

On Energy and Expected Uncertainty Measures in Weighted Distributions

Broderick O. Oluyede

Department of Mathematical Sciences
Georgia Southern University, Statesboro, GA 30460
Boluyede@GeorgiaSouthern.edu

Mekki Terbeche

Department of Mathematics
Faculty of Sciences
University of Oran, Es-Senia, Oran

Abstract

In this note, bounds and inequalities for the comparisons of weighted energy functions, entropy, and discrimination information measures and their unweighted counterparts are presented. Inequalities for weighted expected uncertainty, cross-entropy or discrimination information measures are also presented. A useful result on the convergence of the weighted kernel density informational energy estimates is given and some informational energy applications presented.

Mathematics Subject Classification: 62B10, 62N05

Keywords: Stochastic inequalities, Weighted distribution functions, Bounds.

1 Introduction

The use of weighted distributions in research related to reliability, bio-medicine, ecology and several other areas is of tremendous practical importance in mathematics, probability and statistics. These distributions arise naturally as a result of observations generated from a stochastic process and recorded with some weight function. Several authors have presented important results on length-biased distributions and on weighted distributions in general. Rao [7] unified the concept of weighted distributions. Bhattacharyya et al [1] studied

and compared nonparametric unweighted and length-biased density estimates of fibers. Vardi [10] derived the nonparametric maximum likelihood estimate (NPMLE) of a lifetime distribution in the presence of length bias and established convergence to a pinned Gaussian process with a simple covariance function under mild conditions. For additional and important results on weighted distributions see Patil and Rao [6], Gupta and Keating [4] among others.

Let $\{P_\theta : \theta \in \Theta\}$ be a family of probability distributions associated with a σ -finite measure λ . Assume that the function P_θ is continuous and the mapping $\theta \rightarrow p_\theta$, where $p_\theta = dP_\theta/d\lambda$ is almost surely (a.s.) upper semicontinuous, separable random process, and the energy function $e(p_\theta) = E[p_\theta]$ exists and is finite on the parameter space. Energy functions are measures of dispersion of distributions that varies monotonically with dispersive order.

The purpose of this note is to compare energy, entropy and discrimination information measures involving length-biased distribution functions and weighted distribution functions in general. We also present some inequalities involving certain cross-entropy or discrimination measures for the comparisons of weighted distributions with the parent distributions. Comparisons between weighted and the parent distributions that involves the informational energy function and entropy are also presented. Section 2 contains some basic definitions and important utility notions. In section 3, comparisons and inequalities involving uncertainty and cross-entropy or discrimination measures are presented. In section 4, basic convergence result for the energy function is presented via the application of Schuster's lemma for kernel length-biased distributions. In section 5, some applications are presented.

2 Some Basic Results

In this section, we present some definitions and useful utility notions. Let \mathcal{F} be the set of absolutely continuous distribution functions satisfying

$$F(0) = 0, \lim_{x \rightarrow \infty} F(x) = 1, \sup\{x : F(x) < 1\} = \infty. \quad (1)$$

Note that if the mean of a random variable with a distribution function in \mathcal{F} is finite, it is positive. Let X be a non-negative random variable with distribution function F , survival function \bar{F} and probability density function (pdf) f . In a weighted distribution problem, a realization x of X enters into the investigators record with probability proportional to a weight function $W(x)$. Obviously, the recorded x is not an observation on X , but rather an observation on a weighted random variable X_w . The weighted random variable X_w has a survival or reliability function given by

$$\bar{G}_W(x) = \frac{E_F[W(X)|X > x]}{E_F(W(X))} \bar{F}(x). \quad (2)$$

This survival or reliability function can also be expressed as

$$\overline{G}_W(x) = \overline{F}(x)(W(x) + T_F(x))/E_F(W(X)), \quad (3)$$

where $T_F(x) = \int_x^\infty (\overline{F}(u)W'(u)du)/\overline{F}(x)$, and $W'(u) = dW(u)/du$, assuming that $W(x)\overline{F}(x) \rightarrow 0$ as $x \rightarrow \infty$. The probability density function corresponding to the reliability function given in (2) is referred to as a weighted probability density function (wpdf) with weight function $W(x) \geq 0$. The wpdf is given by

$$g_W(x) = W(x)f(x)/\delta^*, \quad (4)$$

where $0 < \delta^* < \infty$ is a normalizing constant. When $W(x) = x$, the corresponding pdf is called the length-biased probability density function and is given by

$$g_l(x) = xf(x)/\mu_F, \quad (5)$$

where $0 < \mu_F = \int_0^\infty \overline{F}(x)dx < \infty$. The length-biased survival function $\overline{G}_l(x)$ is given by:

$$\overline{G}_l(x) = \overline{F}(x)V_F(x)/\mu_F, \quad (6)$$

where $V_F(x) = E_F(X|X > x)$ is the vitality function. Note that $\overline{G}_l(x)$ and $\overline{F}(x)$ are stochastically ordered, that is $\overline{G}_l(x) \geq \overline{F}(x)$ for all $x \geq 0$.

