Bernoulli Distribution: A member of the exponential family

February 10, 2020

A very special class of probability distributions is the exponential, which have the property that analgebraic form of the distrbution can serve as a prior in a Bayesian theorem. The consequence of this property is that graphical inference can be carried out at the symbolic level. This primer introduces the Bernoulli distribution and develops some of its relevant properties.

1 Bernoulli

Basics

$p(x = 1|\mu) = \mu$

$p(x = 0|\mu) = 1 - \mu$

Check for normalization:

$$\sum_{0,1} p(x = 1|\mu) = 1 + 1 - \mu = 1$$

Mean:

$$\sum_{0,1} xp(x = 1|\mu) = 1\mu + 0(1 - \mu) = \mu$$

Variance:

$$\sum_{0,1} (x - \mu)^2 p(x = 1|\mu) = \mu^2 p(x = 0|\mu) + (1 - \mu)^2 p(x = 1|\mu)$$

$$= \mu^2(1-\mu) + (1-u)^2\mu$$

which simplifies to:

$$= \mu(1-\mu)$$

Maximum likely probability estimates from data sets
Consider a data set D $\{x_1, x_2, \ldots, x_N\}$
We can express the probability of the observations as:

$$p(D|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

The goal of maximum likelihood is to find the value of μ that maximizes this probability. To do this it is easier to work with the logarithm that is a monotonic increasing function. So that

$$\ln(D|\mu) = \sum_{n=1}^{N} p(x_n|\mu) = \sum_{n=1}^{N} \{x_n \ln u + (1-x_n)\ln(1-\mu)\}$$

which has a derivative with respect to μ that is set to zero

$$\frac{\partial(\ln p(D|\mu)}{\partial\mu} = \sum_{n=1}^{N} \left(\frac{x_n}{/\mu} + \frac{1-x_n)(-1)}{1-\mu} \right) = 0$$

that can be manipulated to show

$$\mu = \sum_{n=1}^{N} \frac{x_n}{N}$$

so μ counts the fraction of ones.

Beta distribution Thr motivation for the Beta distribution is to specicify a prior ditsribution that expresses our prior beliefs in the form of the Bernoulli to protect us from a skewed result from a unlikely data set. The Beta distribution, Beta(α,β) -s given by

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\mu^{a-1}(1-\mu^{b-1})$$

From the definitions of the Beta distribution, we know that its normalizing constant is given by:

$$\int_0^1 \mu^{a-1}(1-\mu^{b-1})d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{1}$$

We can se this expression (and $(\Gamma(x+1) = x\Gamma(x)$ to show that

$$E(\mu) = \frac{a}{(a+b)}$$

$$E(\mu) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \mu^{a+1-1}(1-\mu^{b-1})d\mu \tag{2}$$

$$\tag{3}$$

Now taking of advantage of Eq.**??** with $\alpha+1$ instead of α,

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} \tag{4}$$

$$= \frac{a}{(a+b)} \tag{5}$$

Exercise Show the result of combining the Bernouli likelihood distribution with is Beta prior.