1 Normal binomial distribution

1.1 Probability mass function

The probability mass function (p.m.f.) for a random variable X with a 'normal' binomial distribution with probability of success p, and m trials is

$$P(X = k) = \binom{m}{k} p^k (1-p)^{m-k} \tag{1}$$

that is k successes and $m - k$ failures with a coefficient for the number of different ways we can have k successes and $m - k$ failures.

1.2 Likelihood function

If we take n i.i.d. samples from X consisting of n integers between 0 and m, $\{k_1, \ldots, k_n\}$, the probability of these data, aka the likelihood is

$$P(\{k_1, \ldots, k_n\}|p) = L(K|p, m) = \prod_{i=1}^{n} \binom{m}{k_i} p^{k_i} (1-p)^{m-k_i} \tag{2}$$

that is the product of all the samples.

1.3 Maximum likelihood estimation

In order to estimate the value of p from the sample $\{k_1, \ldots, k_n\}$, we maximise the value of the likelihood function with respect to p. Since $\log$ is an increasing function, maximising the log of the likelihood function is equivalent to maximising the likelihood function. But because likelihood functions tend to take the form of large products, maximising the log of the likelihood is often easier than maximising the likelihood. So we will maximise the log of the likelihood function, a.k.a. the 'log likelihood'.

$$l(p, m) = \log L(K|p, m) = \log \left(\prod_{i=1}^{n} \binom{m}{k_i} p^{k_i} (1-p)^{m-k_i} \right) \tag{3}$$

$$= \sum_{i=1}^{n} \log \binom{m}{k_i} + k_i \log p + (m - k_i) \log(1-p) \tag{4}$$

The derivative with respect to p is

$$\frac{\partial l(p, m)}{\partial p} = \sum_{i=1}^{n} \frac{k_i}{p} - \frac{m - k_i}{1-p} \tag{5}$$

Letting the derivative equal 0 gives

$$\sum_{i=1}^{n} \frac{k_i}{\hat{p}} - \frac{m-k_i}{1-\hat{p}} = 0 \tag{6}$$

$$\implies \sum_{i=1}^{n} \frac{k_i(1-\hat{p}) - \hat{p}(m-k_i)}{\hat{p}(1-\hat{p})} = 0 \tag{7}$$

$$\implies \sum_{i=1}^{n} k_i - k_i\hat{p} - m\hat{p} + k_i\hat{p} = 0 \tag{8}$$

$$\implies \hat{p} = \frac{1}{nm}\sum_{i=1}^{n} k_i \tag{9}$$

As you might expect, our maximum likelihood estimate for p is the total number of successes divided by the total number of trials. It's interesting to note that m is technically a parameter of the binomial distribution, and if we didn't know m beforehand, we would be unable to estimate p or m using the maximum likelihood method (I think).

The number of trials m is also a parameter of the binomial distribution. In order to find an estimate for m using maximum likelihood estimation, we take the derivative of $l(p, m)$ with respect to m.

$$\frac{\partial l(p,m)}{\partial m} = \sum_{i=1}^{n} \frac{\partial \log \binom{m}{k_i}}{\partial m} + \log(1-p) \tag{10}$$

Using

$$\frac{\partial \log \binom{m}{k_i}}{\partial m} = H_m - H_{m-k_i} \tag{11}$$

where H_n is the nth Harmonic number [1], and letting the derivative equal 0 gives

$$\sum_{i=1}^{n} H_m - H_{m-k_i} + \log(1-p) = 0 \tag{12}$$

$$\tag{13}$$

Theoretically we could sub in our expression for $\hat{p}$ and solve for m (??? is this true ???). But I don't know how to do this right now.

Now we know how to do maximum likelihood estimation, we move onto a more complicated distribution.

2 Conway-Maxwell binomial distribution

The Conway-Maxwell binomial distribution is similar to the binomial distribution in that we can think of it as a sum of Bernoulli trials. But for the Conway-Maxwell binomial distribution, the Bernoulli variables are associated with each other. Not dependent, not correlated, but associated.

[1] `https://en.wikipedia.org/wiki/Harmonic_number`

2.1 Probability mass function

The p.m.f. of a random variable X with the Conway-Maxwell binomial distribution with probability of success p, dispersion parameter ν, and number of trials m is

$$P(X = k) = \frac{1}{S(p, \nu, m)} \binom{m}{k}^{\nu} p^k (1-p)^{m-k} \tag{14}$$

where

$$S(p, \nu, m) = \sum_{i=0}^{m} \binom{m}{i}^{\nu} p^i (1-p)^{m-i} \tag{15}$$

is a normalising function.

The use of the dispersion parameter ν enables the Conway-Maxwell binomial distribution to 'over-disperse' or 'under-disperse' relative to a binomial distribution i.e., have greater or lesser variance.

When $\nu > 1$ the distribution is under-dispersed relative to a binomial distribution. In the limit that $\nu \to \infty$ all the mass accumulates at $m/2$ for even m, and at $\lfloor m/2 \rfloor$ and $\lceil m/2 \rceil$ for odd m. This corresponds to negatively associated Bernoulli variables. For $\nu < 1$, the distribution is over-dispersed relative to a binomial distribution. In the case where $\nu \to -\infty$ all the mass is distributed at 0 and m. This is the extreme case of positive association, where all the Bernoulli variables have the value. So, ν measures the strength of the negative or positive association between the Bernoulli variables that make up the Conway-Maxwell binomial distribution. Note that if $\nu = 1$ we have the 'normal' binomial distribution described in section 1.

