ON STOCHASTIC COMPARISONS OF ENERGY
FUNCTIONS WITH APPLICATIONS

BRODERICK O. OLUYEDE

(Received 1 March 2001)

ABSTRACT. We develop simple methods for the stochastic comparisons of informational energy functions. We introduce modified informational energy functions and uncertainty of parameter functions are introduced for models with realistic parameter spaces. We present inequalities, comparisons, and applications including test procedures for testing the equality of informational energy functions. Some illustrative examples are also presented.

1. Introduction. The use of informational energy for stochastic comparisons and inferences is of tremendous practical importance. There are several measures of information content of an experiment, among which are the Shannon capacity introduced by Lindley [5], and the generalized information functions given by Nayak [7]. In a recent work, Morales et al. [6] developed test procedures based on entropy and divergent type statistics as an application of statistical information theory. Energy functions are measures of dispersion of distributions, that varies monotonically with dispersive order, and as such testing for the equality of energy functions can be thought of as non- or semi-parametric testing on dispersion.

The main objective of this paper is to investigate and compare informational energies including certain modified version with regards to the notion of affinity concerning several such functions. This is particularly important and is motivated by problems in areas such as quality control or analysis of variance, and in the measurement of the information content of a statistical experiment and the uncertainty of parameter sets.

Let $\{P_\theta : \theta \in \Theta\}$, be a family of probability density functions associated with respect to a σ-finite measure λ. Consider the likelihood $p_\theta = dP_\theta/d\lambda$ and a sequence of observations $X = (X_1, \ldots, X_n)$ from P_{θ_0}, where P_{θ_0} is from a set $\{P_\theta : \theta \in \Theta\}$ of distributions. Assume that the function P_θ is continuous. Also the mapping $\theta \to p_\theta$ is almost surely (a.s.) upper semicontinuous, separable random process, and the energy function $e(\theta) = E[p_\theta(X)]$ exists and is finite on the parameter space.

The purpose of this paper is to obtain inequalities and compare informational energy functions, reliability and uncertainty measures for weighted distributions. Section 2 contains some basic definitions and utility notions. Also, energy functions are compared. In Section 3, some connections and results on likelihood and informational energy are presented. The results are used to construct test for equality of informational energies. Section 4 is concerned with estimates, test statistics and

procedures based on informational energy functions. Some applications and examples
are given in Section 5. This paper concludes with a discussion in Section 6.

2. Some definitions, utility notions, and comparisons. In this section, we present
some definitions and useful notions. Let $\mathcal{F}$ be the set of absolutely continuous distri-
bution function satisfying

$$H(0) = 0, \qquad \lim_{x \to \infty} H(x) = 1, \qquad \sup\{x : H(x) < 1\} = \infty. \tag{2.1}$$

Note that if the mean of a random variable in $\mathcal{F}$ is finite, it is positive.

The informational energy associated with P_θ is given by

$$e(\theta) = \int_R p_\theta^2(x) d\lambda(x). \tag{2.2}$$

DEFINITION 2.1. Let f and g be two probability density functions. The distance
between f and g is

$$D_2(f,g) = \left(\int (f^{1/2} - g^{1/2}) d\lambda \right)^{1/2}. \tag{2.3}$$

The Hellinger type integral of order $1/2$ is given by

$$B_2(f,g) = \int (fg)^{1/2} d\lambda. \tag{2.4}$$

Note that,

$$B_2(f,g) = 1 - \frac{1}{2(D_2^2(f,g))}, \tag{2.5}$$

$$1 - B_2(f,g) \le D_2^2(f,g) \le \left(1 - B_2^2(f,g)\right)^{1/2}.$$

DEFINITION 2.2. Let u and v be two nonnegative bounded real functions on $\mathbb{R}$.
We say u is exponentially dominated by v if for each $\epsilon \in (0,1)$, there exist $A(\epsilon) < \infty$
such that

$$u(x) \le A(\epsilon)v(x)^{1-\epsilon} \quad \forall x. \tag{2.6}$$

If u and v are exponentially dominated by each other, they are said to be exponentially
equivalent.

