Journal of Theoretical Probability, Vol. 18, No. 14, October 2005 (© 2005)
DOI: 10.1007/s10959-005-7541-3

More on a New Concept of Entropy and Information

Murali Rao[1]

Received May 25, 2004; revised November 9, 2004

An alternative notion of entropy called CRE is proposed in [Ra1] Rao *et al.* (IEEE Trans. Inf. Theory **50**, 2004). This preserves many of the properties of Shannon Entropy and possesses mathematical properties, which we hope will be of use in statistical estimates. In this article, we develop some more mathematical properties of CRE, show its relation to the $L \log L$ class, and characterize among others the Weibull distribution.

KEY WORDS: Shannon entropy; information; Weibull distribution.

1. INTRODUCTION AND PRELIMINARIES

In two path-breaking contributions in 1948 and 1949 Claude Shannon introduced the world to his theory of Uncertainty/Information. This has mushroomed into a large body of knowledge revolutionizing many areas especially Communication Engineering. He called this measure Entropy.[(10)]

There have been several other definitions of entropy. Most succesful of these is the notion of relative entropy- or K.L. Divergence. The reader may consult[(4)] for more information.

Shannon began with the following measure of uncertainty in a random variable X assuming a finite number of values $x_i, \dots, x_n$ with probabilities $p_1, \dots, p_n$. He defined the uncertainty/information $H(X)$ in X as

$$-H(X) = p_1 \log(p_1) + \cdots + p_n \log(p_n).$$

Inspite of its enormous success, the following two examples show that this measure may not be appropriate in every situation and that there is room for another approach in some situations.

[1] Department of Mathematics, University of Florida, Gainesville, 32611-8105 FL, USA. E-mail: rao@math.ufl.edu

The import of the two examples below is that a measure of uncertainty/information, which considers only probabilities and ignores the values the random variable takes, may in some situations, not do justice to our intuitive and practical notions of randomness or information.

For any discrete random variable X, the Entropy $H(X)$ of X is computed solely using the probabilities $P(X=t)$ and one interprets the Entropy as a measure of the "randomness" in X. If, X denotes the life time of a machine, the price of a stock, the number of properly functioning components required in a complex system, etc, the appropriate probabilities to consider are $P(X>t)$ and not $P(X=t)$. Insisting on the probabilities $P(X=t)$ leads to some difficulties in the interpretation of $H(X)$ as measuring randomness in X.

Example 1. Suppose X_i are random variables such that $P(X_i=1)=1/2$ and $P(X_i=1+(1/i))=(1/2)$ then for large enough i, practically speaking and in many theoretical situations X_i are regarded as being essentially 1, in particular non-random. However the entropy of X_i is independent of $i: H(X_i)=\log 2$ for every i. The example becomes even more dramatic if X is chosen to assume n values all very close and with probabilities $(1/n)$. In this case $H(X)$ is always equal to $\log n$ regardless of its variance.

For a random variable X another interpretation of $H(X)$ is as a measure of the information provided by an observation of X. Note that $H(X)$ ignores the values taken by X which may not always be appropriate.

Example 2. Consider two fair coins X and Y. Instead of the usual heads and tails coin X has \$0 and \$l while Y has \$0 and \$1000000. Observations of X and Y convey vastly different information.

Such considerations led us to define an alternative measure of uncertainty (termed CRE in Ref. 9). This measure enjoys many of the properties of Shannon entropy $H(X)$ of a discrete random variable X. The definition is the same whether or not the random variable has a discrete or continuous distribution. Since it depends on CDF instead of PDF it has continuity properties not possessed by Shannon, entropy. To our knowledge CDF was introduced for the first time to measure uncertainty in Ref. 9.

