Pergamon

Appl. Math. Lett. Vol. 11, No. 2, pp. 69–74, 1998

0893-9659/98 \$19.00 + 0.00

PII: S0893-9659(98)00013-5

Integral Representations and Asymptotic Expansions for Shannon and Renyi Entropies

C. KNESSL
Department of Mathematics, Statistics and Computer Science (M/C 249)
University of Illinois at Chicago
851 South Morgan Street, Chicago, IL 60607-7045, U.S.A.

(Received May 1997; accepted June 1997)

Communicated by B. J. Matkowsky

Abstract—We derive integral representations for the Shannon and Renyi entropies associated with some simple probability distributions. These include the Poisson, binomial, and negative binomial distributions. Then we obtain full asymptotic expansions for the entropies.

Keywords—Entropy, Information theory, Asymptotic expansions.

1. INTRODUCTION

Given a probability distribution $p_{n,k} = \Pr\{Y(n) = k\}$, the Shannon entropy is defined by

$$E(n) = -\sum_k p_{n,k} \log(p_{n,k}). \tag{1.1}$$

Here $Y(n)$ is a random variable that depends upon n, e.g., $Y(n)$ could be the sum of n i.i.d.s. The Renyi entropy is defined by

$$E(n;\omega) = \frac{1}{1-\omega} \log\left[\sum_k p_{n,k}^{\omega}\right], \tag{1.2}$$

and we clearly have $E(n;\omega) \to E(n)$ as $\omega \to 1$.

A central problem of information theory [1–3] is the computation of these entropy functions, especially asymptotically as $n \to \infty$. In some applications, only very rough bounds are used. For example, for a binomial distribution ($p_{n,k} = \binom{n}{k} p^k (1-p)^{n-k}$), the following bounds hold if $p = 1/2$ (see [4]): $(1/2)\log(\pi n/2) \le E(n) \le (1/2)\log(\pi e n/2)$. Thus, as $n \to \infty$, $E(n) \sim (1/2)\log n$, and this also holds for a large class of discrete distributions. Recently [5], there has been some interest in obtaining more precise asymptotic estimates for these entropies. In [5], the authors showed that for a wide class of distributions and $n \to \infty$,

$$E(n) = \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log\left(2\pi\sigma^2\right) + o(1), \tag{1.3}$$

We would like to thank W. Szpankowski for making available the preprint [5].

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TeX

where $Y(n)$ has variance $n\sigma^2$ and has a normal approximation as $n \to \infty$. For the binomial distribution the $o(1)$ term in (1.3) was characterized as an asymptotic series in powers of n^{-1}, i.e., $\sum_{\ell=0}^{\infty} a_\ell n^{-\ell-1}$. The last result was obtained using a method called "analytical depoissonization", which involves generating functions and the estimation of complex contour integrals.

In this paper, we obtain simple (exact) integral representations for the Shannon entropy of some discrete distributions, such as Poisson, binomial, and negative binomial. From these, full asymptotic series as $n \to \infty$ for $E(n)$ are obtained. The coefficients in these series are expressed in terms of the Taylor coefficients of elementary functions. Then we consider general $Y(n) = X_1 + \cdots + X_n$, where X_j are i.i.d.s. We calculate the first correction to (1.3) and obtain the analogous term for the Renyi entropy $E(n;\omega)$.

2. POISSON DISTRIBUTION

We illustrate our method by first considering the Poisson distribution, where

$$p_{n,k} = e^{-n\lambda} \frac{(n\lambda)^k}{k!}, \qquad k = 0, 1, 2, \dots. \tag{2.1}$$

The Shannon entropy is thus

$$\begin{aligned} E(n) &= -\sum_{k=0}^{\infty} e^{-n\lambda} \frac{(n\lambda)^k}{k!} \left[-n\lambda + k\log(\lambda n) - \log(k!)\right] \\ &= n\lambda - n\lambda \log(n\lambda) + \sum_{k=0}^{\infty} \log(k!) e^{-n\lambda} \frac{(n\lambda)^k}{k!}. \end{aligned} \tag{2.2}$$

