Axioms for Rényi Entropy with Signed Measures*

Adam Brandenburger†

Pierfrancesco La Mura‡

This version: July 23, 2019

1 Introduction

The maximum entropy method was introduced to physics as a way of deriving the Boltzmann distribution of statistical mechanics (Jaynes, 1957). This method has subsequently been widely used in information theory, statistics, and many applications besides physics. In this paper, we modify the definition of entropy to handle signed measures. Elsewhere (Brandenburger, La Mura, and Zoble, 2019), we use this modification to provide an entropy-based characterization of the simplest quantum system, namely, the qubit.

We consider Rényi entropy (Rényi, 1961), which includes Shannnon entropy (Shannon, 1948) as a special case and is used in communication theory, computer science, and quantum information, among other applications. Rényi entropy for ordinary (unsigned) measures was axiomatized in Rényi (1961) and Daróczy (1963). Here, we modify the Rényi axioms so that they retain their intent when signed measures are introduced, and we derive the family of entropy functionals that satisfies our axioms.

2 Axioms for Entropy

Rényi (1961) showed that his definition of entropy satisfied a list of axioms which he conjectured gave a characterization. Daróczy (1963) proved the conjecture. The approach followed

*Financial support from the NYU Stern School of Business, J.P. Valles, NYU Shanghai, and the HHL - Leipzig Graduate School of Management is gratefully acknowledged.

†Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, New York, NY 10012, U.S.A., adam.brandenburger@stern.nyu.edu, http://www.adambrandenburger.com

‡HHL - Leipzig Graduate School of Management, 04109 Leipzig, Germany, plamura@hhl.de

by Rényi and Daróczy was first to axiomatize entropy for a larger class of measures (non-negative measures with total weight less than or equal to one) and then to specialize the construction to probabilities. We proceed in a similar manner by starting with a set of axioms which characterizes a notion of entropy for signed measures, and then specializing the construction to signed probabilities.

Given a finite set $X = \{x_1, \ldots, x_n\}$, a signed measure Q on X is defined by a tuple $Q = (q_1, \ldots, q_n)$ of real numbers. The quantity $w(Q) = |\Sigma_i q_i|$ will be called the weight of Q. We require $w(Q) \neq 0$ but we do not require $w(Q) = 1$ (except when Q is a signed probability measure).

Given two signed measures $P = (p_1, \ldots, p_m)$ and $Q = (q_1, \ldots, q_n)$, we denote by $P * Q$ the signed measure which is the product $(p_1q_1, \ldots, p_1q_n, \ldots, p_mq_1, \ldots, p_mq_n)$ whenever it is well-defined, i.e., whenever $\Sigma_{i,j} p_i q_j \neq 0$. Also, we denote by $P \cup Q$ the signed measure $(p_1, \ldots, p_m, q_1, \ldots, q_n)$ whenever it is well defined, i.e., whenever $\Sigma_i p_i + \Sigma_j q_j \neq 0$. We write (q) for the signed measure consisting of the scalar q. We impose the following axioms on entropy H:

Axiom 1. *(Real-Valuedness) $H(Q)$ is a non-constant real-valued function of Q.*

Axiom 2. *(Symmetry) $H(Q)$ is a symmetric function of the elements of Q.*

Axiom 3. *(Continuity) $H(Q)$ is a continuous function of each of the elements of Q.*

Axiom 4. *(Calibration) $H((\frac{1}{2})) = 1$.*

Axiom 5. *(Additivity) $H(P * Q) = H(P) + H(Q)$ whenever $H(P * Q)$ is well-defined.*

Axiom 6. *(Mean-Value Property) There is a strictly monotone and continuous function $g : \mathbb{R} \to \mathbb{R}$ such that for any P, Q, whenever $H(P \cup Q)$ is well-defined*

$$H(P \cup Q) = g^{-1}[\frac{w(P)g(H(P)) + w(Q)g(H(Q))}{w(P \cup Q)}].$$

Axiom 7. *(Smoothness) $H((q, 1-q))$ is smooth (C^∞) at $q = 0$.*

Some comments on the axioms. The forms of Axioms 2-6 are carried over without essential change from axioms for Rényi entropy with non-negative measures. (Notice that Axiom 2 is built into the set-up.) Axiom 1 ensures that entropy can be viewed as a measure of the amount or quantity of information in a system, and, to this end, states that entropy must be an ordinary (i.e., real) number. This axiom has bite when applied to signed vs. unsigned measures, because simply extending the domain of ordinary Rényi entropy to negative arguments may yield a complex-valued functional. (In particular, if α is an odd integer, then we may get the log of a negative number.) Concerning Axiom 7, Rényi entropy with non-negative measures is smooth in the interior of its domain. Axiom 7 imposes smoothness at $q = 0$, since this is no longer a boundary value of q.