In this article we investigate some more interesting mathematical properties of CRE. In Section 3 we show that CRE dominates $E(|X-E(X)|)$. In Sections 4 and 5 CRE is related to the $L \log L$ class.[8,11] Section 7 shows how to characterize, using CRE, Failure Time Distributions. As the name suggests these distributions play a fundamental role in the analysis of failure time data.[5] Specifically we consider the Weibull distribution. The derivation clearly shows the relationship of CRE and the hazard function thus giving a definitive meaning to max CRE interpretation. Finally we prove a 'meta' theorem- a characterization of very general distributions.

2. PRELIMINARIES

Generally by X we will denote non-negative random variables. Given a random variable X its decreasing distribution function $F(t)$ is defined by

$$F(t) = P(X > t)$$

and we define CRE(X) by

Definition 1.

$$\mathrm{CRE}(X) = \int_0^\infty F(t)|\log F(t)|dt \quad = \int_0^\infty P(X > t)|\log P(X > t)|dt.$$

The following inequality is a simple consequence of Jensen's inequality and indeed equivalent to it. Because it can be used in more varied situations it is also more useful. As in Ref. 1 we will call this the log-sum inequality and will be used repeatedly.

Log-Sum Inequality: Let m be a sigma finite measure. If f and g are positive and m integrable then

$$\int f \log \frac{f}{g} dm \geq \left[\int f\, dm\right] \log \frac{\int f\, dm}{\int g\, dm}.$$

3. AN INEQUALITY

We will find the following inequality useful.

Proposition 1. Let X and Y be non-negative, have the same distribution and be independent, Then

$$E(|X - Y|) \leqslant 2\mathrm{CRE}(X). \tag{1}$$

In particular, for any non-negative variable X,

$$E[|X - E(X)|] \leqslant 2\mathrm{CRE}(X). \tag{2}$$

Proof. Write $F(t) = P(X > t)$. Then

$$\begin{aligned} P\big[(X > t) \cup (Y > t)\big] &= P(X > t) + P(Y > t) - P\big[X > t, Y > t\big] \\ &= 2F(t) - F(t)^2, \end{aligned}$$

so that

$$2F(t) - 2F(t)^2 = P[\max(X, Y) > t] - P[\min(X, Y) > t]. \tag{3}$$

Integrating both sides of (3) from zero to infinity

$$\begin{aligned} 2\int_0^\infty F(t)(1 - F(t))dt &= E[\max(X, Y) - \min(X, Y)] \\ &= E[|X - Y|]. \end{aligned} \tag{4}$$

It can be verified that

$$x(1 - x) \leqslant x|\log(x)|, \quad 0 < x < 1.$$

Using this we have finally

$$\begin{aligned} 2\mathrm{CRE}(X) &= 2\int_0^\infty F(t)|\log(F(t))|dt \geqslant 2\int_0^\infty F(t)(1 - F(t))dt \\ &= E[|X - Y|]. \end{aligned} \tag{5}$$

The last inequality follows from (4). This proves (1). Equation (2) follows easily from (1). Indeed,

$$\begin{aligned} E[|X - Y|] &= \int E[|X - a|]dF_X(a) \geqslant \int |a - E(X)|dF_X(a) \\ &= E[|X - E(X)|]. \end{aligned}$$

This completes the proof. □

Remark 1. Equation (2) has a more general version, whose proof is similar: For any σ field G

$$2\mathrm{CRE}(X/G) \geqslant E[|X - E(X/G)|/G].$$

4. CRE AND L LOG L

The Orlicz Space $L\log^+ L$ plays an important role in Analysis (see Refs. 7, 8, 11).

In this section we show that for a non-negative random variable X, CRE(X) is finite iff X is in $L\log^+ L$, i.e., $E(X\log^+ X) < \infty$.