We shall represent $E(n)$ as an integral, from which the asymptotic series as $n \to \infty$ may be easily derived. We start from the elementary identity [6, p. 378]

$$\log A = \int_0^\infty \frac{e^{-x} - e^{-Ax}}{x}\, dx, \qquad A > 0, \tag{2.3}$$

set $A = j$ and sum from $j = 1$ to $j = \ell$. This yields

$$\log(\ell!) = \sum_{j=1}^{\ell} \log j = \int_0^\infty \left(\ell - \frac{1 - e^{-\ell x}}{1 - e^{-x}}\right) \frac{e^{-x}}{x}\, dx. \tag{2.4}$$

Using (2.4) to represent $\log(k!)$ in (2.2) as an integral and then evaluating the sum over k gives

$$E(n) = n\lambda - n\lambda\log(n\lambda) + \int_0^\infty \frac{e^{-x}}{x} \left[n\lambda - \frac{1 - \exp\left(n\lambda\left(e^{-x} - 1\right)\right)}{1 - e^{-x}}\right] dx. \tag{2.5}$$

We write the integral in (2.5) as $I = \lim_{\epsilon \to 0}(P_1 + P_2 + P_3)$, where

$$\begin{aligned} P_1 &= n\lambda \int_\epsilon^\infty \frac{e^{-x}}{x}\, dx, \qquad P_2 = -\int_\epsilon^\infty \frac{e^{-x}}{1 - e^{-x}} \frac{1}{x}\, dx, \\ P_3 &= \int_\epsilon^\infty \frac{e^{-x}}{x} \frac{1}{1 - e^{-x}} \exp\left[n\lambda(e^{-x} - 1)\right] dx. \end{aligned} \tag{2.6}$$

For small ϵ, we have

$$P_1 = n\lambda\left[-\log\epsilon - \gamma + o_\epsilon(1)\right], \tag{2.7}$$

where γ is the Euler constant and $o_\epsilon(1)$ denotes terms that vanish as $\epsilon \to 0$. Similarly, for P_2, we obtain

$$\begin{aligned} P_2 &= -\int_\epsilon^\infty \frac{e^{-x}}{x} \left[\frac{1}{1 - e^{-x}} - \frac{1}{x} - \frac{1}{2}\right] dx - \int_\epsilon^\infty e^{-x} \left[\frac{1}{x^2} + \frac{1}{2x}\right] dx \\ &= -1 + \frac{1}{2}\log(2\pi) - \frac{1}{\epsilon} - \frac{1}{2}\log\epsilon + 1 - \frac{\gamma}{2} + o_\epsilon(1). \end{aligned} \tag{2.8}$$

Here we have set $\epsilon = 0$ in the first integral, explicitly evaluated it as $1 - \log\sqrt{2\pi}$ (see [6, p. 377]) and integrated by parts in the second integral. In P_3, we change variables with $y = \lambda(1-e^{-x})$ and obtain

$$\begin{aligned} P_3 &= \int_{\lambda(1-e^{-\epsilon})}^{\lambda} e^{-ny} \frac{-1}{y\log(1-y/\lambda)}\, dy \\ &= \int_0^{\lambda} e^{-ny}\left[\frac{-1}{y\log(1-y/\lambda)} - \frac{\lambda}{y^2} + \frac{1}{2y}\right] dy + o_\epsilon(1) + \overline{P} - \int_{\lambda}^{\infty} e^{-ny}\left(\frac{\lambda}{y^2} - \frac{1}{2y}\right) dy, \end{aligned} \tag{2.9}$$

where

$$\overline{P} = \int_{\lambda(1-e^{-\epsilon})}^{\infty} e^{-ny}\left(\frac{\lambda}{y^2} - \frac{1}{2y}\right) dy. \tag{2.10}$$

