April 1990 LIDS-P-1965

CAN ONE DECIDE THE TYPE OF THE MEAN FROM THE EMPIRICAL MEASURE? [1]

Sanjeev R. Kulkarni

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

and

Ofer Zeitouni[2]

Department of Electrical Engineering
Technion, Haifa 32000, Israel

February 22, 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. Using large deviations techniques, an algorithm is proposed which is shown to lead to an a.s. correct decision for a class of A which are not necessarily countable. 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.

[1]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.

[2]This work was partially done while the author visited the Laboratory for Information and Decision Systems, MIT

1 Introduction

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

This problem was considered by Cover in [1], where he 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 $\overline{\mu}_T \in A$, will make (a.s.) only a finite number of mistakes whereas, for measures with $\overline{\mu}_T \in B \backslash N$, the test makes (a.s.) only a finite number of mistakes, where N is a set of Lebesgue measure 0. Various extensions of this result were considered by Koplowitz [3], who showed various properties of sets A which allow for such a decision and gave some characterizations of the set N.

In this note, we extend the result of [1] 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 positive sequence $\epsilon(m) \rightarrow_{m\rightarrow\infty} 0$ such that, for each m, the open blow up $B_m = A_m^{(\sqrt{2\epsilon(m)})} \hat{=} \{x : d(x, A_m) < \sqrt{2\epsilon(m)}\}$ is such that the Lebesgue measure of $B_m \backslash A_m$ is smaller than $1/m^2$. (We will use the fact that the open blow ups B_m satisfy $(d(A_m, B_m^c))^2 \geq 2\epsilon(m) > 0$.)

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, c.f. Lemma 2 and the remarks which follow Theorem 1.

In Section 2, we describe a decision algorithm which changes its decisions 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 $A^c \backslash N$ 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 [3]). In Section 3, the results are extended to allow a (countable) sub-decision inside the set A.

2 The decision rule and proof of the main theorem

We begin by first describing the proposed decision rule. Let $\beta(m)$ be a given sequence, to be defined below. For any input sequence $x_1, \cdots, x_n$, form the subsequences

$$X^m \hat{=} (x_{\beta(m-1)}, \cdots, x_{\beta(m)-1}).$$

Let $\overline{\mu}_X^m$ denote the empirical mean of the sequence X^m. At the end of each parsing, make a decision whether $\overline{\mu}_T \in A$ according to whether $\overline{\mu}_X^m \in B_m$ 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 P_T with $\overline{\mu}_T \in A$, 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 P_T with $\overline{\mu}_T \in A^c \backslash N$, 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 $n > n(\omega)$.

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 $\overline{E}$ denotes the closure of E, whereas E^o denotes the interior of E. Let μ be a probability measure with support in $[0,1]$. The mean of μ is denoted $\overline{\mu}$. Let $M_\mu(\lambda) \hat{=} E_\mu(\exp(\lambda x))$ denote the moment generating function of μ and let $\Lambda(\lambda) \hat{=} \log(M(\lambda))$. Let $I_\mu(x) = \sup_\lambda(\lambda x - \Lambda(\lambda))$ be the Legendre transform of $\Lambda(\lambda)$, and let $H(\nu|\mu)$ denote the relative entropy of ν with respect to μ, i.e. $H(\nu|\mu) = \int_0^1 d\nu(x) \log(\frac{d\nu}{d\mu})$ if $\frac{d\nu}{d\mu}$ exists and ∞ otherwise. It is known that both $I(x)$ and $H(\nu|\mu)$ are convex, lower semicontinuous functions (e.g, see [2]). Further, it is well known that for any open (closed) set C in $[0,1]$,

$$\inf_{x \in C} I_\mu(x) = \inf_{\{\nu : \int_0^1 x d\nu(xn) \in C\}} H(\nu|\mu). \tag{2.1}$$

Next, let $\overline{\mu}_n \hat{=} \frac{1}{n} \sum_{i=1}^n x_i$ denote the empirical mean of the sequence $x_1, x_2, \cdots, x_n$, and let $L_n \hat{=} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ denote the empirical measure of the sequence $x_1, x_2, \cdots, x_n$. By the classical Cramer theorem, one has that, for any closed set C, and any probability measure μ with support in $[0,1]$, (c.f. [2, proof of Lemma 1.2.5]),

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

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

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

Note that by (2.1), one also has that

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

Since $d(A_m, B_m^c)^2 \geq 2\epsilon(m)$, one has that $I_m \geq \epsilon(m)$. Indeed, by [2, Exercise 3.2.24], $2H(\nu|\mu) \geq ||\nu - \mu||_{var}^2 \geq (d(A_m, B_m^c))^2$, where the last inequality holds for $\{\nu : m_\nu \in B_m^c\}$ and $\{\mu : \overline{\mu} \in A_m\}$. Next, let

$$\alpha(m) \hat{=} \frac{\log 2 + 2 \log m}{I_m} \tag{2.5}$$

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

$$P_\mu(\overline{\mu}_{\alpha(m)} \in B_m^c) \leq \frac{1}{m^2} \tag{2.6}$$

Finally, let

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

Proof of Theorem 1

a) Assume $\overline{\mu}_T \in A$. Then there exists an m such that $\overline{\mu}_T \in A_m$. Note however 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,

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

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

b) Let C_m denote the $2\sqrt{2\epsilon(m)}$ blow up of B_m. Let

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

Clearly, the Lebesgue measure of N is zero. Now we may repeat the arguments of part a) in the following way: let $\overline{\mu}_T \in B \backslash N$. For an m_0 large enough, $\overline{\mu}_T \in C_m^c$ for all $m > m_0$. On the other hand, $d(\overline{\mu}_T, B_m)^2 \geq 2\epsilon(m)$ by our construction. Noting that the rate function $\inf_{x \in B_m} I_{P_T}(x) > \epsilon(m)$, the proof follows identically as in part a).

