

AD-A071 885

NAVAL RESEARCH LAB WASHINGTON DC
ACCURACY OF AN INFORMATION-THEORETIC, LIGHT-LOAD APPROXIMATION --ETC(U)
JUL 79 J E SHORE

F/0 9/4

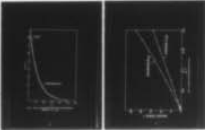
UNCLASSIFIED

NRL-4R-4037

NL

| OF |

AD
A071885



END
DATE
FILMED
8 -79
DDC

12
b.s.

NRL Memorandum Report 4037

**Accuracy of an Information-Theoretic, Light-Load
Approximation for the M/M/1 Busy Period
Probability Density**

JOHN E. SHORE

*Information Systems Staff
Communications Sciences Division*

LEVEL

AD A 071 885

July 11, 1979

DDC
RECEIVED
JUL 27 1979
REGISTRY
C

DDC FILE COPY



NAVAL RESEARCH LABORATORY
Washington, D.C.

Approved for public release; distribution unlimited.

79 07 27 006

14 NRL-MR-4037

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Memorandum Report 4037	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) 6 ACCURACY OF AN INFORMATION-THEORETIC, LIGHT-LOAD APPROXIMATION FOR THE M/M/1 BUSY PERIOD PROBABILITY DENSITY.	5. TYPE OF REPORT & PERIOD COVERED Interim report of a continuing NRL problem.	
	6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s) 10 John E. Shore	8. CONTRACT OR GRANT NUMBER(s)	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375	12 14p.	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NRL Problem B02-35 61153N, RR014-09-41
11. CONTROLLING OFFICE NAME AND ADDRESS 11 11 Ju / 79	12. REPORT DATE July 11, 1979	
14. MONITORING AGENCY NAME & ADDRESS (If different from Controlling Office) 16 RR01409 17 RR0140941	13. NUMBER OF PAGES 13	
	15. SECURITY CLASS. (of this report) UNCLASSIFIED	
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE		
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) M/M/1 Busy Period Entropy Maximization Information Theory Queuing Theory		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An exponential approximation for the M/M/1 busy period probability density is derived by entropy maximization and studied numerically. The approximation is accurate for light loads: For load factors less than about .1, the approximation is accurate to within 10% in the range where the cumulative distribution function is as large as .95.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251 950

JOB

CONTENTS

I. INTRODUCTION AND SUMMARY 1

II. BACKGROUND AND DERIVATION 2

A. The Maximum Entropy Principle and the Minimum Cross-entropy Principle 2

B. Justifying the Principles as General Methods of Inference 3

C. Application to Busy Period Approximations 5

III. ACCURACY OF THE M/M/1 APPROXIMATION 6

IV. DISCUSSION 8

REFERENCES 9

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	<input type="checkbox"/>
By _____	
Distribution/ _____	
Availability Codes	
Dist	Avail and/or special
A	

ACCURACY OF AN INFORMATION-THEORETIC, LIGHT-LOAD
APPROXIMATION FOR THE M/M/1 BUSY PERIOD
PROBABILITY DENSITY

I. INTRODUCTION AND SUMMARY

In M/M/1 queuing systems, customers arrive with independent, exponentially distributed interarrival times at an average rate λ from an infinite customer pool; they wait in an infinite capacity queue; they are served independently by a single server with exponentially distributed service times at an average rate μ ; and they return to the customer pool. If the system is empty and a customer arrives at time t_1 , and if t_2 is the next time at which the system is empty, then the period between t_1 and t_2 is called a busy period. The probability density function for the M/M/1 busy period is known exactly, namely [1, p. 215]

$$q_e(t) = \frac{1}{t\sqrt{\lambda/\mu}} e^{-(\lambda+\mu)t} I_1(2t\sqrt{\lambda\mu}) \quad (1)$$

where I_1 is the modified Bessel function of the first kind (order one).

