Binary Variables

Definition. If x is a random variable and $x \in \{0, 1\}$ then x is a binary variable.

Definition. Let $P(x = 1) = \mu$. Then the *Bernoulli* distribution is given by

$$\mathrm{Bern}(x|\mu) = \mu^x (1 - \mu)^{1-x}.$$

Let $\mathcal{D} = x_1, ..., x_N$, then

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

which forms the likelihood function for $p(\mu|\mathcal{D})$. Then,

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

Since the above function depends on $\mathcal{D}$ only through the sum, $\sum_{n=1}^{N} x_n$ is called a *sufficient statistic*. Maximising the log will result in the sample mean as follows

$$\mu_{\mathrm{ML}} = \frac{1}{N} \sum_{n=1}^{N} x_n.$$

If there is a μ chance of a binary variable being 1, then we can use a binomial distribution to compute the probability of drawing m instances of 1 out of a data set of N elements.

Definition. Given a size N and m, and a mean μ, the *binomial distribution* is defined as follows

$$\mathrm{Bin}(m|N, \mu) = \binom{N}{m} \mu^m (1 - \mu)^{N-m}$$

where

$$\binom{N}{m} = \frac{N!}{m!(N - m)!}.$$

For the binomial distribution, we have the following properties

$$\mathbb{E}[m] = N\mu$$
$$\mathrm{var}[m] = N\mu(1 - \mu).$$

We now write a useful way of generalising the factorial to non-negative reals.

Definition. The *Gamma Distribution* is given by the following

$$\Gamma(x) = \int y^{x-1} e^{-y} \ dy$$

It is easy to show that the Gamma distribution possesses the property $\Gamma(x) = x!$ for any $x \in \mathbb{N}$.

For a Bayesian approach, we need to define a prior for μ. Due to various useful properties, we use the following.

Definition. Given hyper-parameters a and b, the *Beta Distribution* is given by the following.

$$\mathrm{Beta}(\mu|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1} (1-\mu)^{b-1}$$

We define the prior $p(\mu) = \mathrm{Beta}(\mu|a, b)$.

The Beta distribution has the following properties.

$$\int_0^1 \mathrm{Beta}(\mu|a, b) \ d\mu = 1.$$

$$\mathbb{E}[\mu] = \frac{a}{a+b}$$

$$\mathrm{var}[\mu] = \frac{ab}{(a+b)^2 (a+b+1)}$$

The posterior can then be deduced by multiplying the beta function with the binomial likelihood, then normalising.