Entropy Measures

Abhinav Sapra

Abstract

This is a brief report on the paper "Generalised information and entropy measure in physics" by Christian Beck, written for the course "Physics of Quantum Information" offered at CMI. In this report, we will focus on three measures of entropy: namely, the Shannon entropy, the R´enyi entropies, and the Tsallis entropies, the last two of which are generalized entropy measures. We will also discuss the axiomiatic foundations that we expect different notions of information to respect.

1 Prologue

Information measures in physics play a crucial role in quantifying uncertainty, complexity, and the fundamental aspects of the universe. From the entropy in thermodynamics, which characterizes the disorder and energy distribution in a system, to the information entropy in information theory, which measures the amount of uncertainty or randomness in a message, these measures provide insights into the behavior of physical systems. In quantum mechanics, measures such as von Neumann entropy help understand the entanglement and correlations between particles, essential for studying quantum information processing. Moreover, in statistical mechanics, information measures like mutual information aid in analyzing the relationships between different parts of a system, shedding light on collective behavior and phase transitions. Overall, these information measures are indispensable tools for unraveling the intricacies of nature, from the microscopic realm of particles to the macroscopic scale of the cosmos.

2 How to Measure Information

2.1 Basics

Typically information measures only depend on the probability distribution of the possible outcomes in a given experiment. The idea is as follows :
Consider a set, say A, of possible outcomes $\{A_1, \dots, A_n\}$ of an experiment. If the probability that event A_i occurs is p_i, then $\sum_{i=1}^{n} p_i = 1$.

We do not know which event will occur. But suppose that one of these events, say A_j, finally takes place. Then we have clearly gained some information, because before the event occurred we did not know which event would occur.

Assume the probability p_j is close to 1, then we gain very less information as event A_j was most likely to occur, on the other hand if it had a low probability to occur, intuitively we would have gained more information.

But how do we quantify this information . . . ?

Assume there exists a function, say h, on the probabilities of possible outcomes, i.e., it tells the information gain from occurrence of each outcome depending only on the probability distribution. As stated earlier, $h(A_j)$ should be close to Zero if p_j is close to 1.

one example would be, $h(p_j) = log_a(p_j)$, with a suitable basis a. With $a = 2$, h is called a bit-number. There exists other examples for this function and we choose a measure that caters to our needs in a particular case. We emphasis, that an information measure is not a universally fixed quantity.
Keeping the above picture in mind, we get that the average information gain (I) from an experiment, i.e., over a long sequence of trials is,

$$I(\{p_i\}) = \sum_{i=1}^{n} p_i h(p_i)$$

Using this one can define the notion of Entropy (S) as the "Missing Information", i.e.,

$$S = -I$$

If the probability distribution is sharply peaked around one almost certain event j, we gain very little information from our long-term experiment of independent trials: The event j will occur almost all of the time, which we already knew before. However, if all events are uniformly distributed, i.e., $p_i = \frac{1}{n} \; \forall \; i$, we get a large amount of information by doing this experiment, because before we do the experiment we have no idea which events would actually occur, since they are all equally likely. In this sense, it is reasonable to assume that an (elementary) information measure should take on an extremum (maximum or minimum, depending on sign) for the uniform distribution. Moreover, events i that cannot occur ($p_i = 0$) does not influence our gain of information in the experiment at all. In this way we arrive at the most basic principles that an information measure should satisfy.

2.2 Khinchin's Axioms

There are various formal ways to choose which set of axioms our information must satisfy. The ones we present in this section were formulated by Khinchin, which describe the properties that an information measure that yields ordinary Boltzmann-Gibbs type of statistical mechanics should satisfy. We assume,as before, that our experiment has n events and probability distribution $P = \{p_1, \ldots, p_n\}$.

Axiom 1.

$$I = I(p_1, \ldots, p_n)$$

This is to say that the information measure only depends on the probability distribution and nothing else.

Axiom 2.

$$I(\{\frac{1}{n}, \ldots, \frac{1}{n}\}) \leq I(\{p_j\})$$

for any probability distribution $\{p_j\}$ *This is to say that the information measure achieves minima at uniform distribution* $\{\frac{1}{n}, \ldots, \frac{1}{n}\}$ *and thus, the entropy achieves maxima at uniform distribution.*

Axiom 3.

$$I(\{p_1, \ldots, p_n, 0\}) = I(\{p_1, \ldots, p_n\})$$

This is to say that the information does not depend on the events are not possible and it remains same if the sample space is enlarged to have such events.

