Information Distinguishability with Application to Analysis of Failure Data

Ehsan S. SOOFI, Nader EBRAHIMI, and Mohamed HABIBULLAH*

In maximum entropy (ME) modeling, the information discrepancy between two distributions is measured in terms of their entropy difference. In discrimination information statistics the information discrepancy between two distributions is measured in terms of the Kullback–Leibler function (i.e., relative entropy or cross-entropy). This article presents an equivalence between Kullback–Leibler functions and entropy differences involving an ME distribution. Based on this equivalence, the concept of *information discrimination (ID) distinguishability* is introduced as a unifying framework for the two methods of measuring information discrepancy between distributions. Applications of ID distinguishability as diagnostics for examining robustness of parametric procedures and sensitivity of nonparametric statistics across parametric families of distributions is proposed. The equivalence result facilitates estimation of Kullback–Leibler functions in terms of entropy estimates. Application of the ID distinguishability to modeling failure data brings a new dimension into entropy estimation—entropy estimation based on the hazard function. ID statistics for modeling lifetime distributions with increasing failure rates are studied. Two illustrative examples are analyzed.

KEY WORDS: Entropy; Kullback–Leibler; Lifetime distribution; Nonparametric; Reliability.

1. INTRODUCTION

Parametric inference about a vector of unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ begins with postulating a *known model*, $f(x|\boldsymbol{\theta})$, that represents the *unknown true* data generating distribution. Although distinction between the model and the unknown data-generating distribution is seldom made explicit, concerns are often expressed regarding compatibility of data with the postulated model and robustness of the inference against deviations of $f(x|\boldsymbol{\theta})$ from the true data-generating distribution. Nonparametric inference embeds the data-generating distribution in a broad class $\Omega_{\boldsymbol{\theta}}$ of families of distributions indexed by $\boldsymbol{\theta}$ and proceeds with inference suitable for all members of $\Omega_{\boldsymbol{\theta}}$. But interest is sometimes shown in comparing outcomes of a nonparametric procedure on data generated from a few models in $\Omega_{\boldsymbol{\theta}}$.

Maximum entropy (ME) principle of inference (Jaynes 1957; Shore and Johnson 1980) considers a class of distributions

$$\Omega_{\boldsymbol{\theta}} = \{f(x|\boldsymbol{\theta}) : E_f[T_j(X|\boldsymbol{\theta})] = \theta_j, j = 0, 1, \ldots, m\},$$

where T_j's are absolutely integrable functions with respect to f and $\boldsymbol{\theta} = (\theta_1, \ldots \theta_m)$. For the continuous case, the inference is based on the model that maximizes the differential entropy

$$H[f(x|\boldsymbol{\theta})] = -\int f(x|\boldsymbol{\theta}) \ln f(x|\boldsymbol{\theta})\, dx, \quad (1)$$

subject to the information constraints that define $\Omega_{\boldsymbol{\theta}}$. The ME model $f^*(x|\boldsymbol{\theta})$ in $\Omega_{\boldsymbol{\theta}}$, if it exists, is of the form

$$f^*(x|\boldsymbol{\theta}) = C(\boldsymbol{\theta})\exp[\eta_1(\boldsymbol{\theta})T_1(x) + \cdots + \eta_m(\boldsymbol{\theta})T_m(x)], \quad (2)$$

where $C(\boldsymbol{\theta})$ is the normalizing constant (i.e., partition function) and $\eta_1, \ldots, \eta_m$ are Lagrange multipliers. Many well-known distributions are ME subject to various types of constraints. Table 1 presents some examples. Kapur (1989) provided detail for these and many other ME distributions; the quartic exponential was studied by Zellner and Highfield (1988).

The concern in the ME paradigm is how closely $f^*(x|\boldsymbol{\theta})$ approximates the data-generating distribution. This concern stems from the query about whether the data-generating distribution can be satisfactorily described by the information constraints specified about the members of $\Omega_{\boldsymbol{\theta}}$. If this is the case, then the entropy of data distribution is expected to be somewhat close to the ME (Jaynes 1982) given by

$$H[f^*(x|\boldsymbol{\theta})] = -\ln C(\boldsymbol{\theta}) - \eta_1(\boldsymbol{\theta})\theta_1 - \cdots - \eta_m(\boldsymbol{\theta})\theta_m. \quad (3)$$

When the constraints do not reflect the information content of the underlying random mechanism of data-generating process, a nonparametric estimate of the entropy solely based on the data would generally yield an unacceptably lower value than the parametric ME (3) estimated by the data. In such a case the use of $f^*(x|\boldsymbol{\theta})$ would be inadequate, because it would fail to correctly predict the future outcomes. Jaynes, on the basis of a historical development in the quantum theory, concluded that "the principle of maximum entropy is most useful to us in just those cases where it fails to predict the correct experimental facts" (Jaynes 1968, p. 232). If the failure of an ME model "persists on infinite repetition of the experiment, then we will conclude that the physical mechanism of the experiment must contain additional constraints which were not taken into account in the maximum-entropy calculation. The observed deviations then provide a clue as to the nature of these new constraints" (Jaynes 1968, p. 232). The ME approach, therefore, assumes a proactive role for statistics in basic scientific research. This is perhaps the strongest appeal of the ME modeling approach that distinguishes it from the conventional methods.

* Ehsan S. Soofi is Professor, School of Business Administration, University of Wisconsin, Milwaukee, WI 53201. Nader Ebrahimi is Professor, Division of Statistics, Northern Illinois University, DeKalb, IL 60115. Mohamed Habibullah is Associate Professor, Department of Mathematics and Computer Science, University of Wisconsin, Superior, WI 54880. He is presently visiting the Department of Management Science, Northeastern University, Boston, MA 02115. The authors thank two referees and an associate editor for their very extensive and helpful comments on a previous draft. The helpful comments of Timothy Haas and the computing assistance of Muhammad Musa and Sandeep Purao are thankfully acknowledged. Ebrahimi's research was partially supported by U.S. Air Force Office of Scientific Research Grant AFOSR-89-0402. Habibullah's research was funded by a summer faculty development grant from the University of Wisconsin-Superior.

Journal of the American Statistical Association
June 1995, Vol. 90, No. 430, Theory and Methods

Table 1. Examples of Continuous Maximum Entropy Distributions

$f(x) > 0$	$T(X)$	ME Distribution
(a, b)	none	uniform
$(0, 1)$	$\ln X, \ln(1 - X)$	beta
$(0, \infty)$	X	exponential
$(0, \infty)$	$X, \ln X$	gamma
$(0, \infty)$	$X^\beta, \ln X, (\beta \neq 1)$	Weibull
$(a, \infty), a > 0$	$\ln X$	Pareto
$(-\infty, \infty)$	$\lvert X\rvert$	Laplace
$(-\infty, \infty)$	X^2	normal (mean = 0)
$(-\infty, \infty)$	X, X^2	normal
$(-\infty, \infty)$	$\ln(1 + X^2)$	generalized Cauchy
$(-\infty, \infty)$	$X, \ln(1 + e^{-\lambda X})$	generalized logistic
$(-\infty, \infty)$	$X, e^{-\lambda X}$	generalized extreme value
$(-\infty, \infty)$	X, X^2, X^3, X^4	quartic exponential

Entropy-based procedures for exploring constraints that may be operational in a data-generating distribution have been developed for discrete data analysis (Gokhale and Kullback 1978; Soofi 1992, 1994). For the continuous case, however, the existing entropy-based papers (Arizono and Ohta 1989; Chandra, DeWet, and Singpurwalla 1982; Dudewicz and Van der Meulen 1981; Ebrahimi, Habibullah, and Soofi 1992; Gokhale 1983; Vasicek 1976) have remained in the conventional hypothesis testing paradigm. These works have framed the problem as goodness-of-fit statistics and tests of distributional hypotheses and thus have not gone beyond recommending a decision on "significance" of the fit. Consequently, the existing works on entropy-based procedures have not utilized the essence of ME modeling as a guide for further understanding of constraints that govern a continuous data-generating distribution. In contrast, in this article we view the problem as that of *estimation* of information discrepancy between the unknown data-generating distribution and the ME model. This approach is in the same vein as that taken by Akaike (1973), which perhaps "can provide solutions for various important practical problems which have hitherto been treated as problems of statistical hypothesis testing rather than of statistical decision or estimation." This framework of course does not exclude the use of an estimated information discrepancy as a test statistic in the traditional sense.

Information-theoretic quantities proposed for measuring distributional disparities are based either on *entropy distinguishability* (i.e., entropy difference) or on Kullback–Leibler discrimination information (i.e., relative entropy). Based on entropy distinguishability, Vasicek (1976) proposed a test of normality, Dudewicz and Van der Meulen (1981) proposed a test of uniformity, Chandra et al. (1982) proposed a test of exponentiality, and Gokhale (1983) discussed goodness-of-fit tests for ME distributions. Based on the Kullback–Leibler function, Arizono and Ohta (1989) developed a test of normality that turned out to be the same as the test proposed by Vasicek (1976). Ebrahimi et al. (1992) also used the Kullback–Leibler function and developed a test of exponentiality that came out to be the same as the test developed by Gokhale (1983). These coincidental cases have not yet been explicated at a general level.