This paper concerns a new approximation to (1),

$$q_a(t) = (\mu - \lambda)\exp[-(\mu - \lambda)t] , \quad (2)$$

which is quite accurate for light load conditions. Specifically, the average absolute percentage error of (2) satisfies

$$A_{.95} \approx 73\rho , \quad (3)$$

for $0 < \rho < .2$, where $\rho = \lambda/\mu$ and the average is computed over the range of t in which the cumulative probability distribution of $q_e(t)$ goes from zero to .95. Hence, the approximation (1) is accurate to within about 10% for $\rho \leq .1$.

The new approximation is of interest, not only because it is considerably simpler than (1), but also because it was discovered by information-theoretic

Note: Manuscript submitted June 1, 1979.

methods [2] without using the known solution (1). Indeed, I have not yet found a way to derive (2) from (1).

II. BACKGROUND AND DERIVATION

A. The Maximum Entropy Principle and the Minimum Cross-entropy Principle

Suppose you know that a system has a set of possible states x_i with unknown probabilities $q^\dagger(x_i)$, and you then learn constraints on the distribution q^\dagger : either values of certain expectations $\sum_i q^\dagger(x_i) f_k(x_i)$ or bounds on these values. Suppose you need to choose a distribution q that is in some sense the best estimate of q^\dagger given what you know. Usually, there remains an infinite set of distributions that are not ruled out by the constraints. Which one should you choose?

The principle of maximum entropy states that, of all the distributions q that satisfy the constraints, you should choose the one with the largest entropy - $\sum_i q(x_i) \log(q(x_i))$. Entropy maximization was first proposed as a general inference procedure by Jaynes [3]. Since then, it has been applied successfully in a remarkable variety of fields, including traffic networks [4], and queuing theory [2],[5]. For a lengthy list of applications and references, see [6].

The principle of minimum cross-entropy is a generalization that applies in cases when a prior distribution p that estimates q^\dagger is known in addition to the constraints. The principle states that, of the distributions q that satisfy the constraints, you should choose the one with the least cross-entropy $\sum_i q(x_i) \log(q(x_i)/p(x_i))$. Unlike entropy maximization, cross-entropy minimization generalizes correctly for continuous probability

densities. One then minimizes the functional $\int dx q(x) \log(q(x)/p(x))$. The name cross-entropy is due to Good [7]. Other names include expected weight of evidence [8, p. 72], directed divergence [9, p. 7], and discrimination information [9, p. 37]. The principle of minimum cross-entropy was first proposed by Kullback [9, p. 37]. Like entropy maximization, cross-entropy minimization has been applied in many fields (see [6]). When the prior density is uniform, the principle of minimum cross-entropy reduces to the principle of maximum entropy. In this case, one selects the posterior by maximizing the posterior entropy

$$\int dx q(x) \log(q(x)/p(x)) \quad , \quad (4)$$

subject to the constraints provided by the known expected values.

B. Justifying the Principles as General Methods of Inference

Until recently, entropy maximization was justified best on the basis of entropy's unique properties as an uncertainty measure. That entropy has such properties is undisputed: one can prove, up to a constant factor, that entropy is the only function satisfying axioms that are accepted as requirements for an uncertainty measure [10]. Intuitively, the maximum entropy principle follows quite naturally from such axiomatic characterizations. For example, Jaynes states that the maximum entropy distribution "is uniquely determined as the one which is maximally noncommittal with regard to missing information" [3, p. 623], and that it "agrees with what is known, but expresses 'maximum uncertainty' with respect to all other matters, and thus leaves a maximum possible freedom for our final decisions to be influenced by the subsequent sample data" [11, p. 231].

Similar justifications can be advanced for cross-entropy minimization. Like entropy, cross-entropy has various properties that are desirable for an information measure [12],[13], and one can argue [14] that cross-entropy measures the amount of information necessary to change a prior p into a posterior q . The principle of minimum cross-entropy then follows intuitively much like entropy maximization.

To some, entropy's unique properties make it obvious that entropy maximization is the correct way to account for constraint information. To others, such an informal and intuitive justification yields plausibility but not proof --- why maximize entropy; why not some other function? As a result, entropy maximization has remained controversial despite its success.

Recently, R. Johnson and I have obtained a new, formal justification for entropy maximization using a different approach [6]. This approach is based on the observation that previous justifications are weak, not only because they rely on informal, intuitive arguments, but also because they are indirect --- they are based on a formal description of what is required of an information measure rather than on a formal description of what is required of a method for taking new information into account.

