A combinatorial interpretation for Tsallis 2-entropy

JUAN PABLO VIGNEAUX

Institut de Mathématiques de Jussieu–Paris Rive Gauche, Université Paris Diderot

juan-pablo.vigneaux@imj-prg.fr

Objectives

- Improve the understanding of Tsallis entropy.
- Contribute to the microscopical foundations of non-extensive statistical mechanics.

  Non-extensive statistical mechanics is a generalization of usual statistical mechanics where Boltzman-Gibbs-Shannon entropy is replaced by Tsallis entropy. The corresponding equilibrium states (q-Gaussians) present polynomial tails and have been observed in several contexts.

Tsallis entropy describes the growth of q-multinomial coefficients

Let $(p_1, ..., p_s)$ be a probability distribution. It is well known that

$$\lim_n \frac{1}{n} \ln \binom{n}{p_1 n, ..., p_s n} = H_1(p_1, ..., p_s), \tag{1}$$

where H_1 denotes Shannon entropy: $H_1(p_1, ..., p_s) = -\sum_{i=1}^s p_i \ln(p_i)$.

I have proved that

$$\lim_n \frac{1}{n^2} \log_q \begin{bmatrix} n \\ p_1 n, ..., p_s n \end{bmatrix}_q = \frac{H_2(p_1, ..., p_s)}{2}, \tag{2}$$

where H_2 is Tsallis 2-entropy [1], [2]: $H_2(p_1, ..., p_s) = 1 - \sum_{i=1}^s p_i^2$.

While the multinomial coefficient $\binom{n}{k_1, ..., k_m}$ counts words with symbols $S = \{\sigma_1, ..., \sigma_s\}$ such that σ_i appears k_i times, **its q-generalization** $\begin{bmatrix} n \\ k_1, ..., k_s \end{bmatrix}_q$ **counts flags of vector spaces** $V_1 \subset ... \subset V_s = \mathbb{F}_q^n$ **such that** $\dim V_i = \sum_{j=1}^i k_j$. (We suppose everywhere that q is a prime power; $\mathbb{F}_q$ is the field with q elements.)

Non-additivity has a combinatorial counterpart

Let (p_0, p_1) and (q_0, q_1) be probability distributions. **The flags $V_{00} \subset V_{01} \subset V_{10} \subset V_{11} = \mathbb{F}_q^n$ counted by** $\begin{bmatrix} n \\ p_0q_0n, p_0q_1n, p_1q_0n, p_1q_1n \end{bmatrix}_q$ **can be determined by an iterated choice of subspaces, whose dimensions are chosen independently**: pick first a subspace $V_0 \subset \mathbb{F}_q^n$ of dimension $p_0q_0n + p_0q_1n = p_0n$ (there are $\begin{bmatrix} n \\ p_0n, p_1n \end{bmatrix}_q$ of those) and then pick a subspace of $V_{00} \subset V_0$ of dimension $q_0 \dim V_0 = q_0(p_0n)$ and another subspace $V_{10} \subset \mathbb{F}_q^n / V_0$ of dimension $q_0(n - \dim V_0) = q_0(p_1 n)$. This corresponds to the combinatorial identity

$$\begin{bmatrix} n \\ p_0q_0n, p_0q_1n, p_1q_0n, p_1q_1n \end{bmatrix}_q = \begin{bmatrix} n \\ p_0n, p_1n \end{bmatrix}_q \begin{bmatrix} p_0n \\ q_0(p_0n), q_1(p_0n) \end{bmatrix}_q \begin{bmatrix} p_1n \\ q_0(p_1n), q_1(p_1n) \end{bmatrix}_q, \tag{3}$$

Applying $\frac{2}{n^2}\ln(-)$ to both sides and taking the limit $n \to \infty$, we obtain

$$H_2(p_0q_0, p_0q_1, p_1q_0, p_1q_1) = H_2(p_0, p_1) + H_2(q_0, q_1) - H_2(p_0, p_1)H_2(q_0, q_1). \tag{4}$$

Flags as generalized configurations

Classical model: a set $S = \{\sigma_1, ..., \sigma_s\}$ of possibles states of a particle (spins); the configuration of n independent particles or history of a single particle up to time n (trajectory) is represented by a vector $\mathbf{x} \in S^n$.

Proposed generalization (q-systems): the configuration of n particles (or a trajectory at time n) is described by a flag of vector spaces $V_1 \subset ... \subset V_s = \mathbb{F}_q^n$.

The *type* $(k_1, ..., k_s)$ of a classical configuration correspond to the frequencies of appearance of each symbol; their q-analogue is the collection $(k_1, ..., k_s)$ that determines the dimension of the subspaces in the flag. The corresponding multinomial coefficient ($\binom{n}{k_1, ..., k_s}$ or $\begin{bmatrix} n \\ k_1, ..., k_s \end{bmatrix}_q$) counts the number of configurations of type $(k_1, ..., k_s)$.

This allows us to interpret the maximization of $H_2(p_1, ..., p_s)$ as a limiting version of $\max \begin{bmatrix} n \\ k_1, ..., k_s \end{bmatrix}_q$ (taking k_i in such a way that $p_i = \lim_n \frac{k_i}{n}$). Therefore, **we obtain a combinatorial interpretation for the maximum entropy principle involving H_2.**

Dynamics: a stochastic process that generates vector spaces

Let $\{X_i\}_{i \geq 1}$ be a collection of independent random variables, such that $X_i \sim \text{Bernoulli}\left(\frac{\theta q^i}{1 + \theta q^i}\right)$, for each i. (Remember that q is a fixed prime power; $\theta > 0$ is a real parameter.)

