Complete monotonicity of the entropy in the central limit theorem for gamma and inverse Gaussian distributions

Yaming Yu *

Department of Statistics, University of California, Irvine 92697, USA

A R T I C L E I N F O

Article history:
Received 16 June 2008
Received in revised form 11 August 2008
Accepted 14 August 2008
Available online 22 August 2008

A B S T R A C T

Let $H_\alpha$ be the differential entropy of the gamma distribution $\text{Gam}(\alpha, \sqrt{\alpha})$. It is shown that $(1/2) \log(2\pi e) - H_\alpha(\alpha)$ is a completely monotone function of $\alpha$. This refines the monotonicity of the entropy in the central limit theorem for gamma random variables. A similar result holds for the inverse Gaussian family. How generally this complete monotonicity holds is left as an open problem.

1. Introduction

The classical central limit theorem (CLT) states that for a sequence of independent and identically distributed (i.i.d.) random variables $X_i$, $i = 1, 2, \ldots$, with a finite mean $\mu = E(X_1)$ and a finite variance $\sigma^2 = \text{Var}(X_1)$, the normalized partial sum $Z_n = (\sum_{i=1}^{n} X_i - n\mu)/\sqrt{n\sigma^2}$ tends to $N(0, 1)$ in distribution as $n \to \infty$. There exists an information theoretic version of the CLT. The differential entropy for a continuous random variable $S$ with density $f(s)$ is defined as

$$H(S) = -\int_{-\infty}^{\infty} f(s) \log f(s) ds.$$

Barron (1986) proved that if the (differential) entropy of $Z_n$ is eventually finite, then it tends to $(1/2) \log(2\pi e)$, the entropy of $N(0, 1)$, as $n \to \infty$. As a consequence of Pinsker’s inequality (Cover and Thomas, 1991), convergence in entropy implies convergence in distribution in the normal CLT case.

An interesting feature of the information theoretic CLT is that $H(Z_n)$ increases monotonically in $n$ (which reminds us of the second law of thermodynamics); since the entropy is invariant under a location shift, equivalently for a sequence of i.i.d. random variables $X_i$ with a finite second moment, we have

$$H\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right) \leq H\left(\frac{1}{\sqrt{n+1}} \sum_{i=1}^{n+1} X_i\right), \quad n \geq 1. \quad (1)$$

Although the special case $n = 1$ is known to follow from Shannon’s classical entropy power inequality (Stam, 1959; Blachman, 1965), it was not until 2004 before the above inequality was finally proven by Artstein et al. (2004); see also Madiman and Barron (2007) for generalizations.

This paper is concerned with a possible refinement of this monotonicity, at least when the distribution of $X_i$ belongs to some special families. Recall that a function $f(x)$ on $x \in (0, \infty)$ is called completely monotonic if its derivatives of all orders exist and satisfy

$$(-1)^k f^{(k)}(x) \geq 0, \quad x > 0,$$

* Tel.: +1 949 824 7361.
E-mail address: yamingy@uci.edu.

© 2008 Elsevier B.V. All rights reserved.

doi:10.1016/j.spl.2008.08.008
for all \( k \geq 0 \). In particular, a completely monotone function is both monotonically decreasing and convex. A discrete sequence \( f_n \) on \( n = 1, 2, \ldots \) is completely monotonic if
\[
(-1)^k \Delta^k f_n \geq 0, \quad n = 1, 2, \ldots
\]
for all \( k \geq 0 \), where \( \Delta \) denotes the first difference operator, i.e., \( \Delta f_n = f_{n+1} - f_n \). Some basic properties of completely monotone functions can be found in Feller (1966).

**Theorem 1.** Let \( X_i, i = 1, 2, \ldots \), be i.i.d. random variables with distribution \( F \), mean \( \mu \), and variance \( \sigma^2 \in (0, \infty) \). Denote
\[
h(n) = \frac{1}{2} \log(2\pi e) - H \left( \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^{n} (X_i - \mu) \right), \quad n = 1, 2, \ldots
\]
Then \( h(n) \) is a completely monotone function of \( n \) if \( F \) is either a gamma distribution or an inverse Gaussian distribution.