This article is organized as follows. In Section 2 we present a result on the equivalence of entropy differences and Kullback–Leibler discrimination functions involving an ME distribution. This simple result enables us to introduce the concept of *information discrimination (ID) distinguishability* as a unifying framework for the two information-theoretic formulations of measuring compatibility of data with an ME model. The ID-distinguishability framework explicates the entropy-difference statistics in terms of Kullback–Leibler statistics whose foundation is well established (Kullback 1959). We define the *ID index* of distributions based on the Kullback–Leibler discrimination function.

In Section 3 we exploit an analogy between the euclidean distance and the relative entropy (Csiszar 1975) and propose studying robustness of parametric procedures and sensitivity of nonparametric procedures based on the concept of ID distinguishability. We discuss ID distinguishability of some location-scale families and the Student-*t* distribution relative to normality and ID distinguishability of some lifetime distributions relative to exponentiality.

In Section 4 we discuss developing ID statistics. The result on the equivalence between entropy difference and Kullback–Leibler function gives a way to estimate relative entropies via entropy estimates. This approach greatly facilitates developing information diagnostics for model building, model testing, and other statistical purposes. We exploit a known but never-used relationship between entropy and the hazard rate function and study entropy estimation based on the hazard rate function.

In Section 5 we focus on ID distinguishability of data generated from increasing failure rate (IFR) distributions. We develop information diagnostics that are useful for assessing departure of the data-generating distribution from exponentiality in the IFR class. We also illustrate an application of ID index by examining sensitivity of some tests of exponentiality to the parametric family of alternatives. We analyze two examples to illustrate the use of information-theoretic diagnostics tailored for analysis of failure data. In Section 6 we provide some concluding remarks.

2. INFORMATION DISTINGUISHABILITY

The entropy of the ME distribution provides a benchmark for comparing distributions in Ω_θ. Entropy is a measure of concentration of probabilities. Low entropy distributions are more concentrated, and hence more informative, than high entropy distributions. The discrepancy

$$\delta H(f, f^*) = H[f^*(x \mid \theta)] - H[f(x \mid \theta)] \tag{4}$$

measures the additional amount of information contained in f that is not included in f^*.

A number of indices based on $\delta H(f, f^*)$ have been defined for comparison of distributions. Shannon (1948) used the normal distribution as the basis for comparing uncertainty in continuous distributions. Among those distributions with a given variance σ^2, the normal distribution has the ME of $\frac{1}{2}\ln(2\pi e\sigma^2)$. Shannon defined the *entropy power*, $N(f)$, of a distribution $f(x \mid \sigma^2)$ as the variance of any normal density with the same entropy as $H[f(x)]$. In symbols, $N(f) \equiv \exp\{2H[f(x)]\}/2\pi e$, which may be written in terms of (4) as $N(f) = \exp[-2\delta H(f, \phi)]$, where $\phi = N(0, 1)$. Thus $N(f)$ is a measure of closeness of the information content of f to the standard normal density. Vasicek (1976) proposed a test of normality based on $N(f)$.

Dudewicz and Van der Meulen (1981) defined the *entropy power variance ratio* (EPVR) of distributions by the ratio $\text{EPVR}(f) \equiv \exp\{H[f(x)]\}/\sigma(f) = [2\pi e N(f)]^{1/2}/\sigma(f)$. They introduced the concept of *EPVR distinguishability*, $[\text{EPVR}(f_1) \neq \text{EPVR}(f_2)]$, and developed a test of uniformity based on $\text{EPVR}(f)$.

Gokhale (1983) proposed tests of ME distributions by considering the *entropy power fraction* (EPF) of $f(x|\theta)$ in Ω_θ, defined by

$$\text{EPF}(f|\theta) \equiv \frac{\exp\{H[f(x|\theta)]\}}{\exp\{H[f^*(x|\theta)]\}} = \exp\{H[f(x|\theta)] - H[f^*(x|\theta)]\}. \quad (5)$$

Note that $0 < \text{EPF}(f|\theta) \leq 1$. An $\text{EPF}(f|\theta) \approx 1$ indicates that f is very similar to f^*, in information. Two densities, $f_1(x|\theta)$ and $f_2(x|\theta)$, in Ω_θ are said to be *EPF distinguishable* if $\text{EPF}(f_1|\theta) \neq \text{EPF}(f_2|\theta)$.

The most well-known information-theoretic measure of discrepancy between distributions is the Kullback–Leibler function. Distributions in Ω_θ are also compared according to the Kullback–Leibler discrimination function (i.e., relative entropy),

$$K(f:f^*|\theta) \equiv \int f(x|\theta) \ln \frac{f(x|\theta)}{f^*(x|\theta)}\, dx. \quad (6)$$

It is well known that $K(f:f^*|\theta) \geq 0$ and the equality holds if and only if $f(x|\theta) = f^*(x|\theta)$ almost everywhere. Arizono and Ohta (1989) and Ebrahimi et al. (1992) developed tests of normality and exponentiality based on $K(f:f^*|\theta)$ that are the same as the tests developed by Vasicek (1976) and Gokhale (1983).

Now we define an index for comparing distributions based on the relative entropy (6).

Definition 2.1. The ID index of distributions in Ω_θ is defined by

$$\text{ID}(f:f^*|\theta) \equiv 1 - \exp[-K(f:f^*|\theta)]. \quad (7)$$

Two distributions $f_1(x|\theta)$ and $f_2(x|\theta)$ in Ω_θ are *ID distinguishable* if $\text{ID}(f_1:f^*|\theta) \neq \text{ID}(f_2:f^*|\theta)$.

In an analogy of the euclidean geometry (Csiszar 1975), all distributions in Ω_θ on the same information sphere centered at the ME distribution are not ID distinguishable, and those on different spheres centered at f^* are ID distinguishable. This analogy is useful for diagnosing sensitivity of statistics to parametric families in robustness studies and in nonparametric statistics.

The ID transformation defined in (7) is a normalization, so $0 \leq \text{ID}(f:f^*|\theta) \leq 1$. Similar transformations of the Kullback–Leibler function have been proposed in other contexts (Joe 1989). An $\text{ID}(f:f^*|\theta) \approx 0$ indicates that f and f^* are approximately the same. This means that the constraints used in the ME calculations have high *information value* for the distribution, f. Otherwise, the constraints cannot satisfactorily describe the class of distributions to which f belongs. In this case, an additional or a different set of constraints are required for approximating f by an ME distribution. Incremental contribution of additional constraints for approximating f by an ME model is determined by the reduction of the ID index.

Next we present a result that unifies two streams of information-theoretic research: entropy difference and relative entropy methods of measuring disparity between distributions.

Theorem 2.1. Two distributions, $f_1(x|\theta)$ and $f_2(x|\theta)$, in Ω_θ are EPF distinguishable if and only if they are ID distinguishable.

Proof. Expand $K(f:f^*|\theta)$ and use (1) and (2) to obtain

$$\begin{aligned} K(f:f^*|\theta) &= -H[f(x|\theta)] - E_f[\ln f^*(X|\theta)] \\ &= -H[f(x|\theta)] - \ln C(\theta) - \eta_1(\theta)E_f[T_1(X)] \\ &\quad - \cdots - \eta_m(\theta)E_f[T_m(X)]. \end{aligned} \quad (8)$$

For all f in Ω_θ, $E_f[T_j(X)] = \theta_j$, $j = 1, \ldots, m$; thus using (3) in (8) gives

$$K(f:f^*|\theta) = H[f^*(x|\theta)] - H[f(x|\theta)]. \quad (9)$$

Now using (9) in (7) and comparing with (5) gives

$$\begin{aligned} \text{ID}(f:f^*|\theta) &= 1 - \exp\{H[f(x|\theta)] - H[f^*(x|\theta)]\} \\ &= 1 - \text{EPF}(f|\theta). \end{aligned} \quad (10)$$

The relationship (9) was observed by Soofi (1992) for the discrete uniform case. It has been used by Arizono and Ohta (1989) for the normal case and by Ebrahimi et al. (1992) for the exponential case. But none of these works recognized the relationship's generality.

The foregoing result shows that the indices defined based on the difference between the entropy of a distribution in Ω_θ and the ME are legitimate measures of discrepancy between the respective densities. Thus, for example, the entropy power defined by Shannon measures the discrepancy between the density of a standardized random variable and the standard normal density. The relation (9) formalizes the fact that the difference between two entropies vanishes if and only if $f(x|\theta) = f^*(x|\theta)$ almost everywhere. This result explicates the entropy-difference statistics in terms of Kullback–Leibler statistics whose foundation is well established. It also gives a new and easily understood interpretation of the Kullback–Leibler function and greatly facilitates estimation of the relative entropies involving an ME distribution.