Our approach in [6] was to formalize the requirements of inductive inference directly in terms of four consistency axioms that make no reference to information measures or properties of information measures. The four axioms are based on a single fundamental principle: If a problem can be solved in more than one way --- for example, in different coordinate systems --- the results should be consistent. We were then able to prove that the principle of maximum entropy is correct in the following sense: Given information in the form of constraints on expected values, there is only one distribution

satisfying these constraints that can be chosen in a manner that satisfies the axioms; this unique distribution can be obtained by maximizing entropy. This result for entropy maximization was obtained both directly and as a special case (uniform priors) of an analogous, more general result for the principle of minimum cross-entropy.

C. Application to Busy Period Approximations

The approximation (2) is actually a general result for M/G/1 systems (general service time distributions rather than just exponential) that happens to be accurate in the M/M/1 case. Let $s(t)$ be an arbitrary service time probability density with moments s_m . Then it is well known that the m th moment b_m of the exact busy period probability density $q_e(t)$ can be expressed exactly in terms of the first m moments $\{s_1, \dots, s_m\}$ of $s(t)$, for example

$$b_1 = \frac{s_1}{1 - \lambda s_1}, \quad (5)$$

$$b_2 = \frac{s_2}{(1 - \lambda s_1)^3}, \quad (6)$$

etc. [1, pp. 214-215]. The moments b_m provide information about the busy period probability density in a form suitable for applying the principle of minimum cross-entropy. For example, using (5) and assuming a uniform prior density (reasonable provided one believes that the maximum busy period is finite), one chooses an estimate $q_a(t)$ of $q_e(t)$ by maximizing the entropy (4) subject to a constrained first moment b_1 . This is a well-known problem, with the solution [15] $q_a(t) = (1/b_1)\exp(-t/b_1)$, or

$$q_a(t) = (s_1^{-1} - \lambda)\exp[-(s_1^{-1} - \lambda)t] \quad (7)$$

In the M/M/1 case, (7) becomes (2) since $s_1 = 1/\mu$. More information can be used to choose better estimates --- for example, one can compute the first two moments of q_e from s_1 and s_2 using (5)-(6) and obtain a two-moment estimate of q_e by maximizing entropy subject to the constrained moments b_1 and b_2 [2].

How accurate are such "information-theoretic approximations"? About all that can be said in general is that the approximations are the least-biased choices given the information available. To use the language of statistics [7,9], the approximations are the hypotheses that are best supported by the information available. Of course, more can be said in specific cases, e.g. M/M/1, when q_e itself is known. In the next section, I compare the exact M/M/1 density (1) with the one-moment approximation (2).

III. ACCURACY OF THE M/M/1 APPROXIMATION

The exact density (1) and the approximation (2) are plotted in Fig. 1 for the case $\lambda = 1$, $\mu = 10$. Qualitatively, the two results appear to be close --- indeed, this plot stimulated the conjecture that (2) might be a good light load approximation in general [16]. Furthermore, the conjecture was supported by the following argument, which is due to A. E. Ephremides [17]: Equation (2) is identical to the exact M/M/1 residence time probability density [1, p. 202]. Since most busy periods will consist of single customer residences under light load conditions, it makes sense that the busy period should tend to (2).

In order to evaluate the conjecture quantitatively, I chose two figures of merit, both based on the absolute percentage error

$$P(t) = 100 \left| q_e(t) - q_a(t) \right| / q_e(t) \quad (8)$$

Let T_c be the point at which the cumulative distribution of q_e reaches the value c , i.e.,

$$\int_0^{T_c} dt q_e(t) = c \quad (9)$$

Then, let M_c and A_c be the maximum and average percentage error in the region $0 \leq t \leq T_c$, i.e.,