Theorem 1. Let X be a non-negative random variable. Then

$$E(X\log^+ X) \leqslant E(X:X>1)\log(eE(X:X>1) + \mathrm{CRE}(X). \tag{6}$$

Proof.

$$\begin{aligned} E(X\log^+ X) &= \int_1^\infty (1+\log t)P(X>t)dt \\ &= E[(X-1)^+] + \int_1^\infty (\log t)P(X>t)dt. \end{aligned} \tag{7}$$

Now for $t>1$

$$tP(X>t) \leqslant E(X:X>t) \leqslant E(X:X>1)$$

so that

$$t \leqslant \frac{E(X;X>1)}{P(X>t)}$$

implying

$$\log t \leqslant \log E(X:X>1) - \log P(X>t).$$

Using this in (2.2) we get

$$E(X\log^+ X) \leqslant E(X:X>1) + E(X:X>1)\log E(X:X>1) + \mathrm{CRE}(X)$$

as claimed.

On the other hand, by the log-sum inequality, for any $p>1$

$$\int_1^\infty P(X>t)\log[t^p P(X>t)]dt \geqslant E[(X-1)^+]\log\frac{E[(X-1)^+]}{\int_1^\infty t^{-p}\,dt}.$$

Rewriting both sides of the above inequality we get

$$\begin{aligned} &\int_1^\infty P(X>t)\log P(X>t)dt + p\int_1^\infty P(X>t)\log(t)dt \\ &\quad \geqslant E[(X-1)^+]\log[(p-1)E[(X-1)^+]. \end{aligned}$$

That is to say

$$\int_1^\infty P(X>t)\log P(X>t)dt \geqslant -pE(X\log^+ X)+pE[(X-1)^+]$$
$$+E[(X-1)^+]\log(p-1)E[(X-1)^+]$$
$$=-pE(X\log^+ X)+E[(X-1)^+]\log[e^p(p-1)E[(X-1)^+]]. \qquad (8)$$

Since $x|\log(x)|\leqslant e^{-1}$ for $0<x<1$ we finally have

$$\text{CRE}(X)\leqslant e^{-1}+pE(X\log^+ X)-E[(X-1)^+]\log[e^p(p-1)E[(X-1)^+].$$

Thus we can say
CRE(X) is finite iff $X\epsilon L\log^+ L$
That concludes the proof. □

5. A FORMULA FOR CRE

We derive an alternative expression for CRE, which will have applications.

Definition 2. A locally integrable function B is said to be of bounded variation if

$$\sup\int \phi'(x)B(x)dx<\infty. \qquad (9)$$

where the supremum is taken over all C^1 functions ϕ of compact support and not exceeding one in absolute value. This definition is equivalent to the standard definition, except that the function is only almost everywhere defined and an appropriate “version” leads to the standard one. For our purposes this will be most convenient.

If B is a function of bounded variation as defined above there is a unique (signed measure) denoted dB such that

$$\int \phi'(x)B(x)dx=-\int \phi(x)dB(x) \qquad (10)$$

for all C^1 functions ϕ of compact support. Further it is known that,[(2)] if B is of bounded variation and ρ is Lipschitz then so is $\rho(B)$ and

$$d\rho(B)=\rho'(B)dB. \qquad (11)$$

This formula shows the importance of the above definition of functions of bounded variation for integration purposes.

Proposition 2. Let X be non-negative and $F(t)$ its decreasing distribution: $P(X>t)=F(t)$. Then

$$\mathrm{CRE}(X)=-E[X(1+\log F(X))]. \tag{12}$$

Proof. Using (11) with $\rho(t)=t\log(t)$ we have

$$F(t)\log F(t)=-\int_t^{\infty}(1+\log F(s))dF(s). \tag{13}$$

In the above equation we interpret $dF(s)$ as $F(s+ds)-F(s)$. Integrating (13) with respect to t and changing the order we get

$$\begin{aligned}-\mathrm{CRE}(X) &= -\int_0^{\infty} t[1+\log(F(t))]dF(t)\\ &= E[X(1+\log F(X))]. \end{aligned}\tag{14}$$

