JOURNAL OF ALGORITHMS 2, 164–177 (1981)

A Space-Efficient On-Line Method of Computing Quantile Estimates*

JUDEA PEARL

School of Engineering and Applied Science, University of California, Los Angeles 90024

Received June 29, 1980

The task of computing an estimate for the quantile (ζ_q) for an unknown distribution F (i.e., $F(\zeta_q) = q$) is usually performed by the "sample quantile" method, which computes the $\lfloor Nq \rfloor + 1$ smallest element from the set of N observations, and thus requires that all N samples be retained in memory. This paper introduces a recursive method of estimating ζ_q based on the fact that if the terminal nodes of a uniform d-ary tree are assigned random values, independently drawn from a distribution F, then the minimax alue of the root node converges to a specified quantile of F for very tall trees. The new estimate is shown to be almost as precise as that produced by the sample quantile method and, like it, is guaranteed to converge to ζ_q when the sample is large for any arbitrary distribution F. However, in contrast to the sample quantile computation the proposed method requires the retention in storage of at most $\log_2 N$ representative data points, where N is the number of samples observed in the past. Moreover, the estimate can be updated quickly using an average of 4, and a maximum of $2 \log_2 N$, comparisons with each new observation.

1. INTRODUCTION

We address the problem of estimating the q-quantile of an unknown distribution $F_X(x)$ on the basis of N independent samples $X_1, X_2, \ldots, X_N$ drawn from this distribution. In other words, we seek an estimate of the quantity ζ_q that satisfies $F_X(\zeta_q) = q$, where $0 < q < 1$. We will assume that in the neighborhood of $x = \zeta$, the density $f(x) = F'_X(x)$ is continuous and has a continuous derivative $f'(x)$.

The most obvious way of estimating ζ_q is to calculate the quantile of the sample. If Nq is not an integer, and if we arrange the sample values in an

*This work was supported in part by National Science Foundation Grants MCS 78-07468 and MCS 78-18924.

0196-6774/81/020064-14$02.00/0

N is the number of samples observed in the past. Moreover, each new update would require no more than $2\log_2 N$ comparisons.

2. The Minimax Quantile Estimator

The method examined in this paper is based on the observation [3] that if the terminal nodes of a uniform d-ary tree are assigned random values, independently drawn form a distribution F, then the minimax value of the root node converges to a specified quantile of F for very tall trees.

Consider a uniform tree with branching factor d and height h in which the minimax value of each node is computed by the following procedure: the values of the terminal nodes are assigned externally, the value of each nonterminal node at any odd level of the tree (so-called MIN nodes) is given by the minimum value of its d successors, and the value of each nonterminal node at any even level of the tree (so-called MAX nodes) is given by the maximal value of its d successors. Let us examine the distribution of the minimax value of the root node when the terminal values, $V_0(S_1)$, $V_0(S_2),\ldots, V_0(S_{2^h})$, are random variables independently drawn from a common distribution function $F_{V_0}(v) = P\ (V_0 \leq v)$. Without loss of generality we assume that h is an even integer, i.e., $h = 2n$, and that the root is a MAX node.

Denoting the minimax value of the root node by $V_n(S)$, it is easy to show that its distribution function is governed by the recursive relation

$$F_{V_n}(v) = \left[1 - \left[1 - F_{V_{n-1}}(v)\right]^d\right]^d. \tag{1}$$

The function

$$g(x) = \left[1 - (1 - x)^d\right]^d \tag{2}$$

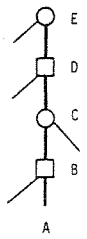
is monotonic increasing over $[0, 1]$ with an unstable fixed point at $x = 1 - \xi$, where ξ is the unique positive root of the equation $x^d + x - 1 = 0$. Consequently, $F_{V_n}(v)$ converges to the limits

$$\begin{aligned} \lim_{n\to\infty} F_{V_n}(v) &= 0, & F_{V_0}(v) &< 1 - \xi, \\ &= 1 - \xi, & F_{V_0}(v) &= 1 - \xi, \\ &= 1, & F_{V_0}(v) &> 1 - \xi. \end{aligned} \tag{3}$$

Equation (3) implies that when the tree is sufficiently tall, the density of $V_n(S)$, $F'_{V_n}(v)$, becomes highly concentrated around those values v^* satisfying

$$F_{V_0}(v^*) = 1 - \xi. \tag{4}$$

nodes until the root value is established. Clearly, there is no need to store the data in a tree structure; it is sufficient to maintain an updated record of the path connecting the root with the currently examined terminal node. For example, if A is the terminal node examined:



then all the terminal nodes to the left of A have been inspected and the minimax values of all subtrees emerging from the path E–D–C–B–A toward the left are known. These values are sufficient for incorporating the new datum A, and propagating its impact upward. Each node along the path should be assigned a counter and a storage register. When a new input arrives from below, the counter is incremented by one and the register changes to the MAX (or MIN) value of the new input and its previous content. When any of the counters exceeds d it is set to 1 and the content of the corresponding register is delivered upward as an input to the node above. Thus, when the data are generated sequentially, one need not keep in storage a complete record of past observations; a condensed summary of only h numbers (together with h counters) would suffice. Moreover, the number of comparisons performed with the arrival of each datum is at worst h, and on the average 2.

