Approximations to Gibbs/Shannon Entropy: Entropic Polynomials

Rama Murthy Garimella and Ravi Kishore

September 5, 2019

Approximations to Gibbs/Shannon Entropy: Entropic Polynomials

Garimella Rama Murthy, Ravi Kishore

Mahindra École Centrale, Hyderabad,India - 500049

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.

1 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. 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 with 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.

This research paper is organized as follows. In section 2, based on higher degree approximation of logarithm function, Shannon entropy is shown to lead to structured polynomials with some properties. The research paper concludes in section 3.

2 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* $H(X) \approx \Big(1 - \sum_{i=1}^{m} p_i^2\Big) \log_2 e$.

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

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

Let $p_i = (1-q_i)$ *with* $0 < p_i < 1$. *Then we have* $0 < q_i < 1$. *This implies*

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

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

$$\begin{aligned} H(X) &\approx -\sum_{i=1}^{m} (1-q_i)(-q_i)\log_2 e \\ &= \sum_{i=1}^{m} (1-q_i)(q_i)\log_2 e \\ &= \sum_{i=1}^{m} p_i(1-p_i)\log_2 e \\ &= \left(1 - \sum_{i=1}^{m} p_i^2\right)\log_2 e \end{aligned}$$

Remark 1: In the above approximation, the error term is $\sum_{j>=2} (-1)^j \frac{q_i^j}{j}$.
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 $q_i^2 + q_i^4 + q_i^6 + \ldots = \frac{q_i^2}{1-q_i^2}$.
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
$H(X) \approx f(p_1, p_2, \ldots, p_m) = \Big(1 - \sum_{i=1}^{m} p_i^2\Big) \log_2 e.$

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. For continuous case i.e., for probability density functions associated with continuous random variables, similar results can easily be derived and are avoided for brevity.

We readily have that:

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

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

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

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.

$$\begin{aligned}
\Big[\frac{(1-p_i)}{2} - \frac{(1-p_i)^2}{3} + \ldots \Big] &= \frac{1}{2} - \sum_{i=3}^{R} (-1)^i \frac{(1-p_i)^{i-2}}{i} \\
&= \frac{1}{2} - \sum_{i=1}^{R} (-1)^i \frac{(1-p_i)^i}{i+2} \\
&= \frac{1}{2} - \sum_{i=1}^{R} \frac{(-1)^i}{i+2} \big[\sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} (p_i)^{i-j} \big] \\
&= \frac{1}{2} - \sum_{j=1}^{R} \sum_{i=j}^{k} \Big[\frac{(-1)^i}{i+2} \binom{i}{j} (-1)^{i-j} (p_i)^{i-j} \Big] - \sum_{i=1}^{R} \Big[\frac{(-1)^i}{i+2} \binom{i}{0} (-1)^i (p_i)^i \Big]
\end{aligned}$$

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

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

$H(p_1, p_2, \dots, p_n) \approx f(p_1, p_2, \dots, p_n)$, where $f(p_1, p_2, \dots, p_n) = \sum\limits_{i=1}^{n} g(p_i)$, where all the polynomials $\{g(p_1), g(p_2), \dots, g(p_n)\}$ 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.

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 $f(p_1, p_2, \dots, p_n) = \sum\limits_{i=1}^{n} g(p_i)$.

2.1 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: $S_q(p) = \frac{1}{(q-1)}(1 - \sum\limits_{i=1}^{n} p_i^q) = S_q(p_1, p_2, \dots, p_n)$.

   Our contribution: $H(p_1, p_2, \dots, p_n) = \sum\limits_{i=1}^{n} g^{(r)}(p_i)$, where $g(.)$ 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.

3 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.

4 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 2] 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 1] G. Rama Murthy, "Gibbs, Shannon Entropy and related measures: Tsallis Entropy" Easy chain Preprint #866.