3. DIAGNOSING STATISTICS ACROSS PARAMETRIC FAMILIES

The ID indices of distributions are useful for planning robustness and power studies. Often it is desirable to examine robustness of parametric procedures according to "departure" of the data-generating distribution from the assumed model. It is natural to expect that the closer the data-generating distribution to the assumed model, the better the performance of the parametric procedure. But "departure" is not usually operationalized according to a general disparity measure between distributions. Usually, data are generated from a few distributions chosen from some families of distributions, and in each family the departure from the assumed model is discussed in terms of some parameters. Sometimes characteristics such as symmetry and kurtosis are used as measures of departure. The ID index provides a general normalized measure for operationalizing departures from an assumed model, both within a parametric family and across the parametric families.

Performance of nonparametric statistics are usually examined in application to a few parametric families of distributions. For example, the powers of tests of distributional hypotheses are routinely examined by considering a few distributions as alternatives. Analogous to studying powers of parametric tests in terms of the euclidian distance between the null and alternative values, one can use ID indices for studying the powers of nonparametric tests across various parametric families of distributions. Statistics that show different behaviors for the same ID index are said to be sensitive to the parametric families.

3.1 ID Distinguishability Relative to Normality

Let $T_1(X) = X^2$ and $\theta_1 = \sigma^2$. Then the ME distribution in $\Omega_\sigma = \{f(x|\sigma^2): E(X^2|\sigma^2) = \sigma^2\}$ is the normal distribution $f^*(x|\sigma^2) = N(0, \sigma^2)$. Consider the following location-scale families: logistic with density

$$f_1(x|\tau_1, \lambda_1) = \lambda_1 \exp\{-\lambda_1(x - \tau_1)\} \times [1 + \exp\{-\lambda_1(x - \tau_1)\}]^2, \quad -\infty < \tau_1 < \infty, \quad \lambda_1 > 0;$$

Laplace with density

$$f_2(x|\tau_2, \lambda_2) = \frac{1}{2} \lambda_2 \exp(-\lambda_2|x - \tau_2|), \quad -\infty < \tau_2 < \infty, \quad \lambda_2 > 0;$$

and Gumbel with density

$$f_3(x|\tau_3, \lambda_3) = \lambda_3 \exp\{-\lambda_3(x - \tau_3)\} \times \exp[-\exp\{-\lambda_3(x - \tau_3)\}], \quad -\infty < \tau_3 < \infty, \quad \lambda_3 > 0;$$

τ_i, and λ_i, $i = 1, 2, 3$, being the location and scale parameters of the distributions.

For $\pi^2/(3\lambda_1^2) = 2/\lambda_2^2 = \pi^2/(6\lambda_3^2) = \sigma^2$, f_1, f_2, and f_3 are members of Ω_σ. The entropy of all location-scale distributions with finite variance σ^2 is in the form of

$$H[f(x|\sigma^2)] = \frac{1}{2}\ln(\sigma^2) + h(f), \tag{11}$$

where $h(f)$ is a constant independent of the parameters of f (see, for example, Zellner 1984). The entropies of the foregoing families are given by (11) with $h(f_1) = \frac{1}{2}\ln(3e^4/\pi^2)$, $h(f_2) = \frac{1}{2}\ln(2e^2)$, and $h(f_3) = \frac{1}{2}\ln(6e^2/\pi^2) + \gamma$, $\gamma = 0.5771 \ldots$ being the Euler constant (see Soofi and Gokhale 1991). Using $H[f(x|\sigma^2)]$ given in (11) and $H[f^*(x|\sigma^2)] = \frac{1}{2}\ln(\sigma^2) + \frac{1}{2}\ln(2\pi e)$ in (10) gives the ID indices of all location scale families relative to normality.

Figure 1 shows the ID indices of these distributions as horizontal lines in the (σ^2, ID) plane. We note that given the variance, the logistic distribution is the closest of the three location-scale families to the normal distribution and the Gumbel is the farthest.

Figure 1 also shows the ID indices of the Student-t distributions with various degrees of freedom, k. Observe that the Student-t family is always ID distinguishable with the three location-scale families considered here, with the ordering depending on the degrees of freedom.

The ID indices of distributions relative to normality are useful for examining the robustness of normal theory statistics and the powers of tests of normality. Figure 1 is useful to plan studies that give insights about the sensitivity of a statistic developed under the normal theory to the parametric families. For example, in Figure 1 we see that t_8 is almost not ID distinguishable with the logistic distribution. Thus by comparing the performance of a test developed under normality on data generated from t_8 and from logistic distributions, we can find out whether the test is sensitive to these families of distributions.

3.2 ID Distinguishability Relative to Exponentiality

Suppose that $T_1(X) = X$ and $\theta_1 = \mu$. Then the ME distribution in $\Omega_\mu = \{f(x|\mu): E(X|\mu) = \mu\}$ is the exponential distribution $f^*(x|\mu) = (1/\mu)\exp(-x/\mu)$. Three well-known families of distributions usually considered as alternatives to exponentiality are gamma with density

$$f_1(x|\lambda_1, \beta) = \lambda_1^\beta x^{\beta-1}\exp(-\lambda_1 x)/\Gamma(\beta), \quad \lambda_1 > 0, \quad \beta > 0;$$

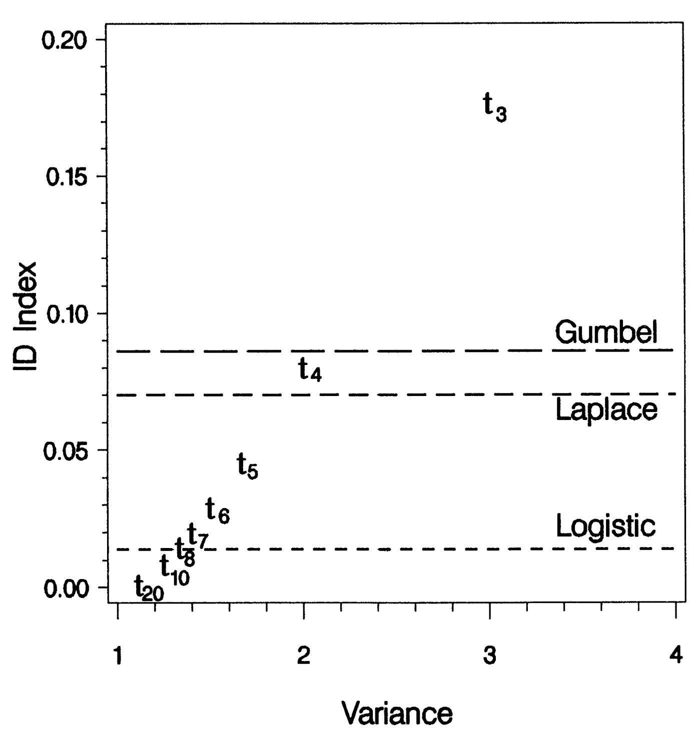

Figure 1. Information Indices of Distributions Relative to Normality.

Weibull with density

$$f_2(x \mid \lambda_2, \beta) = \lambda_2 \beta x^{\beta-1} \exp(-\lambda_2 x^{\beta}), \qquad \lambda_2 > 0, \qquad \beta > 0;$$

and lognormal with density

$$f_3(x \mid \lambda_3, \beta) = \frac{\beta}{x\sqrt{2\pi}} e^{-1/2\beta^2(\ln x - \lambda_3)^2},$$

$$-\infty < \lambda_3 < \infty, \qquad \beta > 0.$$

Here β is the shape parameter and λ_1, λ_2, and e^{λ_3} are the scale parameters for f_1, f_2, and f_3. Note that the lognormal density has been parameterized such that for a given value of β the lognormal shape is comparable with the shapes of gamma and Weibull distributions.

Letting $\lambda_1 = \mu\beta$, $\lambda_2 = [\Gamma(1 + 1/\beta)/\mu]^{\beta}$, and $\lambda_3 = \ln \mu - 1/(2\beta^2)$, for f_1, f_2, and f_3 makes these distributions members of Ω_μ. (In power-comparison studies, setting the mean of each distribution equal to 1 is a routine practice.)

Figure 2 shows the graphs of ID indices for these families as functions of the shape parameter. Observe that for any given $\beta \neq 1$, the gamma and Weibull distributions are ID distinguishable in Ω_μ, with $\mathrm{ID}(f_1 : f^* \mid \beta) \le \mathrm{ID}(f_2 : f^* \mid \beta)$; the equality holds only for $\beta = 1$ in which case, the two distributions are identical and exponential. Also in Figure 2 we note that for any given β the lognormal family is ID distinguishable with gamma and with Weibull. The ID indices of gamma, Weibull, and lognormal are increasing (decreasing) in β for $\beta > 1$ ($\beta < 1$).