Next we present some well known and useful definitions on the hazard rate function, mean residual life function, informational energy function and entropy measure.

Definition 2.1 . A distribution function F is said to have increasing (decreasing) hazard rate or failure rate on $[0, \infty)$, denoted by IHR (DHR) or IFR (DFR), if $F(0-) = 0$, $F(0) < 1$ and $P(X > x + t|X > t) = \overline{F}(x + t)/\overline{F}(t)$ is decreasing (increasing) in $t \geq 0$ for each $x > 0$.

Definition 2.2 . A distribution function F is said to have decreasing (increasing) mean residual life if $F(0-) = 0$, $F(0) < 1$ and $\int_t^\infty \overline{F}(x)dx/\overline{F}(t)$ is decreasing (increasing) in $t \geq 0$.

Note that if F has DHR and $\mu_F = \int_0^\infty \overline{F}(x)dx < \infty$, then F has increasing mean residual life (IMRL).

Definition 2.3 . The informational energy associated with a probability density function (pdf) f in \mathcal{F} is given by

$$e(f) = \int f^2(x)dx, \quad (7)$$

where $f(x) = dF(x)/dx$ and F is the corresponding distribution function.

Definition 2.4 . Let X be a non-negative random variable with finite variance and differentiable probability density function (pdf) f . The uncertainty measure associated with a distribution function F in \mathcal{F} is the differential entropy given by

$$H(f(X)) = -E_f(\log f(X)) = -\int f(x)\log f(x)dx, \quad (8)$$

where $f(x) = dF(x)/dx$.

The cross-entropy (Giuasu [3]) is

$$H(f_1, f_2) = -E_{f_1}(\log(f_1(X)/f_2(X))) = -\int f_1(x)\ln(f_1(x)/f_2(x))dx. \quad (9)$$

$H(f(X))$ is commonly referred to as the Shannon entropy measure (Shannon [9]). It is the expected uncertainty contained in $f(x)$ about the predictability of an outcome X . The following result is well known. Let X and Y be random variables and suppose that the density of Y has bounded derivative and if U is any random variable independent of X and Y , then convolution decreases Fisher information and increases entropy. That is, $I(Y+U) \leq I(Y)$ and $H(X+U) \geq H(X)$. These inequalities follow from the convexity of the functions x^2 and $x\log x$. Note that equality hold if and only if U is a constant, almost surely.

The following definition is due to Ebrahimi and Pellerey [2].

Definition 2.5 . The uncertainty of residual life distribution $H(X; t)$, of a component at time t , is the entropy of the residual life random variable $(X - t|X > t)$, and is given by

$$\begin{aligned} H(X; t) &= -\int_t^\infty \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx \\ &= \log \bar{F}(t) - \{\bar{F}(t)\}^{-1} \int_t^\infty f(x) \log f(x) dx \\ &= 1 - \{\bar{F}(t)\}^{-1} \int_t^\infty f(x) \log \{\lambda_F(x)\} dx, \end{aligned} \quad (10)$$

where $\lambda_F(x) = f(x)/\bar{F}(x)$ is the hazard or failure rate function.

The entropy of the weighted residual life random variable $(X_W - t|X_W > t)$ is given by

$$\begin{aligned} H(X_W; t) &= -\int_t^\infty \frac{f_W(x)}{\bar{F}_W(t)} \log \frac{f_W(x)}{\bar{F}_W(t)} dx \\ &= \log \bar{F}_W(t) - \{\bar{F}_W(t)\}^{-1} \int_t^\infty f_W(x) \log f_W(x) dx \end{aligned} \quad (11)$$

Under length biased sampling $H(X_W; t)$ reduces to

$$H(X_l; t) = 1 - \{\bar{F}(t)V_F(t)\}^{-1} \int_t^\infty x f(x) \log(x \lambda_F(x)/V_F(x)) dx. \quad (12)$$

where $V_F(t) = E(X - t | X > t) + t$, is the vitality function and $E_F(X - t | X > t)$ is the mean residual life function. $H(X; t)$ is the expected uncertainty in the conditional distribution of $X - t$ given $X > t$ about the predictability of the remaining lifetime of the component that has survived for time t .