The Proposition is proved. □

6. AN APPLICATION

Before we come to the application note that the "conjugate" or the Fenchel Transform of the convex function $x\log x$ is $\exp(y-1)$ i.e.,

$$\exp(y-1)=\sup[xy-x\log x: 0<x<\infty]$$

so that for all $x>0$ and $y>0$

$$xy\leqslant x\log x+\exp(y-1). \tag{15}$$

Further if the random variable X has a continuous distribution and F is its decreasing distribution: $F(t)=P(X>t)$ then the random variable $F(X)$ is uniformly distributed so that $E[\log F(X)]=-1$. With these two facts at hand we have

$$\begin{aligned}\mathrm{CRE}(X) &= -E[X(1+\log F(X))]=E(X)E(\log F(X))\\ &\quad -E[X\log F(X)]=E[(X-E(X))(-\log F(X))]\\ &= 2E[(X-E(X))(-\log F(X)^{\frac{1}{2}})]\\ &\leqslant 2E[|(X-E(X)|(-\log F(X)^{\frac{1}{2}})].\end{aligned}$$

Now using (15)

$$\leqslant 2E[|X-E(X)|\log|X-E(X)|]+2E[\exp[-\log F(X)^{\frac{1}{2}}-1]]$$
$$=2E[|X-E(X)|\log|X-E(X)|]+\frac{2}{e}\int_0^1 t^{-\frac{1}{2}}dt$$
$$=2E[|X-E(X)|\log|X-E(X)]+\frac{4}{e}.$$

This gives an upper bound for CRE interms of $|X-E(X)|$.

It is easily possible that $E[|X-E(X)|\log|X-E(X)|]=0$ as the following simple example shows.

Example 3. Let X take values $1-a$ and $2+b$ where a and b are positive and $0<a<1$ with probabilities $(1+b/1+a+b)$ and $(a/1+a+b)$, respectively. Then $E(X)=1$ and $E[|X-E(X)|\log|X-E(X)|]=0$ if and only if $a=(1/1+b)$.

7. MAX CRE DISTRIBUTIONS

Some of the most important highlights of Shannon entropy is the derivation of most of the important useful distributions. Let us describe this briefly. The starting point here is Laplace's principle of insufficient reason. This states: If the only information we have about an experiment is that it has n outcomes, then the only reasonable assumption is that all outcomes are equally likely. A far reaching generalisation of this is the MaxEnt principle or the principle of maximum entropy enunciated in 1957 by E.T.Jaynes. The MaxEnt principle states: Out of all distributions consistent with a given set of constraints choose one that maximizes entropy. This principle has been applied to derive all the most useful probability distributions in terms of some simple moments (for many examples see Ref. 6). For a debate of this method and some answers see Ref. 3.

Examples 4. The uniform distribution on an interval $[a,b]$ is the MaxEnt distribution without constraints. However if the mean m is prescribed then the MaxEnt distribution is the truncated exponential $c\exp(-kx)$ where c and k are chosen to satisfy

$$c\int_a^b \exp(-kx)dx=1 \quad \text{and} \quad c\int_a^b x\exp(-kx)dx=m.$$

In this section, we show how certain distributions have max CRE characteristics. These will be special cases of a "meta theorem".

Definition 3. Let X be a non-negative random variable. We define the normalized CRE- NCRE- of X:

$$\text{NCRE}(X) = E(X)^{-1}\int_0^\infty |F(t)\log F(t)|dt = E(X)^{-1}\text{CRE}(X), \tag{16}$$

where $F(t)$ is the decreasing distribution of X as defined before.

A non-negative random variable $W_{q,\lambda}$ is Weibull distributed if its decreasing distribution $F(t)$ is given by

$$F(t) = \exp(-\lambda^q t^q), \tag{17}$$

where λ and q are positive parameters. We find, if $W_{q,\lambda}$ is Weibull distributed, then for any $p > 0$

$$\begin{aligned} E[W_{q,\lambda}^p] &= \lambda^{-p}\Gamma\left(1+\frac{p}{q}\right), \\ \text{CRE}(W_{q,\lambda}) &= \lambda^{-1}q^{-1}\Gamma\left(1+\frac{1}{q}\right), \\ \text{NCRE}(W_{q,\lambda}) &= \frac{1}{q}. \end{aligned} \tag{18}$$

Now we have the following characterization of the Weibull distribution.

