Statistics & Probability Letters 12 (1991) 323–327
North-Holland

October 1991

Can one decide the type of the mean from the empirical measure?

Sanjeev R. Kulkarni

Laboratory for Information and Decision Systems, M.I.T., Cambridge, MA 02139, USA, and M.I.T. Lincoln Laboratory, Lexington, MA 02173, USA

Ofer Zeitouni *

Department of Electrical Engineering, Technion, Haifa 32000, Israel

Received March 1990
Revised September 1990

Abstract: The problem of deciding whether the mean of an unknown distribution is in a set A or in its complement based on a sequence of independent random variables drawn according to this distribution is considered. We propose an algorithm which leads to an a.s. correct decision for any A in a class of sets satisfying certain structural assumptions. This class includes not only all countable sets, but many uncountable sets as well. A refined decision procedure is also presented which, given a countable decomposition of A, can determine a.s. to which set of the decomposition the mean belongs. This extends and simplifies a construction by Cover.

Keywords: Hypothesis testing, empirical measure, Cramer's theorem.

1. Introduction

Consider the following hypothesis testing problem: Let $x_1, x_2, \ldots$ denote a sequence of i.i.d. random variables with unknown marginal law μ_T, with support [0, 1]. The mean of μ_T, denoted $\bar{\mu}_T$, belongs either to a (known) set A which has measure 0 or to its complement A^c. We want to decide, based on the observation sequence $x_1, x_2, \ldots$, whether $\bar{\mu}_T \in A$ or not.

This problem was considered by Cover (1973), who treated the case of $A = \mathcal{Q}_{[0,1]}$, the set of rationals in [0, 1], and more generally the case of countable A. He proposed there a test which, for any measure with $\bar{\mu}_T \in A$, will make (a.s.) only a finite number of mistakes whereas, for measures with $\bar{\mu}_T \in A^c \setminus N$, the test makes (a.s.) only a finite number of mistakes, where N is a set of Lebesgue measure 0. Koplowitz (1977) refined Cover's result in the case that A is countable. Specifically, he showed that if $\bar{A}$ (the closure of A) is countable then N is empty, while if $\bar{A}$ is uncountable then N is uncountable.

In this note, we extend the result of Cover (1973) by allowing the set A to be uncountable, not necessarily of measure 0, such that it satisfies the following structural assumption:

Assumption. There exists a monotone sequence of sets A_m increasing to A and an appropriate posi-

This work was supported by the U.S. Army Research Office under Contract DAAL03-86-K-0171 and by the Office of Naval Research under Air Force Contract F19628-90-C-0002.
* This work was partially done while this author visited the Laboratory for Information and Decision Systems, M.I.T.

tive sequence $\varepsilon(m) \to_{m\to\infty} 0$ such that, for each m the open blowup

$$C_m = A_m^{2\sqrt{2\varepsilon(m)}} \triangleq \left\{x: d(x, A_m) < 2\sqrt{2\varepsilon(m)}\right\}$$

is such that the Lebesgue measure of $C_m \setminus A_m$ is smaller than $1/m^2$.

We note that this assumption implies that if A has Lebesgue measure zero, it is of the first category (i.e., a countable union of nowhere dense sets). The assumption is satisfied by a class of interesting uncountable sets A, e.g., the Cantor set. Obviously, for countable sets, the assumption is satisfied. For more along these lines, cf. Lemma 2 and the remarks which follow Theorem 1.

In Section 2, we describe a decision algorithm which changes its decision after increasingly longer and longer intervals. Those intervals are chosen using entropy bounds. We prove that this algorithm shares the properties of Cover's decision rule, i.e., it makes a finite number of mistakes a.s. on the set A and on $A^c \setminus N$ for an appropriate set N of Lebesgue measure 0. (A characterization of N follows from our proof and is related to the one given in Koplowitz (1977).) In Section 3, the results are extended to allow a (countable) decision inside the set A, i.e., we allow for multiple hypothesis tests inside A.

