Information theory for Tsallis 2-entropy

Juan Pablo Vigneaux
Institut de Mathématiques de Jussieu
Université Paris Diderot

July 26, 2018

Mathematical Physics

A combinatorial interpretation for Tsallis 2-entropy

Juan Pablo Vigneaux

(Submitted on 13 Jul 2018)

While Shannon entropy is related to the growth rate of multinomial coefficients, we show that Tsallis 2-entropy is connected to their q-version; when q is a prime power, these coefficients count the number of flags in $\mathbb{F}_q^n$ with prescribed length and dimensions ($\mathbb{F}_q$ denotes the field of order q). In particular, the q-binomial coefficients count vector subspaces of given dimension. We obtain this way a combinatorial explanation for non-additivity. We show that statistical systems whose configurations are described by flags provide a frequentist justification for the maximum entropy principle with Tsallis statistics. We introduce then a discrete-time stochastic process associated to the q-binomial distribution, that generates at time n a vector subspace of $\mathbb{F}_q^n$. The concentration of measure on certain "typical subspaces" allows us to extend the asymptotic equipartition property to this setting. We discuss the applications to information theory, particularly to source coding.

Subjects: **Mathematical Physics (math-ph)**; Information Theory (cs.IT); Probability (math.PR)
Cite as: **arXiv:1807.05152 [math-ph]**
(or **arXiv:1807.05152v1 [math-ph]** for this version)

Submission history

From: Juan Pablo Vigneaux [view email]
[v1] Fri, 13 Jul 2018 15:54:24 GMT (27kb)

Outline

Shannon entropy

The multinomial coefficient

$$\binom{n}{k_1, \ldots, k_s} := \frac{n!}{k_1! \cdots k_s!}$$

counts the number of words $w \in \Sigma^n$, with $\Sigma = \{\sigma_1, \ldots, \sigma_s\}$, such that σ_i appears k_i times.

Shannon entropy

The multinomial coefficient

$$\binom{n}{k_1, \ldots, k_s} := \frac{n!}{k_1! \cdots k_s!}$$

counts the number of words $w \in \Sigma^n$, with $\Sigma = \{\sigma_1, \ldots, \sigma_s\}$, such that σ_i appears k_i times.

From the point of view of probability and combinatorics, Shannon entropy $H_1(p_1, \ldots, p_s) = \sum_{i=1}^{s} p_i \ln p_i$ appears naturally in the asymptotic formula

$$\binom{n}{p_1 n, \ldots, p_s n} = \exp(nH_1(p_1, \ldots, p_s) + O(\ln n)) \tag{1}$$

q-multinomials

For given $q \in \mathbb{C} \smallsetminus \{1\}$, define

1. q-integers $[n]_q = \frac{q^n - 1}{q - 1}$,
2. q-factorials: $[n]_q! := [n]_q [n-1]_q \cdots [1]_q$.
3. q-multinomial coefficients by

$$\begin{bmatrix} n \\ k_1, \ldots, k_s \end{bmatrix}_q := \frac{[n]_q!}{[k_1]_q! \cdots [k_s]_q!}, \tag{2}$$

where $k_1, \ldots, k_s$ are such that $\sum_{i=1}^{s} k_i = n$.

q-multinomials

For given $q \in \mathbb{C} \smallsetminus \{1\}$, define

1. q-integers $[n]_q = \frac{q^n-1}{q-1}$,
2. q-factorials: $[n]_q! := [n]_q[n-1]_q \cdots [1]_q$.
3. q-multinomial coefficients by

$$\begin{bmatrix} n \\ k_1, \ldots, k_s \end{bmatrix}_q := \frac{[n]_q!}{[k_1]_q! \cdots [k_s]_q!}, \tag{2}$$

where $k_1, \ldots, k_s$ are such that $\sum_{i=1}^{s} k_i = n$.

Remark

When q is a prime power, $\begin{bmatrix} n \\ k_1, \ldots, k_s \end{bmatrix}_q$ counts the number of flags of vector spaces $V_1 \subset V_2 \subset \ldots \subset V_s = \mathbb{F}_q^n$ such that $\dim V_i = \sum_{j=1}^{i} k_j$.

Tsallis 2-entropy

Proposition

Let $(p_1, \ldots, p_s)$ be a probability. Then,

$$\begin{bmatrix} n \\ p_1 n, \ldots, p_s n \end{bmatrix}_q \sim (q^{-1}; q^{-1})_\infty^{1-s} q^{n^2 H_2(p_1, \ldots, p_s)/2}. \tag{3}$$

For any $\alpha \neq 1$, the function

$$H_\alpha(p_1, \ldots, p_s) := \frac{1}{\alpha - 1}\left(1 - \sum_{i=1}^{s} p_i^\alpha\right) \tag{4}$$

is called Tsallis α-entropy (actually, it was introduced by Havrda and Charvát [3]).

Important things to retain

$\binom{n}{k} := \binom{n}{k, n-k}$ counts the words $w \in \{0,1\}^n$ that have k ones.