Note that when the data size N is not known in advance, the height h of the tree (or stack) could increase dynamically. The highest even-level node evaluated can, at any time, be taken as an estimate of the quantile ζ_q. Note also that in order to adjust the value of q, the desired quantile, one need only adjust d, which determines the conditions for resetting the various counters along the traversing path.

The method described above amounts to an exhaustive evaluation of all nodes in the tree. Although it consumes only minimal storage space, this method is quite wasteful in terms of the number of observations (d^h) required before the root node is fully evaluated. A more efficient method based on the alpha–beta (α–β) pruning technique [4] can cut this number substantially.

The basic idea behind the α–β pruning technique is that in normal circumstances one can ignore the majority of the sampled data without

The updating procedure employed by the α–β method is similar to the one described previously. Again, h nodes along a single path are sufficient to summarize the past N observations. However, before a counter is incremented it is now necessary to test whether the cutoff condition is met. If it is, the counter is reset to 1 and the register's content is sent upward. The shallow cutoff test requires only that the newly updated value of each register be compared to the value stored in the node above it, which only doubles the number of comparisons performed per sample. Deep cutoff tests can also be performed at the expense of a slightly more complex updating sequence or, alternatively, by using two registers at each node. However, the additional savings provided by deep cutoffs is only marginal [4].

In summary, both updating methods employ only $O(\log N)$ samples from the past N observations, plus an equal number of counters, and perform at most $O(\log N)$ comparisons per update. But the α–β method makes better use of the observations obtained. It can reach a given level of estimation precision (governed by the height h) by observing, on the average, only a fraction $N^{-3/10}$ of the samples necessary for the exhaustive minimax method.

In the next section we examine the precision of the minimax estimation method as a function of the height h.

3. Estimation Accuracy

Assume that a minimax evaluation procedure was set up to estimate the quantile $q = 1 - \xi$ and that a sufficient number of samples was examined to fully evaluate the minimax value at the height $h = 2n$. We analyze the accuracy of the estimation by assuming $d_1 = d_2 = d$, suggesting that similar results also hold for the case $d_1 \neq d_2$.

The accuracy of the minimax estimator is governed by the distribution function of V_n,

$$F_{V_n}(v) = \left\{1 - \left[1 - F_{V_{n-1}}(v)\right]^d\right\}^d = g\left[F_{n-1}(v)\right], \tag{5}$$

where

$$g(x) = \left[1 - (1 - x)^d\right]^d$$

and

$$F_{V_0}(v) = F_X(v).$$

Since $F_{V_n}(v)$ tends to a step function as $n \to \infty$ (Theorem 1), the variance of V_n will be determined by the slope of $F_{V_n}(v)$ in the neighborhood of $v = \zeta_q$.

(iii) In order to determine $\phi'(0)$ we use (6) and obtain

$$\phi_n'(y) = g'\big[\phi_{n-1}'(ay)\big]a\phi_{n-1}'(ay),$$

which at $y = 0$ becomes

$$\phi_n'(0) = ag'(\zeta)\phi_{n-1}'(0) = \phi_{n-1}'(0) = \phi_{n-2}'(0) = \cdots = \phi_0'(0) = f(\zeta).$$

Thus

$$\phi'(0) = f(\zeta). \tag{10}$$

(iv) The asymptotic behavior of $\phi(y)$ for large y can be found from the fact that $\phi(y)$, being a distribution function, obeys $\phi(y) \to_{y\to\infty} 1$. Therefore, $g[\phi(y)]$ can be approximated by the local behavior of $g(x)$ near $x = 1$; $g(x) \approx 1 - d(1-x)^d$. Using this expression in (7) yields

$$\phi(y) = 1 - d\big[1 - \phi(ay)\big]^d$$

or, $\psi(y) = 1 - \phi(y)$ is a solution of the functional equation

$$\psi(y) = d\big[\psi(az)\big]^d.$$

Fortunately this equation has a solution in close form:

$$\psi(y) = (d)^{-1/(d-1)}\exp\big(-y^{-\ln d/\ln a}\big).$$

Writing $a = \big[(1-q)/qd\big]^2$, we have $-\ln d/\ln a = 1/2(1-\gamma)$ and, taking $\gamma \approx 0.7$, we obtain

$$\phi(y) \underset{y\to\infty}{\approx} 1 - (d)^{-1/(d-1)}\exp\big(-y^{5/3}\big). \tag{11}$$

Thus, in contrast to the celebrated $\exp(-y^2/2)$ law of the normal distribution, the rate of convergence of $\phi(y)$ is somewhat slower. Yet it is sufficiently fast to render the variance fairly independently on the exact shape of the tails of $\phi(y)$.

