Approximations to Entropy Functions: Entropic Polynomials

Rama Murthy Garimella and Ravi Kishore

May 24, 2020

Approximations to Entropy Functions: Entropic Polynomials

Garimella Rama Murthy
Mahindra École Centrale
Hyderabad, India - 500049
Email: rama.murthy@mechyd.ac.in

Ravi Kishore
Mahindra École Centrale
Hyderabad, India - 500049
Email: ravikishore.vasala@mechyd.ac.in

***Abstract*—In this research paper, it is proved that linear/quadratic approximations to Shannon/Gibbs entropy lead to *Tsallis* entropy, $S_q(p)$ for $q = 2/q = 3$. Based on higher degree approximations of logarithm, entropic polynomials are derived. Linear approximation to Renyi entropy is also determined.**

I. Introduction

Boltzmann introduced the concept of "$entropy$" in an effort to innovate the field of statistical mechanics. In the formulation of Boltzmann, entropy of a uniform probability mass function was defined. Gibbs, Shannon generalized the concept of entropy for an arbitrary probability mass function. Shannon placed "$information\ theory$" on a sound mathematical basis [Ash]. Various other types of entropy such as $Renyi$ entropy were defined and their properties are explored.

In recent years, $Tsallis$ introduced an "$entropy\ measure$" in an effort to generalize statistical mechanics. In [Rama 1], the author showed that a linear approximation to logarithmic function will approximate $Shannon/Gibbs$ entropy, $H(X)$ with $Tsallis$ entropy $S_q(p)$ with $q = 2$. In this research paper, based on higher degree approximation of logarithm, Shannon entropy is approximated by structured polynomials. It is also shown that linear approximation to Renyi entropy also leads to Tsallis entropy under some conditions.

This research paper is organized as follows. In section 2 motivation for approximations of Shannon/Gibbs entropy is discussed. In section 3, based on higher degree approximation of logarithm function, Shannon entropy is shown to lead to structured polynomials with some properties. In section 4, Tsallis entropy is shown to result as a linear approximation to Renyi entropy under some conditions. The research paper concludes in section 5.

II. Motivation for Approximations

From the considerations of statistical physics, Tsallis introduced a new entropy measure. For a long time, it is not clear how such a novel entropy measure is related to Gibbs/Shannon entropy. This research article sheds light on such a question. More interestingly, using higher order approximations, it is reasoned that NOVEL polynomial approximations to Gibbs/Shannon entropy result naturally.

III. Approximations to Shannon and Gibbs Entropy

Lemma 1. *Consider a discrete random variable X with finite support for the probability mass function. Under reasonable assumptions, we have that* $H(X) \approx \left(1 - \sum_{i=1}^{m} p_i^2\right) \log_2 e$.

Proof. From the basic theory of infinite series, for $|x| < 1$, we have that:

$$\log_e(1-x) = -x + \frac{x^2}{2} - \frac{x^3}{3} + \frac{x^4}{4} - \frac{x^5}{5} + \tag{1}$$

$$\ldots + (-1)^n \frac{x^n}{n} + \ldots . \tag{2}$$

$$\tag{3}$$

Let $p_i = (1 - q_i)$ with $0 < p_i < 1$; then we have $0 < q_i < 1$. This implies:

$$\log_e(1-q_i) = -q_i + \frac{q_i^2}{2} - \frac{q_i^3}{3} + \frac{q_i^4}{4} - \frac{q_i^5}{5} + \tag{4}$$

$$\ldots + (-1)^n \frac{q_i^n}{n} + \ldots . \tag{5}$$

Now let us consider the entropy, $H(X)$ of a discrete random variable X which assumes finitely many values. We have that

$$\begin{aligned} H(X) &= -\sum_{i=1}^{m} p_i \log_2 p_i \\ &= -\sum_{i=1}^{m} (1-q_i) \log_2 (1-q_i) \\ &= -\sum_{i=1}^{m} (1-q_i) \log_e (1-q_i) \log_2 e \end{aligned}$$

