Maximum Entropy and the Method of Moments in Performance Evaluation of Digital Communications Systems

MOHSEN KAVEHRAD, SENIOR MEMBER, IEEE, AND MYRLENE JOSEPH

Abstract—The maximum entropy criterion for estimating an unknown probability density function from its moments is applied to the evaluation of the average error probability in digital communications. Accurate averages are obtained, even when a few moments are available. The method is stable and results compare well with those from the powerful and widely used Gauss quadrature rules (GQR) method. For test cases presented in this work, the maximum entropy method achieved results with typically a few moments, while the GQR method required many more moments to obtain the same, as accurately. The method requires about the same number of moments as techniques based on orthogonal expansions. In addition, it provides an estimate of the probability density function of the target variable in a digital communication application.

I. INTRODUCTION

ENTROPY maximization historically has roots in the work of C. E. Shannon [1]. For discrete systems, maximizing the source entropy results in the best source encoding, in the sense of enabling the highest information rate over a fixed capacity channel. Later E. T. Jaynes [2] applied the entropy maximization to some problems in statistical mechanics. Since then, there have been many conceptual ramifications of the rationale of maximum entropy [3]–[6]. For example, it is noted in [5] that most of the well-known distributions in statistics are maximum entropy distributions, given appropriate simple moment constraints. For example, the maximum entropy density under the constraints of zero mean and second moment being σ^2 is a normal density with zero mean and σ^2 variance. In other words, among many zero-mean densities with a second moment equal to σ^2, normal density maximizes the entropy function and is the most unbiased one.

In this work we shall extend the application of maximum entropy to evaluation of average error probability in digital communication. In this application the sampled received signal has to be detected in the presence of Gaussian noise and interference. In general, one can evaluate a conditional error probability that is conditioned on the interference sample. The interference, however, may have an unknown probability density $f(x)$ with a finite number of known moments. Examples found in practice are intersymbol interference (ISI), cross-coupling interference (CCI) or cross-polarization interference (CPI), multiple access interference, etc. Although the interference is a function of random variables with known probability densities, the evaluation of its probability density is not always practical. For example, linear ISI comprises a sum

of weighted random variables with known densities and the weights are samples of a channel impulse response. However, the density of the interference itself is most often unknown, theoretically. Therefore, subject to the available moments, one can estimate the unknown density $f(x)$ by maximizing the entropy function with respect to $f(x)$. We accept the approximate estimated density heuristically, since we have not proved it to be the actual one. In applications described, once the probability density of the interference becomes available, one can average the conditional error probability over the interference.

Other applications of maximum entropy have been treated elsewhere [7], [8]. These are moment problems that arise in physics and spectral estimation. Among the moment problems, estimation of multimodal probability density of states for a harmonic crystal and maximum entropy prediction for the isotropic Heisenberg model have been worked out in [7].

For the application in this paper perhaps it is appropriate at this point to outline briefly some other moment methods for comparison purposes. One possibility is to expand $f(x)$ in some set of orthogonal polynomials. The resulting series is truncated after $N + 1$ terms and the coefficients or weights in the expansion are determined by utilizing the first $N + 1$ moments of the unknown probability density function. This requires solution of a system of $N + 1$ linear equations. Proper choice of weighted orthogonal polynomials leads to fast convergence as N grows. An improper choice, however, can lead to oscillating approximations to $f(x)$, and there is further inaccuracy from lack of positivity at each stage of iterations. A popular choice for orthogonal polynomials are Hermite polynomials [9], [10]. It should be noted that Murphy [11] and Nakhla [12] use Legendre and Chebyshev polynomials, respectively. Nakhla's [12] result is related to Murphy's by way of an approximation in evaluating the coefficients in the series expansion. For a large number of problems, both results exhibit good convergence properties with respect to the number of moments required. Powerful alternatives have been developed [13] over many years. For example, Gauss quadrature rules [14] (GQR) were applied to evaluation of error probability due to ISI in digital communication by Benedetto et al. [15] (also see [16]). Here, the unknown density is defined by the quadrature rule $\{ w_j, \zeta_j \}_{j=1}^{N_0}$, a set of weights and nodes. Using N known moments ($N = 2N_0 + 1$) entails the solution to a set of nonlinear equations by diagonalization of a tridiagonal Jacobi matrix [14]. The corresponding numerical results are stable. It should also be noted that in certain problems [12], convergence properties similar to or even better than the Gauss quadrature rule were observed, using the series expansion method. However, as shown in this work and in other applications [7], the maximum entropy method is also stable and may produce results as accurate as GQR using fewer moments. Following these introductory remarks, in Section II we formulate the maximum entropy problem. Applications in digital communication are described in Section III. Also, numerical results and comparison to GQR results are

reported in the same section. Finally, our conclusions are presented in Section IV.