$$M_c = \text{MAX}[P(t)], \quad t \in (0, T_c) \quad (10)$$

$$A_c = \frac{1}{T_c} \int_0^{T_c} dt P(t) \quad (11)$$

Now, although neither q_e nor q_a can be expressed as functions of only t and $\rho = \lambda/\mu$, it turns out that both M_c and A_c depend only on ρ . To see this, note that both q_e and q_a satisfy the scaling equation $q(\lambda, \mu, t) = f^{-1} q(\lambda f, \mu f, t/f)$, where f is an arbitrary scalar factor. It follows from (8) and (9) that P and T_c satisfy

$$P(\lambda, \mu, t) = P(\lambda f, \mu f, t/f) \quad (12)$$

and

$$T_c(\lambda, \mu) = f T_c(\lambda f, \mu f) \quad (13)$$

By combining (12)-(13) with the definitions (10)-(11), it is easy to see that $M_c(\lambda, \mu) = M_c(\lambda f, \mu f)$ and $A_c(\lambda, \mu) = A_c(\lambda f, \mu f)$ both hold, which shows that M_c and A_c both depend only on the ratio $\rho = \lambda/\mu$.

Based on data computed for twelve values of ρ , Fig. 2 shows M_c and A_c as functions of ρ for $c = .95$. That is, Fig. 2 shows the maximum and average percentage error of $q_a(t)$ in the range where the cumulative distribution of $q_e(t)$ is less than .95. The approximation (2) is accurate to within 10% for

$\rho \leq .1$. In general, the average percentage error A_c is about two thirds of the maximum. Since the data for A_c was surprisingly linear, it seemed useful to compute the best (least mean-squares) linear fit that was constrained to pass through the origin. The result is (3). In fact [18], one can show directly from (1) and (2) that

$$\lim_{\rho \rightarrow 0} \frac{d}{d\rho} A_{.95}(\rho) = 68.4 \quad .$$

IV. DISCUSSION

The derivation of (2) is noteworthy because of the information-theoretic method used --- (2) was generated as a hypothesis by cross-entropy minimization. The value of (2) as an approximation to (1) was subsequently supported both quantitatively and qualitatively (the Ephremides explanation). The results illustrate how information-theoretic techniques can be used in system modeling. In general, one models real systems by abstraction, representing the real system by some but not all information about the real system. If the restricted (abstracted) information is in the form of expected values, then cross-entropy minimization is the only self-consistent method of choosing a probability density to model the real system. That the method can be useful is illustrated by the present results.

Since (2) is much easier to compute than (1), the new approximation may be useful in situations where quantitative results are needed. Accuracy within 10% is often sufficient since queuing models themselves are often only approximate abstractions of real systems. Eq. (2) should also be useful in analytical work, given the celebrated convenience of the exponential form.

References

1. L. Kleinrock, Queueing Systems, Vol I: Theory, New York, John Wiley, 1975.
2. J. E. Shore, "Information Theoretic Approximations for M/G/1 and G/G/1 Queueing Systems," NRL Memorandum Report in publication, Naval Research Laboratory, Washington, D.C. 20375
3. E. T. Jaynes, "Information Theory and Statistical Mechanics I," Phys. Rev. 106, 1957, pp. 620-630.
4. V. E. Benes, Mathematical Theory of Connecting Networks and Telephone Traffic, Academic Press, New York, 1965.
5. J. E. Shore, "Derivation of Equilibrium and Time-Dependent Solutions to M/M/ //N and M/M/ Queueing Systems Using Entropy Maximization, Proceedings 1978 National Computer Conference, AFIPS, 1978, pp. 483-487.
6. J. E. Shore and R. W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Trans. Inf. Theory, to be published.
7. I. J. Good, "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables," Annals Math. Stat. 34, 1963, pp. 911-934.
8. I. J. Good, Probability and the Weighing of Evidence, Charles Griffen, London, 1950.
9. S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.
10. C. E. Shannon, "A Mathematical Theory of Communication," Bell System Tech. Jour., 27, 1948, pp. 379-423.
11. E. T. Jaynes, "Prior Probabilities," IEEE Trans. on Systems Science and Cybernetics SSC-4, 1968, pp. 227-241.
12. A. Hobson and B. Cheng, "A Comparison of the Shannon and Kullback Information Measures," J. Stat. Phys. 7, No. 4, 1973, pp. 301-310.
13. R. W. Johnson, "Axiomatic Characterization of the Directed Divergences and Their Linear Combinations," accepted by IEEE Trans. Inf. Theory.
14. A. Hobson, "A New Theorem of Information Theory," J. Stat. Phys. 1, No. 3, 1969, pp. 383-391.
15. J. P. Noonan, et. al., "On the Inverse Problem of Entropy Maximizations", IEEE Trans. Inf. Theory, Jan. 1976, pp. 120-123.
16. H. Vantilborgh, private communication.
17. A. E. Ephremides, private communication.
18. R. W. Johnson, private communication.