The usefulness of the above definition is in the comparisons of small values of
bounded nonnegative functions u and v, respectively.

Let f and g be two nonnegative functions, possibly probability density functions
that are integrable with respect to a σ-finite measure λ, and define

$$e(f,g) = \int \max(f^2, g^2) d\lambda. \tag{2.7}$$

See Bradt and Karlin [4] for a related comparison of dichotomous experiments. We
have the following properties:

(1) $e(f,g) = e(g,f)$,
(2) $e(f,g) = e(f)$, if and only if $f = g$,

(3) for $c \geq 0$, $e(cf,g) = ce(f,g)$, and $0 \leq ce(f,g) \leq \infty$,

(4) for $\delta_1 \leq \delta_2$, $e(\delta_1 f, g) \leq e(\delta_2 f, g)$.

THEOREM 2.3. *Let f and g be probability density functions (pdf), then*

$$0 \leq \{D_k(f,g)\}^k \leq D_1(f,g), \tag{2.8}$$

where $D_k(f,g) = (\int (f^{1/k} - g^{1/k}) d\lambda)^{1/k}$, $k \geq 1$.

PROOF. Let

$$A = \{x : f < g\}, \qquad B = \{x : f \geq g\}. \tag{2.9}$$

Then we have

$$\int_A (g(x) - f(x)) d\lambda(x) \geq \int_A (g^{1/k}(x) - f^{1/k}(x)) d\lambda(x),$$

$$\int_B (f(x) - g(x)) d\lambda(x) \geq \int_B (f^{1/k}(x) - g^{1/k}(x))^k d\lambda(x). \tag{2.10}$$

Consequently,

$$\int_{\mathbb{R}} (f(x) - g(x)) d\lambda(x) \geq \int_{\mathbb{R}} |(f^{1/k}(x) - g^{1/k}(x))|^k d\lambda(x), \tag{2.11}$$

and the result follows. $\qquad \square$

The next result compares the informational energies $e(f)$ and $e(g)$.

THEOREM 2.4. *Let*

(1) $H_1(f,g,c) = \min(f^2(x)/g^2(x) - c^2, 0)$, *and*

(2) $H_2(f,g,c) = \min(g^2(x)/f^2(x) - c^2, 0)$.

Suppose that $P_f(x : g(x) = 0) = P_g(x : f(x) = 0)$, then $e(cf,g) \leq ce(f,g)$ if and only if $E_{g^2}\{H_1(f,g,c)\} \leq E_{f^2}\{H_2(f,g,c)\}$.

PROOF. Note that

$$e(f,cg) = \int_{\{x:c^2 g^2(x) \leq f^2(x)\}} f^2(x) d\lambda(x) + \int_{\{x:c^2 g^2(x) > f^2\}} g^2(x) d\lambda(x). \tag{2.12}$$

Similarly,

$$e(cf,g) = \int_{\{x:c^2 f^2(x) > g^2(x)\}} c^2 f^2(x) d\lambda(x) + \int_{\{x:c^2 f^2(x) \leq g^2(x)\}} g^2(x) d\lambda(x). \tag{2.13}$$

Note that,

$$e(cf,g) - e(f,cg) = \int_{\{x:c^2 g^2(x) > g^2(x)\}} \{c^2 g^2(x) - f^2(x)\} \lambda(x)$$

$$- \int_{\{x:c^2 f^2(x) > g^2(x)\}} \{c^2 f^2(x) - g^2(x)\} d\lambda(x)$$

$$= \int_{\{x:c^2 g^2(x) > g^2(x)\}} \left\{ \frac{f^2(x)}{g^2(x)} - c^2 \right\} g^2(x) d\lambda(x)$$

$$-\int_{\{x:c^2g^2(x)>g^2(x)\}}\left\{\frac{g^2(x)}{f^2(x)}-c^2\right\}f^2(x)d\lambda(x)$$

$$= E_{g^2}\{H_1(f,g,c)\}-E_{f^2}\{H_2(f,g,c)\}.$$

$$(2.14)$$

Consequently,

$$e(cf,g)\le ce(f,g) \qquad (2.15)$$

if and only if

$$E_{g^2}\{H_1(f,g,c)\}\le E_{f^2}\{H_2(f,g,c)\}. \qquad (2.16)$$
□

Theorem 2.5. *Suppose f is exponentially dominated by g, and $\{f_n\}_{n\ge1}$ and $\{g_n\}_{n\ge1}$, are sequences of bounded functions, then*

(1) $e(f)=\int f^2d\lambda(x)$ *is exponentially dominated by $e(g)$, and*

(2) $\lim_{k\to\infty}\sup\{e(g_n)^{1/k}-e(f_n)^{1/k}\}\le0$.

