CANONICAL DISTRIBUTIONS AND PHASE TRANSITIONS

K.B. ATHREYA
AND
J.D.H. SMITH

Department of Mathematics
Iowa State University
Ames, IA 50011, USA

Abstract

Entropy maximization subject to known expected values is extended to the case where the random variables involved may take on positive infinite values. As a result, an arbitrary probability distribution on a finite set may be realized as a canonical distribution. The Rényi entropy of the distribution arises as a natural by-product of this realization. Starting with the uniform distribution on a proper subset of a set, the canonical distribution of equilibrium statistical mechanics may be used to exhibit an elementary phase transition, characterized by discontinuity of the partition function.

1. Introduction.

Since their use by Gibbs in the formulation of statistical mechanics [1], canonical distributions (2.7) have found wide application through the Principle of Maximum Entropy [2] - [6]. According to this principle, canonical distributions are assigned as the most random distributions - i.e. maximizing the information-theoretic entropy (2.6) – subject to knowledge of the expected value(s) of one or more real valued functions on the underlying space. The present paper addresses two leading issues connected with canonical distributions: what special properties do they have, and can they be used to exhibit phase transitions? The first issue is effectively raised by Tribus [3, p.124]:

> The [canonical] distribution ... has many interesting statistical
>
> properties which were first explored by E.T. Jaynes. It seems
>
> fair to state at this writing that not all of the important properties
>
> have yet been found.

The second issue arises from Grandy's observation [6, p.320]:

> The major impression one has at this point is that the apparently
>
> nonanalytic behavior of thermodynamic functions near a phase
>
> transition is exceedingly difficult to extract from the equations
>
> of conventional equilibrium statistical mechanics.

The current paper deals with both these issues by a subtle twist on the Principle of Maximum Entropy: admitting maximization of entropy subject to knowledge of the expected value of an extended-real valued function. It should be noted that the entropy itself is such an expected value, under any distribution admitting zero probabilities.

Canonical distributions are described in Section 2. For simplicity, attention is restricted to finite probability spaces and specification of a single expected value. The classic notation of statistical mechanics is used, e.g. in the choice of sign conventions and in the focus on the

partition function. Section 3 gives a careful demonstration that the canonical distribution is the unique maximizer for the information-theoretic entropy. (Standard proofs from the literature, particularly those using Lagrange multipliers, are not appropriate in the present context. Even proofs by methods closer to those used here may not be applicable. For instance, [3, p. 123] breaks down if $p_i = 0$ and $f_i \neq 0$.)

Section 4 provides a striking answer to the question about the special properties of canonical distributions: There are none, since every distribution is canonical. Specifically, Theorem 4.1 states that every distribution is canonical for the (extended-real valued) log odds function with parameter $\beta = 1$. A curious by-product of the theorem is that choice of parameter values β other than unity leads to the definition of Rényi entropies [7], [5, p.95], [8].

Section 5 applies Theorem 4.1 to the uniform distribution on a proper subset Y of a set X in order to exhibit an elementary phase transition. The phase transition is characterized technically by discontinuity of the partition function. Physically, Y represents a well (or rather *pan*) of finite energy within a region of infinite energy. As usual in statistical mechanics, the canonical distribution describes the equilibrium state of particles within the given system. At finite temperatures, the particles are confined to the pan Y, within which they are distributed uniformly. The phase transition takes place at infinite temperature: the particles boil over, and uniformly occupy the whole space X. Cooling again to finite temperatures, the particles condense back into the pan.

2. The canonical distribution.

Let X be a finite, non-empty set, and let

$$(2.1) \qquad\qquad E : X \to (-\infty, \infty]$$

be an extended-real valued function on the set X. A *probability distribution* on X is a function

$$(2.2) \qquad\qquad p : X \to [0, 1]$$

with

$$(2.3) \qquad \sum_{x \in X} p(x) = 1.$$

The *expected value* of (2.1) under the distribution (2.2) is the extended real

$$(2.4) \qquad \varepsilon = \sum_{x \in X} p(x) E(x).$$

A summand $p(x)E(x)$ in (2.4) is zero whenever $p(x)$ is zero, even if $E(x)$ is infinite. On the other hand, the sum (2.4) is infinite whenever it includes an infinite summand.