II. FORMULATION AND NUMERICAL PROCEDURE

We would like to maximize the entropy function with respect to the unknown density $f(x)$. Instead we can minimize the following with respect to the same:

$$\text{Minimize } H(f) = \int_a^b f(x) \ln [f(x)] \, dx \qquad (1)$$

$$\text{Subject to: } \mu_i = \int_a^b x^i f(x) \, dx \qquad (i = 1, 2, \cdots, N) \quad (2)$$

with

$$\mu_0 = \int_a^b f(x) \, dx = 1 \qquad (3)$$

and

$$-\infty < a, \ b < \infty.$$

This is an isoperimetric problem from calculus of variations (see, for example, Weinstock [17]). We can solve (1) by introducing Lagrange multipliers λ_k, $k = 0, 1, \cdots, N$. We define

$$H^*(f) = \int_a^b f(x) \ln [f(x)] \, dx + c \int_a^b f(x) \, dx$$

$$+ \sum_{k=1}^N \lambda_k \int_a^b x^k f(x) \, dx \quad (4)$$

and to set the variations, with respect to $f(x)$, of this quantity to zero:

$$\ln [f(x)] + \lambda_0 + \sum_{k=1}^N \lambda_k x^k = 0 \qquad (5)$$

where $\lambda_0 = c + 1$. Solving for $f(x)$ leads to

$$f_N(x) = \exp \left[-\lambda_0 - \sum_{k=1}^N \lambda_k x^k \right] . \qquad (6)$$

Notice that we have used f_N in (6) to point out that it is an estimate of the unknown density $f(x)$, based on its first N moments. Also, to decouple the zeroth moment from the rest of the computations, we have separated λ_0 from other λ's in (4)–(6).

Now we consider (3) where

$$\int_a^b f_N(x) \, dx = 1$$

and, hence,

$$e^{\lambda_0} = \int_a^b \exp \left[-\sum_{k=1}^N \lambda_k x^k \right] dx.$$

Using e^{λ_0} in (6) yields

$$f_N(x) = \exp \left\{ -\sum_{k=1}^N \lambda_k x^k - \ln \left[\int_a^b \exp \left(-\sum_{k=1}^N \lambda_k z^k \right) dz \right] \right\}$$

$$\qquad (7)$$

where z is the integration variable. Subject to the constraints of

(2) we can find the estimated moments as

$$\hat{\mu}_i = \int_a^b x^i e^{-\lambda_0 - \Sigma_{k=1}^N \lambda_k \cdot x^k} \, dx \qquad (i = 1, 2, \cdots, N) \quad (8)$$

where $\hat{\ }$ represents an estimated value, and by using the definition of λ_0 we express the estimated moments as

$$\hat{\mu}_i = \int_a^b x^i \exp \left\{ -\sum_{k=1}^N \lambda_k x^k \right.$$

$$\left. - \ln \left[\int_a^b \exp \left(-\sum_{k=1}^N \lambda_k z^k \right) dz \right] \right\} dx$$

$$\qquad (i = 1, 2, \cdots, N). \qquad (9)$$

Now the estimated moments in (9) have to be equal to the known moments μ_i. A somewhat different derivation of (7) and (9) can be found in [7]. A closed form solution of (9) is not possible except for the simple cases of $N = 1$ and 2. The solution for $N = 2$ can be found in [18]. Numerical methods of solution therefore become important. One such method is the Newton–Raphson method, which is used in finding solutions for systems of equations that, like (9), can be expressed in the form

$$G_i(\lambda_1, \lambda_2, \cdots, \lambda_N)$$

$$= \hat{\mu}_i - \mu_i \qquad (i = 1, 2, \cdots, N)$$

$$= \int_a^b (x^i - \mu_i) \exp \left[-\lambda_0 - \sum_{k=1}^N \lambda_k x^k \right] dx$$

$$= 0. \qquad (10)$$

The method starts with an initial guess at the solution, $\tilde{\lambda}^{(1)} = \{\lambda_1^{(1)}, \lambda_2^{(1)}, \cdots, \lambda_N^{(1)}\}$, and generates iterative approximate solutions $\tilde{\lambda}^{(2)}, \tilde{\lambda}^{(3)}, \cdots$, sequentially. If the initial guess $\tilde{\lambda}^{(1)}$ is close to a solution of (10), if the G_i's are continuously differentiable, and if the Jacobian $[\partial G_i / \partial \lambda_j]$ is nonsingular, then $\tilde{\lambda}^{(r)}$ will converge to the solution in limit as r gets large indefinitely. The method is based on the fact that, for small changes $\Delta \tilde{\lambda}^{(r)}$ in the Lagrange multipliers $\tilde{\lambda}^{(r)}$, we have

$$\tilde{\lambda}^{(r+1)} = \tilde{\lambda}^{(r)} + \Delta \tilde{\lambda}^{(r)}$$

and the approximate equality up to a term of order 0 $(\Delta \tilde{\lambda}^{(r)})$

$$G_i(\tilde{\lambda}^{(r)} + \Delta \tilde{\lambda}^{(r)}) \approx G_i(\tilde{\lambda}^{(r)}) + \sum_{k=1}^N \frac{\partial G_i(\tilde{\lambda}^{(r)})}{\partial \lambda_k^{(r)}} \Delta \lambda_k^{(r)}$$

$$(i = 1, \cdots, N). \quad (11)$$

We therefore take $\Delta \tilde{\lambda}^{(r)}$ to be a solution of the linear equation

