A note on exponential families of distributions



2007 J. Phys. A: Math. Theor. 40 F691

(http://iopscience.iop.org/1751-8121/40/30/F01)

IOP FTC
IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) F691–F695 doi:10.1088/1751-8113/40/30/F01

FAST TRACK COMMUNICATION

# A note on exponential families of distributions

**Dorje C Brody**

Department of Mathematics, Imperial College, London SW7 2BZ, UK

Received 15 May 2007, in final form 19 June 2007
Published 12 July 2007
Online at stacks.iop.org/JPhysA/40/F691

**Abstract**
We show that an arbitrary probability distribution can be represented in an exponential form. In physical contexts, this implies that the equilibrium distribution of any classical or quantum dynamical system is expressible in a grand canonical form.

PACS numbers: 05.30.−d, 05.30.Ch, 45.20.Jj

Exponential families of probability distributions play central roles in information theory [1], statistics [2] and statistical mechanics [3]. Thus, there arises the interesting question of whether a given system of probability distributions admits a representation in an exponential form. Recent work has shown that, in the case of finite-dimensional quantum systems, the time average of the density matrix can be expressed as a grand canonical state, which assumes an exponential form [4]. Motivated by this result, in the present paper we derive a general theorem stating that an arbitrary system of discrete or continuous probability densities admits a representation in the form of an exponential family. This is surprising in that even power-law distributions are thereby representable in an exponential form.

The paper is organized as follows. We first establish the result for discrete and finite probability densities. An example of this result has been demonstrated in [4]; the purpose here is to provide a simpler derivation of the general result. We then proceed to consider the exponential representation for an arbitrary smooth positive probability density function $\pi(x)$, and show that an expression of the form $\pi(x) = \exp\left(-\sum_k \beta_k x^k\right)$ is always possible.

1. We begin our analysis in the case of a finite-dimensional discrete probability distribution. Let $H$ be a random variable assuming distinct values $\{E_i\}_{i=0,1,\ldots,n}$ with probabilities $\{\pi_i\}_{i=0,1,\ldots,n}$. Then, there is a linearly independent family of $n$ random variables, including $H$ itself, such that any of these random variables can be expressed as a function of $H$. There is a freedom in the choice of the family; here, for simplicity, we choose the powers of $H$; thus, our family of independent random variables is just the set $\{1, H, H^2, H^3, \ldots, H^n\}$. The linear independence of these random variables, i.e. the fact that the matrix of powers $\left\{E_i^k\right\}$ is nonsnigular, follows from the elementary fact that an $n$th order polynomial vanishing at $n+1$ distinct points must be identically zero. Moreover, the powers $H^m$ for all $m > n$ are obviously expressible as linear combinations of powers $\{H^k\}_{k=0,1,\ldots,n}$. We define the moments

 

$\{\mu_k\}_{k=0,1,\ldots,n}$ of $H$ by

$$\mu_k = \sum_{i=0}^{n} \pi_i E_i^k, \tag{1}$$

where $\mu_0 = 1$. To establish the existence of an exponential representation for $\{\pi_i\}$ two further ingredients are needed; the first is the logarithmic entropy of Shannon and Wiener, defined by

$$S = -\sum_{i=0}^{n} \pi_i \ln \pi_i. \tag{2}$$

The second is the family of variables $\{\beta_k\}_{k=1,\ldots,n}$ conjugate to the moments $\{\mu_k\}$ with respect to the entropy $S$ in the sense that

$$\beta_k = \frac{\partial S}{\partial \mu_k}. \tag{3}$$

We then have the following result.

**Proposition 1.** *The family of probabilities* $\{\pi_i\}_{i=0,1,\ldots,n}$ *introduced above can be expressed in the exponential form*

$$\pi_i = \exp\left(-\sum_{k=1}^{n} \beta_k E_i^k - \ln Z(\beta)\right), \tag{4}$$

*where* $Z(\beta) = \sum_{i=0}^{n} \exp\left(-\sum_{k=1}^{n} \beta_k E_i^k\right)$.

Since the matrix $\left\{E_i^k\right\}$ is nonsingular, equations (1) can be solved to express the $\{\pi_i\}$ as linear functions of the moments $\{\mu_k\}$. That is, we can write

$$\pi_i = \sum_{j=0}^{n} c_{ij} \mu_j, \tag{5}$$

where the constant coefficient matrix $\{c_{ij}\}$ is just the inverse of the matrix $\left\{E_i^k\right\}$. Since the entropy is a concave function of the moments $\{\mu_k\}$, the conjugate variables $\{\beta_k\}$ introduced in (3) are in one-to-one correspondence with $\{\mu_k\}$. In other words, (3) defines a Legendre transform [5]. Thus, in principle we can express the moments $\{\mu_k\}$ in terms of the conjugate variables $\{\beta_k\}$, substitute the results in (5) and express the probabilities $\{\pi_i\}$ in terms of the variables $\{\beta_k\}$. The proposition above states that the result of this nonlinear transform can be expressed analytically, and is given by an exponential family of distributions.

