{n-1} m + \nu_n + |D_{0,n}| = m^n. \tag{9}$$

where ν_i - number of the smallest bans of length of i. From this ratio it is clear that if we increase number of the smallest bans the sizes of supports $D_{0,n}$ decrease. It means that instead of the initial measure P_0 with a set of the smallest bans $\bigcup_{n=1}^{\infty} \Lambda_n$ we for determination of the null hypotheses consider some measure P'_0. For this measure the support Δ'_0 can be less, than the support of the initial measure P_0, i.e. $\Delta'_0 \subseteq \Delta_0$. Therefore having checked a condition of existence of CSC depending on bans, for P_0 we guarantee execution of this condition for a measure P'_0 with a set of bans $\bigcup_{n=1}^{\infty} \Lambda'_n$.

It is obvious that $\Delta_0 \backslash \Delta'_0$ is a measurable set. Let's denote $P_0(\Delta_0 \backslash \Delta'_0) = \mu$. Then P'_0 can be defined as follows.

$$P'_0(A) = \frac{1}{1 - \mu} P_0(A \cap \Delta'_0). \tag{10}$$

Let CSC depending on bans, for testing $H'_{0,n}$ against $H_{1,n}$ (above we proved that such sequence exists) is defined by critical sets S_n, $n = 1, 2, ...$ Then

$$P_{0,n}(S_n) = P_0(S_n \times X^\infty) =$$
$$= P_0((S_n \times X^\infty) \cap (\Delta_0 \backslash \Delta'_0)) +$$
$$+ P_0((S_n \times X^\infty) \cap (\Delta_0 \cap \Delta'_0)). \tag{11}$$

The second item is equal to zero since a critical set S_n in a measure $P'_{0,n}$ has probability 0.

Using determination of μ, we receive an estimation

$$P_0((S_n \times X^\infty) \cap (\Delta_0 \backslash \Delta'_0)) \le \mu. \tag{12}$$

From this it follows that the sequence of criteria with critical sets S_n in case of testing $H_{0,n}$ against $H_{1,n}$ has significance value μ. Thus for $\forall\theta \in \Theta$

$$P_{\theta,n}(S_n) \to 1. \tag{13}$$

Let's recall that μ is a random variable in the probability scheme connected to statistical determination of the smallest bans. We showed that in case of $N \to \infty$ with the probability tends to 1 (in the scheme of an estimation of the smallest bans), $\Lambda'_n = \Lambda_n$. In this case $\mu = 0$ and $P_0 = P'_0$. Thus, we receive that in case of $N \to \infty$ CSC, constructed for testing $H'_{0,n}$ against $H_{1,n}$, will be also consistent for testing $H_{0,n}$ against $H_{1,n}$.

CONCLUSION

Knowledge of a condition $P_\theta(\Delta_0) = 0$ for $\forall\theta \in \Theta$ is enough to define the sequence of criteria depending of bans which asymptotically (in some sense) will be consistent.

Acknowledgements
This work was supported by the Russian Foundation for Basic Research (grant 13-01-00215).

REFERENCES

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

Grusho, A. and Timonina E. (2011). "Prohibitions in discrete probabilistic statistical problems". *Discrete Mathematics and Applications*. 21(3), 275-281.

Grusho, A. and Timonina E. (2011). "Statistical Tests Based on Bans". In *Proceedings of 1st International Symposium and 10th Balkan Conference on Operational Research*, (September 22-24, Thessaloniki, Greece), Vol. 1, 234-241.

Lehmann, E. L. (1997). *Testing Statistical Hypotheses (Springer Texts in Statistics).* 2nd ed. Springer.

Neveu , J. (1997). *Bases mathematiques du calcul des probabilites*. Paris: Masson.

Prokhorov, U. V., and Rozanov, U. A. (1993). *Theory of probabilities*. Moscow: Science.

AUTHOR BIOGRAPHIES

ALEXANDER J. GRUSHO was born 28.08.1946. He is Dr. Sc., Professor, Leading Scientist of IPI RAN. His email is `grusho@yandex.ru`.

NICK A. GRUSHO was born 15.06.1982. He is PhD., Senior Scientist of IPI RAN. His email is `info@itake.ru`.

ELENA E. TIMONINA was born 27.02.1952. She is Dr. Sc., Professor, Leading Scientist of IPI RAN. Her email is `eltimon@yandex.ru`.