Integrating by parts in (2.10) and using $1 - e^{-\epsilon} = \epsilon - \epsilon^2/2 + O(\epsilon^3)$, we obtain

$$\overline{P} = \frac{1}{\epsilon} + \left(\lambda n + \frac{1}{2}\right)\log(\lambda\epsilon) + \frac{1}{2} - n\lambda + (\log n + \gamma)\left(\lambda n + \frac{1}{2}\right) + o_\epsilon(1). \tag{2.11}$$

Using (2.7)–(2.11), we see that $P_1 + P_2 + P_3$ approaches a finite limit as $\epsilon \to 0$, which when used in (2.5) yields the following.

THEOREM 1. *For the Poisson distribution, we have*

$$\begin{aligned} E(n) &= \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log(2\pi\lambda) \\ &\quad + \int_0^{\lambda} e^{-ny}\left[\frac{-1}{y\log(1-y/\lambda)} - \frac{\lambda}{y^2} + \frac{1}{2y}\right] dy - \int_{\lambda}^{\infty} e^{-ny}\left(\frac{\lambda}{y^2} - \frac{1}{2y}\right) dy. \end{aligned}$$

This is exact for all $n \geq 1$. Now we evaluate $E(n)$ for $n \to \infty$. The second integral is exponentially small ($O(e^{-n\lambda}/n)$), and the first has the form $\int_0^a e^{-ny}F(y)\,dy$, where $F(y)$ is analytic at $y = 0$. By Watson's Lemma [7], we obtain the following.

THEOREM 2. *For $n \to \infty$, for the Poisson distribution,*

$$E(n) \sim \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log(2\pi\lambda) + \sum_{\ell=0}^{\infty} \mathcal{F}_\ell \frac{\ell!}{n^{\ell+1}},$$

where $\mathcal{F}_\ell$ are obtained from

$$\mathcal{F}(y) = \frac{-1}{y\log(1-y/\lambda)} - \frac{\lambda}{y^2} + \frac{1}{2y} = \sum_{\ell=0}^{\infty} \mathcal{F}_\ell y^\ell = \sum_{\ell=0}^{\infty} \frac{\mathcal{F}^{(\ell)}(0)}{\ell!} y^\ell,$$

and, in particular,

$$\mathcal{F}_0 = \frac{-1}{12\lambda}, \qquad \mathcal{F}_1 = \frac{-1}{24\lambda^2}.$$

3. BINOMIAL AND NEGATIVE BINOMIAL DISTRIBUTIONS

We consider $0 < p, q < 1$ with $p + q = 1$ and

$$p_{n,k} = \binom{n}{k} p^k q^{n-k}, \qquad 0 \leq k \leq n, \quad n \geq 1 \quad \text{(binomial)}, \tag{3.1}$$

$$p_{n,k} = \binom{k}{n} p^{n+1} q^{k-n}, \qquad k \geq n, \qquad n \geq 0 \quad \text{(negative binomial)}. \tag{3.2}$$

We again represent $\log(k!)$ and $\log[(n-k)!]$ as integrals. For the binomial case, this leads to

$$\begin{aligned} E(n) &= -n\,(p\log p + q\log q) \\ &\quad + \int_0^{\infty} \frac{e^{-x}}{x}\frac{1}{1-e^{-x}}\left[\left(q + pe^{-x}\right)^n + \left(p + qe^{-x}\right)^n - 1 - e^{-nx}\right] dx. \end{aligned} \tag{3.3}$$

By using ideas similar to the Poisson case, we again obtain an exact integral representation for $E(n)$, from which the asymptotics are easily obtained. Below, we give only the final results; the details may be found in [8].