We define certain weighted cross-entropy or discrimination information measures as follows:

$$H^*(f, g_w; t) = \int_t^\infty f(x) \log(f(x)/g_w(x)) dx \quad (13)$$

and

$$H_*(f, g_w; t) = \int_0^t f(x) \log(f(x)/g_w(x)) dx. \quad (14)$$

Similarly, define

$$H^*(g_w, f; t) = \int_t^\infty g_w(x) \log(g_w(x)/f(x)) dx \quad (15)$$

and $H_*(g_w, f; t)$. Also, define

$$H(f, g_w; t) = \int_t^\infty \frac{f(x)}{\bar{F}(t)} \log\left\{ \frac{f(x)/\bar{F}(t)}{g_w(x)/\bar{G}_w(t)} \right\} dx. \quad (16)$$

Note that if the weight function $W(x)$ is non-decreasing in $x > 0$, then $e(g_w) \geq e(f)$.

3 Inequalities for Uncertainty Measures

In this section, we present some results including inequalities and comparisons of cross-entropy and uncertainty measures for weighted and unweighted or parent distributions. Let X and Y be two random variables, and assume that the density of Y has a bounded derivative. If U is any random variable independent of X and Y , then it is well known that convolution increases entropy. This leads to the following question: If X_w and X are the weighted and unweighted random variables respectively, can X_w be expressed as $X_w = X + U$, where U is independent of X and X_w ? In the next result, we show that the measure of uncertainty present with respect to the value of a random variable X is greater than or equal to the measure of uncertainty present with respect to the value of the corresponding weighted random variable X_w .

Theorem 3.1 . *If the weight function $W(x)$ is non-decreasing for all $x \geq 0$, then $H(g_w(X)) \leq H(f(X))$.*

Proof: Note that $g_w(x)/f(x) = W(x)/\delta^*$ is non-decreasing in x . Now, for $0 < \beta < f(x)/g_w(x)$, and using the expansion of $\log(t)$ about $t = 1$, we have

$$\begin{aligned} \int_0^\infty g_w(x) \log(f(x)/g_w(x)) &\leq \int_0^\infty g_w(x) \left\{ \left(\frac{f(x)}{g_w(x)} - 1 \right) - \left(\frac{f(x)}{g_w(x)} - 1 \right)^2 \right\} \frac{1}{2\beta^2} dx \\ &= \int_0^\infty f(x) dx - \int_0^\infty g_w(x) dx \\ &\quad + \int_0^\infty \frac{(f(x) - g_w(x))^2}{g_w(x)^2} \frac{1}{2\beta^2} dx \\ &\leq 0. \end{aligned} \tag{17}$$

Note also that, since $W(x)$ is non-decreasing, we have

$$g_w(x) \log(g_w(x)) \geq g_w(x) \log f(x) \geq f(x) \log f(x) \geq 0, \tag{18}$$

for all $x > 0$. Therefore,

$$\int_0^\infty g_w(x) \log(g_w(x)) dx \geq \int_0^\infty g_w(x) \log(f(x)) dx \geq \int_0^\infty f(x) \log(f(x)) dx. \tag{19}$$

Equivalently,

$$E_{g_w}(-\log(g_w(X))) \leq E_f(-\log f(X)). \tag{20}$$

Consequently, $H(g_w(X)) \leq H(f(X))$.

Theorem 3.2 . Let $W(x)$ be a non-decreasing weight function with $W(x) > 0$. Then

- (i) $H(f, g_w; t) \geq H^*(f, g_w; t)$, for all $t \geq 0$,
- (ii) $H^*(f, g_w; 0) = E_f[\log(W(X))] + C$, where $C = \log(\delta^*)$.

Proof: (i) Note that $\overline{G}_w(x) \leq \overline{F}(x)$, so that $\overline{G}_w(x)/\overline{F}(x) \geq 1$ and $\log(\overline{G}_w(x)/\overline{F}(x)) \geq 0$, for all $x \geq 0$. Now,

$$\begin{aligned} H(f, g_w; t) &= (\overline{F}(t))^{-1} \int_t^\infty f(x) \log(f(x)/g_w(x)) dx + \log \frac{\overline{G}_w(x)}{\overline{F}(x)} \\ &= (\overline{F}(t))^{-1} H^*(f, g_w; t) + \log \frac{\overline{G}_w(x)}{\overline{F}(x)} \\ &\geq (\overline{F}(t))^{-1} H^*(f, g_w; t) \\ &\geq H^*(f, g_w; t), \end{aligned} \tag{21}$$

for all $t \geq 0$.