Note that in Theorem 1, \( h(n) \) is the also the relative entropy, or Kullback–Leibler divergence, between \( (1/\sqrt{n\sigma^2}) \sum_{i=1}^{n} (X_i - \mu) \) and the standard normal. Theorem 1 indicates that the behavior of relative entropy in the classical CLT is (at least for gamma and inverse Gaussian families) highly regular in a sense.

In Sections 2 and 3, we derive Theorem 1 for the gamma and inverse Gaussian families respectively. Part of the reason these two families are chosen is that they are analytically tractable. It seems intuitive that Theorem 1 should hold for a wide class of distributions, but the proofs presented here depend heavily on properties of the gamma and inverse Gaussian distributions, and are unlikely to work for the general situation. How Theorem 1 can be extended is an open problem.

2. The gamma case

Let \( H_g(\alpha, \beta), \alpha > 0, \beta > 0 \), be the entropy of a Gam(\( \alpha, \beta \) random variable whose density is specified by
\[
p(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha-1} e^{-\beta x}, \quad x > 0,
\]
where \( \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \) is Euler’s gamma function. If \( X_i \)'s are i.i.d. with a Gam(\( \alpha, \beta \)) distribution, then the entropy of the standardized sum is
\[
H \left( \frac{\beta}{\sqrt{n\alpha}} \sum_{i=1}^{n} \left( X_i - \frac{\alpha}{\beta} \right) \right) = H_g \left( n\alpha, \sqrt{n\alpha} \right).
\]

With a slight abuse of notation let \( H_g(\alpha) = H_g(\alpha, \sqrt{\alpha}) \). It is easy to see that Theorem 2 below implies Theorem 1 if the common distribution of \( X_i \) belongs to the gamma family. \( \square \)

**Theorem 2.** The function \((1/2) \log(2\pi e) - H_g(\alpha)\) is completely monotonic on \( \alpha \in (0, \infty) \).

**Proof.** Direct calculation gives
\[
H_g(\alpha) = \log \Gamma(\alpha) + \alpha - \frac{1}{2} \log(\alpha) + (1 - \alpha) \psi(\alpha)
\]
(2)
where \( \psi(\alpha) \) is the digamma function defined by \( \psi(\alpha) = \Gamma'(\alpha) / \Gamma(\alpha) \). It is clear that \((1/2) \log(2\pi e) > H_g(\alpha)\), because the standard normal achieves maximum entropy among continuous distributions with unit variance. Moreover,
\[
H_g'(\alpha) = (1 - \alpha) \psi'(\alpha) + 1 - \frac{1}{2\alpha},
\]
(3)
By Leibniz’s rule, for \( k \geq 2, \)
\[
(-1)^k H_g^{(k)}(\alpha) = (-1)^k \left[ (1 - \alpha)\psi^{(k)}(\alpha) - (k - 1)\psi^{(k-1)}(\alpha) + \frac{(-1)^k (k - 1)!}{2\alpha^k} \right].
\]
(4)
The function \( \psi^{(k)}(\alpha), k \geq 1 \), admits an integral representation (Abramowitz and Stegun, 1964, p. 260)
\[
\psi^{(k)}(\alpha) = (-1)^{k+1} \int_0^\infty \frac{t^k e^{-\alpha t}}{1 - e^{-t}} dt.
\]
(5)
Lemma 1. If \( k \geq 2 \) then
\[
\alpha \psi^{(k)}(\alpha) + k \psi^{(k-1)}(\alpha) = (-1)^k \int_0^\infty \frac{t^k e^{-at}}{(1-e^{-t})^2} dt.
\] (6)