**Proof.** Since the row vectors $|i\rangle = \left(E_i^0, E_i^1, E_i^2, \ldots, E_i^n\right)$ for $i = 0, 1, \ldots, n$ are linearly independent, we can express the vector $-\ln \pi_i \in \mathrm{R}^{n+1}$ in the form

$$-\ln \pi_i = \sum_{k=0}^{n} \beta_k E_i^k \tag{6}$$

for some coefficients $\{\beta_k\}$. Substituting (6) in (2), we obtain

$$S = \sum_{k=0}^{n} \beta_k \mu_k, \tag{7}$$

from which we deduce (3) *a posteriori*. Finally, solving (6) for $\pi_i$ we obtain the desired form (5), where the normalization condition for $\{\pi_i\}$ implies that $\beta_0 = \ln Z(\beta)$. □

By the above result, the nonlinear transform (3) can be inverted analytically in the form

$$\mu_k = \sum_{i=0}^{n} E_i^k \exp\left(-\sum_{l=0}^{n} \beta_l E_i^l\right). \tag{8}$$

2. An exponential representation can also be derived in the case of an arbitrary smooth probability density function. In the continuous case, however, the moments of the distribution need not exist in general. Therefore, some of the preceding constructions involving entropy and moments must be altered. We state the main result first.

**Proposition 2.** *Let $\pi(x)$ be a probability density function on the real line such that $\ln \pi(x)$ is quadratically integrable with respect to the Gaussian measure $\mathrm{e}^{-x^2}\mathrm{d}x$. Then $\pi(x)$ can be expressed in the exponential form*

$$\pi(x) = \exp\left(-\sum_{k=1}^{n} \beta_k x^k - \ln Z(\beta)\right), \tag{9}$$

*where $Z(\boldsymbol{\beta}) = \int_{-\infty}^{\infty} \exp\left(-\sum_{k=1}^{n} \beta_k x^k\right) \mathrm{d}x$ and where the value of $n$ may be infinite. The parameters $\{\beta_k\}$ are uniquely determined by $\pi(x)$.*

The statement of proposition 2 is perhaps surprising, because the representation (9) applies, for example, to power-law distributions such as the Cauchy distribution $1/[\pi(1+x^2)]$ for which none of the moments exists. The proof goes as follows.

**Proof.** Since by assumption $\ln \pi(x)$ is quadratically integrable with respect to the Gaussian measure, one can expand $\ln \pi(x) \in \mathcal{L}^2(\mathbb{R}, \mathrm{e}^{-x^2}\mathrm{d}x)$ in terms of the Hermite polynomials $\{H_k(x)\}$, that is,

$$\ln \pi(x) = -\sum_k \gamma_k H_k(x), \tag{10}$$

where

$$\gamma_k = -\frac{1}{\sqrt{\pi} 2^k k!} \int_{-\infty}^{\infty} \ln \pi(x) H_k(x)\, \mathrm{e}^{-x^2}\, \mathrm{d}x. \tag{11}$$

The infinite series on the right side of (10) converges almost everywhere, since the squared Hilbert space norm $\sum_k \gamma_k^2$ converges by assumption. Next, define a set of numbers $\{\beta_k\}$ by the prescription

$$\sum_k \beta_k x^k = \sum_k \gamma_k H_k(x). \tag{12}$$

Substituting this in (10) and solving the result for $\pi(x)$, we deduce (9), where the normalization condition implies that $\beta_0 = \ln Z$.

□

Can we establish a relation analogous to (3) for a general probability density function? To this end we introduce what might appropriately be called the 'Gaussian moments' of $\pi(x)$ by defining

$$\mu_k = \int_{-\infty}^{\infty} x^k \pi(x)\, \mathrm{e}^{-x^2}\, \mathrm{d}x. \tag{13}$$

Similarly, we define the 'Gaussian entropy' of $\pi(x)$ by

$$S = -\int_{-\infty}^{\infty} \pi(x) \ln \pi(x)\, \mathrm{e}^{-x^2}\, \mathrm{d}x. \tag{14}$$

The coefficients $\{\beta_k\}$ appearing in (9) are then related to the Gaussian moments $\{\mu_k\}$ defined by (13) via relation (3), provided we use the Gaussian entropy (14).