Let us define a stochastic process $\{V_i\}_{i \geq 0}$ such that each V_i is a vector subspace of $\mathbb{F}_q^i$, as follows: $V_0 = 0$ and, at step n, the dimension of V_{n-1} increases by 1 if and only if $X_n = 1$; in this case, V_n is picked at random (uniformly) between all the n-dilations of V_{n-1}. When $X_n = 0$, one sets $V_n = V_{n-1}$.

The n-dilations of $w \subset \mathbb{F}_q^{n-1}$ (subspace) are

$$\text{Dil}_n(w) = \{\, v \subset \mathbb{F}_q^n \mid \dim v - \dim w = 1,\ w \subset v \text{ and } v \not\subset \mathbb{F}_q^{n-1} \,\}. \tag{5}$$

One fixes a sequence of linear embeddings $\mathbb{F}_q^1 \hookrightarrow \mathbb{F}_q^2 \hookrightarrow ...$ beforehand, and identifies $\mathbb{F}_q^{n-1}$ with its image in $\mathbb{F}_q^n$.

I have proved that:

- If $v \subset \mathbb{F}_q^n$ has dimension k, then $\mathbb{P}(V_n = v) = \frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}$.
- $\mathbb{P}(\dim V_n = k) = \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}$.

In other words, $Y_n := \dim V_n = \sum_{i=1}^n X_i$ **follows a q-binomial law.**

We have used here the q-Pochhammer symbols $(a; q)_n := \prod_{i=0}^{n-1}(1 - aq^i)$, with $(a; q)_0 = 1$. The Gaussian binomial formula [3, Ch. 5] implies that $(-\theta; q)_n = \sum_{i=0}^n \begin{bmatrix} n \\ k \end{bmatrix}_q \theta^k q^{k(k-1)/2}$.

The q-binomial distribution concentrates around its mean

As n tends to infinity, the probability distribution of Y_n concentrates strongly around the expected dimension $k_n := \mathbb{E} Y_n = \sum_{i=1}^n \frac{\theta q^{i-1}}{1 + \theta q^{i-1}}$ (see Figure 1). Therefore, there is a notion of "typical subspace", whose dimension is close to k_n: with high probability, V_n gives one of these subspaces.

In the "classical case", one considers random binary sequences $(Z_1, ..., Z_n) \in \{0, 1\}^n$ such that each $Z_i \sim \text{Bernoulli}(p)$. Then $W_n = \sum_{i=1}^n Z_i$ follows a binomial distribution with parameters (p, n), which concentrates around its expected value pn as $n \to \infty$. The "typical sequences" of length n have approximately pn ones, with high probability. This idea is very important in the work of Shannon and was later refined by the so-called type theory [4, Ch. I].

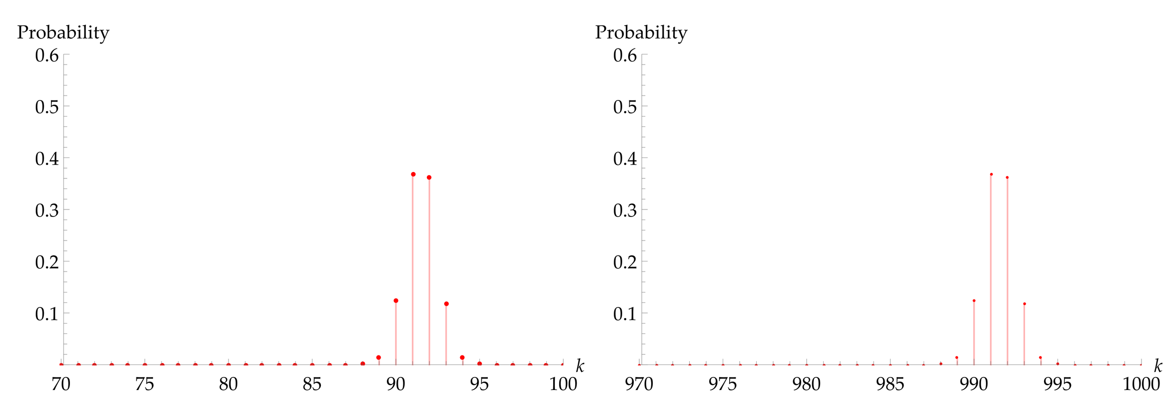

Figure 1: Probability $\mathbb{P}(Y_n = k)$ as a function of k, for $n = 100$ (left) and $n = 1000$ (right), when $q = 3$ and $\theta = 5 \times 10^{-5}$.

References

[1] J. Havrda and F. Charvát, "Quantification method of classification processes. concept of structural a-entropy," *Kybernetika*, vol. 3, no. 1, pp. 30–35, 1967.

[2] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of statistical physics*, vol. 52, no. 1-2, pp. 479–487, 1988.

[3] V. Kac and P. Cheung, *Quantum Calculus*, ser. Universitext. Springer New York, 2001, ISBN: 9780387953410.

[4] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*, ser. Probability and Mathematical Statistics. Academic Press, 1981, ISBN: 9780121984502.

Acknowledgements

I am very grateful to Matilde Marcolli, who pointed out the combinatorial meaning q-multinomial coefficients during a conversation we had at CIRM. I also want to thank Daniel Bennequin for his attentive listening and multiple suggestions.