

Exponential Family Entropy and KL divergence

Philip Schulz and Wilker Aziz

last modified: November 8, 2017

Abstract

This note shows how to compute the entropy and KL divergence (and by extension many other information-theoretic quantities) of exponential family distributions. We use the univariate Gaussian distribution as a running example and encourage the reader to apply the learned techniques to other distributions in exercises. We expect the reader to be familiar with exponential families and some of their properties (such as moment computation).

1 Exponential Family Form

Most exponential family distributions are standardly written in what is called their canonical form. In order to apply results proved for exponential families to them, it is often convenient to rewrite them in their natural parametrisation. The general way of doing this is to apply the identity function $\exp(\log(\cdot))$ to them and reorder terms. Before we do so for the Gaussian, let us introduce the notation we will use for exponential families throughout.

$$p(x | \underbrace{\theta}_{\text{canonical parameters}}) = \underbrace{h(x)}_{\text{base measure}} \exp \left(\underbrace{\eta(\theta)^\top}_{\text{natural parameters}} \times \underbrace{t(x)}_{\text{sufficient statistics}} - \underbrace{A(\eta(\theta))}_{\text{log-normalizer}} \right) \quad (1)$$

By default we assume that all terms above are vector-valued. Notice that the canonical parameters may be a function of any conditioning context and thus Equation (1) covers marginal as well as conditional distributions.

Gaussian Distribution We first write the univariate Gaussian distribution with mean μ and variance σ^2 in its canonical form. In that case we have $\theta^\top = [\mu \ \sigma^2]$.

$$p(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right) \quad (2)$$

Now we apply the identity function to transform this density into its exponential family form.

$$p(x|\mu, \sigma^2) = \exp \left(\log \left(\frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right) \right) \right) \quad (3a)$$

$$= \exp \left(\log \left(\frac{1}{\sqrt{2\pi}} \right) - \log \sigma - \frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right) \quad (3b)$$

$$= \frac{1}{\sqrt{2\pi}} \exp \left(-\log \sigma - \frac{x^2 - 2x\mu + \mu^2}{2\sigma^2} \right) \quad (3c)$$

$$= \frac{1}{\sqrt{2\pi}} \exp \left(\frac{2x\mu - x^2}{2\sigma^2} - \log \sigma - \frac{\mu^2}{2\sigma^2} \right) \quad (3d)$$

$$= \frac{1}{\sqrt{2\pi}} \exp \left(\begin{bmatrix} \frac{\mu}{\sigma^2} & -\frac{1}{2\sigma^2} \end{bmatrix} \times \begin{bmatrix} x \\ x^2 \end{bmatrix} - \log \sigma - \frac{\mu^2}{2\sigma^2} \right) \quad (3e)$$

We recognise this as the exponential family form of the Gaussian with

$$\begin{aligned} h(x) &= \frac{1}{\sqrt{2\pi}} & t(x) &= \begin{bmatrix} x \\ x^2 \end{bmatrix} \\ A(\eta(\theta)) &= \log \sigma + \frac{\mu^2}{2\sigma^2} & \eta(\theta) &= \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} . \end{aligned} \quad (4)$$

Exercise Derive the exponential family form of the multivariate Gaussian, exponential and binomial distributions.

2 Entropy of Exponential Families

First, recall the definition of entropy:

$$\mathbb{H}(X) = -\mathbb{E}[\log(X)] \quad (5)$$

A convenient property of exponential families is that the only terms depending on X are the base measure and the sufficient statistics. This means we can split up the expectation term.

$$\mathbb{H}(X) = -\mathbb{E} \left[\log(h(X)) + \eta(\theta)^\top t(X) - A(\eta(\theta)) \right] \quad (6a)$$

$$= -\mathbb{E}[\log(h(X))] - \eta(\theta)^\top \mathbb{E}[t(X)] + A(\eta(\theta)) \quad (6b)$$

In order to compute the expected sufficient statistics we exploit another convenient property of exponential families.

$$\mathbb{E}[t(X)] = \frac{d}{d\eta(\theta)} A(\eta(\theta)) \quad (7)$$

Gaussian Entropy We now compute the entropy of the Gaussian using the above properties. First, we compute the vector of partial derivatives to get the expected sufficient statistics.

$$\frac{\partial}{\partial \frac{\mu}{\sigma^2}} \left\{ \log \sigma + \frac{\mu^2}{2\sigma^2} \right\} = \frac{\partial}{\partial \mu} \sigma^2 \frac{\mu^2}{2\sigma^2} = \mu \quad (8)$$

$$-\frac{\partial}{\partial \frac{1}{2}\sigma^{-2}} \left\{ \log \sigma + \frac{\mu^2}{2\sigma^2} \right\} = -\frac{\partial}{\partial \sigma^{-2}} 2 \log \left(\sigma^{-2 \times (-\frac{1}{2})} \right) - \frac{\partial}{\partial \sigma^{-2}} \frac{2\mu^2}{2\sigma^2} \quad (9a)$$

$$= \frac{\partial}{\partial \sigma^{-2}} \log (\sigma^{-2}) - \frac{\partial}{\partial \sigma^{-2}} \mu^2 \sigma^{-2} = \sigma^2 - \mu^2 \quad (9b)$$

In the case of the Gaussian we can easily verify that these derivatives are correct by computing the expected sufficient statistics directly.

$$\mathbb{E}[X] = \mu \quad \mathbb{E}[X^2] = \sigma^2 + \mu^2 \quad (10)$$

For the second term we have used the equality $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Plugging the expected sufficient statistics into the general entropy term (Equation (6b)), we get

$$\mathbb{H}(X) = -\log \left(\frac{1}{\sqrt{2\pi}} \right) - \left[\frac{\mu}{\sigma^2} \quad -\frac{1}{2\sigma^2} \right] \times \left[\sigma^2 \frac{\mu}{\sigma^2} + \mu^2 \right] + \log \sigma + \frac{\mu^2}{2\sigma^2} \quad (11a)$$

$$= -\log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{\mu^2}{\sigma^2} + \frac{1}{2} + \frac{\mu^2}{2\sigma^2} + \log \sigma + \frac{\mu^2}{2\sigma^2} \quad (11b)$$

$$= -\log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{\mu^2}{\sigma^2} + \frac{2\mu^2}{2\sigma^2} + \frac{1}{2} + \log \sigma \quad (11c)$$

$$= -\log \left(\frac{1}{\sqrt{2\pi}} \right) + \frac{1}{2} + \log \sigma \quad (11d)$$

$$= \frac{1}{2} \log(2\pi) + \frac{1}{2} + \frac{1}{2} \log(\sigma^2) \quad (11e)$$

$$= \frac{1}{2} (\log(2\pi\sigma^2) + 1) . \quad (11f)$$

As we would expect, the Gaussian entropy only depends on the variance but not on the mean.

Exercise Follow the steps above to compute the entropy of the multivariate Gaussian, exponential and binomial distributions.