The family of density functions for which $\ln\pi(x)$ is quadratically integrable with respect to the Gaussian measure is fairly large and includes, in particular, all the power-law distributions. However, this family is not exhaustive. Nevertheless, the representation (9) can be established for a much wider class of density functions. The idea is to extend the formulation based on the Gaussian measure into the class $\mathfrak{S}$ of positive Schwartz functions (by this we mean functions that have infinite numbers of derivatives, each of which decays faster than any inverse polynomial). This class forms a convex cone which includes, in particular, the Gaussian function $\mathrm{e}^{-x^2}$. Let $s(x) \in \mathfrak{S}$ be a positive Schwartz function such that $s(x)\ln\pi(x)$ is quadratically integrable with respect to the Lebesgue measure. We then construct orthonormal polynomials $\{J_k(x)\}$ in $\mathcal{L}^2(\mathbb{R}, s^2(x)\mathrm{d}x)$ by means of the Gram–Schmidt procedure. Approximating by integration over a finite interval, we can then apply the Weierstrass approximation theorem to establish the completeness of the set $\{J_k(x)\}$. The function $\ln\pi(x)$ can therefore be expanded in a form analogous to (10), with almost everywhere convergence. The coefficients $\{\gamma_k\}$ depend upon the choice of the Schwartz function $s(x)$, whereas the expansion coefficients $\{\beta_k\}$ defined in a manner analogous to (12) are basis independent.

To show that the representation (9) is valid for all smooth density functions we proceed as follows. First, we observe that since $\pi(x)$ is nonnegative, $[1+(\ln\pi(x))^2]^{-1}$ is less than or equal to one for all $x$. Therefore, the function $f(x) = s(x)/[1+(\ln\pi(x))^2]$, for any $s(x) \in \mathfrak{S}$, decays faster than any inverse polynomial. Thus, for an arbitrary smooth density function $\pi(x)$, the logarithm $\ln\pi(x)$ is by construction quadratically integrable in $\mathcal{L}^2(\mathbb{R}, f(x)\mathrm{d}x)$. Of course, the density function could be so perverse that $f(x)$ does not belong to $\mathfrak{S}$, i.e. the derivatives of $f(x)$ need not decay faster than any inverse polynomial. However, the behaviour of these derivatives is immaterial for our construction, since we merely require that all polynomials are quadratically integrable with respect to the measure $f(x)\,\mathrm{d}x$. Consequently, the above exponential representation is indeed valid for all smooth density functions.

In statistics, the exponential family of distributions is generally defined as the totality of density functions that admit representations of the form $\exp\left(-\sum_{k=0}^{n}\beta_k T_k(x)\right)$ for a set of functions (sufficient statistics) $\{T_k(x)\}$, where $n$ is usually assumed finite. The foregoing result thus implies that *the exponential family of distributions is dense in the totality of probability distributions*. Thus, the study of probability distributions could, in principle, be restricted to the exponential type. In specific applications, the practicality of this depends upon the density function $\pi(x)$ and the choice of the Schwartz function $s(x)$, since the rate of convergence depends upon these ingredients.

From the physical point of view, the result established here also leads to an interesting observation concerning equilibrium properties of generic dynamical systems. We note that if a dynamical system is in equilibrium, then the associated equilibrium distribution is necessarily an energy distribution, since steady-state solutions to the Liouville equation (or the Heisenberg equation in the case of a quantum system) are given by functions of the Hamiltonian. Thus, we conclude that *if a dynamical system is in equilibrium, then the relevant equilibrium distribution is necessarily expressible in a grand canonical form*. We emphasize that this result applies not only to thermal equilibrium, but to any form of equilibrium state of a dynamical system.

## Acknowledgment

The author acknowledges support from The Royal Society.

## References

[1] Cover T M and Thomas J Y 1991 *Elements of Information Theory* (Ney York: Wiley)
[2] Brown L D 1986 *Fundamentals of Statistical Exponential Families (Lecture Notes—Monograph Series)* (Hayward, CA: Institute of Mathematical Statistics)
[3] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620
[4] Brody D C, Hook D W and Hughston L P 2007 Unitarity, ergodicity, and quantum thermodynamics *J. Phys. A: Math. Gen.* **40** F503–9
[5] Čencov N N 1982 *Statistical Decision Rules and Optimal Inference* (Providence, RI: American Mathematical Society)