For \( k = 1 \) we have
\[
\alpha \psi'(\alpha) = 1 + \int_0^\infty \frac{1 - e^{-t} - te^{-t}}{(1-e^{-t})^2} e^{-at} dt.
\] (7)

Proof. Using (5) and integration by parts, we get
\[
k \psi^{(k-1)}(\alpha) = (-1)^k \int_0^\infty \frac{kt^{k-1} e^{-at}}{1-e^{-t}} dt
\]
\[
= (-1)^{k+1} \int_0^\infty t^k \left[ -\alpha e^{-at} - \frac{e^{-at}}{1-e^{-t}} \right] dt
\]
\[
= -\alpha \psi^{(k)}(\alpha) + (-1)^k \int_0^\infty \frac{t^k e^{-at}}{(1-e^{-t})^2} dt
\]
which proves (6). The proof of (7) is similar. \( \square \)

In view of (4)–(6), we have \( k \geq 2 \)
\[
(-1)^k H_s^{(k)}(\alpha) = (-1)^k \left[ \psi^{(k)}(\alpha) + \psi^{(k-1)}(\alpha) + (-1)^{k+1} \int_0^\infty \frac{t^k e^{-at}}{(1-e^{-t})^2} dt + \frac{(-1)^k(k-1)!}{2\alpha^k} \right]
\]
\[
= \int_0^\infty \left[ \frac{1-t}{1-e^{-t}} - \frac{te^{-t}}{(1-e^{-t})^2} + \frac{1}{2} \right] t^{k-1} e^{-at} dt
\]
\[
= \int_0^\infty \left[ \frac{1}{1-e^{-t}} - \frac{t}{(1-e^{-t})^2} + \frac{1}{2} \right] t^{k-1} e^{-at} dt.
\] (8)

A parallel calculation using (7) reveals that the expression (8) is also valid for \( k = 1 \).

Lemma 2. The function
\[
u(t) = \frac{1}{1-e^{-t}} - \frac{t}{(1-e^{-t})^2} + \frac{1}{2}
\]
is negative on \( t \in (0, \infty) \).

Proof. Put \( s = 1 - e^{-t} \). Then
\[
u(t) = \frac{s + s^2/2 + \log(1-s)}{s^2} < 0,
\]
given the elementary inequality \( \log(1-s) < -s - s^2/2 \) for \( 0 < s < 1 \). \( \square \)

Lemma 2 immediately yields
\[
(-1)^k H_s^{(k)}(\alpha) < 0, \quad k \geq 1,
\]
thus completing the proof of Theorem 2.

As an illustration, Fig. 1 displays the functions \( H_s^{(k)}(\alpha) \) on \((1, 10)\) for \( k = 0, 1, 2, 3 \). The calculations are based on (2)–(4). The monotone behavior is in clear agreement with Theorem 2.

### 3. The inverse Gaussian case

The inverse Gaussian distribution \( IG(\mu, \lambda), \mu > 0, \lambda > 0 \), has density
\[
p(x; \mu, \lambda) = \left( \frac{\lambda}{2\pi x^3} \right)^{1/2} \exp \left[ -\frac{\lambda(x-\mu)^2}{2\mu^2 x} \right], \quad x > 0.
\] (9)

The mean is \( \mu \) and the variance is \( \mu^3/\lambda \). Some well-known properties are listed below; further details can be found in Seshadri (1993).
The function $H_\alpha(g)$, $g = 0, 1, 2, 3$, on $\alpha \in (1, 10)$.

**Lemma 3.** The inverse Gaussian family is closed under scaling and (i.i.d.) convolution. Specifically,
- if $X \sim IG(\mu, \lambda)$ and $a > 0$ is a constant, then $aX \sim IG(a\mu, a\lambda)$;
- if $X_1, \ldots, X_n$ are i.i.d. random variables distributed as $IG(\mu, \lambda)$, then $\sum_{i=1}^n X_i \sim IG(n\mu, n^2\lambda)$.