2. The decision rule and proof of the main theorem

We begin by first describing the proposed decision rule. Let $B_m = A_m^{\sqrt{2\varepsilon(m)}}$. We will use in the sequel the fact that the open blowup B_m satisfies $(d(A_m, B_m^c))^2 \geqslant 2\varepsilon(m) > 0$, and that $(d(B_m, C_m^c))^2 \geqslant 2\varepsilon(m)$. Let $\beta(m)$ be a given sequence, to be defined below. For any input sequence $x_1, x_2, \ldots,$ form the subsequences

$$X^m \triangleq (x_{\beta(m-1)}, \ldots, x_{\beta(m)-1}).$$

The endpoints of these subsequences X^m form a parsing of the original sequence $x_1, x_2, \ldots.$ Let $\bar{\mu}_{X^m}$ denote the empirical mean of the sequence X^m. At the end of each parsing, make a decision whether $\bar{\mu}_T \in A$ according to whether $\bar{\mu}_{X^m} \in B_m$ or not. Between parsings, don't change the decision.

For the sequence $\beta(m)$ defined below in equation (2.7), we claim:

Theorem 1. (a) *For any measure μ_T with $\bar{\mu}_T \in A$, the decision rule will make (a.s.) only a finite number of mistakes, i.e., for a.e. ω there exists an $n(\omega)$ such that the decision is A for all $n > n(\omega)$.*

(b) *For any measure μ_T with $\bar{\mu}_T \in A^c \setminus N$, where N is a set of Lebesgue measure 0, the decision rule will make (a.s.) only a finite number of mistakes, i.e., for a.e. ω there exists an $n(\omega)$ such that the decision is A^c for all $n > n(\omega)$.*

Before proving the theorem, we introduce some notation and define the sequence $\beta(m)$. For a set $E \subset [0, 1]$, E^c denotes the complement of E and $\bar{E}$ denotes the closure of E, whereas E° denotes the interior of E. Let μ be a probability measure with support in [0, 1]. The mean of μ is denoted $\bar{\mu}$. Let

$$M_\mu(\lambda) \triangleq E_\mu(\exp(\lambda x))$$

denote the moment generating function of μ and let

$$\Lambda(\lambda) \triangleq \log(M(\lambda)).$$

Let

$$I_\mu(x) = \sup_\lambda (\lambda x - \Lambda(\lambda))$$

be the Legendre transform of $\Lambda(\lambda)$, and let $H(\nu \mid \mu)$ denote the relative entropy of ν with respect to μ, i.e.,

$$H(\nu \mid \mu) = \int_0^1 \mathrm{d}\nu(x) \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)$$

if $\mathrm{d}\nu/\mathrm{d}\mu$ exists and ∞ otherwise. It is known that both $I(x)$ and $H(\nu \mid \mu)$ are convex, lower semicontinuous functions in the Euclidean and weak topologies, respectively (e.g., see Deuschel and Stroock, 1989). Further, it is well known that for any open (closed) set C in [0, 1],

$$\inf_{x\in C} I_\mu(x) = \inf_{\{\nu: \bar{\mu} \in C\}} H(\nu \mid \mu). \tag{2.1}$$

Next, let $\mu_n \triangleq n^{-1}\Sigma_{i=1}^n \delta_{x_i}$ denote the empirical measure of the sequence $x_1, x_2, \ldots, x_n$, and let the empirical mean $n^{-1}\Sigma_{i=1}^n x_i$ be denoted $\bar{\mu}_n$. By the classical Cramer theorem, one has, for any closed set C, and any probability measure μ with support in [0, 1] (cf. Deuschel and Stroock, 1989, Proof of Lemma 1.2.5),

$$P_\mu(\bar{\mu}_n \in C) \leqslant 2 \exp\Big(-n \inf_{x \in C} I_\mu(x)\Big). \tag{2.2}$$