$\begin{bmatrix} n \\ k \end{bmatrix}_q := \begin{bmatrix} n \\ k, n-k \end{bmatrix}_q$ counts vector subspaces v of $\mathbb{F}_q^n$ such that $\dim(v) = k$.

Outline

1. Entropies
2. Generalized information theory
3. Some algebra

Types (following Csiszár): Terminology

Consider the surjection that counts the number of ones

$$\begin{array}{rccc} \pi : & \{0,1\}^n & \to & [\![0,n]\!] \\ & (x_1,\dots,x_n) & \mapsto & \sum_{i=1}^{s} x_i \end{array}$$

1. If $\pi(w) = k$, we say that w is of type k.
2. Each set $T_k^n := \pi^{-1}(k)$ is called a *type class*. Note that

$$|T_{pn}^n| = \binom{n}{pn} = \exp(nH_1(p, 1-p) + o(n)).$$

Types (following Csiszár): Typicality

Suppose the sequences are generated by independent coin tosses: $(Z_1, ..., Z_n) \sim \mathrm{Ber}(p)^{\otimes n}$. Then $\pi(Z_1, ..., Z_n) =: W_n \sim \mathrm{Bin}(n, p)$.

Note that $\mathbb{E}(W_n) = pn$. Chebyshev's inequality implies that $W_n \in I_{n,\xi} := [\![np - n^{\frac{1}{2}+\xi}, np + n^{\frac{1}{2}+\xi}]\!]$ with high probability (here $0 < \xi << \frac{1}{2}$).

We can define the *typical sequences* to be $\pi^{-1}(I_{n,\xi})$. Then, $(X_1, ..., X_n)$ is typical w.h.p.

Remark

Typical sequences have qn ones, for q that satisfies $|q - p| \leq n^{\xi - \frac{1}{2}} \to 0$. Then,

$$|\pi^{-1}(I_{n,\xi})| = \sum_{qn \in I_{n,\xi}} |T^n_{qn}| = \exp(nH_1(p, 1-p) + o(n)).$$

Generalization

Let $\mathrm{Gr}(n,k)$ denote the set of all subspaces v of $\mathbb{F}_q^n$ such that $\dim(v)=k$ (grassmannian). Set $\mathrm{Gr}(n)=\cup_{i=0}^{n}\mathrm{Gr}(n)$. Consider the surjection

$$\begin{array}{rccc} \pi: & \mathrm{Gr}(n) & \to & [\![0,n]\!] \\ & v & \mapsto & \dim(v) \end{array}$$

1. If $\pi(v)=k$, we say that v is of type k.
2. Each set $T_k^n := \pi^{-1}(k) = \mathrm{Gr}(n,k)$ is called a *type class*.

$$|T_{pn}^n| = \begin{bmatrix} n \\ pn \end{bmatrix}_q = C(q)q^{n^2 H_2(p,1-p)/2}.$$

Probabilistic model

To talk about "typical subspaces", we need a stochastic process that generates at time n a generalized message $V_n \in \mathrm{Gr}(n)$. Moreover, we want V_n to contain in certain sense V_{n-1} (because this would be the analog of $(Z_1, \dots, Z_n)$).

A clue: there exists a probability distribution $\mathrm{Bin}_q(n,\theta)$ on $[\![0,n]\!]$, called q-binomial with parameters $n \in \mathbb{N}$ and $\theta > 0$, such that

$$\mathrm{Bin}_q(k|n,\theta) = \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_n}.$$

A variable $Y \sim \mathrm{Bin}_q(n,\theta)$ can be written as a sum $X_1 + \dots + X_n$ such that $X_i \sim \mathrm{Ber}\left(\frac{\theta q^{i-1}}{1+\theta q^{i-1}}\right)$.

Probabilistic model

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

Probabilistic model

Fix a a sequence of linear embeddings $\mathbb{F}_q^1 \hookrightarrow \mathbb{F}_q^2 \hookrightarrow \dots$, and identify $\mathbb{F}_q^{n-1}$ with its image in $\mathbb{F}_q^n$.

Set $V_0 = 0$ and, at time n,

1. if $X_n = 0$, do nothing $V_n = V_{n-1}$;
2. if $X_n = 1$, increase dimension: pick V_n at random, uniformly, from $\mathrm{Dil}_n(V_{n-1})$.

Probabilistic model

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

Set $V_0 = 0$ and, at time n,

1. if $X_n = 0$, do nothing $V_n = V_{n-1}$;
2. if $X_n = 1$, increase dimension: pick V_n at random, uniformly, from $\mathrm{Dil}_n(V_{n-1})$.