2.2 Likelihood function

If we take n i.i.d. samples from a random variable X with the Conway-Maxwell binomial distribution with parameters p, ν, and m this gives us n integers between 0 and m, $\{k_1, \dots, k_n\}$. These probability or 'likelihood' of these data is

$$P(\{k_1, \dots, k_n\} | p, \nu, m) = L(K | p, \nu, m) = \prod_{i=1}^{n} \frac{\binom{m}{k_i}^{\nu} p^{k_i} (1-p)^{m-k_i}}{S(p, \nu, m)} \tag{16}$$

$$= \frac{\prod_{i=1}^{n} \binom{m}{k_i}^{\nu} p^{k_i} (1-p)^{m-k_i}}{S(p, \nu, m)^n} \tag{17}$$

Again, just the product of the individual probabilities of each sample.

2.3 Maximum likelihood estimation

Once again, we maximise the log-likelihood function, which is

$$l(p, \nu, m) = \log L(K | p, \nu, m) = \log \frac{\prod_{i=1}^{n} \binom{m}{k_i}^{\nu} p^{k_i} (1-p)^{m-k_i}}{S(p, \nu, m)^n} \tag{18}$$

$$= -n \log S(p, \nu, m) + \sum_{i=1}^{n} \log \binom{m}{k_i}^{\nu} + k_i \log p + (m - k_i) \log(1-p) \tag{19}$$

Maximising will involve calculating the partial derivative of $\log S(p,\nu,m)$ with respect to the parameters.

$$\frac{\partial \log S(p,\nu,m)}{\partial p} = \frac{1}{S(p,\nu,m)}\frac{\partial S(p,\nu,m)}{\partial p} \tag{20}$$

For the partial derivative without the $\log$, we have

$$\frac{\partial S(p,\nu,m)}{\partial p} = \frac{\partial \sum_{i=0}^{m} \binom{m}{i}^{\nu} p^i(1-p)^{m-i}}{\partial p} \tag{21}$$

$$= \sum_{i=0}^{m} \binom{m}{i}^{\nu} \left[ip^{i-1}(1-p)^{m-i} - (m-i)p^i(1-p)^{m-i-1}\right] \tag{22}$$

which gives

$$\frac{\partial \log S(p,\nu,m)}{\partial p} = \sum_{i=0}^{m} \frac{\binom{m}{i}^{\nu} \left[ip^{i-1}(1-p)^{m-i} - (m-i)p^i(1-p)^{m-i-1}\right]}{\binom{m}{i}^{\nu} p^i(1-p)^{m-i}} \tag{23}$$

$$= \sum_{i=0}^{m} ip^{-1} - (m-i)(1-p)^{-1} \tag{24}$$

$$= \sum_{i=0}^{m} \frac{i}{p} - \frac{m-i}{1-p} \tag{25}$$

$$= \sum_{i=0}^{m} \frac{i(1-p) - (m-i)p}{p(1-p)} \tag{26}$$

$$= \sum_{i=0}^{m} \frac{i - mp}{p(1-p)} \tag{27}$$

$$= \frac{m(m+1) - 2m(m+1)p}{2p(1-p)} \tag{28}$$

$$= \frac{m(m+1)(1-2p)}{2p(1-p)} \tag{29}$$

using $\sum_{i=0}^{m} i = m(m+1)/2$.

So the partial derivative of the log-likelihood function with respect to p is

$$\frac{\partial l(p,\nu,m)}{\partial p} = -n\frac{\partial \log S(p,\nu,m)}{\partial p} + \sum_{i=1}^{n} \frac{k_i}{p} - \frac{m-k_i}{1-p} \tag{30}$$

$$= -n\frac{\partial \log S(p,\nu,m)}{\partial p} + \sum_{i=1}^{n} \frac{k_i - mp}{p(1-p)} \tag{31}$$

$$= \frac{-nm(m+1)(1-2p)}{2p(1-p)} + \sum_{i=1}^{n} \frac{k_i - mp}{p(1-p)} \tag{32}$$

$$= \frac{-nm(m+1)(1-2p) - 2nmp + 2\sum_{i=1}^{n} k_i}{2p(1-p)} \tag{33}$$

$$= \frac{-nm\left[(m+1)(1-2p) + 2p\right] + 2\sum_{i=1}^{n} k_i}{2p(1-p)} \tag{34}$$

$$= \frac{-nm\left[m - 2mp + 1 - 2p + 2p\right] + 2\sum_{i=1}^{n} k_i}{2p(1-p)} \tag{35}$$

$$= \frac{-nm\left[m(1-2p) + 1\right] + 2\sum_{i=1}^{n} k_i}{2p(1-p)} \tag{36}$$

References

[1] Joseph B. Kadane, *Sums of possibly associated Bernoulli variables: The Conway-Maxwell binomial distribution*. Bayesian Analysis 11, 403 - 420, (2016)