$$G_i(\tilde{\lambda}^{(r)}) = \hat{\mu}_i - \mu_i = \sum_{k=1}^N \left[-\frac{\partial G_i(\tilde{\lambda}^{(r)})}{\partial \lambda_k^{(r)}} \right] \Delta \lambda_k^{(r)}$$

$$(i = 1, \cdots, N). \quad (12)$$

This method is straightforward, but somewhat prosaic. A more elegant application of Newton's iterative method has been adopted in [7], as follows. Define the potential function

$$\Gamma = \lambda_0 + \sum_{k=1}^N \mu_k \lambda_k. \qquad (13)$$

Evaluate the Hessian matrix H_{ki} as

$$H_{ki} = \frac{\partial^2 \Gamma}{\partial \lambda_k \partial \lambda_i} = \hat{\mu}_{k+i} - \hat{\mu}_k \cdot \hat{\mu}_i \qquad (14)$$

which can be proven to be positive definite. Then, starting with some initial value for $\tilde{\lambda}^{(r)}$, find $\tilde{\lambda}^{(r+1)}$ such that $\Delta \lambda_k^{(r)}$ is a solution of

$$\sum_{k=1}^{N} H_{ik} \Delta \lambda_k^{(r)} = \hat{\mu}_i - \mu_i. \qquad (15)$$

Note that (15) is similar to (12) and the Hessian matrix in the former is an analog for the differential term in the RHS of the latter. Alternative methods can also be developed by considering the convexity of $H(f)$ in (1) and using convex programming optimization techniques.

A. Symmetrical Unknown Densities

In practice we often encounter a random variable ω, distributed over $(-b, b)$ with an unknown density $g(\omega)$ with zero odd moments. This is obviously a probability density function that has even symmetry around its zero mean. For this case, in order to use the formulas developed based on maximum entropy, we make the following change of variable:

$$x = \omega^2. \qquad (16)$$

This is for computational simplicity. The change of variable is done by taking advantage of the even symmetry in $g(\omega)$. The new probability density is

$$f_N(\omega^2) = \frac{1}{2\omega} [g(\omega) + g(-\omega)]$$

$$= \frac{1}{\omega} g(\omega). \qquad (17)$$

Although ω is distributed over $(-b, b)$, x can only take on positive values over $(0, b^2)$; hence, $f_N(x)$ is not evenly symmetric around $x = 0$ and it can be expressed by the general formula in (7) as

$$f_N(\omega^2) = \frac{1}{\omega} g(\omega) = \exp \left\{ - \sum_{k=1}^{N'} \lambda_k \omega^{2k} \right.$$

$$\left. - \ln \left[\int_0^b 2\Omega \exp \left[- \sum_{k=1}^{N'} \lambda_k \Omega^{2k} \right] d\Omega \right] \right\}$$

or equivalently

$$g(\omega) = \omega \exp \left\{ - \sum_{k=1}^{N'} \lambda_k \omega^{2k} \right.$$

$$\left. - \ln \left[\int_0^b 2\Omega \exp \left(- \sum_{k=1}^{N'} \lambda_k \Omega^{2k} \right) d\Omega \right] \right\} \qquad (18)$$

where N' is the number of even moments. Similarly, the estimated moments in (9) can be redefined as

$$\hat{\mu}_{2i} = \int_0^b 2\omega^{2i+1} \exp \left\{ - \sum_{k=1}^{N'} \lambda_k \omega^{2k} \right.$$

$$\left. - \ln \left[\int_0^b 2\Omega \exp \left(- \sum_{k=1}^{N'} \lambda_k \Omega^{2k} \right) d\Omega \right] \right\} d\omega. \qquad (19)$$

B. Numerical Algorithm

In general, having assumed some initial values for $\tilde{\lambda}^{(r)}$, e.g., all zeros, using (9), (14), and (15), we can compute the update $\Delta \tilde{\lambda}^{(r)}$ for the next Newton iteration. Since the Hessian matrix in (15) is symmetric, the linear system in the equation was solved by the PORT library [19] routine DSYLE (used with double-precision arithmetic). Notice that the Hessian matrix is also positive definite; hence, a routine for solution of symmetric and positive definite systems might be more time efficient to use here. Finally, the numerical integrals involved in (9) are carried out by the PORT library [19] routine DQUAD which implements an adaptive quadrature integration in double precision.

The Newton iterations stop when the updated moments agree with the actual moments to a desired accuracy. Starting by providing the computer program with one or two moments, the computer program requests the desired number of Newton iterations. Based on all zero initial $\tilde{\lambda}^{(r)}$, the computer will print out the updated λ's ($\tilde{\lambda}^{(r+1)}$) as well as estimated moments for a cross-check against the actual moments. The user then enters one or two additional moments and the process continues until the desired accuracy is reached. In all our computations described in the following section, on the average, accurate results are obtained after 15 iterations. The computational algorithm described here is similar to that of [7].