Theorem 1. *Axioms 1-7 hold if and only if*

$$H(Q) := H_{2k}(Q) = -\frac{1}{2k-1}\log_2\left(\frac{\sum_i \mid q_i \mid^{2k}}{\mid \sum_i q_i \mid}\right), \tag{1}$$

where $k = 1, 2, \ldots$ is a free parameter.

Note that when Q is a signed probability measure, i.e., $\Sigma_i q_i = 1$, equation (2.1) reduces to

$$H_{2k}(Q) = -\frac{1}{2k-1}\log_2\left(\sum_i {q_i}^{2k}\right),$$

where we have also omitted the absolute value in the numerator, since $2k$ is an even integer.

The theorem follows from three lemmas.

Lemma 1. *Under Axioms 1, 3, 4, and 5, if $q \neq 0$, then $H((q)) = -\log_2 |q|$.*

Proof. Let $h(q) := H((q))$. Axioms 1 and 3 imply that h is real-valued and continuous. Axiom 5 implies that $h(pq) = h(p) + h(q)$ whenever $p, q \neq 0$. This is a version of Cauchy's logarithmic functional equation (Aczél and Dhombres, 1989, Equation (7) and Theorem 3) with general solution $h(q) = c\log_2 |q|$, where c is a real constant. Axiom 4 fixes $c = -1$. □

Lemma 2. *Under Lemma 1 and Axioms 5 and 6, we have $g(x) = -dx + e$ (linear) or $g(x) = d2^{(1-\alpha)}x + e$ (exponential), where $d \neq 0$, e, and $\alpha \neq 1$ are arbitrary constants.*

Proof. We extend the argument in Daróczy (1963) to signed measures. If Q is a signed measure, then from Lemma 1 and induction on Axiom 6 we obtain

$$H(Q) = H((q_1) \cup \cdots \cup (q_n)) = g^{-1}\left[\frac{\sum_j w((q_j))\, g(H((q_j)))}{w((q_1) \cup \cdots \cup (q_n))}\right] = g^{-1}\left[\frac{\sum_j |q_j|\, g(-\log_2 |q_j|)}{|\sum_j q_j|}\right]. \tag{2}$$

From this and Axiom 5, we have for signed measures P and Q, provided $\Sigma_{i,j} p_i q_j \neq 0$

$$g^{-1}\left[\frac{\sum_{i,j} |p_i q_j|\, g(-\log_2 |p_i q_j|)}{|\sum_{i,j} p_i q_j|}\right] = g^{-1}\left[\frac{\sum_i |p_i|\, g(-\log_2 |p_i|)}{|\sum_i p_i|}\right] + g^{-1}\left[\frac{\sum_j |q_j|\, g(-\log_2 |q_j|)}{|\sum_j q_j|}\right].$$

Define $f : \mathbb{R}_{++} \to \mathbb{R}$ by $f(t) = g(-\log_2 t)$. Substituting, we get

$$f^{-1}[\frac{\sum_{i,j}\ |p_i q_j|\, f(|p_i q_j|)}{|\sum_{i,j}\ p_i q_j|}] = f^{-1}[\frac{\sum_i\ |p_i|\, f(|p_i|)}{|\sum_i\ p_i|}] \times f^{-1}[\frac{\sum_j\ |q_j|\, f(|q_j|)}{|\sum_j\ q_j|}].$$

Setting $Q = (q)$ (where $q \neq 0$), this becomes

$$\frac{1}{|q|}\, f^{-1}[\frac{\sum_i\ |p_i|\, f(|p_i q|)}{|\sum_i\ p_i|}] = f^{-1}[\frac{\sum_i\ |p_i|\, f(|p_i|)}{|\sum_i\ p_i|}].$$