THEOREM 3. *For the binomial distribution,*

$$E(n) = \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log(2\pi pq) - \log\left(\frac{\Gamma(n)}{\sqrt{2\pi}} n^{-n} e^{n} \sqrt{n}\right)$$
$$+ \int_0^{-\log(q)} e^{-n\xi}\left[\frac{1}{e^{\xi}-1}\frac{-1}{\log\left[(e^{-\xi}-q)/p\right]} - \frac{p}{\xi^2} + \frac{1}{2\xi}\right] d\xi - \int_{-\log(q)}^{\infty} e^{-n\xi}\left[\frac{p}{\xi^2} - \frac{1}{2\xi}\right] d\xi$$
$$+ \int_0^{-\log(p)} e^{-n\xi}\left[\frac{1}{e^{\xi}-1}\frac{-1}{\log\left[(e^{-\xi}-p)/q\right]} - \frac{q}{\xi^2} + \frac{1}{2\xi}\right] d\xi - \int_{-\log(p)}^{\infty} e^{-n\xi}\left[\frac{q}{\xi^2} - \frac{1}{2\xi}\right] d\xi,$$

where $\Gamma(\cdot)$ is the gamma function. As $n \to \infty$,

$$E(n) \sim \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log(2\pi pq) + \sum_{\ell=0}^{\infty} G_\ell \frac{\ell!}{n^{\ell+1}},$$

where $\sum_{\ell=0}^{\infty} G_\ell \xi^\ell = G(\xi)$ with

$$G(\xi) = \frac{1}{e^{\xi}-1}\left[\frac{-1}{\log\left[(e^{-\xi}-q)/p\right]} + \frac{-1}{\log\left[(e^{-\xi}-p)/q\right]} - \frac{1}{\xi}\right] + \frac{1}{2\xi}.$$

In particular,

$$G_0 = \frac{1}{3} - \frac{1}{12}\left(\frac{1}{p} + \frac{1}{q}\right), \qquad G_1 = \frac{1}{2}\left(\frac{1}{p} + \frac{1}{q}\right) - \frac{1}{12} - \frac{1}{24}\left(\frac{1}{p^2} + \frac{1}{q^2}\right),$$
$$G_2 = \frac{1}{180} - \frac{1}{24}\left(\frac{1}{p} + \frac{1}{q}\right) + \frac{1}{16}\left(\frac{1}{p^2} + \frac{1}{q^2}\right) - \frac{19}{720}\left(\frac{1}{p^3} + \frac{1}{q^3}\right).$$

By setting $\Delta = 1 - e^{-\xi}$ and defining c_ℓ from the Taylor series $-\Delta/\log(1-\Delta) = \sum_{\ell=0}^{\infty} c_\ell \Delta^\ell$, we can expand $G(\xi)$ in terms of c_ℓ and powers of Δ. This yields the alternate representation

$$E(n) \sim \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log(2\pi pq)$$
$$+ \sum_{m=0}^{\infty} \frac{m!(n-1)!}{(m+n)!}\left\{\frac{1}{2}c_{m+1} - c_{m+1}\left[p^{-m} + q^{-m} - 1\right] + c_{m+2}\left[p^{-m-1} + q^{-m-1} - 1\right]\right\},$$

as $n \to \infty$. We have $c_0 = 1$, and for $m \geq 1$, $c_m = -\sum_{j=0}^{m-1} c_j/(m+1-j)$.

Analogous results for the negative binomial case are given below.

THEOREM 4. *For the negative binomial distribution (with (3.2)),*

$$E(n) = \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log(2\pi q p^{-2}) + 1 - n\log\left(1 + \frac{1}{n}\right) + \log\left(\frac{\Gamma(n)}{\sqrt{2\pi}} n^{-n} e^{n} \sqrt{n}\right)$$
$$+ \int_0^{-\log(p)} e^{-ny}\left[\frac{1}{e^{y}-1}\frac{-1}{\log\left[(1-pe^{y})/q\right]} - \frac{qe^{-y}}{py^2} + \frac{e^{-y}}{2y}\right] dy$$
$$+ \int_0^{\infty} e^{-ny}\left[\frac{1}{e^{y}-1}\frac{-1}{\log(q+pe^{y})} + \frac{e^{-y}}{py^2} + \frac{e^{-y}}{2y}\right] dy - \int_{-\log(p)}^{\infty} e^{-(n+1)y}\left(\frac{q}{py^2} - \frac{1}{2y}\right) dy,$$