In Figure 2 we observe that for any given value of β, discrimination between the gamma and exponential should be more difficult than between the Weibull and exponential. Thus for a given β, tests of exponentiality can be expected to show more power against the Weibull than the gamma. This explicates the results of many studies (e.g., Ebrahimi et al. 1992; Kochar and Gupta 1988; Lin and Mudholkar 1980) indicating that for a given value of the shape parameter β, tests of exponentiality show more power against the Weibull as compared with the gamma. But comparing powers of tests against gamma and Weibull based on the shape parameters does not reveal how sensitive a test is to the parametric family of the alternative. Note that gamma ($\beta = 3$) and Weibull ($\beta = 1.8$) are almost not ID distinguishable, $\mathrm{ID}(f_1 : f^* \mid \beta = 3) \approx \mathrm{ID}(f_2 : f^* \mid \beta = 1.8)$. Thus tests of exponentiality that are not sensitive to the parametric family of the alternatives can be expected to show about the same power against gamma ($\beta = 3$) and Weibull ($\beta = 1.8$). Comparing the powers of nonparametric statistics such as tests of exponentiality in terms of the ID indices of these distributions will reveal sensitivity of each testing procedure to the parametric families; see Section 5.2.

4. ID STATISTICS

In general, estimation of Kullback–Leibler function when f is an unknown continuous distribution is a difficult problem. An immediate consequence of the result (9) is that one can estimate the discrimination function between members of Ω_θ and f^* by estimating the individual entropies in (9). In (3) we have already seen that the ME is a function of $\boldsymbol{\theta}$. The entropies of many well-known distributions were given

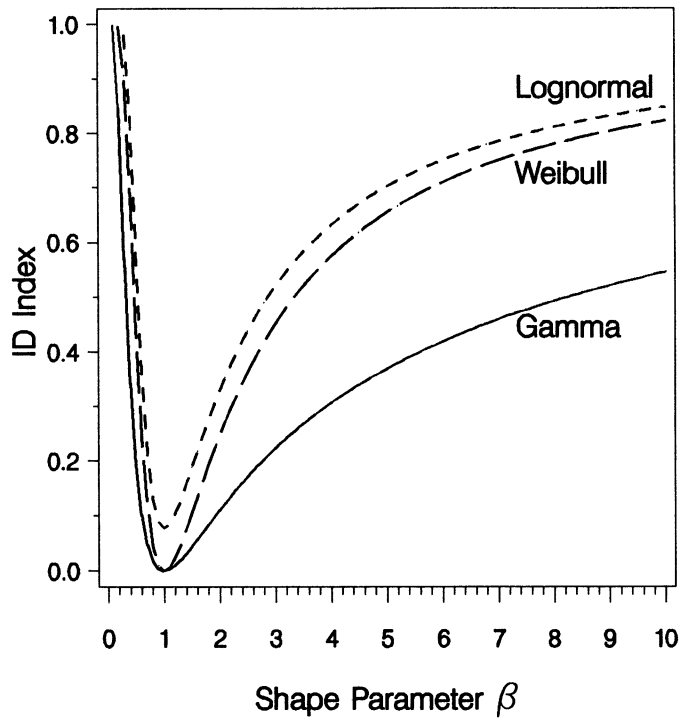

Figure 2. Information Indices of Lifetime Distributions Relative to Exponentiality: ----, lognormal; – – – –, Weibull; ——, Gamma.

in terms of their parameter(s) by Verdugo Lazo and Rathie (1978). When the value of $\boldsymbol{\theta}$ is determined *externally* according to the nature of the problem, or when it is hypothesized to be a particular value $\boldsymbol{\theta}_0$, then $H[f^*(x \mid \boldsymbol{\theta})]$ is known. In most statistical problems, however, $\boldsymbol{\theta}$ is determined *internally*, based on the data, and the estimation of $H[f^*(x \mid \boldsymbol{\theta})]$ becomes a parametric estimation problem. Given a sample of size n, $H[f^*(x \mid \boldsymbol{\theta})]$ may be estimated using a Bayesian or a sampling theory procedure. For example, plugging the maximum likelihood estimate (MLE) $\boldsymbol{\theta}_n$ of $\boldsymbol{\theta}$ under the ME model into (3) gives $H[f^*(x \mid \boldsymbol{\theta}_n)]$ as the MLE of $H[f^*(x \mid \boldsymbol{\theta})]$. The ME modeling approach requires estimating the entropy of the data-generating distribution according to a nonparametric procedure and comparing it with the entropy of $f^*(x \mid \boldsymbol{\theta})$. Various (e.g., Bayesian, MLE) nonparametric entropy estimation procedures are available (see Ebrahimi, Pflughoeft, and Soofi 1994 and Mazzuchi, Soofi, and Soyer 1993, 1994 for the latest developments). Using a parametric estimate $H[f(x \mid \boldsymbol{\theta}_n)]$ and a nonparametric entropy estimate $H_n[f]$ in (10) gives an ID statistic,

$$\mathrm{ID}_n(f : f^* \mid \boldsymbol{\theta}_n) = 1 - \exp\{H_n[f] - H[f^*(x \mid \boldsymbol{\theta}_n)]\}. \quad (12)$$

$\mathrm{ID}_n(f : f^* \mid \boldsymbol{\theta}_n)$ estimates the information discrepancy between the unknown underlying distribution of the data-generating process f and the information-theoretic estimated model $f^*(x \mid \boldsymbol{\theta}_n)$ in Ω_θ. If the nonparametric entropy estimate is the entropy of a nonparametric density estimate that satisfies the information constraints of Ω_θ, then $0 \le \mathrm{ID}_n(f : f^* \mid \boldsymbol{\theta}_n) \le 1$. It is easily seen that using a consistent estimator of $\boldsymbol{\theta}$ and a consistent estimator of $H[f]$ gives a consistent estimator of $\mathrm{ID}(f : f^* \mid \boldsymbol{\theta})$.

If an $\mathrm{ID}_n(f : f^* \mid \boldsymbol{\theta}_n)$ is not near zero, then a search for additional or different types of constraints is in order. Suppose that instead of the constraints $\{E[T_j(X|\theta)] = \theta_j, j = 1, \ldots m_1\}$, we use $\{E[T_k(X|\zeta)] = \zeta_k, k = 1, \ldots m_2\}$ in the ME calculation and find that $\mathrm{ID}(f{:}f^*|\zeta_1, \ldots, \zeta_{m2}) \le \mathrm{ID}(f{:}f|\theta_1, \ldots, \theta_{m1})$. Then, analogous to the information indices developed by Soofi (1992, 1994) for the discrete case, we can assess the incremental contribution of the new constraints by computing the *relative information* (RI) value,

$$\mathrm{RI}_n(\zeta_{1,n}, \ldots, \zeta_{m2,n}|\theta_{1,n}, \ldots, \theta_{m1,n}) = 1 - \frac{\mathrm{ID}_n(f{:}f^*|\zeta_{1,n}, \ldots, \zeta_{m2,n})}{\mathrm{ID}_n(f{:}f^*|\theta_{1,n}, \ldots, \theta_{m1,n})}. \quad (13)$$

If there is a substantial incremental contribution, then the new constraint should be used. When the original set of constraints $\{E[T_j(X)] = \theta_j, j = 1, \ldots m_1\}$ is included in the new set of constraints $\{E[T_k(X|\zeta)] = \zeta_k, k = 1, \ldots m_2\}$, then (13) may be interpreted as the *partial information* value of the additional constraints used in the ME computation.

A computed information index may be judged large or small in a number of ways. For a descriptive purpose, one may use a calibration scheme to judge the computed value. The graphs of the ID indices of a few distributions in Ω_θ are useful for calibration of an estimated ID index. For example, Figure 1 may be used to calibrate an ID index estimated for members of Ω_σ. It may also be used for calibrating ID indices computed for the location-scale families with a given variance, for the symmetric distributions with a given variance, and so on. For example an estimated ID index of .05 may be interpreted as the data-generating distribution being about as close to the normality as t_5, or an index of .17 may be interpreted as the distribution being about as far apart from normality as t_3. Similarly, Figure 2 may be used for calibrating ID indices estimated from lifetime data; see Example 5.2. Also, McCulloch (1989) provided an interesting calibration of the Kullback–Leibler function in terms of discrepancy between two Bernoulli distributions with parameters .5 and q. One may easily develop a similar calibration scheme for the ID index.

For inferential purposes, uncertainty associated with an ID_n index may be accounted for either in Bayesian paradigm or in sampling theory paradigm. In the Bayesian framework, the uncertainty associated with an ID index is induced via the variation of $\boldsymbol{\theta}$ and the class of densities used in the nonparametric entropy estimation (see Mazzuchi et al. 1993). In the frequentist framework, the source of uncertainty in ID_n is the sampling variations of the statistics θ_n and H_n. The sampling distribution of an ID_n may be examined using a bootstrap scheme. In the traditional sampling theory approach, $\mathrm{ID}_n(f{:}f^*)$ provides a test statistic for significance testing. Both Bayesian and the sampling theory approaches are computer intensive.