We next define the sequence $\beta(m)$: for any m, let B_m be the open cover of the set A_m described above. For any m, compute

$$I_m \triangleq \inf_{\{\mu: \bar{\mu} \in A_m\}} \inf_{x \in B_m^c} I_\mu(x). \tag{2.3}$$

Note that by (2.1), one also has that

$$I_m = \inf_{\{\mu: \bar{\mu} \in A_m\}} \inf_{\{\nu: \bar{\nu} \in B_m^c\}} H(\nu \mid \mu). \tag{2.4}$$

Since $d(A_m, B_m^c)^2 \geqslant 2\varepsilon(m)$, one has that $I_m \geqslant \varepsilon(m)$. Indeed, by Deuschel and Stroock (1989, Exercise 3.2.24),

$$2H(\nu \mid \mu) \geqslant \| \nu - \mu \|_{\text{var}}^2 \geqslant \big(d(A_m, B_m^c)\big)^2,$$

where the last inequality holds for $\{\nu: \bar{\nu} \in B_m^c\}$ and $\{\mu: \bar{\mu} \in A_m\}$. Next, let

$$\alpha(m) \triangleq \frac{\log 2 + 2 \log m}{I_m}. \tag{2.5}$$

Note that, by (2.2), for any μ such that $\bar{\mu} \in A_m$,

$$P_\mu\big(\bar{\mu}_{\alpha(m)} \in B_m^c\big) \leqslant \frac{1}{m^2}. \tag{2.6}$$

Finally, let

$$\beta(m) = \sum_{i=1}^{m} \alpha(i), \qquad \beta(0) = 0. \tag{2.7}$$

Proof of Theorem 1. (a) Assume $\bar{\mu}_\text{T} \in A$. Then there exists an m such that $\bar{\mu}_\text{T} \in A_m$. Note that the event of making an error infinitely often is equivalent to the event of making an error at the parsing intervals infinitely often. However,

$$\sum_{m=1}^{\infty} \text{Prob}\{\text{error in } m\text{th parsing}\} \leqslant \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty$$

where we have used (2.6) above. Therefore, part (a) of the theorem follows by the Borel–Cantelli lemma.

(b) Let

$$N = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} C_m \backslash A.$$

Note that

$$|N| \leqslant \lim_{n \to \infty} \sum_{m=n}^{\infty} |C_m \backslash A_m| \leqslant \lim_{n \to \infty} \sum_{m=n}^{\infty} \frac{1}{m^2} = 0$$

where $|\cdot|$ denotes the Lebesgue measure. Therefore, the Lebesgue measure of N is zero. Now we may repeat the arguments of part (a) in the following way: let $\bar{\mu}_\text{T} \in A^c \backslash N$. For an m_0 large enough, $\bar{\mu}_\text{T} \in C_m^c$ for all $m > m_0$. On the other hand, $d(\bar{\mu}_\text{T}, B_m)^2 \geqslant 2\varepsilon(m)$ by our construction. Noting that the rate function $\inf_{x \in B_m} I_{\mu_\text{T}}(x) > \varepsilon(m)$, the proof follows identically as in part (a). □

Remarks. (1) The theorem could have been proved by obtaining (2.6) using more traditional bounds but with a slower decision procedure (i.e., larger $\alpha(m)$).

(2) It is interesting to note that the Cantor set satisfies the assumption. Indeed, the covering sets B_m are just the intervals associated with the Cantor partition.

(3) By modifying the structure of the decision rule, one may also make a hypothesis test inside A. This is pursued in Section 3.