(ii) Let $W(x)$ be a non-decreasing weight function with $W(x) > 0$. Then

$$\begin{aligned}
 H^*(f, g_w; 0) &= \int_0^\infty \frac{f(x)}{\bar{F}(0)} \log \frac{f(x)/\bar{F}(0)}{g_w(x)/\bar{G}_w(0)} dx \\
 &= \int_0^\infty f(x) \log \frac{f(x)}{g_w(x)} dx \\
 &= \int_0^\infty f(x) \log \frac{f(x)}{W(x)f(x)/\delta^*} dx \\
 &= - \int_0^\infty f(x) \log(W(x)) dx + \log(\delta^*) \\
 &= -E_F(\log W(X)) + C,
 \end{aligned} \tag{22}$$

where $C = \log(\delta^*)$.

4 Kernel Length-Biased Estimates

In this section, kernel length-biased density estimates are presented and convergence result on the energy functions is established in Theorem 4.2. Let $f_n(x)$ be the kernel estimate based on n independent observations. The corresponding length-biased estimate is given by

$$g_n(x) = x f_n(x) / \mu_n, \tag{23}$$

where

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^n K((x - X_i)/h_n), \tag{24}$$

μ_n is an estimate of μ_F , and h_n is a sequence of positive constant tending to zero, and K is a bounded density function satisfying $\lim_{u \rightarrow \infty} |uK(u)| = 0$. Parzen (1962) showed that $E(f_n(x)) \rightarrow f(x)$ as $n \rightarrow \infty$, $E(f_n(x) - f(x))^2 \rightarrow 0$ as $n \rightarrow \infty$, $\frac{f_n(x) - E(f_n(x))}{\sigma(f_n(x))}$ converges in distribution to the standard normal distribution and $nh_n \sigma^2(f_n(x)) \rightarrow f(x) \int K^2(u) du$ as $n \rightarrow \infty$. Using the estimate $\mu_n = n / \sum_{i=1}^n X_i^{-1}$, we have

$$g_n(x) = x(n^2 h_n)^{-1} \sum_{i=1}^n K((x - X_i)/h_n) / \sum_{i=1}^n X_i^{-1}. \tag{25}$$

The length-biased pdf can also be estimated by using the fact that $\mu_n \leq \mu_n^* = \sum_{i=1}^n X_i/n$, to get

$$g_n(x) = x(nh_n \mu_n^*)^{-1} \sum_{i=1}^n K((x - X_i)/h_n). \tag{26}$$

Lemma 4.1 . (Schuster (1969)). If f and its first $r + 1$ derivatives are bounded and $\{\delta_n\}$ is a sequence of positive numbers such that $b_n = o(\delta_n)$, then there exists positive constants C_1 and C_2 , such that

$$P\{Sup_x |f_n^{(r)}(x) - f^{(r)}(x)| > \delta_n\} \leq C_1 \exp\{-C_2 n \delta_n^2 b_n^{2r+2}\}, \quad (27)$$

for sufficiently large n .

Theorem 4.2 . Let $g_n(x)$ be given by equations (23) or (25). Then for $\delta_n > 0$,

$$P\{|\int g_n^2(x)dx - \int g^2(x)dx| \geq \delta_n\} \rightarrow 0, \quad (28)$$

as $n \rightarrow \infty$.

Proof: We note that

$$|\int g_n^2(x)dx - \int g^2(x)dx| \leq \int |g_n^2(x) - g^2(x)|dx \leq Sup_x |g_n^2(x) - g^2(x)|.$$

An application of Schuster's Lemma leads to

$$\begin{aligned} P\{|\int g_n^2(x)dx - \int g^2(x)dx| \geq \delta_n\} &\leq P\{|Sup_x |g_n^2(x) - g^2(x)| \geq \frac{\delta_n}{2}\} \\ &\leq C_1 \exp\{-C_2 n \delta_n^2 b_n^2\}. \end{aligned}$$

Applying Borel-Cantelli Lemma, it follows that $C_1 \exp\{-C_2 n \delta_n^2 b_n^2\} \rightarrow 0$ as $n \rightarrow \infty$, by an appropriate choice of the sequence b_n , and the fact that the series $\sum_{n=1}^{\infty} \exp\{-C_2 n \delta_n^2 b_n^2\} < \infty$.