Remark 4.1. ID statistics may be developed for members of Ω_θ with a particular characteristic (see Sec. 5) or for a particular parametric family in Ω_θ. For example, to estimate the ID index between the gamma family and the exponential model in Ω_μ, we must first estimate the shape parameter β. Then, using the estimated value of β in Figure 2, we find an ID statistic. If the ID statistic is large, then we conclude that the data discriminate between the gamma and the exponential distributions. The uncertainty for such an estimated index may be accounted for in Bayesian or sampling theory framework. If the MLE of β is used, then the result will be the MLE of ID index between the gamma family and the ME model in Ω_μ.

An interesting situation arises when developing ID statistics for lifetime data. Let $F(x)$ represent the distribution of the lifetime, $S(x) = 1 - F(x)$, and denote the associated hazard function or failure rate by $r(x) = f(x)/S(x)$. Noting that $E[\ln S(X)] = -1$, we have

$$H_r[f] = 1 - E_f[\ln r(X)]. \quad (14)$$

Teitler, Rajagopal, and Ngai (1986) were first to observe this relationship between the entropy and the hazard rate. Thus we may estimate the entropy of the lifetime distribution via either the hazard rate function formula (14) or the usual density formula (1); that is, $H_f[f] = -E_f[\ln f(X)]$.

Let f_n denote an estimate of a density function and let r_n denote an estimate of the hazard rate function based on a set of observations $x_1, \ldots, x_n$. Then we may estimate the entropy of lifetime distribution via either

$$H_{r_n} = 1 - E_n[\ln r_n(X)] \quad (15)$$

or

$$H_{f_n} = -E_n[\ln f_n(X)], \quad (16)$$

where E_n is expectation with respect to f_n. Note that although $H_f[f] = H_r[f]$, the two procedures defined in (15) and (16) do not necessarily yield the same estimate. The following example demonstrates this point.

Example 4.1. Foldes, Rejto, and Winter (1980) proposed the hazard rate function estimator,

$$r_n(x) = \frac{f_n(x)}{S_n(x) + 1/n}, \quad \text{for} \quad x \ge 0. \quad (17)$$

That is, for any pair of $r_n(x)$ and $f_n(x)$ related by (17), we have

$$\begin{aligned} H_{r_n} &= 1 - E_n\{[\ln f_n(X) - \ln[S_n(X) + 1/n]\} \\ &= 1 + H_{f_n} + E_n\{\ln[S_n(X) + 1/n]\}. \end{aligned}$$

Thus

$$\begin{aligned} \Delta_n = H_{r_n} - H_{f_n} &= 1 + E_n\{\ln[S_n(X) + 1/n]\} \\ &\ge 1 + E_n\{\ln[S_n(X)]\}. \end{aligned}$$

Because $E_n\{\ln[S_n(X)\} = -1$, $\Delta_n \ge 0$. Therefore, the hazard rate function formula yields a larger entropy for the data than the density formula. Note that $\Delta_n \to 0$ as $n \to \infty$.

Nonparametric entropy estimators usually underestimate entropy. This is because the support of f is infinite but the estimates are computed on the finite support, generally over the range of the data. Usually, entropy estimators that yield larger estimates are less biased than those giving smaller estimates.

5. APPLICATION TO FAILURE DATA

In this section we show how $\mathrm{ID}_n(f{:}f^*)$ can be used for modeling failure data. Many parametric distributions (e.g., exponential, gamma, Weibull) used for modeling lifetime data are ME subject to various types of constraints; see Table 1. Nonparametric estimators of the hazard rate function and the density are available for many classes of lifetime distributions, such as increasing failure rate (IFR), decreasing failure rate (DFR), increasing failure rate in average (IFRA), and new better than used (NBU); see Boyles and Sameniago (1984) or Chang and Rao (1993) for the latest developments. In the rest of this article we focus on the exponential distribution and the IFR class. The procedures that we present here may be adapted for other ME models and other nonparametric classes of distributions for which estimators of the density and the hazard rate function are available.

5.1 ID Statistics for IFR Distributions

A distribution is said to be IFR if $r(x)$ is a nondecreasing function of x for all $x \geq 0$. The exponential distribution is the minimal IFR distribution. Let $T_1(X) = X$ and $\theta_1 = \mu$ and consider $\Omega_\mu = \{f(x\,|\,\mu): E(X) \leq \mu\}$. The ME distribution in Ω_μ is the exponential $f^*(x\,|\,\mu) = (1/\mu)\exp(-x/\mu)$. Barlow, Bartholomew, Bremner, and Brunk (1973) found the MLE's of the density and the hazard rate of IFR distributions as follows:

$$
\begin{aligned}
f_{n\mathrm{IFR}}(x) &= 0 && x < x_{(1)} \\
&= a_i(x) && x_{(i)} \leq x \leq x_{(i+1)}, \quad i = 1, 2, \ldots, n-1, \\
&= 0 && x_{(n)} < x
\end{aligned}
$$

and

$$
\begin{aligned}
r_{n\mathrm{IFR}}(x) &= 0 && x < x_{(1)} \\
&= r_i && x_{(i)} \leq x \leq x_{(i+1)}, \quad i = 1, 2, \ldots, n-1. \\
&= \infty && x_{(n)} < x
\end{aligned}
$$

Here

$$
\begin{aligned}
a_i(x) &= r_i \exp\left\{-\int_0^x r_n(u)\,du\right\} \\
&= r_i \exp\{-[(x - x_{(i)})r_i + s_i]\};
\end{aligned}
$$

$x_{(1)} < \cdots < x_{(n)}$ are the order statistics for a sample of size n; $s_i = \sum_{j=1}^{i} d_{j-1} r_{j-1}$;

$$
d_i = x_{(i+1)} - x_{(i)}; \tag{18}
$$

and

$$
r_i = \min_{i \leq t \leq n-1} \max_{1 \leq s \leq i} \frac{t - s + 1}{\sum_{j=s}^{t} (n-j) d_j}.
$$

The entropy of the MLE of IFR distributions may be computed via either the hazard rate formula (15) by

$$
H_{r_{n\mathrm{IFR}}} = 1 - \sum_{i=1}^{n-1} \ln r_i [\exp\{-s_i\} - \exp\{-(s_i + d_i r_i)\}]
$$

or the density formula (16) by

$$
\begin{aligned}
H_{f_{n\mathrm{IFR}}} = &-\sum_{i=1}^{n-1} [\ln r_i - s_i + x_{(i)} r_i] \\
&\times [\exp\{-s_i\} - \exp\{-(s_i + r_i d_i)\}] \\
&+ \sum_{i=1}^{n-1} [(r_i x_{(i)} + 1)\exp\{-s_i\} \\
&\qquad - (r_i x_{(i+1)} + 1)\exp\{-(s_i + r_i d_i)\}].
\end{aligned}
$$

Both $H_{r_{n\mathrm{IFR}}}$ and $H_{f_{n\mathrm{IFR}}}$ provide estimates for the entropy of IFR distributions.

The entropy of the ME model is $H[f^*(x\,|\,\mu)] = \ln \mu + 1$. Under IFR assumption, the entropy of the ME model is estimated by

$$
H_n[f^*(x\,|\,\mu)] = H[f^*(x\,|\,\mu_{n\mathrm{IFR}})] = \ln \mu_{n\mathrm{IFR}} + 1, \tag{19}
$$

where $\mu_{n\mathrm{IFR}}$ is the mean of the MLE of IFR distributions given by

$$
\begin{aligned}
\mu_{n\mathrm{IFR}} &= \int_0^\infty S_{n\mathrm{IFR}}(x)\,dx \\
&= x_{(1)} + \sum_{i=1}^{n-1} (1/r_i)[\exp\{-s_i\} - \exp\{-(s_i + r_i d_i)\}].
\end{aligned}
$$

In (19) two types of information are used: IFR property determined externally and the first moment $E(X) = \mu_{n\mathrm{IFR}}$ determined internally.