Axiom 4.

$$I(\{p_{i,j}^{A,B}\}) = I(\{p_i^A\}) + \sum_i p_i^A \cdot I(\{p^B(B_j|A_i)\})$$

This is to say that the information gain in joint system, made out of two sub-systems A and B, with possible outcomes $\{A_i\}_{i=1}^n$ and $\{B_j\}_{j=1}^m$ respectively, is given by the above identity, where p_i^A are probabilities related to sub-system A and p_j^B are probabilities related to sub-system B and the joint system is described by $p_{i,j}^{A,B} = p_i^A \cdot p^B(B_j|A_i)$, where $p^B(B_j|A_i)$ is the conditional probability of outcome B_j in system B given A_i occurred in system A.

If the two sub-systems are independent then we get, $p_{i,j}^{A,B} = p_i^A \cdot p_j^B$. Moreover,

$$I(\{p_{i,j}^{A,B}\}) = I(\{p^A{}_i\}) + \sum_i p_i^A \cdot I(\{p^B(B_j|A_i)\}) = I(\{p^A{}_i\}) + \sum_i p_i^A \cdot I(\{p_j^B\})$$

$$= I(\{p^A{}_i\}) + I(\{p_j^B\}) \cdot \sum_i p_i^A = I(\{p_i^A\}) + I(\{p_j^B\})$$

2.3 Shannon Entropy

Consider an information measure given by the function $I(\{p_i\}_{i=1}^n) = k \cdot \sum_{i=1}^n p_i \log(p_i)$, for some arbitrary k in $\mathbb{R}$.
Then the entropy is given by $S = -I = -k \cdot \sum_{i=1}^n p_i \log(p_i)$. Moreover, we can verify that shannon entropy is a concave function of probabilities,

$$\frac{\partial(S)}{\partial(p_i)} = -\ln p_i - 1$$

$$\frac{\partial^2(S)}{\partial(p_i) \cdot \partial(p_j)} = -\frac{1}{p_i}\delta_{ij} \leq 0$$

thus, as a sum of concave functions of single probabilities the shannon entropy is also a concave function.
Let us check if the above measure satisfies Khinchin's axioms:

Axioms 1 and Axioms 3 follow directly from the definition. For Axiom 2, consider the function $f(x) = -x \ln x$. Observe that f is a concave function as $\frac{\partial^2 f}{\partial^2 x} = -\frac{1}{x}$. Thus using Jensen's inequality we get,

$$f(\sum \frac{1}{n} p_i) \geq \sum \frac{1}{n} f(p_i)$$

$$f(\frac{1}{n}) \geq \frac{1}{n} \sum f(p_i)$$

$$n f(\frac{1}{n}) \geq \sum f(p_i)$$

$$S(\frac{1}{n}, \ldots, \frac{1}{n}) \geq S(p_1, \ldots, p_n)$$

For Axiom 4, we have

$$
\begin{aligned}
I(\{p_{i,j}^{A,B}\}) &= I(\{p_i^A p^B(B_j|A_i)\}) \\
&= \sum_i \sum_j p_i^A p^B(B_j|A_i) \ln p_i^A p^B(B_j|A_i) \\
&= \sum_i \sum_j p_i^A p^B(B_j|A_i) \ln p_i^A + \sum_i \sum_j p_i^A p^B(B_j|A_i) \ln p^B(B_j|A_i) \\
&= \sum_j p^B(B_j|A_i) \sum_i p_i^A \ln p_i^A + \sum_i p_i^A \sum_j p_j^B(B_j|A_i) \ln p^B(B_j|A_i) \\
&= \sum_j p^B(B_j|A_i) \cdot I(\{p_i^A\}) + \sum_i p_i^A I(\{p_j^B(B_j|A_i)\}) \\
&= I(\{p_i^A\}) + \sum_i p_i^A \cdot I(\{p^B(B_j|A_i)\})
\end{aligned}
$$

same calculations go through if we introduce a factor of k in the definition. Thus, we have successfully verified axioms 2 and 4, therefore the Shannon Entropy in a valid entropy measure according to Khinchin's axioms.

continuous shannon entropy:

For continuous probability distribution, we define the shannon entropy as function of variable u with normalization $\int p(u)du = 1$ as,

$$S = -\int_{-\infty}^{\infty} p(u) \ln(\sigma(p(u)))du$$

where, σ is a scaling factor to make the term inside the logarithm a dimensionless quantity.