Consequently, $P\{|\int g_n^2(x)dx - \int g^2(x)dx| \geq \delta_n\} \rightarrow 0$, as $n \rightarrow \infty$.

5 Applications

In this section, we present examples on energy functions involving the normal and Rayleigh distributions respectively.

1. **Normal Distribution.** Let $f(x; \mu, \sigma)$ and $g_w(x; \mu, \sigma)$ denote the pdf for the normal and weighted normal distributions respectively. Let $W(x)$ be a non-decreasing weight function. In particular, we let $W(x) = x$. The energy function associated with the normal pdf $f(x; \mu, \sigma)$ is given by

$$e(f(x; \mu, \sigma)) = \int_{-\infty}^{\infty} f^2(x; \mu, \sigma)dx = \pi^{-1/2}(2\sigma)^{-1}. \quad (29)$$

The energy function $e(f(x; \mu, \sigma))$ is a bijective function of σ . If σ_f and σ_{g_w} are the standard deviations of the normal distribution functions F and G_w respectively, then $\sigma_{g_w} \geq \sigma_f$ if and only if $e(f(x; \mu, \sigma)) \geq e(g_w(x; \mu, \sigma))$.

2. Rayleigh Distribution. Let

$$f(x; \beta) = 2\beta^{-1}xe^{-\beta x^2}, \quad \text{if } x > 0 \text{ and } \beta > 0. \quad (30)$$

Then $\mu_F = \beta^{-1/2}\pi^{1/2}/2$ and the corresponding weighted pdf with $W(x) = x$ is given by

$$g_W(x; \beta) = 4\pi^{-1/2}\beta^{3/2}x^2e^{-\beta x^2}, \quad \text{if } x > 0 \text{ and } \beta > 0. \quad (31)$$

Note that

$$\begin{aligned} \int_0^\infty (g_W(x; \beta))^2 dx &= \int_0^\infty 16\pi^{-1}x^4\beta^3e^{\beta x^2} dx \\ &\geq \int_0^\infty 4\pi^{-1/2}x^2\beta^{3/2}e^{\beta x^2} dx \\ &= \int_0^\infty f^2(x; \beta) dx. \end{aligned} \quad (32)$$

Consequently $e(g_W) \geq e(f)$ for all $x > 0$.

ACKNOWLEDGEMENTS. The authors are grateful to the editor and the referee for taking the time to review this paper.

References

- [1] Bhattacharyya, B. B., Franklin, L. A., and Richardson, G. D., A Comparison of Nonparametric Unweighted and Length-Biased Density Estimation of Fibers, *Communications in Statistics-Theory and Methods*, **17(11)** (1988), 3629-3644.
- [2] Ebrahimi, N. and Pellerey, F., New Partial Ordering of Survival Functions Based on the Notion of Uncertainty, *Journal of Applied Probability*, **32** (1995), 202-211.
- [3] Guiasu, S., *Information Theory with Applications*, New York, McGraw-Hill, 1977.
- [4] Gupta, R.C., and Keating, J.P., Relations for Reliability Measures Under Length-Biased Sampling, *Scandinavian Journal of Statistics*, **13** (1985), 49-56.
- [5] Parzen, E., On Estimation of a Probability Density Function and Mode, *Annals of Mathematical Statistics*, **33** (1962), 1065-1076.

- [6] Patil, G. P., and Rao, C. R., Weighted Distributions and Size-Biased Sampling with Applications to Wildlife and Human Families, *Biometrics*, **34** (1978), 179-189.
- [7] Rao, C. R., On Discrete Distributions Arising Out of Methods of Ascertainment, *In Classical and Contagious Discrete Distributions*, G. P. Patil (ed.), Calcutta, India, Pergamon Press and Statistical Publishing Society, (1965), 320-332.
- [8] Schuster, E. F., Estimation of a Probability Density Function and its Derivatives, *Annals of Statistics*, **40** (1969), 1187-1195.
- [9] Shannon, C. E., Mathematical Theory of Communication, *Bell System Technical Journal*, **27** (1948), 379-423 and 623-656.
- [10] Vardi, Y., Nonparametric Estimation in the Presence of Bias, *Annals of Statistics*, **10(2)** (1982), 616-620.

Received: June 7, 2006