The use of $H[f^*(x\,|\,\mu_{n\mathrm{IFR}})]$ and the IFR entropy estimates $H_{r_{n\mathrm{IFR}}}$ or $H_{f_{n\mathrm{IFR}}}$ in (12) gives two ID statistics:

$$
\begin{aligned}
\mathrm{ID}_{n1} &= \mathrm{ID}(f{:}f^*\,|\; r_{n\mathrm{IFR}}, \mu_{n\mathrm{IFR}}) \\
&= 1 - \exp\{H_{r_{n\mathrm{IFR}}} - H[f^*(x\,|\,\mu_{n\mathrm{IFR}})]\}
\end{aligned} \tag{20}
$$

and

$$
\begin{aligned}
\mathrm{ID}_{n2} &= \mathrm{ID}(f{:}f^*\,|\; f_{n\mathrm{IFR}}, \mu_{n\mathrm{IFR}}) \\
&= 1 - \exp\{H_{f_{n\mathrm{IFR}}} - H[f^*(x\,|\,\mu_{n\mathrm{IFR}})]\}
\end{aligned} \tag{21}
$$

An ID_{n1} (ID_{n2}) statistic near zero indicates that the mean constraint reflects the information content of the data-generating mechanism, so that the underlying distribution may be satisfactorily represented by the exponential distribution, the least IFR case. A relatively large value of ID_{n1} (ID_{n2}) indicates that other constraint(s) are operational in the data-generating process, and thus the departure of the data-generating IFR distribution from exponentiality is pronounced. An ID_{n1} (ID_{n2}) = 1 corresponds to the degeneracy, the most IFR case (Barlow and Mendel 1992).

The ID statistics defined in (20) and (21) are consistent estimators of the information index of IFR distributions. The asymptotic values of these estimates under gamma and Weibull distributions are shown by the graphs of $\mathrm{ID}(f_1{:}f^*)$ and $\mathrm{ID}(f_2{:}f^*)$ in Figure 2 with $\beta \geq 1$. Thus when the sample is large, one may use Figure 2 for calibrating the estimated index.

Remark 5.1. Traditionally, the MLE of $H[f^*(x\,|\,\theta)]$ under the ME model $f^*(x\,|\,\theta)$ is used for estimating the entropy

Table 2. Monte Carlo Biases and Root Mean Squared Errors of Entropy Estimators and Information Discrimination Statistics for IFR Distributions (n = 20)

		Entropy estimators				ID statistics			
		$H_{r_{nIFR}}$		$H_{f_{nIFR}}$		ID_{n1}		ID_{n2}	
ID	H	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
				Exponential					
0	1.00	−.07	.24	−.15	.27	.05	.09	.11	.14
				Gamma					
.10	.89	−.16	.26	−.23	.31	.11	.14	.16	.19
.20	.79	−.17	.25	−.25	.32	.12	.15	.17	.19
.30	.65	−.17	.26	−.25	.32	.10	.14	.15	.17
				Weibull					
.10	.89	−.16	.25	−.24	.30	.13	.16	.19	.20
.20	.79	−.18	.24	−.27	.32	.12	.15	.18	.19
.30	.65	−.20	.26	−.29	.33	.10	.14	.15	.18

of the ME distributions (Arizono and Ohta 1989; Dudewicz and Van der Meulen 1981; Ebrahimi et al. 1992; Gokhale 1983; and Vasicek 1976). The MLE of $H[f^*(x|\mu)]$ under exponentiality without using the IFR property is given by $H[f^*(x|\bar{x})] = \ln \bar{x} + 1$. Combining $H[f^*(x|\bar{x})] = \ln \bar{x} + 1$ and IFR entropy estimates, $H_{r_{n\mathrm{IFR}}}$, $H_{f_{n\mathrm{IFR}}}$, gives two other ID statistics,

$$\begin{aligned}\mathrm{ID}_{n3} &= \mathrm{ID}(f{:}f^* \mid r_{n\mathrm{IFR}}, \bar{x}) \\ &= 1 - \exp\{H_{r_{n\mathrm{IFR}}} - H[f^*(x|\bar{x})]\}\end{aligned}$$

and

$$\begin{aligned}\mathrm{ID}_{n4} &= \mathrm{ID}(f{:}f^* \mid f_{n\mathrm{IFR}}, \bar{x}) \\ &= 1 - \exp\{H_{f_{n\mathrm{IFR}}} - H[f^*(x|\bar{x})]\}.\end{aligned}$$

Because $\bar{x}$ could be less than $\mu_{n\mathrm{IFR}}$, $\ln \bar{x} + 1$ may turn out to be less than the entropy of MLE of the IFR distributions. So the estimated information indices could turn out to be negative. In a Monte Carlo study, 32% of the simulation runs resulted in ID_{n3} being negative under exponentiality.

5.2 Monte Carlo Results

We performed a Monte Carlo study of the IFR entropy estimates and ID statistics based on the ID indices of the gamma and the Weibull with respect to exponentiality. We generated 2,000 samples of size n = 10, 20, 30, 40, 50 from gamma and Weibull distributions with ID indices ranging from 0 to .35. All computations were done using MINITAB on a mainframe. A GAUSS program also was used to check the computational accuracy for a number of the simulation parameters. Both programs produced virtually identical results. The MINITAB macros are available upon request from the authors.

Table 2 shows the bias and the root mean squared error (RMSE) of the entropy estimates over the simulation runs for n = 20 and ID = 0, .10, .20, .30. The pattern is evident: Both statistics are relatively insensitive to the two parametric families under consideration and also to departure from exponentiality. The hazard rate function procedure shows less bias and better MSE performance. Note that the difference between the estimated IFR entropies is

$$\Delta_{n\mathrm{IFR}} = 1 - \sum_{i=1}^{n-1} [(1 + s_i)\exp\{-s_i\} + (1 - s_i)\exp\{-(s_i + r_i d_i)\}],$$

which depends on the data. The simulation results indicate that $\Delta_{n\mathrm{IFR}} \geq 0$, suggesting that the hazard rate–based entropy estimator performs better than the corresponding density-based estimator.

Also shown in Table 2 are the bias and RMSE of the ID_{n1} and ID_{n2} statistics. Both statistics seem to be insensitive to the parametric families considered as well as to the departure from exponentiality. The sampling behavior of ID_{n1} appears to be better than that of ID_{n2}.

Table 3 gives the upper percentile points of ID_{n1} and ID_{n2} for selected values of n. These percentiles may be used for significance testing when ID_{n1} or ID_{n2} are viewed as traditional test statistics.

We also examined sensitivity of the powers of a number of tests of exponentiality to the parametric family of the

Table 3. Upper α Percentiles of Information Indices ID_{n1} and ID_{n2} Obtained by Monte Carlo Simulations

	α					
	.10		.05		.01	
n	ID_{n1}	ID_{n2}	ID_{n1}	ID_{n2}	ID_{n1}	ID_{n2}
3	.734	.847	.809	.889	.902	.994
5	.587	.702	.633	.741	.728	.810
10	.324	.426	.374	.467	.477	.573
15	.224	.302	.280	.356	.362	.444
20	.191	.257	.227	.292	.293	.351
30	.132	.180	.164	.204	.216	.270
50	.085	.113	.107	.131	.142	.162

alternative in the IFR class. Next we report a portion of this study. We report on ID_{n1} defined in (20) and on the following two tests:

$$T_{n1} = \sum_{i=1}^{n} - \ln \ln\left[1 - \frac{i}{n+1}\right] \frac{(n-i+1)d_i}{D}$$

and

$$T_{n2} = \sum_{i=1}^{n} \frac{(n+1)^3 i - 3(n+1)^2 i^2 + 2(n+1)i^3}{6} \cdot \frac{(n-i+1)d_i}{D},$$

where d_i is defined in (18) and $D = \sum_{i=1}^{n} (n-i+1)d_i$. Bickel and Doksum (1969) studied T_{n1}, and Klefsjo (1983) proposed T_{n2}. Both T_{n1} and T_{n2} reject exponentiality for large observed values.

Figure 3 shows the power functions of the ID_{n1}, T_{n1}, and T_{n2} against the gamma and the Weibull families as functions of the ID index for $n = 20$ and $\alpha = .10$. The upper 10th percentile of the Monte Carlo distribution of each statistic under exponentiality (ID = 0) was used as the critical value of the test. We observe that the power functions of ID_{n1} against the gamma and the Weibull are almost identical functions of the ID index. The power functions of T_{n1} against these alternatives are also almost identical functions of the ID index. Thus ID_{n1} and T_{n1} are not sensitive to the gamma and Weibull families of alternatives in the IFR class, whereas T_{n2} appears to be sensitive to these families. The power of T_{n2} against Weibull is higher than against gamma.

Note that ID_{n1} performs relatively well against both alternatives. Furthermore, ID_{n1} performs better in terms of the power than does a previously proposed entropy-based test of exponentiality that did not use the IFR property.