III. APPLICATIONS—NUMERICAL RESULTS AND COMPARISON TO OTHER METHODS

A. Elementary Examples

To ensure the algorithm was operating properly, we estimated two known elementary density functions. The estimated density as well as the actual density function are illustrated in Fig. 1(a) and (b). Asterisks represent the actual and a continuous line represents the estimated probability density function. Fig. 1(a) depicts

$$f(x) = x + 0.5 \qquad 0 \le x \le 1 \qquad (20)$$

with moments

$$\mu_n = \frac{1.5n + 2}{n^2 + 3n + 2} \qquad (n = 0, 1, 2, \cdots) \qquad (21)$$

along with the estimated $f_N(x)$. Also, Fig. 1(b) shows

$$f(x) = 2x \qquad 0 \le x \le 1 \qquad (22)$$

with actual moments

$$\mu_n = \frac{2}{n + 2} \qquad (n = 0, 1, 2, \cdots). \qquad (23)$$

The reason we chose two linear functions of x was to verify whether a simple monomial can be approximated well by an exponential of polynomials. In both cases a total of five moments were used. As can be observed, the estimation is accurate.

B. Evaluation of Error Probability in Digital Communication

As a first example, in the context of cross-polarization interference [20], [21], suppose we have two M-state quadrature amplitude modulated (M-QAM) signals, each having a square constellation. The signals are orthogonally polarized and transmitted over a nondispersive fading channel. The channel cross-couples the two signals by a constant coupling ξ and a random, uniformly distributed phase ϕ. A sample of the received signal, x, on the in-phase rail of the reference polarization is

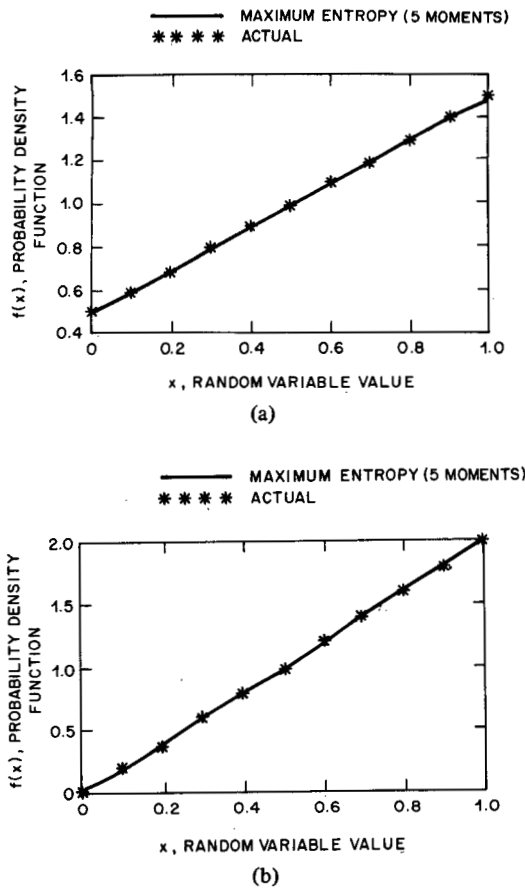$$x = \delta_1 + \xi [\delta_2 \cos \phi - \beta_2 \sin \phi] + n \qquad (24)$$

Fig. 1.   (a) $[f(x) = x + 0.5]$ simple test example. (b) $[f(x) = 2x]$ simple test example.

where (δ_2, β_2) are the interfering signal amplitude levels from the in-phase and quadrature rails, respectively. The desired signal amplitude level is δ_1 and is detected on the in-phase rail of the reference polarization. The sample of zero-mean Gaussian noise n has a variance σ^2. The amplitude levels of both polarization signals, i.e., (δ_1, β_1) and (δ_2, β_2) are uniformly distributed and take on values from $\{ \pm 1, \pm 3, \cdots, \pm (\sqrt{M} - 1)\}$, independently.

The interference term in (24) is

$$I = \xi \, [\delta_2 \, \cos \, (\phi) - \beta_2 \, \sin \, (\phi)]. \qquad (25)$$

We can derive the conditional error probability as

$$P_{e|I} = \frac{\sqrt{M} - 1}{\sqrt{M}} \, \text{erfc} \, \left[\frac{1 - I}{\sigma \sqrt{2}} \right] \qquad (26)$$

where

$$\text{erfc} \, (x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt. \qquad (27)$$

Clearly, the average error probability can be evaluated from

$$P(e) = \int_I P_{e|I}(I) f_N(I) \, dI \qquad (28)$$

where $f_N(I)$ is the estimated interference density. Now, using the moments of I with the maximum entropy method, we will find the average error probability in (28).

The moments of interference in this case are evaluated as follows. The odd moments are all zero due to the even symmetry in the probability density of the variables involved. Hence, we can use (18) and (19) to compute the average error

probability that is defined here as

$$P(e) = \int P_{e|I}(\omega) g(\omega) \, d\omega$$

where the numerical integration involved here is carried out in our program by the PORT library [19] routine DQUAD implementing an adaptive quadrature integration in double precision. The even moments are obtained from

$$\mu_n = E\{I^{2n}\} = \xi^{2n} \frac{(2n-1)!!}{(2n)!!} \sum_{j=0}^n \binom{n}{j}$$
$$\cdot E\{\delta_2^{2j}\} \cdot E\{\beta_2^{2(n-j)}\} \qquad (29)$$

where $E\{\cdot\}$ stands for statistical average,

$$(2n-1)!! = 1 \cdot 3 \cdot 5 \cdots (2n-1) \qquad (30)$$

and

$$(2n)!! = 2 \cdot 4 \cdots (2n).$$

The first 15 even moments for two cases of $\xi = 0.1$ and $\xi' = 0.17783$ for 16-QAM signals are presented in Table I(a) and (b), respectively.