Proof. (1) Let $f^*=f^2(x)$ and $g^*=g^2(x)$, and apply Jensen's inequality to the concave function $y\mapsto y^{1-\epsilon}$ to obtain

$$\begin{aligned}
e(f)&=\int f^*(x)d\lambda(x)\\
&\le\int B(\epsilon)(g^*)^{1-\epsilon}d\lambda(x)\\
&\le B(\epsilon)\left(\int(g^*)(x)d\lambda(x)\right)^{1-\epsilon}\\
&=B(\epsilon)(e(g))^{1-\epsilon},
\end{aligned} \qquad (2.17)$$

where $B(\epsilon)=A^2(\epsilon)$ and $A(\epsilon)$ is given in Definition 2.2.

(2) Note that $\{f_n^*\}_{n\ge1}$ and $\{g_n^*\}_{n\ge1}$, are bounded sequences, so there exists a convergent subsequence such that $e(f)=\lim_{j\to\infty}e(f_{n_j})$ and $e(g)=\lim_{j\to\infty}e(g_{n_j})$.

Let $e(g_n)^{1/k}=N_k$ and $e(f_n)^{1/k}=M_k$, then

$$\lim_{k\to\infty}e(g_n)^{1/k}=\lim_{k\to\infty}N_k=N, \qquad \lim_{k\to\infty}e(f_n)^{1/k}=\lim_{k\to\infty}M_k=M. \qquad (2.18)$$

Consequently, $N\le C(\epsilon)M^{1-\epsilon}$ for every $\epsilon\in(0,1)$, and the result follows. □

3. Informational energy and likelihood. In this section, the connection between likelihood function and the informational energy function is established. Consider the function given by

$$g_n(X,\theta)=\frac{1}{n}\sum_{i=1}^n p_\theta(X_i), \qquad g_n(X,A)=\inf_{\theta\in A}g_n(X,\theta), \quad A\in\Theta. \qquad (3.1)$$

Also, let D_θ be the set of all compact sets $A\subset\Theta$ containing θ in their interior. Furthermore, we assume that for every $\theta\in\Theta$, there exists $A\subset D_\theta$ such that for at least one n on the set energy rate

$$e_n(A)=E_{\theta_0}(g_n(X,A)) \qquad (3.2)$$

is finite. Note that $e_n(A)\le e(\theta)<\infty$ for every $\theta\in A$.

If $e_1(\theta) > 0$, then as $n \to \infty$

$$e_n(A) \uparrow e(A) \equiv \sup_n e_n(A), \tag{3.3}$$

$$g_n(X, A) \longrightarrow e(A) \quad \text{a.s.} \tag{3.4}$$

Clearly, $e(A)$ is the informational energy about the unknown parameter in the set A.

It is clear that these results can be formulated to give the set entropy function. Consider the loglikelihood function

$$h_n(X, \theta) = -\frac{1}{n} \sum_{i=1}^{n} \log\left(p_\theta(X_i)\right), \tag{3.5}$$

then

$$E_{\theta_0}\left(h_n(X, A)\right) = H_n(A), \tag{3.6}$$

where $H_n(A)$ is the set entropy function defined for an open or compact set $A \subset \Theta$, and

$$h_n(X, A) = \inf_{\theta \in A} h_n(X, \theta), \quad A \in \Theta. \tag{3.7}$$

It follows that if $H_1(\theta) > -\infty$ and $n \to \infty$, then $H_n(A) \uparrow H(A) \equiv \sup_n H_n(A)$ and $h_n(X, A) \to H(A)$ a.s., where $H(A)$ is the uncertainty as to whether the unknown parameters are in the set A.