We conclude this section by a (partial) characterization of the sets A of measure 0 which satisfy the assumption:

Lemma 1. *A set A which is of measure* 0 *and which satisfies the assumption is of the first category* (*i.e., A is a countable union of nowhere dense sets*). *In the other direction, a* closed *set A of Lebesgue measure zero satisfies the assumption.* (*Note that such a set is automatically nowhere dense and therefore of the first category.*)

Proof. ($\Rightarrow$) From the assumption, $A = \bigcup_m A_m$. We need only show that each A_m is nowhere dense. But this follows immediately from the existence of a sequence of open blowups of A_m with arbitrarily small Lebesgue measure (namely, C_k for $k \geqslant m$). Indeed, if that were not the case, any blowup of A_m would have included a common nonempty open set, and therefore its Lebesgue measure couldn't possibly converge to zero with the blowup size.

($\Leftarrow$) If A is closed and of measure zero, take $A_m = A$. One can cover A_m by open balls with with total measure smaller than $1/2m^2$. Since A is closed, it is compact and therefore one may extract a finite cover, with the smallest ball of radius $r_m > 0$. Choosing now $\varepsilon(m) < \frac{1}{8} r_m^2$, it follows that C_m is covered by the r_m-blowup of those balls, whose total measure is smaller than $1/m^2$. □

We note that an example in Halmos (1974, Exercise 4, p. 66) suggests that it is unlikely that one could in general dispense with the requirement that A be closed in the converse of Lemma 1. Indeed, in Halmos (1974) a set F of nonzero Lebesgue measure is constructed which is nowhere dense. Naively, by taking A_m to be a dense subset of F (maybe uncountable), and $A = \bigcup_{m=1}^{\infty} A_m$, the blowup of A_m doesn't satisfy the assumption (for it is of measure larger than $|F|$). However, there may exist another choice of sequence A'_m which does satisfy the assumption.

3. Countable hypothesis testing

In this section, we refine the decision rule to allow for deciding among a countable set of hypotheses. In addition to deciding whether or not $\bar{\mu}_T \in A$, we also make a hypothesis test inside A. Suppose that A is written as $A = \bigcup_{i=1}^{\infty} S_i$ where the S_i are disjoint. We are interested not only in whether $\bar{\mu}_T \in A$, but if so, to which of the S_i does $\bar{\mu}_T$ belong. Specifically, we wish to decide among the following countable set of hypotheses:

$$\mathrm{H}_i: \quad \bar{\mu}_T \in S_i, \quad i = 1, 2, \ldots,$$

$$\mathrm{H}_0: \quad \bar{\mu}_T \notin A.$$

For the theorem below, restrictions must be placed on the decomposition of A. Namely, we assume that the S_i are pairwise positively separated meaning that $d(S_i, S_j) > 0$ for every $i \neq j$. (Note that, as before, A is required to satisfy the structural assumption of the introduction.)

We modify our previous decision rule as follows. At the end of each parsing (defined by the sequence $\beta(m)$), find the least index k (if one exists) such that $\bar{\mu}_{X^m}$ is contained in the $\sqrt{2\varepsilon(m)}$ open blowup of $S_k \cap A_m$. If such a k exists, then decide that $\bar{\mu}_T \in S_k$. Otherwise (if $\bar{\mu}_{X^m} \notin (S_i \cap A_m)^{(\sqrt{2\varepsilon(m)})}$ for all i) decide that $\bar{\mu}_T \notin A$. Alternatively, we can think of this decision procedure as first deciding whether or not $\bar{\mu}_T \in A$ as before. Then, if the decision is that $\bar{\mu}_T \in A$, make a refinement by deciding that $\bar{\mu}_T \in S_k$ where k is the least index i such that $\bar{\mu}_{X^m} \in (S_i \cap A_m)^{(\sqrt{2\varepsilon(m)})}$.