Now using the above infinite series and neglecting the terms $\frac{q_i^2}{2}, \frac{q_i^3}{3}, \frac{q_i^4}{4}, \ldots$ we have

$$\begin{aligned} H(X) &\approx -\sum_{i=1}^{m} (1-q_i)(-q_i) \log_2 e \\ &= \sum_{i=1}^{m} (1-q_i)(q_i) \log_2 e \end{aligned}$$

$$= \sum_{i=1}^{m} p_i(1-p_i)\log_2 e$$

$$= \Big(1 - \sum_{i=1}^{m} p_i^2\Big)\log_2 e$$

Remark 1: In the above approximation, the error term is $\sum\limits_{j>=2} (-1)^j \frac{q_i^j}{j}$. Which is same as $q_i^2(\frac{1}{2} - \frac{q_i}{3}) + q_i^4(\frac{1}{4} - \frac{q_i}{5}) + q_i^6(\frac{1}{6} - \frac{q_i}{7}) + \ldots$

It can be upper bounded by a geometric series $q_i^2 + q_i^4 + q_i^6 + \ldots = \frac{q_i^2}{1-q_i^2}$. Thus, the approximations is good under some conditions.

Remark 2: Thus, the square of the $L^2 - norm$ of the vector corresponding to the probability mass function (of a discrete random variable) is utilized to approximate the entropy of the discrete random variable. In summary, we have that

$H(X) \approx f(p_1, p_2, \ldots, p_m) = \Big(1 - \sum\limits_{i=1}^{m} p_i^2\Big)\log_2 e.$

Thus, an approximation to Gibbs-Shannon entropy naturally leads to the scaled Tsallis entropy for the real parameter $q = 2$. The quantity $H(X)$ with the above approximation is rounded-off to the nearest integer [Rama 4]. For continuous case i.e., for probability density functions associated with continuous random variables, similar results can easily be derived.

We readily have that:

$$H(X) \approx -\sum_{i=1}^{m} p_i\Big[-(1-p_i) + \frac{(1-p_i)^2}{2} - \frac{(1-p_i)^3}{3} + \ldots\Big].$$

$$H(X) \approx -\sum_{i=1}^{m} p_i(1-p_i)\Big[-1 + \frac{(1-p_i)}{2} - \frac{(1-p_i)^2}{3} + \ldots\Big].$$

$$H(X) \approx \big(1 - \sum_{i=1}^{m} p_i^2\big) - \sum_{i=1}^{m} p_i(1-p_i)^2\big[\frac{1}{2} - \frac{(1-p_i)}{3} + \frac{(1-p_i)^2}{4} + \ldots\big].$$

It can be readily seen that quadratic approximation to $log(.)$ leads to Tsallis entropy for $q = 3$, i.e., $S_3(p)$. Now we provide higher order approximation.

Suppose we truncate the infinite series at R. Let us specifically consider the quantity $\Big[\frac{(1-p_i)}{2} - \frac{(1-p_i)^2}{3} + \ldots\Big]$. Using Binomial theorem, we can express the quantity as follows.

$$\Big[\frac{(1-p_i)}{2} - \frac{(1-p_i)^2}{3} + \ldots\Big] = \frac{1}{2} - \sum_{i=3}^{R}(-1)^i \frac{(1-p_i)^{i-2}}{i}$$

$$= \frac{1}{2} - \sum_{i=1}^{R}(-1)^i \frac{(1-p_i)^i}{i+2}$$

$$= \frac{1}{2} - \sum_{i=1}^{R} \frac{(-1)^i}{i+2}\big[\sum_{j=0}^{i}\binom{i}{j}(-1)^{i-j}(p_i)^{i-j}\big]$$

$$= \frac{1}{2} - \sum_{j=1}^{R}\sum_{i=j}^{k}\Big[\frac{(-1)^i}{i+2}\binom{i}{j}(-1)^{i-j}(p_i)^{i-j}\Big] - \sum_{i=1}^{R}\Big[\frac{(-1)^i}{i+2}\binom{i}{0}(-1)^i(p_i)^i\Big]$$