The conditional error probability can be expressed from (26) as

$$P_{e|I} = \frac{(\sqrt{M} - 1)}{\sqrt{M}} \, \text{erfc} \, \left[\sqrt{\frac{3E_b \, \log_2 \, M}{2(M-1)(N_0)}} \, (1 - I) \right] \qquad (31)$$

in which

$$\frac{E_b}{N_0} = \frac{M-1}{3\sigma^2} \, \frac{1}{\log_2 \, M} \qquad (32)$$

is the bit energy-to-noise density ratio. After performing the averaging operation in (28), in Fig. 2(a) and (b) we plotted the average error probability versus E_b/N_0 for $\xi = 0.1$ and $\xi = 0.17783$, respectively. The plotted points marked by asterisks correspond to results using the GQR method with a total of 31 moments (including the zeroth moment) and the continuous line corresponds to the maximum entropy method results using a total of seven moments (including the zeroth moment) in averaging over the interference term. As can be seen, surprisingly, the results are the same, although the maximum entropy method requires only seven moments as opposed to 31 moments required by the GQR. Similar behavior was observed for other values of ξ. The dashed lines in Fig. 2(a) and (b) correspond to GQR results when a total of seven moments (including the zeroth moment) were applied. As can be observed, for large values of E_b/N_0, that is, when the interference dominates, GQR loses its accuracy. This is more pronounced in Fig. 2(b) for a higher interference level. For low values of E_b/N_0, the two methods seem to produce similar results. Note that the GQR results with 31 moments are quite accurate. This was verified in [20], for some limiting cases, by asymptotic evaluation of average error probability.

As a second example, in the context of ISI, we envision having a two-level pulse amplitude modulated signal over a 10 percent rolloff raised-cosine shaped channel [22]. This is one of the examples considered by Murphy [11]. The overall impulse response is denoted

$$h(t) = \frac{\sin \, [\pi(1 - \alpha)t/T] + \sin \, [\pi(1 + \alpha)t/T]}{2\pi t[1 - (2\alpha t/T)^2]/T} \qquad (33)$$

where T is the signal symbol duration and α is the rolloff factor. The received signal is sampled every T seconds. When there is a timing offset τ in the sampling, the received signal

TABLE I
MOMENTS OF INTERFERENCE FROM (29). (a) FOR $\xi = 0.1$. (b) FOR $\xi = 0.17783$.

ORDER	MOMENTS	ORDER	MOMENTS
2	.500E-01	2	.158E-00
4	.495E-02	4	.495E-01
6	.613E-03	6	.194E-01
8	.854E-04	8	.855E-02
10	.129E-04	10	.407E-02
12	.203E-05	12	.203E-02
14	.331E-06	14	.105E-02
16	.551E-07	16	.551E-03
18	.929E-08	18	.294E-03
20	.158E-08	20	.158E-03
22	.271E-09	22	.857E-04
24	.467E-10	24	.467E-04
26	.808E-11	26	.255E-04
28	.140E-11	28	.140E-04
30	.244E-12	30	.771E-05

(a)   (b)

———— MAXIMUM ENTROPY (7 MOMENTS)
✳✳✳ GQR (31 MOMENTS)
— — — GQR (7 MOMENTS)

BIT ENERGY-TO-NOISE DENSITY $\frac{E_b}{N_o}$, dB

(a)

———— MAXIMUM ENTROPY (7 MOMENTS)
✳✳✳ GQR (31 MOMENTS)
— — GQR (7 MOMENTS)
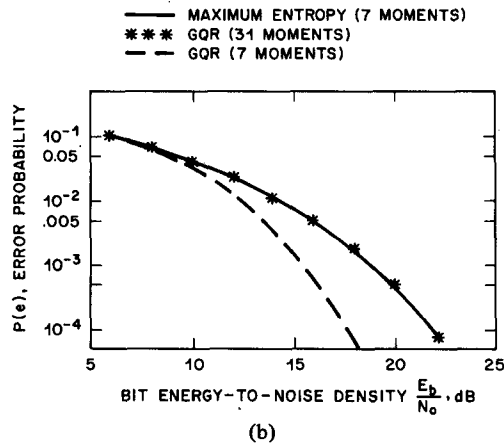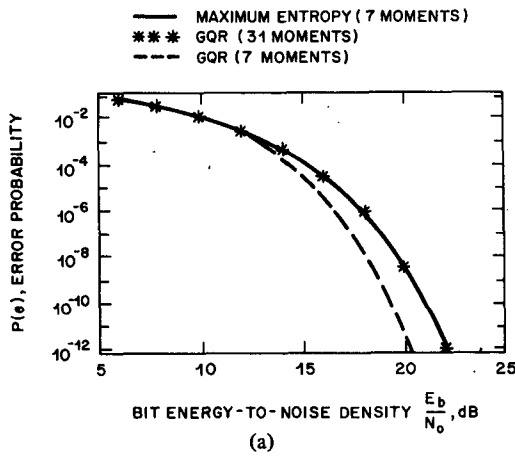
BIT ENERGY-TO-NOISE DENSITY $\frac{E_b}{N_o}$, dB

(b)

Fig. 2. Probability of bit error versus E_b/N_0 for 16-QAM dual signals. (a) $\xi = 0.1$. (b) $\xi = 0.17783$.

contains ISI. At the reference sample time, the received signal sample in Gaussian noise is

$$r_0 = h_0 a_0 + \sum_i{}' h_i a_{-i} + n. \qquad (34)$$

In (34), a_i's represent the independent uniformly distributed data symbols taking on values from $\{\pm 1\}$, h_i is a sample of channel impulse response, n is the zero-mean Gaussian noise sample and the prime on the summation denotes omission of the zero-index term.