For a probability distribution (2.2), the function value $p(x)$ gives the probability of a point x in X. This probability may be zero. The reciprocal $p(x)^{-1}$, which may be infinite, gives the betting odds "$p(x)^{-1}$ to one" for the point x. Define the *log odds function*

$$(2.5) \qquad X \to [0, \infty]; x \mapsto \log p(x)^{-1}.$$

Then the expected value of the log odds function is the (*information-theoretic*) *entropy*

$$(2.6) \qquad H = - \sum_{x \in X} p(x) \log p(x)$$

of the distribution (2.2) [9, p.151] [10].

For a function (2.1) with at least one finite value, and for a non-negative real number β, the *canonical distribution*

$$(2.7) \qquad q : X \to [0, 1]$$

with parameter β is the probability distribution given by

$$(2.8) \qquad - \log q(x) = \log Z(\beta) + \beta E(x)$$

with *partition function* or *Zustandsumme*

$$(2.9) \qquad Z(\beta) = \sum_{x \in X} e^{-\beta E(x)}.$$

If β is zero, (2.9) reduces to the cardinality $|X|$ of the set X, and (2.7) reduces to the uniform distribution on X. Otherwise, if $E(x)$ is infinite, then both $e^{-\beta E(x)}$ and $q(x)$ reduce to zero. Define

$$(2.10) \qquad\qquad X' = \{x \in X | E(x) < \infty\}.$$

Then

$$(2.11) \qquad\qquad \forall\, \beta > 0,\ Z(\beta) = \sum_{x \in X'} e^{-\beta E(x)}.$$

Since $\sum_{x \in X'} e^{-\beta E(x)}$ is a continuous function of β, one has

$$(2.12) \qquad\qquad \lim_{\beta \to 0+} Z(\beta) = |X'|.$$

Proposition 2.1. *If $|X'| \geq 1$, the partition function*

$$(2.13) \qquad\qquad Z : [0, \infty) \to (0, \infty)$$

is logarithmically convex.

Proof. For $\beta > 0$ (2.11) yields

$$(2.14) \qquad\qquad \frac{d}{d\beta} \log Z(\beta) = -\sum_{x \in X'} q(x) E(x) = -\sum_{x \in X} q(x) E(x)$$

and $\dfrac{d^2}{d\beta^2} \log Z(\beta) = \dfrac{Z''(\beta)}{Z(\beta)} - \left[\dfrac{Z'(\beta)}{Z(\beta)}\right]^2 = \left[\sum_{x \in X'} q(x) E(x)^2\right] - \left[\sum_{x \in X'} q(x) E(x)\right]^2$

$= \sum_{x \in X'} q(x)[E(x) - \sum_{y \in X'} q(y) E(y)]^2 \geq 0$. Thus $\log Z(\beta)$ is convex on the open interval $(0, \infty)$. The convexity of $\log Z(\beta)$ on all of $[0, \infty)$ then follows, using (2.12), by $\log Z(0) = \log |X| \geq \log |X'| = \lim_{\beta \to 0+} \log Z(\beta)$. $\square$

Corollary 2.2. *If $|E(X')| > 1$, the function*

$$(2.15) \qquad\qquad \log Z : (0, \infty) \to (-\infty, \infty)$$

is strictly convex.

Proof. If $|E(X')| > 1$, the proof of Proposition 2.1 shows that $\dfrac{d^2}{d\beta^2} \log Z(\beta)$

$= \sum_{x \in X'} q(x) \left[E(x) - \sum_{y \in X'} q(y) E(y)\right]^2 > 0$. $\square$

3. Entropy maximization.

Consider a function (2.1) with at least one finite value, and a parameter β. The expected value of (2.1) under the canonical distribution (2.7) is

$$(3.1) \qquad \varepsilon = \sum_{x \in X} q(x) E(x).$$

It will be shown that, amongst all distributions (2.2) satisfying (2.4) with ε given by (3.1), the entropy (2.6) is maximized only by the canonical distribution.