3 More Information Measures

3.1 Rényi Entropies

If we replace Axiom 4 of the Khinchin axioms with the additivity rule for independent sub-systems, we get measures known as Rényi Entropies, given by

$$S_q^{(R)} = \frac{1}{q-1} \ln \sum_i p_i^q$$

for an arbitrary real parameter q.

Let us verify if this parameterized measure satisfies Khinchin's Axioms:

For relaxed Axiom 4, consider

$$
\begin{aligned}
S_q^{(R)}(\{p_{i,j}^{A,B}\}) &= S_q^{(R)}(\{p_i^A p_j^B\}) \\
&= \frac{1}{q-1} \ln \sum_i \sum_j (p_i^A)^q (p_j^B)^q \\
&= \frac{1}{q-1} \ln \sum_i (p_i^A)^q \sum_j (p_j^B)^q \\
&= \frac{1}{q-1} \ln \sum_i (p_i^A)^q + \frac{1}{q-1} \ln \sum_i (p_j^B)^q \\
&= S_q^{(R)}(\{p_i^A\}) + S_q^{(R)}(\{p_j^B\})
\end{aligned}
$$

For Axiom 2, consider $S_q^{(R)}$, which is a non-increasing function of q, hence $S_q^{(R)} \leq S_0^{(R)} = \ln(n) = S_q^{(R)}(\{1/n, \ldots, 1/n\})$.Thus, it satisfies both Axioms(with relaxed additivity property). It can also be shown that these entropies uniquely satisfy these conditions (up to a multiplicative constant).

Shannon entropy as limit of Rényi entropy :
Consider, $\lim_{q\to 1} S_q^{(R)}$, then by L'Hopital rule, we get

$$\begin{aligned}\lim_{q\to 1} S_q^{(R)} &= \lim_{q\to 1} \frac{\sum_i (p_i^q)^{-1} \sum_i p_i^q \ln p_i}{-1}\\ &= \lim_{q\to 1} \frac{\sum_i p_i \ln p_i}{\sum_i p_i}\\ &= -\sum_i p_i \ln p_i\\ &= S\end{aligned}$$

But unlike Shannon entropy we do not have straight forward way to express Rényi entropy of a joint system in terms of Rényi entropy of sub-systems. Moreover, unlike Shannon entropy, the Rényi entropies do not have definte convexity.

3.2 Tsallis Entropies

The Tsallis entropies are also defined for arbitrary real parameters as

$$S_q^{(T)} = \frac{1}{q-1}(1 - \sum_{i=1}^{n} p_i^q)$$

It is easy to see that this satisfies Axiom 1 and Axiom 3.

For Axiom 2, consider the function

$$f(p_1, \ldots, p_n) = \frac{1}{q-1}(\sum_{i=1}^{n} p_i^q)$$

where, $\sum p_i = 1$.

Now to maximize this we use Lagrange multipliers, so consider the function $L(p, \lambda) = \frac{1}{q-1}(1 - \sum p_i^q) + \lambda(\sum p_i - 1)$. In order to maximize f we want $\frac{\partial L}{\partial p} = 0$ and $\frac{\partial L}{\partial \lambda} = 0$, i.e.,

$$\frac{\partial L}{\partial p_i} = \frac{-q p_i^{q-1}}{q-1} + \lambda = 0$$

this gives that $p_i = (\lambda(q-1)/q)^{1/q-1}$ for all i. This $p_i s$ are equal and sum upto 1, thus $p_i = 1/n$ for all i. Therefore,$S_q^{(T)}$ is maximized at uniform distribution.
Axiom 4 for Tsallis entropies :
Consider a joint system of two independent sub-systems, i.e., $p_{i,j}^{A,B} = p_i^A p_j^B$, then we can write