Let $h(p_i) = \sum\limits_{i=1}^{R} \frac{1}{i+2}(p_i)^i$ and $k = i - j$. Then the following holds.

$$\Big[\frac{(1-p_i)}{2} - \frac{(1-p_i)^2}{3} + \ldots\Big] = \frac{1}{2} - h(p_i) - \sum_{j=1}^{R}\Big[\sum_{k=0}^{R-j}\Big(\frac{(-1)^{2k+j}}{k+j+2}\binom{k+j}{j}(p_i)^k\Big)\Big]$$

$H(p_1, p_2, \ldots, p_n) \approx f(p_1, p_2, \ldots, p_n)$, where $f(p_1, p_2, \ldots, p_n) = \sum\limits_{i=1}^{n} g(p_i)$, where all the polynomials $\{g(p_1), g(p_2), \ldots, g(p_n)\}$ have the same coefficients, that add upto ONE. These polynomials are structured ones in the spirit of Euler, Bernoulli polynomials. Like Euler/Bernoulli numbers, the coefficients of such structured polynomials can be studied for interesting properties [Rama 2].

Remark 3: The sequence of polynomials approximating Shannon entropy are SAME for any random variable. Tsallis entropy is a special case where only the constant coefficients and q^{th} coefficient in q_i are considered and all other coefficients are zero.

Tsallis entropy: $S_q(p) = \frac{1}{(q-1)} - \frac{1}{(q-1)}\sum\limits_{i=1}^{n} p_i^q$. Our approximation $f(p_1, p_2, \ldots, p_n) = \sum\limits_{i=1}^{n} g(p_i)$.

A. Algebraic Interpretations of Entropy Functions

1) Shannon/Gibbs entropy $H(X) = -\sum\limits_{i=1}^{n} p_i \log_e p_i$.
2) Tsallis entropy for real parameter q: $S_q(p) = \frac{1}{(q-1)}(1 - \sum\limits_{i=1}^{n} p_i^q) = S_q(p_1, p_2, \ldots, p_n)$.

   Our contribution: $H(p_1, p_2, \ldots, p_n) = \sum\limits_{i=1}^{n} g^{(r)}(p_i)$, where $g^{(r)}(.)$ is a polynomial (in p_i) (for any arbitrary polynomial) and r is the degree at which $\log(1 - q_i)$ (where $q_i = 1 - p_i$) is truncated. i.e., We have a sequence of polynomials as r increased providing a better approximation to Shannon entropy of any random variable . That is, Coefficients of polynomials are independent of the Probability Mass Function (PMF). i.e., sequence of polynomials providing better approximation have same coefficients for any PMF.

□

We are currently exploring properties of such entropic polynomials.

IV. Relationship between Renyi Entropy and Tsallis Entropy

We now reason that Renyii entropy is approximated by Tsallis Entropy under some conditions [Rama 3].

Definition 1 (Renyi entropy)**.** *Renyi entropy (of a discrete random variable X) of order* α*, where* $\alpha \geq 0$ *and* $\alpha \neq 1$ *is defined as* $H_\alpha(X) = \frac{1}{1-\alpha} \log\big(\sum_{i=1}^{n} p_i^\alpha\big)$.

We can rewrite $H_\alpha(X)$ as follows.
$H_\alpha(X) = \frac{1}{1-\alpha} \log\Big(1 - \big(1 - \sum_{i=1}^{n} p_i^\alpha\big)\Big)$.
Letting $1 - \sum_{i=1}^{n} p_i^\alpha = r$, we have that $H_\alpha(X) = \frac{1}{1-\alpha}\log(1-r)$. However, from the basic theory of infinite series [Kno], we have that: $\log(1-r) = -r + \frac{r^2}{2} - \frac{r^3}{3} + \ldots$ for $|r| < 1$. We consider non-degenerate probability mass functions (PMF's). For such PMF's it readily follows that for $\alpha \geq 1$ and $0 < r < 1$. Thus, if we truncate the infinite series for $\log(1-r)$, we have that $\log(1-r) \approx -r$ for $|r| < 1$.