The conditional error probability in terms of unperturbed signal-to-noise ratio, E_b/N_0, is defined as

$$P_{e|x} = \frac{1}{2} \text{erfc} \left[\sqrt{\frac{E_b}{N_0}} (1 - z) \right] \qquad (35)$$

where $z = \sum_i{}' a_{-i} h_i / h_0$, $E_b/N_0 = h_0^2/2\sigma^2$, and σ^2 is the noise variance.

Now, the average error probability can be evaluated by averaging $P_{e|z}$ over intersymbol interference, z, using different moment methods. A method for calculation of moments of a sum of weighted random variables, as in the ISI term, can be found in Prabhu [23]. The odd moments are again zero in this case because of the even symmetry of the symbol's probability density function. The ISI in this example is caused by sample timing offset and we evaluate the performance for I) timing offset $= 0.1T$, and II) timing offset $= 0.2T$. The first case corresponds to one of the examples in Murphy [11]. At a signal-to-noise ratio of 15 dB we can compare our result to the exact result in Murphy's Fig. 6 [11]. As mentioned earlier, in both cases we use a 10 percent rolloff factor in the raised-cosine shaping filter. In Fig. 3(a) we show the average bit error probability versus E_b/N_0, for a timing offset of $0.1T$. Exact results were achieved by GQR using 27 moments. Maximum entropy required 11 moments to obtain nearly the exact bit error probability while GQR with 11 moments resulted in a bit error probability value that was almost two orders of magnitude smaller than the exact value, at 18 dB signal-to-noise ratio. Note that at 15 dB signal-to-noise ratio, the error probability on the curve shown by asterisks is the same as that in Fig. 6 of Murphy [11]. Fig. 3(b) shows similar results for a timing offset value of $0.2T$. Note that the exact value curves in Fig. 3(b) correspond to GQR using 21 moments and maximum entropy using nine moments. The Gauss quadrature rule using nine moments yields values that could be misleading in this practice because, as seen, the accurate curves reach an irreducible error rate value at a signal-to-noise ratio of 16 dB while GQR using nine moments continues to show a waterfall effect. Notice that in these two cases a similar behavior as in Fig. 2(a) and (b) can be observed. That is, for example, Fig. 3(b) corresponding to a larger interference peak value indicates that GQR with a smaller number of moments is less accurate than the same in Fig. 3(a) corresponding to a smaller interference peak level.

Another test in this same context is when the rolloff $\alpha = 0$, the timing offset $\tau = 0.2T$ and an 11 pulse truncation is adopted. For a signal-to-noise ratio of 16 dB this is an example in [12]. As seen in Table II the exhaustive search method [15] results in $P_e = 2.76 \times 10^{-3}$. For 15 moments (including the zeroth moment), GQR results in $P_e = 2.74 \times 10^{-3}$. Maximum entropy for five moments provides $P_e = 2.51 \times 10^{-3}$ while GQR for five moments results in $P_e = 3.77 \times 10^{-5}$. In Table II we also provide results of using Murphy's method. It appears that Murphy's method converges as well in this example. Therefore, the maximum entropy method produces results as accurate as Murphy's method using approximately the same number of moments in the ISI example. However, as stated in the Introduction, methods using density function expansion in terms of orthogonal polynomials, such as Murphy's, can lead to oscillating approximations in some problems [24]. It would be interesting to see if our method produced stable results in these problems.

In this paper our chief purpose is to find out whether maximum entropy can be used as a tool in calculating average error probability in a digital communication system. However, as a by-product the estimated interference probability density function is provided. Fig. 4 depicts the estimated interference density function for Case I of the above example ($0.1T$ timing offset). It is obvious that the number of moments is vital to the
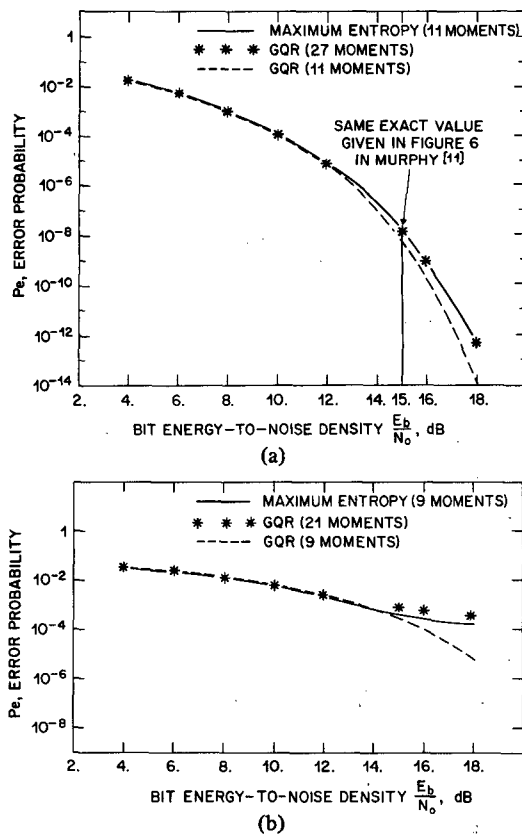
Fig. 3. Probability of bit error versus E_b/N_0. (a) For ISI Case I with timing offset = $0.1T$. (b) For ISI Case II with timing offset = $0.2T$.