Theorem 2. Among all positive random variables X with given $(p+1)$ st moment the Weibull distribution $W_{q,\lambda}$ has the maximal NCRE. Here the parameters q and λ are given by

$$\frac{1}{q} = \frac{c_p^p}{(p+1)}\frac{E(X^{p+1})}{E(X)^{p+1}}$$

and

$$\lambda^{-p-1} = \frac{E(X^{p+1})}{\Gamma(1+\frac{p+1}{q})}, \tag{19}$$

where

$$c_p = \Gamma(1+\frac{1}{p}). \tag{20}$$

Proof. For any randomvariable $X \geqslant 0$, by the log-sum inequality:

$$\int_0^\infty F(t) \log\left[\frac{F(t)}{\exp(-\mu^p t^p)}\right] \geqslant E(X) \log\left[\frac{E(X)}{\mu^{-1} c_p}\right], \tag{21}$$

where $F(t)$ is the decreasing distribution of X and c_p is defined in (20).

Rewriting (21) we get

$$\mathrm{CRE}(X) \leqslant \frac{\mu^p}{(p+1)} E(X^{p+1}) - E(X) \log \frac{E(X)}{\mu^{-1} c_p}.$$

Or recalling the definition of NCRE

$$\mathrm{NCRE}(X) \leqslant \frac{\mu^p}{(p+1)} \frac{E(X^{p+1})}{E(X)} - \log\left[\frac{E(X)}{\mu^{-1} c_p}\right]. \tag{22}$$

Choosing μ so that $\mu^{-1} c_p = E(X)$, we get

$$\mathrm{NCRE}(X) \leqslant \frac{c_p^p}{(p+1)} \frac{E(X^{p+1})}{E(X)^{p+1}}. \tag{23}$$

Finally let $q > 0$ be defined by

$$\frac{1}{q} = \frac{c_p^p}{(p+1)} \frac{E(X^{p+1})}{E(X)^{p+1}}$$

and then λ be defined by

$$\lambda^{-p-1} \Gamma\left(1 + \frac{p+1}{q}\right) = E(X^{p+1}).$$

For λ and q thus defined

$$E[W_{q,\lambda}^{p+1}] = E(X^{p+1})$$

and from (23)

$$\mathrm{NCRE}(X) \leqslant \frac{1}{q} = \mathrm{NCRE}(W_{q,\lambda}).$$

Because of (18) the proof is finished. □

A slightly different and more general approach is the following: let X be a non-negative random variable and $r_1, \ldots, r_n$ be functions on $(0, \infty)$ into itself. Put

$$F(t) = P(X > t),$$
$$R_i(t) = \int_0^t r_i(s) ds. \tag{24}$$

Suppose $\alpha_i = E(R_i(X))$ are given. Interms of F we can write

$$\int_0^\infty F(s) r_i(s) ds = \alpha_i. \tag{25}$$

Write

$$h(t) = \exp(-\Sigma_1^n \lambda_i r_i(t)), \tag{26}$$

where $\lambda_i \geqslant 0$ will be chosen momentarily.

By the log-sum inequality

$$\int_0^\infty F(t) \log\left[\frac{F(t)}{h(t)}\right] \geqslant E(X) \log\left[\frac{E(X)}{\int_0^\infty h(s) ds}\right], \tag{27}$$

which is the same as (use (25))

$$\text{CRE}(X) \leqslant \Sigma_1^n \lambda_i \alpha_i - E(X) \log\left[\frac{E(X)}{\int_0^\infty h(s) ds}\right]. \tag{28}$$

If r_i are increasing as functions of t and if

$$\lim(\max_i r_i(t)) = \infty,$$

$h(t)$ is decreasing and defines an "improper" decreasing distribution, i.e., a sub-probability on $[0, \infty)$, a proper distribution, i.e., a probability distribution if $r_i(0) = 0$ for $1 \leqslant i \leqslant n$.