Proposition 3.1 [11]. *Let p and q be probability distributions on X. Then*

$$(3.2) \qquad \forall\ x \in X,\ p(x) \log q(x) - p(x) \log p(x) \le q(x) - p(x)$$

and

$$(3.3) \qquad \forall\ x \in X,\ p(x) \log q(x) - p(x) \log p(x) = q(x) - p(x) \Leftrightarrow p(x) = q(x).$$

Proof. If $p(x) = 0$, (3.2) reduces to $0 \le q(x)$. Equality holds iff $q(x) = 0 = p(x)$, verifying (3.3) in this case. Otherwise, if $q(x) = 0$, (3.2) reduces to $-\infty \le -p(x)$. Equality cannot hold in this case, so (3.3) is verified again. Otherwise, both $p(x)$ and $q(x)$ are non-zero. Division by $p(x)$ shows that (3.2) is equivalent to

$$(3.4) \qquad \log \frac{q(x)}{p(x)} \le \frac{q(x)}{p(x)} - 1.$$

Consider the auxiliary function

$$(3.5) \qquad \varphi : (0, \infty) \to \mathbb{R}; y \mapsto y - 1 - \log y.$$

Then $\varphi'(y) = 1 - y^{-1}$ is negative on $(0, 1)$ and positive on $(1, \infty)$. Thus φ has a unique global minimum of 0 at $y = 1$. Setting $y = q(x)/p(x)$ verifies (3.4), equality holding iff $p(x) = q(x)$. $\square$

Theorem 3.2 [11]. *Consider the problem of maximizing the entropy* (2.6) *subject to the constraints* (2.2)–(2.4), *with ε given by* (3.1). *Then the maximization is given by a unique distribution, the canonical* (2.7).

Proof. Consider a distribution p satisfying the constraints, and let the canonical distribution be q. By (3.2), one has

$$(3.6) \qquad \sum_{x \in X} p(x) \log q(x) - \sum_{x \in X} p(x) \log p(x) \leq 0.$$

By (3.3), equality holds in (3.6) iff $p = q$. Now (3.6) may be rewritten via (2.8) as

$$-\sum_{x \in X} p(x) \log p(x) \leq -\sum_{x \in X} p(x) \log q(x)$$

$$= \sum_{x \in X} p(x) \left[\log Z(\beta) + \beta E(x)\right]$$

$$= \log Z(\beta) + \beta\varepsilon = -\sum_{x \in X} q(x) \log q(x).$$

Thus the maximum entropy is that of the canonical distribution, and the canonical distribution is the only one yielding this maximum. □

Corollary 3.3. *Suppose $|E(X')| > 1$. Let the interval J be the interior of the convex hull of $E(X')$. Then for ε in J, there is a unique parameter value β, given by*

$$(3.7) \qquad \frac{d}{d\beta} \log Z(\beta) = -\varepsilon,$$

such that the canonical distribution with parameter β yields expected value ε and entropy

$$(3.8) \qquad H(\varepsilon) = \log Z(\beta) + \beta\varepsilon.$$

Proof. Corollary 2.2 shown that (3.7) has a unique solution. Equation (3.8) then follows from (2.8) as in the proof of Theorem 3.2.

Corollary 3.4. *In the context of Corollary 3.3,* (3.8) *yields a concave function*

$$(3.9) \qquad H : J \to (0, \infty)$$

as the Legendre transform of the convex function (2.15).

Proof. Compare [6, §2C]. □

4. Every distribution is canonical.

Let (2.2) be an arbitrary distribution on X. It is readily shown that there is a function (2.1) and a parameter β such that (2.2) is canonical.

Theorem 4.1. *Each probability distribution* (2.2) *on* X *is canonical for the log odds function* (2.5) *and the parameter* $\beta = 1$.

Proof. For $\beta = 1$ and $E(x) = \log p(x)^{-1}$, (2.9) yields

$$(4.1) \qquad Z(1) = \sum_{x \in X} \exp(-\log p(x)^{-1}) = 1.$$

Equation (2.8) then shows that $\log q(x)$ coincides with $\log p(x)$. $\square$

Within the context of Theorem 4.1, it is interesting to consider canonical distributions q with parameter values different from 1. Using (3.8), the entropy of such a distribution is $H(\varepsilon) = \log Z(\beta) + \beta \varepsilon$. However, the expected value ε for the log odds function may itself be interpreted as an entropy H. Solving (3.8) formally as

$$(4.2) \qquad H = \log Z(\beta) + \beta H$$

for H then yields the *Rényi entropy*

$$(4.3) \qquad H = \frac{1}{1 - \beta} \log \sum_{x \in X} p(x)^{\beta}$$

for the parameter $\beta \neq 1$ [7], [5, p.95], [8].