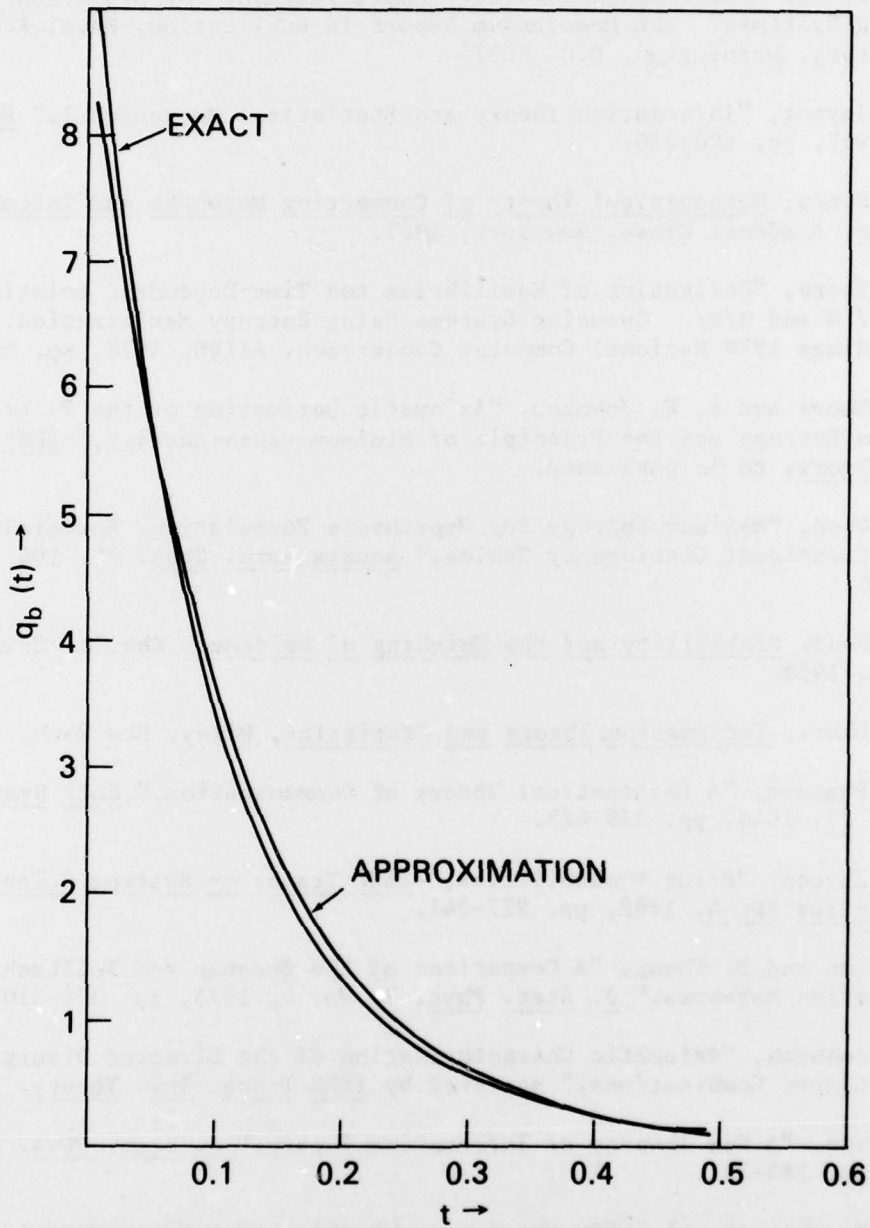


Fig. 1 — Exact and approximate M/M/1 busy period probability densities ($\gamma = 1, \mu = 10$)