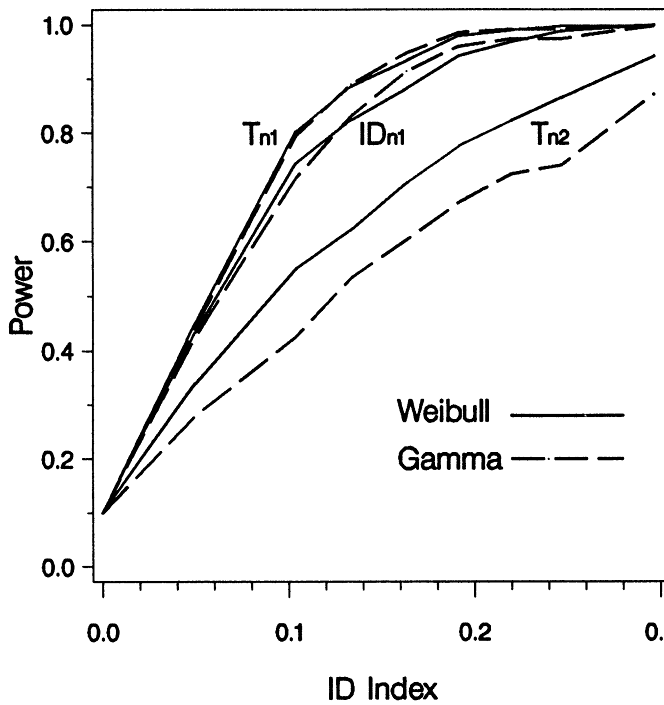

Figure 3. Sensitivity of Powers of Tests of Exponentiality to Parametric Families of the Alternatives in IFR Class.

5.3 Illustrative Examples

We now present information analyses of two well-known data sets.

Example 5.3: Information Analysis of Insulating Fluid Failure Data. The data shown in Table 4 are times (in minutes) to break down of an insulating fluid under various levels of voltage stress. Nelson (1975) indicated that "the purpose of the experiment was to estimate the distribution of time to breakdown at 20 kV. Also, the test was to assess whether time to breakdown has an exponential distribution." The analysis was done by *assuming* a Weibull model and estimating the common shape parameter. Nelson, in several papers, and Lawless (1982) have applied various estimation procedures to the data under the Weibull assumption and have estimated the shape parameter to be close to 1. This means that under the assumption that a model in the Weibull family is suitable for the data-generating distribution, exponentiality is plausible. Mazzuchi and Soyer (1992) used Nelson's result and analyzed the data under the *assumption of exponentiality* in the Bayesian framework.

Table 4 summarizes information quantities for this data obtained based on IFR assumption. The ID statistics are near zero for 32 kV, 34 kV, and 36 kV, clearly confirming exponentiality at these stress levels. At the other stress levels, the ID statistics are not so low, but because of the small number of observations we cannot use the graphs shown in Figure 2 as calibration for determining the level of departure from exponentiality. Making firm conclusions about the distribution at 26 kV and 28 kV is difficult with so few observations. Accounting for the sampling variations and comparing these ID statistics with percentiles shown in Table 3 indicate that the departures from exponentiality are not statistically significant. Our interpretation is that the first mean constraint seems satisfactory for the data-generating distribution, at least for a few stress levels. If the mean constraint is meaningful in the context of the data-generating mechanism, then the exponential distribution would adequately represent the data-generating distribution. An assessment of suitability of the mean (total life) constraint should be simpler and a more realistic task than determining plausibility of the Weibull family on an a priori ground.

Example 5.2: Information Analysis of Rat Survival Data. The data shown in Table 5 (p. 667) are survival times in weeks of 20 male rats that were exposed to a high level of radiation. This data has been analyzed under the *gamma assumption* by a number of authors, including Lawless (1982).

The information statistics for this data are shown in Table 5. We begin with the use of the ID statistic as a descriptive measure and use Figure 2 for calibration of the computed value .65. A comparison with the graphs in Figure 2 indicates that the computed index is about the same as the ID index of a Weibull distribution with $\beta \approx 5$ and is larger than the ID index of a gamma distribution with $\beta = 10$. We conclude

Table 4. Information Analysis of Insulating Fluid Failure Data

	26 kV	28 kV	30 kV	32 kV	34 kV	36 kV	38 kV
			Data				
	5.79	68.85	7.74	.27	.19	.35	.09
	1,579.52	108.29	17.05	.40	.78	.59	.39
	2,323.70	110.59	20.46	.69	.96	.96	.47
		426.07	21.02	.79	1.31	.99	.73
		1,067.60	22.66	2.75	2.78	1.69	.74
			43.40	3.91	3.16	1.97	1.13
			47.30	9.88	4.15	2.07	1.40
			139.07	13.95	4.67	2.59	2.38
			141.12	15.93	4.85	2.71	
			175.88	27.80	6.50	2.90	
			194.90	53.24	7.35	3.67	
				82.85	8.01	3.99	
				89.29	8.27	5.35	
				100.58	12.06	13.77	
				215.10	31.75	25.50	
					32.52		
					33.91		
					36.71		
					72.89		
			Mean				
Arithmetic	1,303.00	356.28	75.51	41.16	14.36	4.61	.92
MLE (IFR)	1,529.54	405.84	84.91	43.76	15.39	4.89	.92
			Nonparametric IFR entropy				
Hazard Rate Formula	6.70	6.51	5.15	4.75	3.71	2.51	.86
Density Formula	6.15	6.28	5.00	4.71	3.68	2.49	.65
			Parametric entropy				
Exponential	8.33	7.01	5.44	4.78	3.73	2.59	.98
			Information index relative to				
Exponential	.80	.39	.25	.03	.02	.07	.11

that the ID statistic relative to exponentiality is rather large and the discrepancy is pronounced. In order to account for the sampling variations, we use Table 3. We find that ID_{n1} as a test statistic is highly significant. Our interpretation is that perhaps there are other types of constraints operational in the data-generating distribution than the mean constraint imposed by Ω_μ.

At this point we include $E(\ln X)$ in addition to $E(X)$ in the ME calculation. In the class of distributions with constraints $E(X)$ and $E(\ln X)$, the ME model is gamma distribution (see Table 1). Based on the arithmetic mean and the geometric mean of the data, the gamma shape parameter is estimated as $\hat{\beta} \approx 10$. This is in the range of the estimates obtained by other authors using a number of methods. The ID index of data-generating IFR distribution relative to gamma is estimated as $\mathrm{ID}_{n1}(f:f_1^*) \approx .22$. Using (13), the incremental contribution of the additional constraint is RI $\approx 65\%$. Thus the additional constraint has substantially reduced the ID statistic relative to exponentiality, but the ID statistic relative to gamma is not yet near zero. Note that from Figure 2 we also obtain the parametric ID statistic between gamma and exponentiality as $\mathrm{ID}_n(f_1:f^*) \approx .54$, which indicates that the data can clearly discriminate between the two distributions.

Alternatively, we may consider a different set of constraints that does not include $E(X)$. We use $E(X^\beta)$, $\beta \neq 1$, and $E(\ln X)$, for which the Weibull distribution is the ME model (see Table 1). An estimate of β is needed for the ME calculations. Using the MLE $\hat{\beta} \approx 3.6$ gives the ID statistic for the data-generating IFR distribution relative to Weibull as $\mathrm{ID}_{n1}(f:f_2^*) \approx .23$. Again, the reduction in ID index due to the new constraint is RI = 65%. From Figure 2, we see that the parametric ID statistic between Weibull and exponentiality is $\mathrm{ID}_n(f_2:f^*) \approx .52$, which indicates that the data can clearly discriminate between the two distributions.

The estimated information index with respect to the Weibull ($\hat{\beta} \approx 3.6$) is about the same as the estimated information index with respect to the gamma ($\hat{\beta} \approx 10$). We conclude that a Weibull seems to be as plausible as a gamma for modeling the rat survival data. Which of the two types of constraints is more suitable, or what other types of constraints are operational in the data-generating process, ought to be determined by the scientists in the field.

6. CONCLUSIONS

In this article we have adopted the viewpoint that the probability law of data-generating process (i.e., likelihood function) may be much more complicated than can be completely described by one of the known parametric families of distributions on a priori grounds. In the ME modeling, the fact that a model is just an approximation to the underlying data generating distribution is *explicit,* and the sense

Table 5. Information Analysis of Rat Survival Data

Data									
152	152	115	109	137	88	94	77	160	165
125	40	128	123	136	101	62	153	83	69

	Mean	
MLE (IFR)	115.89	
Arithmetic	113.45	
Geometric	107.07	
	Nonparametric IFR entropy	
Hazard Rate Formula	4.71	
Density Formula	4.60	
	Parametric entropy	
Exponential	5.75	
Gamma ($\hat{\beta} \approx 10$)	4.96	
Weibull ($\hat{\beta} \approx 3.6$)	4.97	
	Nonparametric information index relative to:	
Exponential	.65	Relative Reduction
Gamma ($\hat{\beta} \approx 10$)	.22	RI_n = 55%
Weibull ($\hat{\beta} \approx 3.6$)	.23	RI_n = 55%
	Parametric information index relative to exponentiality	
Gamma ($\hat{\beta} \approx 10$)	.54	
Weibull ($\hat{\beta} \approx 3.6$)	.52	

of the approximation is also made explicit in the statement of the problem. The ME model approximates "typical" distributions in the class Ω_θ of the distributions that satisfy the constraints in the sense that the entropies of "typical" distributions in Ω_θ are "close" to the unique upper bound given by the maximum entropy $H[f^*(x|\theta)]$. From a practical viewpoint, it should be a simpler task to identify some types of moment constraints that the data-generating distribution may satisfy than to specify a complete formula for the data-generating distribution. The ME distributions include but are not limited to many of the known parametric families of distributions.