TABLE II
PROBABILITY OF ERROR VERSUS HIGHEST ORDER MOMENT (INCLUDING
THE ZEROTH-ORDER MOMENT). IDEAL BAND-LIMITED PULSE, 11 PULSE
TRUNCATION APPROXIMATION, SNR = 16 dB, TIMING OFFSET = $0.2T$.
EXHAUSTIVE METHOD PRODUCES AN ERROR PROBABILITY OF
2.7614×10^{-3}.

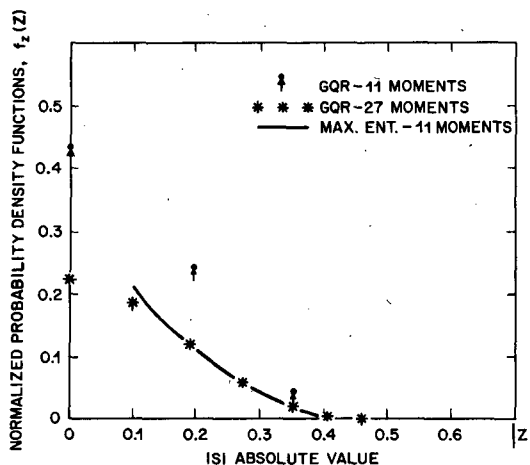| HIGHEST ORDER MOMENT	MURPHY	GQR	MAX ENTROPY
5	2.79×10^{-3}	3.77×10^{-5}	2.51×10^{-3}
7	2.77×10^{-3}	8.86×10^{-4}	2.79×10^{-3}
9	2.76×10^{-3}	2.42×10^{-3}	2.76×10^{-3}
11	2.76×10^{-3}	2.96×10^{-3}	2.76×10^{-3}
13	2.76×10^{-3}	2.82×10^{-3}	2.76×10^{-3}
15	2.76×10^{-3}	2.74×10^{-3}	2.76×10^{-3}

Fig. 4. Normalized estimated probability density functions of ISI for $0.1T$ timing offset value.

convergence and accuracy of the GQR. For example, GQR with 11 moments provides five weights (density function samples) and because of even symmetry, only three weights contain information on tail distribution (as shown in Fig. 4). Therefore, the probability of error evaluation has to rely only on these three samples to obtain appropriate numerical results. On the other hand, maximum entropy with a few moments seems to follow the exact tail distribution (for GQR with 27 moments) closely, as seen in Fig. 4. This opens up a new area for future work to explain the reasons. We observe in Fig. 4 that maximum entropy provides a very robust tail distribution, and this is what is important in error probability computations.

These observations all lead us to the fact that most distributions dealt with in communications theory are indeed maximum entropy distributions or have a target density which is smooth. Therefore, the maximum entropy estimate appears efficient. Hence, if the stability and accuracy of the maximum entropy estimate is tested and verified in other applications of digital communication, it can be accepted as an efficient way of computing average error probability.

Clearly, maximum entropy, like any other approximation method, has its own limitations and is not prescribed as the best solution in all moment problems that may arise in practice. In particular, if the unknown density is spiky, the maximum entropy estimate becomes inefficient. Fortunately, as state earlier, most interference densities that we deal with in communication theory belong to the smooth class.

IV. CONCLUSIONS

In this work we implemented a maximum entropy method to estimate an unknown density from knowledge of a finite number of its moments. We applied the technique to evaluation of average probability of error, a widely used performance measure in digital communication. Our assessment of the method, in a few examples studied here and in [25], is that

1) rather accurate averages can be obtained, even from a few known moments;

2) average are stable, i.e., do not oscillate around the actual value, and compete well with results from the well-developed and widely used Gauss quadrature rules (GQR) method.

Merits of the presented approach should still be searched for in other successful applications. While we have treated a limited number of problems, the extracted features seem appealing.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communications," Bell Syst. Tech. J., vol. 27, pp. 379, 623, 1948.
[2] E. T. Jaynes, "Information theory and statistical mechanics—I," Phys. Rev., pp. 620–630, 1957.
[3] ——, "On the rationale of maximum entropy methods," Proc. IEEE, vol. 70, pp. 939–952, Sept. 1982.
[4] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," NRL Memo. Rep. 3898, Off. Naval Res., Dec. 1978.
[5] J. M. Van Campenhout and T. M. Cover, "Maximum entropy and conditional probability," IEEE Trans. Inform. Theory, vol. IT-27, pp. 483–489, July 1981.
[6] H. J. Landau, personal communication.
[7] L. R. Mead and N. Papanicolaou, "Maximum entropy in the problem of moments," J. Math. Phys., vol. 25, pp. 2404–2417, Aug. 1984.