5. An elementary phase transition.

To obtain an elementary model of a phase transition, the method of Theorem 4.1 may be applied to a uniform distribution on a subset of a finite set. Let the cardinality of the set X be N, and let the cardinality of a proper, non-empty subset Y be n. Consider the uniform distribution

$$(5.1) \qquad p(x) = \textbf{ if } \ x \in Y \textbf{ then } \ n^{-1} \textbf{ else } \ 0$$

on Y, with corresponding log odds function

(5.2) $$E(x) = \textbf{ if } \ x \in Y \textbf{ then } \ \log n \textbf{ else } \ \infty.$$

For the function (5.2), the partition function (2.9) is discontinuous. Indeed,

(5.3) $$Z(0) = \sum_{x \in X} 1 = N,$$

while (2.12) yields

(5.4) $$\lim_{\beta \to 0+} Z(\beta) = |Y| = n < N.$$

In this simple model, it is the discontinuity of the partition function at $\beta = 0$ that is characteristic of a phase transition. As noted in Section 2, the canonical distribution with parameter 0 is uniform on all of X. For $\beta > 0$, (2.11) yields

(5.5) $$Z(\beta) = n^{1-\beta},$$

and (2.8) then shows that the canonical distribution with positive parameter β is uniform on the proper subset Y.

The phase transition admits a more directly physical interpretation. Suppose that physical space is partitioned into N cells, and that X is the set of cells. The subset Y corresponds to a certain confined region of space. The function (5.2) may be interpreted as an energy distribution, under which Y is a well of finite energy within a sea of infinite energy. (In this context, words like "caisson" or "pan" might be more appropriate than the usual word "well".) Since values of E now carry the units of energy, while (2.8) forces the $\beta E(x)$ to be dimensionless, one may write

(5.5) $$\beta = 1/\kappa T$$

with κ as Boltzmann's constant and T as temperature. The canonical distribution with parameter β then describes the equilibrium distribution of a system of particles subject to the energy distribution (5.2) at the temperature T corresponding to β via (5.5) [6, §3A]. At finite temperatures, the particles are confined to a uniform distribution within the pan Y. Under the phase transition at infinite temperature, the particles then literally boil over and uniformly occupy the whole space X.

REFERENCES

1. J.W. Gibbs, *Elementary Principles in Statistical Mechanics Developed with Especial Reference to the Rational Foundation of Thermodynamics*, Yale University Press, New Haven, CT, 1902.

2. E.T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev **106** (1957), 620.

3. M. Tribus, *Rational Descriptions, Decisions and Designs*, Pergamon, Elmsford, NY, 1969.

4. R.D. Levine and M. Tribus (eds.), *The Maximum Entropy Formalizm*, MIT Press, Cambridge, MA, 1979.

5. J.H. Justice (ed.), *Maximum Entropy and Bayesian Methods in Applied Statistics*, Cambridge University Press, Cambridge, 1986.

6. W.T. Grandy, Jr., *Foundations of Statistical Mechanics, Volume I*, Reidel, Dordrecht, 1987.

7. A. Rényi, *On measures of entropy and information,* in *"Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability 1960, Volume 1"*, University of California Press, Berkeley, CA, 1961, p. 547.

8. J. Aczél, *Measuring information beyond information theory – why some generalized information measures may be useful, others not*, Aequationes Math. **27** (1984), 1.

9. W. Pauli, *Die allgemeinen Prinzipien der Wellenmechanik,* in *"Handbuch der Physik, Bd. 24/1 Quantentheorie"*, Springer, Berlin, 1933, pp. 83.

10. C. Shannon, *A mathematical theory of communication*, Bell Systems Technical Journal **23** (1948), 349.

11. K.B. Athreya, *Entropy maximization*, IMA Preprint Series # 1231, Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN, 1994.