In all cases let us write

$$\text{CRE}(h) = -\int_0^\infty h(t) \log h(t) dt = \Sigma_1^n \lambda_i \int_0^\infty r_i(t) h(t) dt. \tag{29}$$

Suppose we can find $\lambda_i > 0$ such that

$$\int_0^\infty r_i(t) h(t) dt = \alpha_i, \quad 1 \leqslant i \leqslant n,$$
$$\int_0^\infty h(t) dt \leqslant E(X), \tag{30}$$

then from (28) and (29)

$$\text{CRE}(X) \leqslant \text{CRE}(h).$$

Thus we have proved

Theorem 3. Let $X \geqslant 0$ and r_i $1 \leqslant i \leqslant n$ be non-negative functions. Put $R_i(t) = \int_0^t r_i(s)ds$ Suppose

$$\alpha_i = E(R_i(X)), \quad 1 \leqslant i \leqslant n.$$

Let $h(t)$ be defined by (26), where we assume λ_i have been chosen to satisfy

$$\int_0^\infty r_i(t)h(t)dt = \alpha_i, \quad 1 \leqslant i \leqslant n,$$
$$\int_0^\infty h(t)dt \leqslant E(X).$$

Then

$$\mathrm{CRE}(X) \leqslant \mathrm{CRE}(h).$$

Example 5. Let $r_1(t) = 1$ and $r_2(t) = t^p$, $p > 0$. Then

$$\int_0^\infty r_1(t)F(t)dt = \int_0^\infty r_1(t)P(X > t)dt = E(X)$$

and

$$\int_0^\infty t^p F(t)dt = \frac{1}{p+1}E(X^{p+1}).$$

Thus if $E(X)$ and $E(X^{p+1})$ are given then

$$\mathrm{CRE}(X) \leqslant \mathrm{CRE}(h),$$

where $h(t) = \exp(-\lambda_1 - \lambda_2 t^p)$.
Here λ_1 and λ_2 are chosen so that

$$\begin{aligned} E(X) &= \int_0^\infty h(t)dt = \exp(-\lambda_1)\int_0^\infty \exp(-\lambda_2 t^p)dt \\ &= \lambda_2^{-\frac{1}{p}} \exp(-\lambda_1)\Gamma(1 + \frac{1}{p}) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{p+1}E(X^p) &= \int_0^\infty t^p F(t)dt = \exp(-\lambda_1)\int_0^\infty t^p \exp(-\lambda_2 t^p)dt \\ &= \frac{\exp(-\lambda_1)}{p\lambda_2^{1+\frac{1}{p}}}\Gamma(1 + \frac{1}{p}). \end{aligned}$$

The last two equalities give

$$\frac{1}{p+1}E(X^p)=\frac{1}{p\lambda_2}E(X).$$

This gives the value of λ_2:

$$\lambda_2=\frac{(p+1)}{p}\frac{E(X)}{E(X^p)}.$$

Then λ_1 is given by:

$$\exp(-\lambda_1)\Gamma(1+\frac{1}{p})=\lambda_2^{\frac{1}{p}}E(X).$$

Now we prove a general result giving MAX–CRE characterizations of very general distributions. Let us start with some preparations.

Definition 4. A decreasing function f with $f(\infty)=0$ and $\int_0^\infty f(t)dt<\infty$ will be called a distribution. Given a distribution f we define

$$\mathrm{CRE}(f)=-\int_0^\infty f(t)\log[f(t)]dt. \tag{31}$$

Write $f(t)=-\int_t^\infty df(s)$. We can then write (31) as

$$\begin{aligned}-\int_0^\infty f(t)\log[f(t)]dt &= \int_0^\infty \log[f(t)]dt\int_t^\infty df(s)\\ &= \int_0^\infty df(s)\int_0^s \log[f(t)]dt.\end{aligned}$$

Thus

$$\mathrm{CRE}\ (f)=\int_0^\infty F(s)df(s), \tag{32}$$

where

$$F(s)=\int_0^s \log\ f(t)dt. \tag{33}$$

In other words CRE(f) is nothing but the integral of F relative to the measure df. We can now state the following general theorem:

Theorem 4. **(MAX–CRE theorem).** Let f be a distribution and F be defined by (33). Then for all distributions g satisfying

$$\int_0^\infty g(s)ds \geqslant \int_0^\infty f(s)ds,$$
$$\int_0^\infty F(s)dg(s)ds = \int_0^\infty F(s)df(s)ds,$$

we have

$$\text{CRE }(g) \leqslant \text{CRE }(f).$$

Proof. Applying the log-sum inequality we have

$$\int_0^\infty g(s)\log\left[\frac{g(s)}{f(s)}\right]ds \geqslant \left(\int_0^\infty g(s)ds\right)\log\left[\frac{\int_0^\infty g(s)ds}{\int_0^\infty f(s)ds}\right] \geqslant 0.$$

In particular

$$\int_0^\infty g(s)\log g(s)ds \geqslant \int_0^\infty g(s)\log f(s)ds$$
$$= -\int_0^\infty F(s)dg(s) = -\int_0^\infty F(s)df(s) = -\text{CRE }(f).$$

This concludes the proof. □

Example 6. Let $0 \leqslant a < b$ and define

$$\begin{aligned} f(t) &= 1, \quad 0 \leqslant t \leqslant a, \\ &= \frac{b-t}{b-a}, \quad a \leqslant t \leqslant b, \\ &= 0, \quad b \leqslant t, \end{aligned}$$

f corresponds to the uniform distribution on $[a, b]$, i.e., X is uniformly distributed on $[a, b]$ then

$$f(t) = P(X > t) 0 \leqslant t < \infty.$$

Applying the above theorem, and a little algebra we get

The uniform distribution on $[a, b]$ has max-CRE among all distributions g satisfying

$$g(a) = 1, \quad g(b) = 0,$$
$$\int_0^\infty t\,dg(t) = \int_0^\infty t\,df(t),$$
$$\int_a^b (b-t)\log(b-t)dg(t) = \int_a^b (b-t)\log(b-t)df(t).$$

Example 7. Let

$$f(t) = \exp(-t^p), \quad t \geqslant 0, \quad p > 0.$$

This corresponds to the Weibull distribution with parameter p. Applying the theorem we find

Among all distributions whose first and $(p+1)st$ moments coincide with those of the Weibull, the latter has the max CRE.

REFERENCES

1. Cover T. M., and Thomas J. A. (1991). *Elements of Information Theory*, Wiley, New York.
2. Federer H. (1969). *Geometric Measure Theory*, Springer, Berlin.
3. Grendar M. Jr., and Grendar M. (2000). What is the Question that MaxEnt answers? Bayesian Inference And Maximum Entropy Methods In Science and Engineering. 20th International Workshop Gif-sur-Yvette, France. Ed. Ali Mohammad-Djafari American Institute of Physics Conference Proceedings 568.
4. Jumarie G. (1990). *Relative Information*, Springer.
5. Kalbfleisch J. D., and Prentice R. L. (1980). *The Statistical Analysis of Failure Time Data Wiley Series in Probability*.
6. Kapur J. N. (1989). *Maximum Entropy Methods in Science and Engineering*, Wiley, New York.
7. Krasnoselskli, M. A., and Rutickii, Ya.B. (1961). *Convex Functions and Orlicz Spaces*, Noordhoff, Leiden.
8. O'Neil, R. (1966). Les fonctiones conjugues et les integrales fractionnaires de la classe $L\log^+, L$. *C.R. Acad.Sci.Paris* **263**, 463–466.
9. Rao M., Chen Y., Vemuri B., and Fei W. (2004). Cumulative residual entropy : a new measure of information. *IEEE Trans.Inf.Theory* **50**.
10. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech.J.* **27**, 379–423, 623–656.
11. Stein E. M. (1969). Note on the Class L log L. *Studia Math.* **32**, 305–310.