[8] A. Papoulis, "Maximum entropy and spectral estimation: A review," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-29, pp. 1176-1186, 1981.

[9] E. Y. Ho and Y. S. Yeh, "A new approach for evaluating the error probability in the presence of intersymbol interference and Gaussian noise," *Bell Syst. Tech. J.,* vol. 49, pp. 2249-2265, Nov. 1970.

[10] O. Shimbo and M. J. Celebiler, "The probability of error due to intersymbol interference and Gaussian noise in digital communication systems," *IEEE Trans. Commun. Technol.,* vol. COM-19, pp. 113-119, Apr. 1971.

[11] J. V. Murphy, "Binary error rate caused by intersymbol interference and Gaussian noise," *IEEE Trans. Commun.,* vol. COM-19, pp. 1039-1046, Sept. 1973.

[12] M. Nakhla, "A new method for error rate calculation and its application to fiber optic communication systems," presented at Nat. Telecommun. Conf., Washington, DC, Nov. 1979.

[13] G. A. Baker and P. Graves-Morris, *Padé Approximations.* Reading, MA: Addison-Wesley, 1980.

[14] G. H. Golub and J. H. Welsch, "Calculations of Gauss quadrature rules," *Math. Comput.,* vol. 26, pp. 221-230, Apr. 1969.

[15] S. Benedetto *et al.,* "Error probability in the presence of intersymbol interference and additive noise for multilevel digital signals," *IEEE Trans. Commun.,* vol. COM-21, pp. 181-190, Mar. 1973.

[16] M. H. Meyers, "Computing the distribution of a random variable via Gaussian quadrature rules," *Bell Syst. Tech. J.,* vol. 61, pp. 2245-2261, Nov. 1982.

[17] R. Weinstock, *Calculus of Variations with Applications to Physics and Engineering.* New York: Dover, 1974.

[18] C. R. Rao, *Linear Statistical Inference and its Applications,* 2nd ed. New York: Wiley, 1973, pp. 162-163.

[19] *The PORT Mathematical Subroutine Library,* AT&T Bell Labs., Comput. Inform. Services, Murray Hill, NJ, 1976.

[20] M. Kavehrad, P. J. McLane, and C.-E. Sundberg, "Combined quadrature amplitude modulation and convolutional codes for cross-coupled multidimensional channels," in *Proc. Int. Conf. Commun.,* Toronto, Ont., Canada, June 1986.

[21] M. Kavehrad, "Performance of cross-polarized M-ary QAM signals over nondispersive fading channels," *AT&T Bell Lab. Tech. J.,* vol. 63, pp. 499-521, Mar. 1984.

[22] R. Lucky *et al., Principles of Data Communication.* New York: McGraw-Hill, 1968.

[23] V. K. Prabhu, "Some considerations of error bounds in digital systems," *Bell Syst. Tech. J.,* vol. 50, pp. 3127-3152, Dec. 1971.

[24] A. Gillett, "Calculation of error probabilities in a radio system with two-path fading," TeleCom Australia Res. Lab., Victoria, Australia, Rep. 7451, Sept. 1981.

[25] M. Kavehrad and M. Joseph, "Maximum entropy and the method of moments in evaluation of probability of error in digital communications systems," in *Proc. Inform. Sci. Syst. Conf.,* Princeton Univ., Princeton, NJ, Mar. 1986.

★
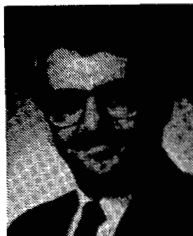
Mohsen Kavehrad (M'78-SM'86) was born on January 1, 1951. He received the B.S. degree from Tehran Polytechnic Institute, Tehran, Iran, in 1973, the M.S. degree from Worcester Polytechnic Institute, Worcester, MA, in 1975, and the Ph.D. degree from the Polytechnic Institute of New York, Brooklyn, in 1977.

During 1977-1978, he worked at Fairchild Industries. He then joined GTE (Satellite and Laboratories) and stayed until 1981. Since then he has been with AT&T Bell Laboratories, where he works in the Communications Systems Research Lab at Crawford Hill, Holmdel, NJ. He has worked on satellite communications, packet switching networks, digital radio modulation and coding studies over line-of-sight multipath channels, local wireless communication, and portable radios. Currently he is working on optical local area network problems. He has taught part-time at Northeastern University, Boston, MA, and Lowell University, Lowell, MA.

Dr. Kavehrad is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and a former Technical Editor for the IEEE COMMUNICATIONS MAGAZINE. He established and was the Chairman of the IEEE Communications Chapter of New Hampshire in 1984. He is a member of Sigma Xi.

★
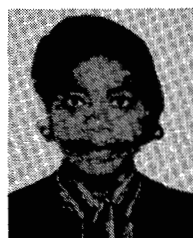
Myrlene Joseph is a computer science major in her senior year at the City College of the City University of New York.

Since her sophomore year, she has been highly motivated in software development and mathematics. She plans to pursue a career in systems analysis.