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

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

Concentration of measure

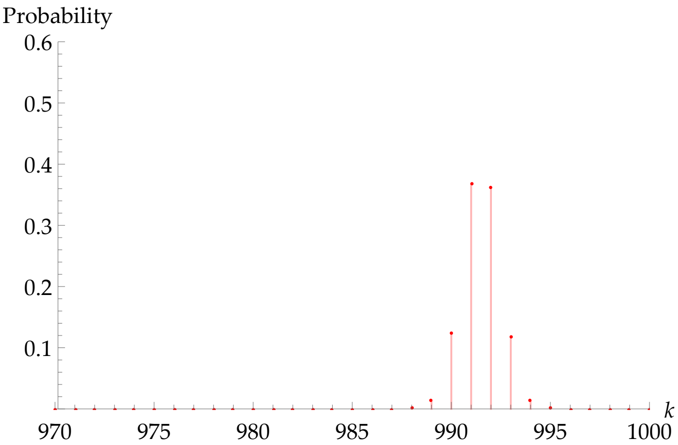

Figure: $\mathrm{Bin}_q(k|n,\theta)$ for $q=3$, $n=1000$, $\theta=5\times10^{-5}$.

The asymptotic formulas allow us to prove that

$$\mathrm{Bin}_q(n-d|n,\theta) = \mathbb{P}(V_n \in \mathrm{Gr}(n-d,n)) = \begin{bmatrix} n \\ n-d \end{bmatrix}_q \frac{\theta^{n-d} q^{(n-d)(n-d-1)/2}}{(-\theta;q)_n} \to \mu(d),$$

and μ defines a probability distribution on $\mathbb{N}$.

We define a function $\Delta : [0,1) \to \mathbb{N}$ such that
$\Delta(p) =$ minimum d such that $\mu([\![0,d]\!]) \geq 1-p$.

Theorem

For any $\delta \in (0,1)$ *and* $\varepsilon > 0$ *be such that* $p_\varepsilon := 1-\varepsilon$ *is a continuity point of* Δ, *let* $A_n = \bigcup_{k=0}^{d(A_n)} \mathrm{Gr}(n-k,n)$ *be the smallest set of the form* $\bigcup_{k=0}^{d} \mathrm{Gr}(n-k,n)$ *such that* $\mathbb{P}(V_n \in A_n^c) \le \varepsilon$.
Then, there exists $n_0 \in \mathbb{N}$ *such that, for every* $n \ge n_0$,

1. $A_n = \bigcup_{k=0}^{\Delta(p_\varepsilon)} \mathrm{Gr}(n-k,n)$;
2. *for any* $v \in A_n$ *such that* $\dim v = k$,

$$\left| \frac{\log_q(\mathbb{P}(V_n = v)^{-1})}{n} - \frac{n}{2} H_2(k/n) \right| \le \delta. \tag{6}$$

The size of A_n *is optimal, up to the first order in the exponential: let* $s(n,\varepsilon)$ *denote* $\min\{|B_n| : B_n \subset \mathrm{Gr}(n)$ *and* $\mathbb{P}(V_n \in B_n) \ge 1-\varepsilon\}$; *then*

$$\lim_n \frac{1}{n} \log_q |A_n| = \lim_n \frac{1}{n} \log_q s(n,\varepsilon) = \lim_n \frac{n}{2} H_2(\Delta(p_\varepsilon)/n) = \Delta(p_\varepsilon). \tag{7}$$

Outline

Recurrence

Combinatorics says

$$\binom{n}{k_1, k_2, k_3} = \binom{n}{(k_1 + k_2), k_3}\binom{k_1 + k_2}{k_1, k_2}$$

Recurrence

Combinatorics says

$$\binom{n}{k_1, k_2, k_3} = \binom{n}{(k_1 + k_2), k_3}\binom{k_1 + k_2}{k_1, k_2}$$

Set $k_i = p_i n_i$ and apply $\lim \frac{1}{n} \ln(-)$, to obtain

$$H_1(p_1, p_2, p_3) = H_1(p_1 + p_2, p_3) + (p_1 + p_2) H_1(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}).$$

Some references

3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Fig. 6. At the left we have three possibilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. On the right we first choose between two possibilities each with probability $\frac{1}{2}$, and if the second occurs make another choice with probabilities $\frac{2}{3}$, $\frac{1}{3}$. The final results have the same probabilities as before. We require, in this special case, that

$$H(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}) = H(\tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{2}H(\tfrac{2}{3}, \tfrac{1}{3})$$

The coefficient $\frac{1}{2}$ is because this second choice only occurs half the time.

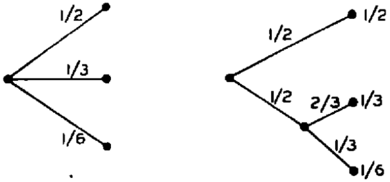

Fig. 6—Decomposition of a choice from three possibilities.

P. Baudot and D. Bennequin, *The homological nature of entropy*, Entropy, 17 (2015), pp. 3253–3318.

I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*, Probability and mathematical statistics, Academic Press, 1981.

J. Havrda and F. Charvát, *Quantification method of classification processes. concept of structural a-entropy*, Kybernetika, 3 (1967), pp. 30–35.

J. P. Vigneaux, *The structure of information: from probability to homology*, ArXiv e-prints, (2017).

J. P. Vigneaux, *A combinatorial interpretation for tsallis 2-entropy*, arXiv preprint arXiv:1807.05152, (2018).