and as $n \to \infty$

$$E(n) \sim \frac{1}{2}\log n + \frac{1}{2} + \frac{1}{2}\log\left(2\pi q p^{-2}\right) + \sum_{\ell=0}^{\infty} H_\ell \frac{\ell!}{n^{\ell+1}},$$

where $\sum_{\ell=0}^{\infty} H_\ell y^\ell = H(y)$ with

$$H(y) = \frac{1}{e^{y}-1}\left[\frac{-1}{\log\left[(1-pe^{y})/q\right]} + \frac{-1}{\log(q+pe^{y})} + \frac{1}{y}\right] + \frac{1}{2y}.$$

In particular, $H_0 = 1/6 - p^2/(12q) = -(q + q^{-1} - 4)/12$.

We have thus obtained the full asymptotic series in a simple form. This method should also apply to other discrete distributions that involve factorial and/or binomial coefficient factors.

4. LEADING TERM FOR GENERAL i.i.d.s

We now consider general discrete distributions and obtain the first two terms in the expansion of $E(n) - (1/2)\log n$, and a corresponding result for the Renyi entropy.

Let X_1 be a random variable with density $f(x)$ and moment generating function $F(\theta) = E[e^{\theta X_1}]$. Then for i.i.d. X_j, we set $Y = Y(n) = X_1 + X_2 + \cdots + X_n$. The density of Y is

$$g(y;n) = \frac{1}{2\pi i}\int_{\mathrm{Br}} e^{-\theta y}\,[F(\theta)]^n\,d\theta, \tag{4.1}$$

where Br is a vertical contour in the complex θ-plane. Assuming the first few moments of X_1 are finite, we set $E[X_1^k] = m_k$, $y = nm_1 + \sqrt{n}z$, and expand (4.1) for θ small. Scaling $\theta = O(n^{-1/2})$ leads to the "central limit theorem" approximation (see, also, [9,10])

$$g(y;n) = \frac{1}{\sqrt{2\pi\sigma^2 n}}\exp\left[-\frac{z^2}{2\sigma^2}\right]\left\{1 + \frac{f_*}{\sqrt{n}} + \frac{g_*}{n} + \cdots\right\}, \tag{4.2}$$

where $\sigma^2 = m_2 - m_1^2$ is the variance of X_1 and

$$f_* = \frac{\kappa_3}{6}\frac{1}{\left(\sqrt{2}\sigma\right)^3}H_3\left(\frac{z}{\sqrt{2}\sigma}\right), \qquad g_* = \frac{\kappa_4}{24}\frac{1}{\left(\sqrt{2}\sigma\right)^4}H_4\left(\frac{z}{\sqrt{2}\sigma}\right) + \frac{\kappa_3^2}{72}\frac{1}{\left(\sqrt{2}\sigma\right)^6}H_6\left(\frac{z}{\sqrt{2}\sigma}\right). \tag{4.3}$$

Here $H_k(\cdot)$ is the k^{th} Hermite polynomial and κ_j is the j^{th} cummulant of X_1. Thus, $H_3(x) = 8x^3 - 12x$, $H_4(x) = 16x^4 - 48x^2 + 12$, and

$$\kappa_3 = m_3 - 3m_1m_2 + 2m_1^3, \qquad \kappa_4 = m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4. \tag{4.4}$$

For $n \to \infty$, we use (4.2) to evaluate (1.1), noting that the sum may be approximated by $\sqrt{n}$ times an integral from $z = -\infty$ to $z = +\infty$, with an error that is exponentially small. We thus obtain the following.