Theorem 2. *If $A = \bigcup_{i=1}^{\infty} S_i$ satisfies the assumption and the S_i are pairwise positively separated then*: (a) *For any measure μ_T with $\bar{\mu}_T \in S_i$ for some i, the decision rule will make (a.s.) only a finite number of mistakes, i.e., for a.e. ω there exists an $n(\omega)$ such that the decision is S_i for all $n > n(\omega)$.*

(b) *For any measure μ_T with $\bar{\mu}_T \in A^c \backslash N$, where N is a set of Lebesgue measure 0, the decision rule will make (a.s.) only a finite number of mistakes, i.e. for a.e. ω there exists an $n(\omega)$ such that the decision is A^c for all $n > n(\omega)$.*

Proof. (a) Suppose that $\bar{\mu}_T \in S_i$. By the same considerations that led to (2.6), for any μ such that $\bar{\mu} \in S_i \cap A_m$ we have

$$P_\mu\left(\bar{\mu}_{X^m} \notin (S_i \cap A_m)^{(\sqrt{2\varepsilon(m)})}\right) \leqslant \frac{1}{m^2}. \tag{3.1}$$

Since $\bar{\mu}_T \in S_i \subseteq A$, for sufficiently large m, $\bar{\mu}_T \in A_m$. Also, since the S_j are pairwise positively separated and i is finite, for large enough m the sets $(S_j \cap A_m)^{(\sqrt{2\varepsilon(m)})}$ and $(S_i \cap A_m)^{(\sqrt{2\varepsilon(m)})}$ are disjoint for all $j < i$. That is, for sufficiently large m, denoted $m_0(i)$, as long as $\bar{\mu}_{X^m} \in (S_i \cap A_m)^{(\sqrt{2\varepsilon(m)})}$ we have $\bar{\mu}_{X^m} \notin (S_j \cap A_m)^{(\sqrt{2\varepsilon(m)})}$ for all $j < i$. Hence, for all $m > m_0(i)$, i is the least index satisfying the requirements of the decision procedure (so that a correct decision is made) iff $\bar{\mu}_{X^m} \in (S_i \cap A_m)^{(\sqrt{2\varepsilon(m)})}$. Therefore

$$\sum_{m=1}^{\infty} \operatorname{Prob}\{\text{error in } m\text{th parsing}\}$$

$$\leqslant m_0(i) + \sum_{m=m_0(i)+1}^{\infty} P\left(\bar{\mu}_{X^m} \notin (S_i \cap A_m)^{(\sqrt{2\varepsilon(m)})}\right)$$

$$\leqslant m_0(i) + \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty$$

so that part (a) follows by the Borel–Cantelli Lemma.

(b) This part is identical to part (b) of Theorem 1. □

Remarks. (1) Cover's result on countable hypothesis testing is a special case of this result since every countable set A clearly satisfies the assumption and can be written as the union of pairwise positively separated sets.

(2) If one is willing to allow the test to fail for some points in A, then the requirement that the S_i be pairwise positively separated can be dropped. The sets $N_1 \subset A^c$ and $N_2 \subset A$ on which the test fails in the general case can be characterized, and conditions on the S_i for which $N_1 \cup N_2$ is a null set can be obtained. These results can be obtained using the techniques presented here, or via a density argument using the Lebesgue density theorem (cf. Kulkarni, 1991).

Acknowledgments

We would like to thank J. Koplowitz for helpful discussions and for his interest in this work.

References

Cover, T.M. (1973), On determining the irrationality of the mean of a random variable, *Ann. Statist.* **1**, 862–871.

Deuschel, J.D. and D.W. Stroock (1989), *Large Deviations* (Academic Press, Boston, MA).

Halmos, P.R. (1974), *Measure Theory* (Springer, Berlin).

Koplowitz, J. (1977), Abstracts of papers, *Int. Symp. on Info. Theory, Cornell Univ., Ithaca, NY, October 10–14, 1977*, p. 64.

Koplowitz, J. (1990), Private communication.

Kulkarni, S.R. (1991), Problems of computational and information complexity in machine vision and learning, Ph.D. Thesis, Dept. of Electr. Eng. and Comput. Sci., M.I.T. (Cambridge, MA).