4. Test procedures based on informational energy. In this section, statistical inference via informational energy function is developed. Estimates and test procedures are presented. Let $X_{11}, X_{12}, \ldots, X_{1n_i}$, be independent random samples with distribution functions F_i, $i = 1, 2$, respectively. An estimate of the informational energy function $e(F_j) = \int (f_j^2(x)) dx$ proposed by Bhattacharyya and Roussas [3] is given by

$$\tilde{e}(F_j) = \int \left(\hat{f}_j^2(x)\right) dx, \tag{4.1}$$

where

$$\hat{f}_j(x) = [n_j h_j]^{-1} \sum_{i=1}^{n_j} K\left(\frac{x - X_{ji}}{h_j}\right), \tag{4.2}$$

where h_j is a bandwidth, and K is a known symmetric and bounded function probability density function such that $\lim_{y \to \infty} y K(y) = 0$. Ahmad [1] proposed the estimate

$$e(\hat{F}_j) = [n_j^2 h_j]^{-1} \sum_{r=1}^{n_j} \sum_{s=1}^{n_j} K\left(\frac{X_{jr} - X_{js}}{h_j}\right). \tag{4.3}$$

Bhattacharyya and Roussas' estimate [3] is a special case of Ahmad's estimate [1], since

$$\tilde{e}(F_j) = [n_j^2 h_j]^{-2} \sum_{r=1}^{n_j} \sum_{s=1}^{n_j} K^{(2)}\left(\frac{X_{jr} - X_{js}}{h_j}\right), \tag{4.4}$$

where $K^{(2)}(y)$ is the convolution of $K(y)$ with itself. See Ahmad and Kochar [2] for details.

A test statistics for testing $H_0 : e(F_1) = e(F_2)$ is given by

$$T(F_1, F_2) = \int \left(e(\hat{F}_1) - e(\hat{F}_2) \right)^2 dx. \tag{4.5}$$

In the case $h_1 = h_2 = h$, $f_1 = f$ is a fixed probability density function and g is a function such that $f_2 = f + \gamma g$ is a probability density function for sufficiently small $|\gamma|$, the α-level test rejects H_0 if $T(F) = T(f) > t_f$, where

$$P_{H_0}(T(f) > t_f) = \alpha, \tag{4.6}$$

and t_f is the α-level critical point of the distribution of $T(F)$ under the null hypothesis $H_0 : \gamma = 0$, that is, $e(F_1) = e(F_2)$.

Let $H_1 = H_1(\gamma)$ denote the alternative hypothesis that $\gamma = \delta/\sqrt{nh}$, $\delta \neq 0$, then

$$\pi(\delta) = \lim_{n \to \infty} P_{H_1}(T(f) > t_f) \longrightarrow 1, \tag{4.7}$$

as $|\delta| \to 0$ provided $0 < |\delta| < \infty$. Also, $\alpha < \pi(\delta) < 1$ for $0 < |\delta| < \infty$.

THEOREM 4.1. *Let $\hat{\theta}$ be the maximum likelihood estimator of θ. If $B = (b_1, b_2, \ldots, b_k)^T$ and $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^T$, where $b_i = \partial e(\theta)/\partial \theta_i$ and $\sigma^2(\theta) = B^T I_F^{-1}(\theta) B > 0$, then*

$$\sqrt{n}(e(\hat{\theta}) - e(\theta)) \xrightarrow{L} N(0, B^T I_F^{-1} B), \tag{4.8}$$

as $n \to \infty$, where $I_F(\theta)$ is the Fisher information matrix.

