UDC 519.234.3

Adaptive hypothesis testing with unknown alternative distribution

V.O. Vasilyev

Moscow Institute of Physics and Technology (State University)

V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences

In this paper, Bayesian version of the mixture model with independent observations is considered. Formally the mixture model corresponds to the mixture density $f(x)$ that represents the probability density composed by weighted sum of local densities $f_i(x)$, $i = 1,2$ where a sum of weights p_i is equal to 1. Mixture models are used to make statistical inferences about the properties of the subpopulations given only observations on the pooled population. In particular, the problems of testing hypotheses H_i are solved when one makes a decision which class the observed sample x_n belongs to. These problems can be solved using different criteria. For example, according to the Neiman-Pearson or Bayesian criterion one has to construct statistics in the form of likelihood ratio and compare it with some threshold. However, the optimal solution requires an accurate information about all distributions and prior probabilities [1]. If all the local distributions are specified parametrically, then the problem is solved by the EM method (see [2], [3], [4]). If at least one of the local distributions is completely unknown and there is no training sample from the distribution of this class, then the famous EM method is not applicable.

In this article, a decision of the hypothesis testing problem using Bayesian criterion is proposed for the case when: (1) the distribution under the hypothesis H_0 is known, (2) the distribution under the alternative H_1 is completely unknown, and (3) the prior probabilities of hypotheses are also unknown. This statement of the problem is motivated by the following considerations. In many applications of signal processing the noise distribution, corresponding to the hypothesis H_0, are either known or can be accurately restored using characteristics of a measuring device. At the same time, the distribution of a useful signal is unknown because it never can be observed without noise and it is very difficult to gather any information about its characteristics. Therefore, the distribution of signal + noise, corresponding to the hypothesis H_1, is supposed to be unknown. The third condition is justified again by practice, where the useful signal in a pure form is not observed.

The formal statement of the problem is reduced to defining the two-component process (S_n, X_n), where (S_n) is unobservable component and (X_n) is observable one. Random variables $\{S_i\}_{i=1}^n$ are i.i.d. and $S_i = 0$ with a probability p and $S_i = 1$ with a probability $(1-p)$

correspondingly, where $0 \le p \le 1$. Let $F_0(x)$ and $F_1(x)$ be such conditional cdf, that $X_i \sim F_{s_i}, \forall i \in \{1,2\}$, where s_i is realization of S_i.

It is necessary to estimate an unobservable realization $\{s_i\}_{i=1}^n$ by an observed realization $\{x_i\}_{i=1}^n$ knowing only a pdf $f_0(x)$ under the hypothesis H_0. The prior probability p and a pdf $f_1(x)$, corresponding to the hypothesis H_1, are unknown. Developed approach is compared with Bayesian criterion: $s_i = 1$ if $f_1(x_i) / f_0(x_i) > p/(1-p)$, else 0, where all probability characteristics are known.

The main idea of the novel method reduces to the following. For Bayesian statement of the problem the marginal pdf of the observable component X_n is $f(x) = f_0(x)p + f_1(x)(1-p)$. Therefore one can rewrite the Bayesian criteria like: $s_i = 1$ if $f(x_i) / f_0(x_i) > 2p$. The unknown marginal pdf $f(x)$ and the probability p will be estimated below. For the density $f(x)$ the kernel density estimator $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i)$ is suggested, where $K_h(x) = K(x/h)/h$, $K(\cdot)$ is the kernel function; h is the bandwidth, $h > 0$; $\{x_i\}_{i=1}^n$ is the sample drawn from distribution with the pdf $f(x)$. To calculate h the unbiased cross-validation method (UCV) [5] is applied. The variable bandwidth kernel density estimation (VBKDE) [6] is used to receive a more accurate estimator $\hat{f}(x)$. For the unknown probability p following estimator

$$\hat{p} = \arg\max_{0 \le \tilde{p} \le 1} \frac{|\hat{S}''(\tilde{p})|}{(1 + (\hat{S}'(\tilde{p})^2)^{3/2}}$$

is proposed, where $\hat{S}(\tilde{p}) = \int (f_0(x)\tilde{p} - \hat{f}(x))^+ dx$ and $(a)^+ = \max(0, a)$.

This work is supported by Russian Foundation for Basic Research (RFBR 13-08-00744a).

References

1. *Lehmann E.* Testing statistical hypotheses. – New York: Wiley, 1959.
2. *Baum L., Petrie T.* Statistical inference for probabilistic functions of finite state Markov chains. – The Annals of Mathematical Statistics. – 1966. – V. 37, N 6. – P. 1554–1563.
3. *Dempster A., Laird N., Rubin D.* Maximum likelihood from incomplete data via the EM algorithm. – JSTOR. – 1977. – V. 39, N 1. – P. 1–38.
4. *Lindgren G.* Markov regime models for mixed distributions and switching regressions. – Scand J. Statist. – 1978. – V. 5. – P. 81–91.
5. *Rudemo M.* Empirical choice of histograms and kernel density estimators. – Scandinavian Journal of Statistics. – 1982. – V. 9. – P. 65–78.
6. *Salgado-Ugarte I.H., Pérez-Hernández M.A* Exploring the use of variable bandwidth kernel density estimators. – Stata Journal. – 2003. – V. 3, N 2. – P. 133–147.