**Lemma 4.** If $X \sim IG(\mu, \lambda)$ then $\lambda(X - \mu)^2/\mu^2X$ has a $\chi^2$ distribution with one degree of freedom.

Let $H_\alpha(\mu, \lambda)$ denote the entropy of an $IG(\mu, \lambda)$ random variable. By **Lemma 3**, for an i.i.d. sequence $X_i \sim IG(\mu, \lambda)$, the entropy of the standardized sum is

$$H\left((n\lambda/\mu)^{1/2}, (n\lambda/\mu)^{3/2}\right).$$

It is then clear that **Theorem 3** implies **Theorem 1** in the inverse Gaussian case.

**Theorem 3.** The function $(1/2) \log(2\pi e) - H_\alpha(\theta^{1/2}, \theta^{3/2}), \theta \in (0, \infty)$, is a completely monotone function of $\theta$.

**Proof.** Let $f(\theta) = (1/2) \log(2\pi e) - H_\alpha(\theta^{1/2}, \theta^{3/2})$. As in the proof of **Theorem 2**, we know $f(\theta) > 0$ and only need to show $(-1)^k f^{(k)}(\theta) \geq 0$ for all $k \geq 1$. Write $\theta = \theta^{1/2}$ for notational convenience and let $X \sim IG(\mu, \mu^3)$. Direct calculation using (9) gives

$$f' = \frac{1}{2} \log(2\pi e) + E \log p(X; \mu, \mu^3)$$

$$= \frac{1}{2} - \frac{3}{2} E \log(X/\mu) - \frac{\mu(X - \mu)^2}{2X}$$

$$= -\frac{3}{2} E \log(X/\mu)$$

where **Lemma 4** is used in the last equality. To calculate $E \log(X/\mu)$, let us recall the moment generating function of $X$:

$$E e^{tX} = \exp \left[ \mu^2 \left( 1 - \sqrt{1 - 2t/\mu} \right) \right].$$

Using this and the integral identity

$$\log(a) = \int_0^\infty \frac{e^{-t} - e^{-at}}{t} dt, \quad a > 0,$$
we may express $E \log(X/\mu)$ as

$$E \log(\mu X/\mu^2) = -2 \log(\mu) + E \int_0^\infty \frac{e^{-t} - e^{-\mu X}}{t} \, dt$$

$$= -2 \log(\mu) + \int_0^\infty \frac{e^{-t} - e^{t(1-\sqrt{1+2t})}}{t} \, dt$$

$$= - \log(\theta) + \int_0^\infty \frac{e^{-t} - e^{\theta(1-\sqrt{1+2t})}}{t} \, dt.$$ 

Thus

$$\frac{dE \log(X/\mu)}{d\theta} = -\frac{1}{\theta} + \int_0^\infty \frac{\sqrt{1+2t} - 1}{t} e^{\theta(1-\sqrt{1+2t})} \, dt.$$ 

By a change of variables $s = \sqrt{1+2t} - 1$ in the above integral, we obtain

$$\frac{dE \log(X/\mu)}{d\theta} = -\frac{1}{\theta} + \int_0^\infty \frac{s(s+1)}{s^2/2 + s} e^{-\theta s} \, ds$$

$$= -\int_0^\infty e^{-\theta s} \, ds + \int_0^\infty \frac{2(1+s)}{s+2} e^{-\theta s} \, ds$$

$$= \int_0^\infty \frac{s}{s+2} e^{-\theta s} \, ds.$$ 

It is then clear that

$$(-1)^k \frac{d^k E \log(X/\mu)}{d\theta^k} = -\int_0^\infty \frac{s^k}{s+2} e^{-\theta s} \, ds < 0$$

for all $k \geq 1$; in other words

$$(-1)^k f^{(k)}(\theta) > 0, \quad k \geq 1. \quad \square$$

**Acknowledgment**

The author would like to thank an anonymous referee for his/her valuable comments.

**References**