Hence, it readily follows that with such approximation, we have:

$$\begin{aligned} H_\alpha(X) &\approx \frac{1}{1-\alpha}\Big(-\big(1 - \sum_{i=1}^{n} p_i^\alpha\big)\Big) \\ &= \frac{1}{\alpha-1}\Big(\big(1 - \sum_{i=1}^{n} p_i^\alpha\big)\Big) \\ &= S_\alpha(p). \end{aligned}$$

Where $S_\alpha(p)$ is Tsallis entropy with α is the real parameter. Now, we bound the error term in approximating $\log(1-r)$ by $-r$. The error term is $\frac{r^2}{2} - \frac{r^3}{3} + \frac{r^4}{4} - \frac{r^5}{5} + \ldots = r^2(\frac{1}{2} - \frac{r}{3}) + r^4(\frac{1}{4} - \frac{r}{5} + \ldots)$ with $|r| < 1$. Thus, the error term can be bounded by the following geometric series. i.e., $r^2 + r^4 + r^6 + \ldots = \frac{r^2}{1-r^2}$.
Remark 4: The above approach of approximating entropy (such as Shannon entropy) was first proposed in [Rama1], [Rama4]. Specifically Shannon entropy is approximated by Tsallis entropy for a linear approximation i.e. $\log(1-r) \approx -r$ for $|r| < 1$. It is shown that higher order approximations are different from Tsallis entropy except in the case of approximation. $\log(1-r) \approx -r + \frac{r^2}{2}$ for $|r| < 1$.

The higher order polynomial approximation of $\log(1-r)$ leads to interesting entropy functions. We are currently deriving those polynomials approximating Renyi entropy.

Note: We can consider higher order approximations in association with $\log(1-r)$ and arrive at better approximations of Renyi entropy of order α. We now consider second order approximation: $\log(1-r) \approx -r + \frac{r^2}{2}$ for $|r| < 1$.

Using this approximation, we have that

$$H_\alpha(X) \approx \frac{1}{1-\alpha}\Big(\big(-1 + \sum_{i=1}^{n} p_i^\alpha\big) + \frac{1}{2}\big(1 - \sum_{i=1}^{n} p_i^\alpha\big)^2\Big).$$

Expanding $\big(1 - \sum_{i=1}^{n} p_i^\alpha\big)^2$ and simplifying, we have that $H_\alpha(X) \approx \frac{1}{2(\alpha-1)}\Big(1 - \sum_{i=1}^{n}\sum_{j=1}^{n} p_i^\alpha p_j^\alpha\Big)$. It also can be rewritten as $H_\alpha(X) \approx \frac{1}{2(\alpha-1)}\Big(\big(\sum_{i=1}^{n} p_i\big)^\alpha - \sum_{i=1}^{n}\sum_{j=1}^{n} p_i^\alpha p_j^\alpha\Big)$. In the above expression, multinomial theorem can be used for further simplification. Simplifying the above, we have that $H_\alpha(X) \approx \frac{1}{2(\alpha-1)}\Big(1 - \sum_{k=1}^{n} p_k^{2\alpha} - \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} p_i^\alpha p_j^\alpha\Big)$.

Using the definition of Tsallis entropy, we have $H_\alpha(X) \approx \frac{2\alpha-1}{2\alpha-2} S_{2\alpha}(p) - \frac{1}{2(\alpha-1)}\Big(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} p_i^\alpha p_j^\alpha\Big)$.