$\square$

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

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). Conversely, a **closed** set A of Lebesgue measure zero satisfies the Assumption if A is of the first category.

Proof

($\Longrightarrow$) From the Assumption, $A = \cup_m A_m$. We need only show that each A_m is nowhere dense. But this follows immediately from the existence of a sequence of open blow ups of A_m with arbitrarily small Lebesgue measure (namely, B_k for $k \geq m$).

($\Longleftarrow$) If A is of the first category then $A = \cup_i S_i$ where each S_i is nowhere dense. Let $A_m = \cup_{i=1}^m S_i$. Clearly, the A_m monotonically increase to A. Also, since A_m is nowhere dense, and A is closed, $|A_m^{(\delta)}| \to 0$ as $\delta \to 0$ where $|\cdot|$ denotes Lebesgue measure and $A_m^{(\delta)} = \{x : d(x, A_m) < \delta\}$ is the (open) δ-neighborhood of A_m. For each m, choose any $\delta_m > 0$ such that $|A_m^{(\delta_m)}| < 1/m^2$. Then the Assumption is satisfied with $B_m = A_m^{(\delta_m)}$ and $\epsilon(m) = \delta_m^2/2$.

□

We note that, by a counter example based on [4, Exercise 4, pg. 66], one cannot in general dispense of the requirement that A be closed in the converse direction of Lemma 2. Indeed, in [4] a set F of nonzero Lebesgue measure is constructed which is nowhere dense. To get a contradiction, it now suffices to take a countable dense subset of this set F to be any of the sets A_m.

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 $\overline{\mu}_T \in A$, we also make a hypothesis test inside A. Suppose that A is written as $A = \cup_{i=1}^{\infty} S_i$ where the S_i are disjoint. We are interested not only in whether $\overline{\mu}_T \in A$, but if so to which of the S_i does $\overline{\mu}_T$ belong. Specifically, we wish to decide among the following countable set of hypotheses:

$$H_i : \overline{\mu}_T \in S_i, \quad i = 1, 2, \ldots$$

$$H_0 : \overline{\mu}_T \notin A$$

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 $\overline{\mu}_{\alpha(m)}$ is contained in the $\sqrt{2\epsilon(m)}$ open blow up of $S_k \cap A_m$. If such a k exists, then decide that $\overline{\mu}_T \in S_k$. Otherwise (if $m_{\alpha(m)} \notin (S_i \cap A_m)^{(\sqrt{2\epsilon(m)})}$ for all i) decide that $\overline{\mu}_T \notin A$. Alternatively, we can think of this decision procedure as first deciding whether or not $\overline{\mu}_T \in A$ as before. Then, if the decision is that $\overline{\mu}_T \in A$, make a refinement by deciding that $\overline{\mu}_T \in S_k$ where k is the least index such that $m_{\alpha(m)} \in (S_i \cap A_m)^{(\sqrt{2\epsilon(m)})}$.

Theorem 2 If $A = \cup_{i=1}^{\infty} S_i$ satisfies the Assumption and the S_i are pairwise positively separated then

a) For any measure P_T with $\overline{\mu}_T \in S_i$ 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 P_T with $\overline{\mu}_T \in A^c \backslash N$, 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 $\overline{\mu}_T \in S_i$. By the same considerations that led to (2.6), for any μ such that $\overline{\mu} \in S_i \cap A_m$ we have

$$P_\mu(\overline{\mu}_{\alpha(m)} \notin (S_i \cap A_m)^{(\sqrt{2\epsilon(m)})}) \leq \frac{1}{m^2} \tag{3.8}$$

Since $\overline{\mu}_T \in S_i \subseteq A$, for sufficiently large m, $\overline{\mu}_T \in A_m$. Also, since the S_j are pairwise positively separated and i is finite, for large enough m the sets $(S_j \cap A_m)^{(\sqrt{2\epsilon(m)})}$ and $(S_i \cap A_m)^{(\sqrt{2\epsilon(m)})}$ are disjoint for all $j < i$. That is, for sufficiently large m, denoted $m_0(i)$, as long as $\overline{\mu}_{\alpha(m)} \in (S_i \cap A_m)^{(\sqrt{2\epsilon(m)})}$ we have $\overline{\mu}_{\alpha(m)} \notin (S_j \cap A_m)^{(\sqrt{2\epsilon(m)})}$ 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 $\overline{\mu}_{\alpha(m)} \in (S_i \cap A_m)^{(\sqrt{2\epsilon(m)})}$. Therefore,

$$\sum_{m=1}^{\infty} \text{ Prob error in m-th parsing} \leq m_0(i) + \sum_{m=m_0(i)+1}^{\infty} P(\overline{\mu}_{\alpha(m)} \notin (S_i \cap A_m)^{(\sqrt{2\epsilon(m)})})$$

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

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 set $N_2 \subset A$ on which the test fails in the general case can be characterized, and presumably conditions on the S_i for which N_2 is a null set could be obtained.

Acknowledgments

We would like to thank J. Koplowitz for [3] and for his interest in this work.

References

[1] Cover, T.M., On determining the irrationality of the mean of a random variable, *The annals of Statistics*, Vol 1 (1973), pp. 862-871.

[2] Deuschel, J.D. and Stroock, D.W., *Large deviations*, Academic Press, Boston, 1989.

[3] Koplowitz, J., Abstracts of papers, *Int. Symp. on Info. Theory*, Cornell Univ., Ithica, NY, Oct. 10-14, 1977, p. 64, and Private Communication.

[4] Halmos, P.R., *Measure Theory*, Springer, 1974.