The explication of difference between the ME and the entropy of any distribution in Ω_θ, in terms of the Kullback–Leibler divergence between the respective densities, provides a formal justification for the practice of measuring information discrepancy in terms of the entropy difference. This result provides the basis for introducing the concept of information distinguishability as a theme that unifies the sparse literature on the topic. The graphs of ID indices of known families of distributions with respect to normality and to exponentiality are helpful for visual comparison of information disparities between distributions and for developing intuition about distributions that are and those that are not ID distinguishable. These types of graphs provide useful diagnostics for planning robustness and power studies and performing sensitivity analysis over the space of densities in Ω_θ in a systematic manner. These graphs are also useful for calibration of statistics that measure information discrepancy.

The equivalence between the Kullback–Leibler function involving an ME and the entropy difference greatly facilitates developing ID statistics for assessing suitability of specified constraints for describing a data-generating distribution. Relative information indices of a set of new constraints are instrumental in a search for undiscovered constraints. Taking uncertainty about the information indices in the sampling theory approach is straightforward but computer intensive. Developing Bayesian methods for inference on the magnitudes of ID statistics is delicate but feasible. Preliminary results reported by Mazzuchi et al. (1993, 1994) in a number of Bayesian meetings are quite promising.

The application to analysis of failure data drew our attention to entropy estimation based on the hazard rate function. The superior sampling performance of ID statistics computed based on the MLE of the hazard rate function of the IFR distributions underscores the usefulness of tailoring information statistics for specific types of data. The Monte Carlo study on the sensitivity of tests of exponentiality in the IFR class based on ID indices clearly identified the tests that are and the tests that are not sensitive to the parametric families of alternative. This type of studies are helpful for exploring how nonparametric a nonparametric method really is.

The information analysis of the insulating fluid failure data based on the simple assumption of the mean constraint confirms the result found in the literature based on the more stringent assumption of the Weibull model for the data-generating distribution and then inferring that the shape parameter is unity. The information analysis of the rat survival data shows that the mean constraint does not satisfactorily describe the data-generating distribution. But addition of the $E(\ln X)$ constraint to the $E(X)$ constraint (i.e., gamma ME model) and consideration of a new set of constraints, $E(\ln X)$ and $E(X^\beta)$, $\beta \neq 1$, (i.e., Weibull ME model) produced almost equal ID statistics. This analysis casts doubt on the use of gamma distribution for the data-generating distribution without further justification.

[Received December 1992. Revised June 1994.]

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *2nd International Symposium on Information Theory,* eds. B. N. Petrov and F. Csaki, Budapest: Akademiai Kiado.

Arizono, I., and Ohta, H. (1989), "A Test for Normality Based on Kullback–Leibler Information," *The American Stastician,* 34, 20–23.

Barlow, R. E., Bartholomew, J. M., Bremner, J. M., and Brunk, H. D. (1973), *Statistical Inference Under Order Restrictions,* New York: John Wiley.

Barlow, R. E., and Mendel, M. B. (1992), "De Finetti-Type Representations for Life Distributions," *Journal of the American Statistical Association,* 87, 1116–1122.

Bickel, P. J., and Doksum, K. A. (1969), "Tests for Monotone Failure Rate Based on Normalized Spacings," *Annals of Mathematical Statistics,* 40, 1216–1235.

Boyles, R. A., and Sameniago, F. J. (1984), "Estimating a Survival Curve When New is Better Than Used," *Operations Research,* 32, 732–740.

Chandra, M., De Wet, T., and Singpurwalla, N. D. (1982), "On the Sample Redundancy and a Test for Exponentiality," *Communication in Statistics, Part A—Theory and Methods,* 11, 429–438.

Chang, M. N., and Rao, P. V. (1993), "Improved Estimation of Survival Functions, in the New-Better-Than-Used Class," *Technometrics,* 35, 192–204.

Csiszar, I. (1975), "I-Divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability,* 3, 146–158.

Dudewicz, E. J., and Van Der Meulen, E. C. (1981), "Entropy-Based Tests of Uniformity," *Journal of the American Statistical Association,* 76, 967–974.

Ebrahimi, N., Habibullah, M., and Soofi, E. S. (1992), "Testing Exponentiality Based on Kullback–Leibler Information," *Journal of the Royal Statistical Society,* Ser. B., 54, 739–748.

Ebrahimi, N., Pflughoeft, K., and Soofi, E. S. (1994), "Two Measures of Sample Entropy," *Statistics and Probability Letters,* 20, 225–234.

Foldes, A., Rejto, L., and Winter, B. B. (1980), "Strong Consistency of Nonparametric Estimators for Randomly Censored Data," in *Handbook of Statistics 7,* ed. C. P. R. Krishiniahia, Amsterdam: North-Holland.

Gokhale, D. V. (1983), "On Entropy-Based Goodness-of-Fit Tests," *Computational Statistics and Data Analysis,* 1, 157–165.

Gokhale, D. V., and Kullback, S. (1978), *The Information in Contingency Tables,* New York: Marcel Dekker.

Jaynes, E. T. (1982), "On the Rationale of Maximum-Entropy Methods," *Proceedings of the IEEE,* 70, 939–952.

——— (1968), "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics,* SSC-4, 227–241.

——— (1957), "Information Theory and Statistical Mechanics," *Physical Review,* 106, 620–630.

Joe, H. (1989), "Relative Entropy Measures of Multivariate Dependence," *Journal of the American Statistical Association,* 84, 157–164.

Kapur, J. N. (1989), *Maximum-Entropy Models in Science and Engineering,* New York: John Wiley.

Klefsjo, B. (1983), "Some Tests Against Aging Based on the Total Time on Test Transform," *Communication in Statistics—Theory Method,* 12, 907–927.

Kochar, S. C., and Gupta, R. P. (1988), "A Monte Carlo Study of Some Asymptotically Optimal Tests of Exponentiality Against Positive Aging," *Communication in Statistics—Simulations,* 17, 803–811.

Kullback, S. (1959), *Information Theory and Statistics,* New York: John Wiley.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data,* New York: John Wiley.

Lin, C. C., and Mudholkar, G. S. (1980), "A Test of Exponentiality Based on the Bivariate F Distribution," *Technometrics,* 22, 79–82.

Mazzuchi, T. A., Soofi, E. S., and Soyer, R. (1993), "Information Diagnostics for Bayesian Reliability Analysis," in *Proceedings of the Section on Bayesian Statistical Science, American Statistical Association,* pp. 173–178.

——— (1994), "Lifetime Modeling Via Information Distinguishability," in *Proceedings of the Section on Bayesian Statistical Science, American Statistical Association,* (to appear).

Mazzuchi, T. A., and Soyer, R. (1992), "A Dynamic General Linear Model for Inference From Accelerated Life Tests," *Naval Research Logistics,* 39, 757–773.

McCulloch, R. E. (1989), "Local Model Influence," *Journal of the American Statistical Association,* 84, 473–478.

Nelson, W. B. (1975), "Analysis of Accelerated Life Test Data—Least Squares Methods for the Inverse Power Law Model," *IEEE Transactions on Reliability,* R-24, 103–107.

Shannon, C. E. (1948), "A Mathematical Theory of Communication," *Bell System Technical Journal,* 27, 379–423; 623–656.

Shore, J. E., and Johnson, R. W. (1980), "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory,* IT-26, 26–37.

Soofi, E. S. (1994), "Capturing the Intangible Concept of Information," *Journal of the American Statistical Association,* 89, 1243–1254.

——— (1992), "A Generalizable Formulation of Conditional Logit With Some Diagnostics," *Journal of the American Statistical Association,* 87, 812–816.

Soofi, E. S., and Gokhale, D. V. (1991), "Minimum Discrimination Information Estimator of the Mean With Known Coefficient of Variation," *Computational Statistics & Data Analysis,* 11, 165–177.

Teitler, S., Rajagopal, A. K., and Ngai, K. L. (1986), "Maximum Entropy and Reliability Distributions," *IEEE Transactions on Reliability,* R-35, 391–395.

Verdugo Lazo, A. C. G., and Rathie, P. N. (1978), "On the Entropy of Continuous Probability Distributions," *IEEE Transactions on Information Theory,* IT-24, 120–122.

Vasicek, O. (1976), "A Test for Normality Based on Sample Entropy," *Journal of the Royal Statistical Society,* Ser. B, 38, 54–59.

Zellner, A. (1984), *Basic Issues in Econometrics,* Chicago: University of Chicago Press.

Zellner, A., and Highfield, R. A. (1988), "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions," *Journal of Econometrics,* 37, 195–209.