We now obtain an equivalent expression for Renyi Entropy. Letting $t_i = p_i^\alpha$, we arrive at the vector $\hat{t} = (t_1, t_2, \ldots, t_n)^T)$. In terms of that vector, the following approximation based on quadratic form is readily obtained $H_\alpha(X) \approx \frac{1}{2(\alpha-1}(1 - \hat{t}^T \hat{B} \hat{t})$, where $\hat{B} = \hat{e}\hat{e}^T$ with $\hat{e}$, a column vector of $1's\cdot$ Now, we consider a specific value of α, i.e., $\alpha = 2$ and arrive at an expression for approximating the Renyi entropy:
$H_2(X) \approx \frac{1}{2}\Big(\big(\sum_{i=1}^{n} p_i\big)^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} p_i^2 p_j^2\Big)$.

Initial simplification of the above expression leads to $H_2(X) \approx \frac{1}{2}\Big(S_2(p)\big(1 - S_2(p)\big) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} p_i p_j\Big)$.

On further simplification, we arrive at the following approximation for $H_2(X)$ in terms of Tsallis entropy $S_2(p)$. $H_2(X) \approx \Big(S_2(p) - \frac{1}{2}\big(S_2(p)\big)^2\Big)$.

We now briefly consider the approximation of Shannon entropy of a continuous random variable $h[f] = -\int_{-\infty}^{+\infty} f(x)\log f(x) dx$.

Note: We consider probability density functions which are bounded by 1. For example, suitably normalized exponential density i.e., we consider $0 < f(x) < 1$ fro all x. Let $g(x) = 1 - f(x)$, $0 < g(x) < 1$. Hence, $\log\big(1 - g(x)\big) = -g(x) + \frac{\big(g(x)\big)^2}{2} - \frac{\big(g(x)\big)^3}{3} + \ldots$.

Considering linear approximation, we have that $\log\big(1 - g(x)\big) \approx -g(x)$. Thus, $h[f] \approx -\int_{-\infty}^{+\infty} f(x)\big(-g(x)\big)dx \approx \int_{-\infty}^{+\infty} f(x)\big(1 - f(x)\big)dx = 1 - \int_{-\infty}^{+\infty} f^2(x)dx = S_q(f)$ for $q = 2$. Thus, Shannon entropy is approximated by Tsallis entropy.

V. Significance of Results

In Shannon's information theory, the concept of entropy plays most crucial role. In fact, mutual information definition, effectively is
$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$, where X, Y are input and output random variables of a Discrete memoryless channel(DMC). Also $H(X|Y)$ and $H(Y|X)$ are conditional entropy values. The approximations provided in this research paper potentially provide estimation of quantities like channel capacity effectively. We expect the results derived

in this research paper to be of utility in information theoretic research.

VI. Conclusion

In this research paper, based on approximating logarithmic power series, structured polynomial approximation to Shannon/Gibbs entropy are proposed. Properties of such polynomials are proved. Also, Renyi entropy leads to Tsallis entropy with a linear approximation to logarithmic function(under some conditions).

VII. References

[Ash] R.B.Ash, "Information Theory" Dover Publications, Inc, New York.

[Kno] K. Knopp, "Theory and Applications of Infinite Series" Dover Publications, Inc, New York.

[Rama 1] G. Rama Murthy, "Weakly Short Memory Stochastic Process: Signal Processing Perspectives", Proceedings of International Conference on Frontiers of Interface between Statistics and Sciences, December 20, 2009 to January 02, 2010.

[Rama 2] G. Rama Murthy, "Gibbs, Shannon Entropy and related measures: Tsallis Entropy" Easy chair Preprint #866/2018.

[Rama 3] G. Rama Murthy, "Variance Laplacian: Quadratic Forms in Statistics" Easy chair Preprint #821/2019.

[Rama 4] G. Rama Murthy and T. Jagannadh Swamy, "Gibbs-Shannon Entropy and Related Measures:Tsallis Entropy", proceedings of International Conference on Communication, Signal Processing and VLSI (IC2SV2019), Springer, October 2019.