PROOF. By the asymptotic normality of $\sqrt{n}(\hat{\theta} - \theta)$ and a Taylor's expansion of $e(\theta)$ around θ, we obtain the desired result. □

THEOREM 4.2. *If $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)^T$, then for every $\theta \in \Theta$*

$$\lim_{n \to \infty} g_n(X, \theta) = e(\theta) \quad a.s. \tag{4.9}$$

PROOF. The result follows from (3.4). □

The results above can be used for statistical inference. Now consider the hypothesis, $H_0 : e(\theta) = e(\theta_0)$ against $H_0 : e(\theta) > e(\theta_0)$, where $e(\theta_0)$ is a specified value of the population informational energy. An appropriate test statistics for testing the hypothesis is given by

$$T^* = T^2 I_{(0,\infty)}(T), \tag{4.10}$$

where

$$T = \frac{\sqrt{n}(e(\hat{\theta}) - e(\theta))}{\sigma^2(\hat{\theta})}. \tag{4.11}$$

The statistic T has in the limit the standard normal distribution so that T^2 has a chi-square distribution with one degree of freedom.

A size α-test will reject H_0 if $T^* > \chi^2_{1,2\alpha}$. This follows from the fact that

$$\lim_{n\to\infty} P_{H_0}(T^* > C) = \lim_{n\to\infty} P_{H_0}(T^2 I_{(0,\infty)}(T) > C) = \frac{1}{2} P_{H_0}(\chi^2_{(1)} > C) \qquad (4.12)$$

if $C > 0$, and 1 if $C \le 0$, where $\chi^2_{(1)}$ denotes a random variable having a chi-square distribution with one degree of freedom.

A test of equality of several informational energies, that is,

$$H_0 : e(\theta_1) = e(\theta_2) = \cdots = e(\theta_k), \qquad (4.13)$$

rejects H_0 if $T_1 > C$, where

$$T_1 = \sum_{i=1}^k \left\{ \frac{e(\hat{\theta}_i) - \delta(\hat{\theta}_i)}{\sqrt{\sigma^2(\hat{\theta}_i)/n_i}} \right\}^2, \qquad (4.14)$$

$\delta(\hat{\theta}_i) = (\sum_{i=1}^k e(\hat{\theta}_i)/[\sigma^2(\hat{\theta}_i)/n_i])/\sum_{i=1}^k [\sigma^2(\hat{\theta}_i)/n_i]$, and $\sigma^2(\theta) = B^T I_F^{-1}(\theta)B > 0$, and C is chosen such that

$$P_{H_0}(T_1 > C) = \alpha. \qquad (4.15)$$

The statistic T_1 has in the limit as $n = \sum_{i=1}^k n_i$ goes to infinity the chi-square distribution with $k-1$ degrees of freedom. Consequently, the null hypothesis is rejected at level α if $T > \chi^2_{k-1;\alpha}$.

5. Applications. Let $e(f)$ and $e(g)$ be the informational energies associated with the distribution functions F and G, respectively. In this section, we present some applications and some examples of the results presented in earlier sections.

Confidence intervals for $e(\theta)$ can be readily obtained and is given by

$$e(\hat{\theta}) \pm \frac{c_{\alpha/2}\sigma(\hat{\theta})}{n^{1/2}}, \qquad (5.1)$$

and a nonconservative sample size for a prescribed error ϵ and a risk α is

$$n^* = \left[\frac{\sigma^2(\hat{\theta})c_{\alpha/2}^2}{\epsilon^2} \right] + 1, \qquad (5.2)$$

where $\sigma^2(\hat{\theta})$ is given in Section 4, where $c_{\alpha/2}$ is the critical point of the standard normal distribution at the significance level $\alpha/2$ and [] the greatest integer function.

(1) Normal Distribution. The informational energy for the normal distribution is given by

$$e(f(\mu,\sigma)) = \int_{\mathbb{R}} f^2(x;\mu,\sigma)dx = (\pi)^{-1/2}(2\sigma)^{-1}. \qquad (5.3)$$

Clearly, $e(f)$ is a bijective function of σ. If σ_F and σ_G are the standard deviations of the distribution functions F and G, respectively, then $e(f) \ge e(g)$ if and only if $\sigma_G \ge \sigma_F$.

(2) Let

$$f(x;\beta) = \begin{cases} 2(\pi)^{-1/2}\beta^{-1}e^{-(x/\beta)} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

The corresponding weighted pdf $g(x;\beta) = W(x)f(x;\beta)/E(W(X))$ with $W(x) = x$ is given by

$$g(x;\beta) = \begin{cases} 2x\beta^{-2}e^{-(x/\beta)} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

On applying Theorem 2.4, for $\beta \geq c \geq 1$, we obtain $e(cf,g) \leq e(cg,f)$.