(v) Using a similar argument, the asymptotic behavior of $\phi(y)$ for $y \to -\infty$ can be shown to obey

$$\phi(y) \underset{y\to-\infty}{\approx} (d)^{-d/(d-1)}\exp\big(-y^{5/3}\big). \tag{12}$$

(vi) *Numerical computation of* $\phi(y)$. Although (6) may be used for an iterative computation of $\phi(y)$, it is more convenient to compute y as a function of ϕ using the following transformation. From $\phi_n(y) =$

Numerical computations show that (14) provides an extremely close approximation. Thus, the variance of V_n can be adequately approximated by

$$\sigma_{V_n}^2 = \frac{a^{2n}}{2\pi}\frac{1}{f^2(\zeta)} = \frac{1}{2\pi}\left(\frac{1-q}{qd}\right)^{4n}\frac{1}{f^2(\zeta)}. \tag{15}$$

We summarize these results by stating:

THEOREM 2. *The limiting distribution of the variable* $y_n = (V_n - \zeta_q)/a^n$ *is given by* $\phi(y)$, *the solution to the functional equation* $\phi(y) = g\left[\phi(ay)\right]$, *where* $a = \left[(1-q)/qd\right]^2$ *and* q *is the positive solution of* $(1-q)^d - q = 0$. *The limiting variance of* y_n *is roughly given by* $\left[2\pi f^2(\zeta_q)\right]^{-1}$.

4. Comparisons between the Sample-Quantile and the Minimax Methods

The accuracies of the sample-quantile and the minimax methods are reflected in the variances of the variables z_q and V_n, respectively:

$$\sigma_{z_q}^2 = \frac{q(1-q)}{N_s f^2(\zeta_q)} \tag{16}$$

and

$$\sigma_{V_n}^2 = \frac{1}{2\pi}\frac{1}{f^2(\zeta_q)}\left(\frac{1-q}{qd}\right)^{2h}. \tag{17}$$

Note that in (17), d depends implicitly on q while the expected number of samples $\bar{N}_x$ depends on n and d. In order to compare the two estimation methods under equal conditions, let us assume that both are required to operate with the same accuracy, η, where

$$\eta = f^2(q)\sigma_{z_q}^2 = f^2(q)\sigma_{V_n}^2. \tag{18}$$

The choice of η dictates both the required number of samples N_s for the sample-quantile method and the height h for the minimax method,

$$N_s = \frac{q(1-q)}{\eta}, \tag{19}$$

$O(N^{-0.428})$ behavior established for the minimax method (see Eq. (21)). This selection method, with $d = 3$, is conceptually very appealing but, lacking an efficient pruning scheme, may require a larger sample for achieving a given precision standard. To obtain a standard deviation comparable to that of the minimax method, d must exceed 21, which introduces an additional complication: All 21 samples must be stored at each node before their median can be selected. In general, a total of $d \times \log_d N$ memory registers would be required by this method instead of the $10/7 \log_d N$ registers required by the minimax method.

In a more general "tree-selection" scheme each node may select one of its d successors in accordance with a fixed, but arbitrary, selection rule. The identity of the selected member may be an arbitrary function of the rank positions of all the d values (e.g., if $X_1 > X_2$ select $\max(X_1, X_2, X_3)$, else choose $\min(X_1, X_2, X_3)$). It can be shown that for all rank-dependent selection rules, the distribution of the selected member is a polynomial P in F_x satisfying $P(0) = 0$ and $P(1) = 1$. If P has only one interior unstable fixed point at $F_x = q$ (i.e., $0 < q < 1, P(q) = q, P'(q) > 1$), then the value selected by the root node would converge to the q-quantile of F_x. The task of estimating the q-quantile of F_x could then be performed by the given selection rule using $O(\log N)$ storage registers. The figure of merit of the selection rule (assuming no pruning) is given by the ratio $\alpha = [\log P'(q)]/\log d$ which governs the behavior of the estimation variance via $\sigma = O(N^{-\alpha})$.

If P has a stable interior fixed point (i.e., $P(q) = q$, $P'(q) \leq 1$) then the root's value would not converge to a unique quantity but rather would continue to fluctuate from level to level as a multivalued random variable. In such a case the computation task performed by the tree-schema can no longer be regarded as providing a consistent estimate of a computable property of F_x. In the third possibility, where the only fixed points are the boundaries $F_x = 0$ and $F_x = 1$, the root node would tend toward the lowest or highest support of F_x.

These considerations highlight the basic limitation of recursive "tree-selection" schema; *the only property of F_x computable by such schema is the quantile*. It would still be interesting to explore the computational capabilities of tree architectures where each node is permitted to perform a more general computational function involving both selection and arithmetic operations (such as $1/2\left[\min(x_1, x_2, x_3) + \max(x_1, x_2, x_3)\right]$.

ACKNOWLEDGMENTS

The author wishes to thank David Cantor, Thomas Ferguson, and D. Stott Parker for their stimulating discussions and valuable suggestions.