$$\sum (p_i^A)^q = 1 - (q-1) S_q^A \quad and \quad \sum (p_j^B)^q = 1 - (q-1) S_q^B$$

$$\begin{aligned}1 - (q-1) S_q^{A,B} = \sum_{i,j} (p_{i,j}^{A,B})^q &= \sum_{i,j} (p_i^A)^q (p_j^B)^q\\ &= \sum_i (p_i^A)^q \sum_j (p_j^B)^q\\ &= (1 - (q-1) S_q^A)(1 - (q-1) S_2^B)\\ &= 1 - (q-1) S_q^A - (q-1) S_q^B + (q-1)^2 S_q^A S_q^B\end{aligned}$$

Thus, we get, $S_q^{A,B} = S_q^A + S_q^B - (q-1)S_q^A S_q^B$ for Tsallis entropies.
Relation between Tsallis and Rényi entropies :

$$\sum p_i^q = 1 - (q-1)S_q^{(T)} = \exp{(q-1)S_q^{(R)}}$$

Therefore, we get

$$S_q^{(T)} = \frac{1}{q-1}(1 - \exp{(q-1)S_q^{(R)}})$$

Thus, Tsallis entropy is a monotonous function of Rényi entropy.
Shannon entropy as a limit of Tsallis entropy : Just consider, $\lim_{q\to 1} \frac{1}{q-1}(1 - \sum_i p_i^q) = \lim_{q\to 1} -\sum p_i^q \ln p_i = -\sum p_i^q \ln p_i = S$
But Tsallis entropy is a better measure for statistical mechanics than Rényi due to its concavity

$$\frac{\partial}{\partial p_i} S_q^{(T)} = -\frac{q}{q-1} p_i^{q-1} \tag{1}$$

$$\frac{\partial^2}{\partial p_i \partial p_j} S_q^{(T)} = -q p_i^{q-2} \delta_{ij}. \tag{2}$$

Hence, Tsallis entropy is a concave function as it is a sum of concave functions.
Another property that makes Tsallis entropy a better candidate for statistical mechanics is Lesche-stability.

4 Selecting a suitable information measure

4.1 Axiomatic foundations

The Khinchin axioms apparently are the right axioms to obtain the Shannon entropy in a unique way, but it is very restrictive if one wants to describe general complex systems. Abe has showed that the Tsallis entropy follows uniquely (up to an arbitrary multiplicative constant) from the following generalized version of the Khinchin axioms. Axioms 1–3 are kept, and Axiom 4 is replaced by the following more general version:

New Axiom 4

$$S_q^{A,B} = S_q^A + S_q^{B|A} - (q-1)S_q^I S_q^{B|A} \tag{3}$$

Here $S_q^{B|A}$ is the conditional entropy formed with the conditional probabilities $p(B_j|A_i)$ and averaged over all states i using the so-called escort distributions P_i:

$$S_q^{B|A} = \sum_i P_i S_q(\{p(B_j|A_i)\}). \tag{4}$$

Escort distributions P_i are defined for any given probability distribution p_i by

$$P_i = \frac{p_i^q}{\sum_i p_i^q}. \tag{5}$$

For $q = 1$, the new axiom 4 reduces to the old Khinchin axiom 4, i.e. $S_q^{A,B} = S_q^A + S_q^{B|A}$. For independent systems A and B, the new axiom 4 reduces to the pseudo-additivity property.
The meaning of the new axiom 4 is quite clear. It is a kind of minimal extension of the old axiom 4: If we collect information from two subsystems, the total information should be the sum of the information collected from system A and the conditional information

from system B, plus a correction term. This correction term can a priori be anything, but we want to restrict ourselves to information measures where

$$S^{A,B} = S^A + S^{B|A} + g(S^A, S^{B|A}), \tag{6}$$

where $g(x, y)$ is some function. The property that the entropy of the composed system can be expressed as a function of the entropies of the single systems is sometimes referred to as the composability property. Clearly, the function g must depend on the entropies of both subsystems, for symmetry reasons. The simplest form one can imagine is that it is given by