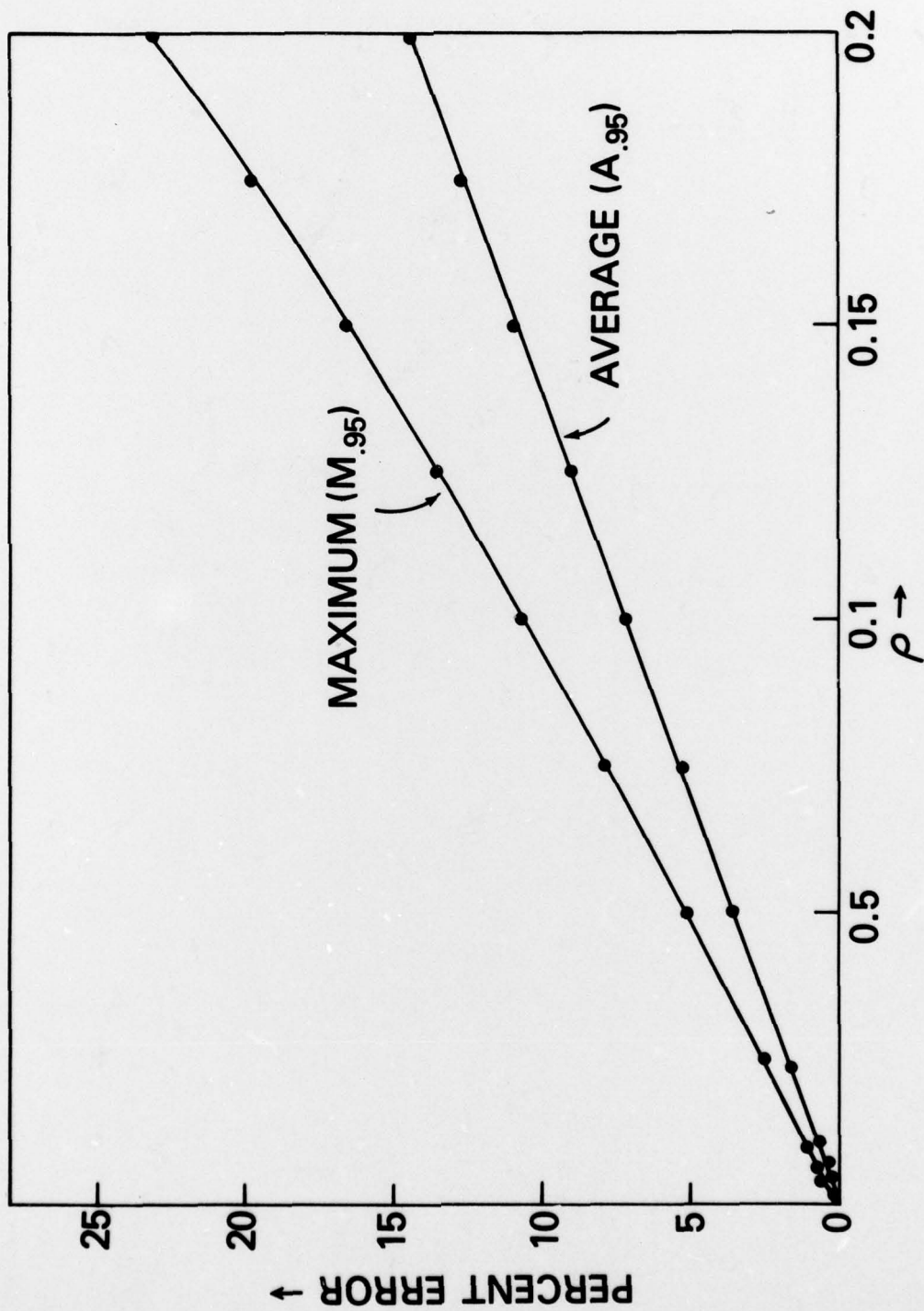


Fig. 2 — Maximum and average percent error vs. ρ