Define $h_q : \mathbb{R}_{++} \to \mathbb{R}$ by $h_q(t) = f(|q|t)$. Then

$$h_q^{-1}[\frac{\sum_i\ |p_i|\, h_q(|p_i|)}{|\sum_i\ p_i|}] = f^{-1}[\frac{\sum_i\ |p_i|\, f(|p_i|)}{|\sum_i\ p_i|}].$$

This shows that the maps h_q and f generate the same means when restricting the p_i to be non-negative. By a theorem on mean values (Hardy, Littlewood, and Pólya, 1952, Theorem 83), this implies that

$$h_q(t) = a(q)f(t) + b(q),$$

where $a(q)$ and $b(q)$ are independent of t, and $a(q) \neq 0$. Substituting, we get

$$f(|q|t) = a(q)f(t) + b(q).$$

This functional equation (restricting q to be non-negative) has the solution

$$f(t) = d \log_2 t + e,$$

or

$$f(t) = dt^{\alpha-1} + e,$$

where $d \neq 0$, e, and $\alpha \neq 1$ are arbitrary constants (Hardy, Littlewood, and Pólya, 1952, Theorem 84). Recalling the definition of f, we then find that either

$$g(x) = -dx + e, \tag{3}$$

or

$$g(x) = d2^{(1-\alpha)x} + e, \tag{4}$$

as required. □

Lemma 3. *Under Lemma 2 and Axioms 1 and 7, we have* $g(x) = d2^{(1-2k)x}$*, where* k *is a positive integer.*

Proof. If g is linear as in equation (2.3), then from equation (2.2) we get

$$-d \cdot H(Q) + e = d \cdot \frac{\sum_i |q_i| \log_2 |q_i|}{|\sum_i q_i|} + e \cdot \frac{\sum_i |q_i|}{|\sum_i q_i|}. \tag{5}$$

If g is exponential as in equation (2.4), then from equation (2.2) we get

$$d \cdot 2^{(1-\alpha)H(Q)} + e = d \cdot \frac{\sum_i |q_i|^\alpha}{|\sum_i q_i|} + e \cdot \frac{\sum_i |q_i|}{|\sum_i q_i|}. \tag{6}$$

Now use Axiom 7. Setting $Q = (q, 1-q)$ in equation (2.5) we find that $H((q, 1-q))$ is not C^1 at $q = 0$. Setting $Q = (q, 1-q)$ in equation (2.6) we find that $H((q, 1-q))$ is C^1 at $q = 0$ only if $e = 0$. If $\alpha < 0$, then $H((0, 1))$ is unbounded (negative), violating real-valuedness in Axiom 1. Thus $\alpha \geq 0$. If $\alpha = 0$, then $H(Q) = 1$ for all Q, violating non-constancy in Axiom 1. Next, suppose α is not an integer and let k be the least integer with $k > \alpha$. Then $\partial H((q, 1-q))/\partial q = \frac{\phi(q)}{\psi(q)}$ where $\phi(0) \neq 0$ and $\psi(0) = 0$. Thus α must be an integer. If α is an odd integer then $H((q, 1-q))$ is eventually not differentiable at 0. It follows that α is an even positive integer. □

The sufficiency direction of Theorem 1 is finished by noting that equation (2.6) reduces to equation (2.1) when $e = 0$ and $\alpha = 2k$. The necessity of Axioms 1-7 is a straightforward calculation.

References

Aczél, J., and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 1989, 26-27.

Brandenburger, A., P. La Mura, and S. Zoble, "Rényi Entropy, Signed Probabilities, and the Qubit, unpublished, July 2019.

Daróczy, Z., "Über die gemeinsame Charakterisierung der zu den nicht vollständigen Verteilungen gehörigen Entropien von Shannon und von Rényi," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1, 1963, 381-388.

Hardy, G., J. Littlewood, and G. Pólya, *Inequalities*, 2nd edition, Cambridge University Press, 1952.

Rényi, A., "On Measures of Information and Entropy," in Neymann, J., (ed.), *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1961, 547-561.

Shannon, C., “A Mathematical Theory of Communication,” *Bell System Technical Journal*, 27, 1948, 379-423 and 623-656.