(3) The following result establishes the relation between informational energy and dispersive ordering of distributions. Let X and Y be two random variables with distribution functions F and G, respectively, and corresponding quantile functions F^{-1} and G^{-1}. The distribution function F is said to be less dispersive than G, (Parzen [8]) denoted by $F \overset{d}{<} G$ if

$$F^{-1}(u) - G^{-1}(v) \leq G^{-1}(u) - G^{-1}(v), \tag{5.6}$$

for $0 < v < u < 1$. When F^{-1} and G^{-1} are differentiable, this definition is equivalent to

$$g(G^{-1}(u)) \leq f(F^{-1}(u)), \quad 0 < u < 1. \tag{5.7}$$

Consequently, $F \overset{d}{<} G$ implies that $e(f) \geq e(g)$, whenever the densities exit.

6. Discussion. In this paper, inequalities and the use of the informational energy for statistical comparisons and inferences in terms of uncertainty of parameters and parameter sets is developed. For the purpose of comparisons an intuitive grasp of notions involving informational energy functions follows by noting that the scale parameters for the distributions are ordered. Non-parametric and parametric estimates are presented. See references therein. Procedures for testing for homogeneity of informational energy are obtained and implemented.

In the discrete setting where X and Y are random variables with joint probability distribution p_{ij}, $i = 1, 2, \ldots, r$, and $j = 1, 2, \ldots, c$, the informational energy and mutual information of order γ concerning X and Y are given by

$$e^{\gamma}(X,Y) = \sum_i \sum_j p_{ij}^{\gamma}, \qquad I^{\gamma}(X,Y) = 1 - \sum_i p_{i+}^{\gamma} - \sum_j p_{+j}^{\gamma} + \sum_i \sum_j p_{ij}^{\gamma}, \tag{6.1}$$

where p_{i+} and p_{+j} are the marginal distributions of X and Y, respectively.

Comparisons of these informational functions and statistical inference concerning parameters and parameter sets can be obtained for both discrete and continuous distributions.

References

[1] I. A. Ahmad, *On asymptotic properties of an estimate of a functional of a probability density*, Scand. Actuar. J. **1976** (1976), no. 3, 176-181. MR 57#1753. Zbl 354.62036.

[2] I. A. Ahmad and S. C. Kochar, *Testing for dispersive ordering*, Statist. Probab. Lett. **7** (1988), no. 3, 179-185. MR 90f:62147. Zbl 662.62044.

[3] G. K. Bhattacharyya and G. G. Roussas, *Estimation of a certain functional of a probability density function*, Skand. Aktuarietidskr **1969** (1969), 201-206. MR 42#7011.

[4] R. N. Bradt and S. Karlin, *On the design and comparison of certain dichotomous experiments*, Ann. Math. Statist. **27** (1956), 390-409. MR 19,332e. Zbl 072.36102.

[5] D. V. Lindley, *On a measure of the information provided by an experiment*, Ann. Math. Statist. **27** (1956), 986-1005. MR 18,783g. Zbl 073.14103.

[6] D. Morales, L. Pardo, M. Salicrú, and M. L. Menéndez, *Asymptotic properties of divergence statistics in a stratified random sampling and its applications to test statistical hypotheses*, J. Statist. Plann. Inference **38** (1994), no. 2, 201-221. MR 94j:62012. Zbl 801.62019.

[7] T. K. Nayak, *On diversity measures based on entropy functions*, Comm. Statist. A—Theory Methods **14** (1985), no. 1, 203-215. MR 86g:62009. Zbl 561.62004.

[8] E. Parzen, *Nonparametric statistical data modeling*, J. Amer. Statist. Assoc. **74** (1979), no. 365, 105-131. MR 81b:62053. Zbl 407.62001.

Broderick O. Oluyede: Department of Mathematics and Computer Science, Georgia Southern University, Statesboro, GA 30460, USA

E-mail address: boluyede@gasou.edu