$$g(x, y) = \text{const} \cdot xy, \tag{7}$$

i.e. it is proportional to both the entropy of the first system and that of the second system. Calling the proportionality constant $q - 1$, we end up with the new axiom 4.
The Tsallis entropies are composable in a very simple way. Suppose the two systems A and B are not independent,

$$\begin{aligned}
S_q^A + \sum_i p_i^q S_q^{B|A} &= \frac{1}{q-1}(1 - \sum_i p_i^q) + \frac{1}{q-1} \sum_i p_i^q (1 - \sum_i p^q(B_j|A_i)) \\
&= \frac{1}{q-1}(1 - \sum_i p_i^q \sum_j (p(B_j|A_i))^q \\
&= \frac{1}{q-1}(1 - \sum_{i,j} (p_i p(B_j|A_i))^q) \\
&= \frac{1}{q-1}(1 - \sum (p_{i,j})^q) \\
&= S_q^{A,B}
\end{aligned}$$

For $q = 1$, the above equation reduces to Khinchin's Axiom 4.

4.2 Lesche stability

Physical systems contain noise. A necessary requirement for a generalized entropic form $S[p]$ to make physical sense is that it must be stable under small perturbations. This means a small perturbation of the set of probabilities $p := \{p_i\}$ to a new set $p' = \{p'_i\}$ should have only a small effect on the value S_{max} of $S_q[p]$ in the thermodynamic state that maximizes the entropy. This should in particular be true in the limit $n \to \infty$ The stability condition can be mathematically expressed as follows :

Stability condition

For every $\epsilon > 0$ there is a $\delta > 0$ such that

$$||p - p'||_1 \leq \delta \Longrightarrow \left| \frac{S[p] - S[p']}{S_{max}} \right| < \epsilon \tag{8}$$

for arbitrarily large W. Here $||A||_1 = \sum_{i=1}^{W} |A_i|$ denotes the L_1 norm of an observable A. Sumiyoshi Abe has proved that the Tsallis entropies are Lesche-stable q, whereas the Rényi entropies are not stable for any $q \neq 1$ (for a discrete set of probabilities p_i with $n \to \infty$). This is an important criterion to single out generalized entropies that may have potential physical meaning. According to the stability criterion, the Tsallis entropies are stable and thus may be associated with physical states.

5 Quantum Rènyi entropies

Let S be the set of sub-normalized quantum states, i.e. $\rho \in S$ is positive semi-definite and has $Tr[\rho] \in (0, 1]$. We are interested in a functional $H(\cdot) : S \to \mathbb{R}$ satisfying the folllowing properties:

- Continuity
- Unitary invariance $H(\rho) = H(U\rho U^\dagger)$
- Normalization H(1/2) =log 2
- Additivity $H(\rho \otimes \tau) = H(\rho) + H(\tau), \ \forall \rho, \tau \in S$
- Arithmetic Mean $H(\rho \oplus \tau) = (\frac{Tr[\rho]}{Tr[\rho+\tau]})H(\rho)+(\frac{Tr[\tau]}{Tr[\rho+\tau]})H(\tau), \ \forall \rho, \tau \geq 0 \, with \ Tr[\rho+\tau] \leq 1$

Von Neumann entropy satisfies all the above properties. Rènyi proved that Shannon entropy is the unique functional satisfying the above properties in a classical setting (satisfying classical specialization of the properties) and Von Neumann entropy is the unique functional in the quantum setting.
Rènyi replaced the last property (arithmetic mean) to something called General Mean :

There exists a continuous and strictly monotonic function g such that, for $\rho, \tau \geq 0$ with $Tr[\rho + \tau] \leq 1$,

$$H(\rho \oplus \tau) = g^{-1}\left(\frac{Tr[\rho]}{Tr[\rho+\tau]} \cdot g(H(\rho)) + \frac{Tr[\tau]}{Tr[\rho+\tau]} \cdot g(H(\tau))\right)$$

Rènyi showed that the Quantum Rènyi entropy of order α satisfies first four properties and the General Mean property with $g_\alpha(x) = \exp((1-\alpha)x)$, where Quantum Rènyi entropy is given by

$$H_\alpha(\rho) := \frac{1}{1-\alpha} \log \frac{Tr[\rho^\alpha]}{Tr[\rho]}$$

with $\alpha \in (0, 1) \cup (1, \infty)$.