THEOREM 5. *For a general discrete distribution with finite moments, we have as $n \to \infty$,*

$$S(n) \equiv E(n) - \frac{1}{2}\log n - \left[\frac{1}{2} + \frac{1}{2}\log\left(2\pi\sigma^2\right)\right] \sim -\frac{\kappa_3^2}{12\sigma^6}\frac{1}{n}.$$

If the third cummulant vanishes (e.g., binomial case with $p = q = 1/2$), then $S(n) \sim -\kappa_4^2/(48\sigma^8)n^{-2}$. If $\kappa_3 = \kappa_4 = \cdots = \kappa_N = 0$ but $\kappa_{N+1} \neq 0$, then $S(n) \sim -\kappa_{N+1}^2/[2(N+1)!\,\sigma^{2N+2}]n^{1-N}$.

A full asymptotic series for $S(n)$ may be obtained by first deriving the full series in (4.2) (as was done in [9,10] for related problems) and using the result in (1.1).

Theorem 5 also applies to continuous distributions, then $E(n) = -\int g(y;n)\log[g(y;n)]\,dy$. For example, if X_1 is exponential with density $e^{-x}(x \geq 0)$, then we obtain (exactly) $E(n) = \log\Gamma(n) + n - (n-1)\psi(n)$, where $\Gamma(\cdot)$ is the gamma function and $\psi(n) = \Gamma'(n)/\Gamma(n)$. Now $S(n) \sim -1/(3n)$. For Gaussian X_1, we clearly have $S(n) = 0$. If X_1 is geometric, i.e., $f(x) = \sum_{k=1}^{\infty} pq^{k-1}\delta(x-k)$, then $g(y;n) = \sum_{k=n}^{\infty}\binom{k-1}{n-1}p^n q^{k-n}\delta(y-k)$ and Theorem 5 gives $S(n) \sim -(q + q^{-1} + 2)/(12n)$. This agrees with Theorem 4 if we note that the entropy $\widetilde{E}(n)$ of this distribution is related to the entropy $E(n)$ of (3.2) by $\widetilde{E}(n+1) = E(n)$.

Using (4.2), we obtain the Renyi entropy.

THEOREM 6. *Under the same assumptions as Theorem 5,*

$$S(n;\omega) \equiv E(n;\omega) - \left[\frac{1}{2}\log n + \frac{1}{2}\log\left(2\pi\sigma^2\right) + \frac{\log\omega}{2(\omega-1)}\right]$$
$$\sim \frac{1}{\omega}\left[\frac{\kappa_4}{8\sigma^4}(1-\omega) + \frac{\kappa_3^2}{12\sigma^6}(\omega-2)\right]\frac{1}{n}.$$

As $\omega \to 1$, this reduces to Theorem 5 for $\kappa_3 \neq 0$.

REFERENCES

1. A. Renyi, On measures of entropy and information, In *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, pp. 547–561, (1961).
2. A. Renyi, On the foundations of information theory, *Rev. Inst. Intern. Statist.* **33**, 1–14, (1965).
3. I. Csiszar, Generalized cut-off rates and Renyi information measures, *IEEE Trans. Information Theory* **41**, 26–34, (1995).
4. S. Chang and E. Weldon, Coding for T-user multiple-access channels, *IEEE Trans. Information Theory* **25**, 684–691, (1979).
5. P. Jacquet and W. Szpankowski, Toward analytical information theory: Entropy computations, (preprint).
6. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*, Fifth Edition, Academic Press, Boston, MA, (1994).
7. N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals*, Dover, New York, (1986).
8. C. Knessl, University of Illinois at Chicago, handwritten notes (available upon request).
9. J.L. Jensen, Saddlepoint expansions for sums of Markov dependent variables on a continuous state space, *Prob. Theory and Related Fields* **89**, 181–199, (1991).
10. H.D. Miller, A convexity property in the theory of random variables on a finite Markov chain, *Ann. Math. Statist.* **32**, 1260–1270, (1961).