Fisher Information for Various Binomial Parameterizations

by Spencer Graves

This document records the computation of the Fisher information for various parameterizations of the binomial distribution. Recent internal PDF simulations by Sundar Dorai-Raj have shown that with yields close to 0 or 1, the actual coverage probabilities using the standard binomial are quite poor, less than 85% for a nominal 95% confidence interval with yield less than 2% or greater than 98%, and the logit and complementary log intervals are much better. This is consistent with recent theoretical work by Brown, Cai and DasGupta (2001, 2002), who carefully documented interesting oscillations with variations in yield in the actual coverage probabilities of alternative confidence interval procedures for the binomial theoretical yield parameter; Brown, Cai and DasGupta considered the standard interval and the logit, among others, but did not consider the complementary log procedure included here and considered in the simulations by Sundar Dorai-Raj.

y = number of yielding die out of n, with theoretical average p.

$$f(y) = \binom{n}{y} p^{y} (1-p)^{n-y}$$

log likelihood

$$l = c + y\log(p) + (n-y)\log(1-p)$$

Standard Parameterization

Fisher Score

$$S_p = \frac{\partial l}{\partial p} = \frac{y}{p} - \frac{(n-y)}{(1-p)} = \frac{y-np}{p(1-p)}$$

Observed information

$$I_p = \left(-\frac{\partial S_p}{\partial p}\right) = \left[\frac{-1}{p(1-p)}\right]\left\{-n + (y-np)\left[-\frac{1}{p} + \frac{1}{(1-p)}\right]\right\}$$
$$= \left[\frac{1}{p(1-p)}\right]\left\{n - (y-np)\left[\frac{2p-1}{p(1-p)}\right]\right\}$$

Fisher Information

$$E(I_p) = \left[\frac{n}{p(1-p)}\right]$$

In this case, the variance of the maximum likelihood estimate $\hat{p} = y/n$ is exactly the inverse of the Fisher information:

$$\text{var}(\hat{p}) = \left[\frac{p(1-p)}{n}\right]$$

However, the normal approximation for the distribution of $\hat{p}$ is not very good if p is close to 0 or 1.

Logistic

$$p = \frac{e^{\zeta}}{1+e^{\zeta}}, \text{ so } \zeta = \log\left(\frac{p}{1-p}\right) \text{ and } \frac{\partial p}{\partial \zeta} = p(1-p)$$

Fisher Score

$$S_\zeta = \frac{\partial l}{\partial \zeta} = \frac{\partial l}{\partial p}\frac{\partial p}{\partial \zeta} = \left[\frac{y-np}{p(1-p)}\right]p(1-p) = (y-np)$$

Observed information

$$I_\zeta = \left(-\frac{\partial S_\zeta}{\partial \zeta}\right) = np(1-p)$$

Fisher Information = observed information. When $y = 0$ or n, $\hat{p} = 0$ or 1, and the maximum likelihood estimate for the logit parameter is $\hat{\zeta} = \mp\infty$. These outcomes occur with probability $(1-p)^n$ and p^n, respectively. Fortunately, the standard asymptotics still apply here, with the result that the distribution of $\hat{\zeta}$ admits a normal approximation with variance parameter given by the inverse of the Fisher information:

$$\frac{1}{np(1-p)}.$$

Since $\hat{\zeta} = \mp\infty$ with nonzero probability, the actual variance of $\hat{\zeta}$ is, of course, infinite. However, as long as n is large enough to make the probability of such events sufficiently small, the normal approximation to the distribution of $\hat{\zeta}$ can still be useful.

Probit

$$p = \Phi(z), \text{ so } \frac{\partial p}{\partial z} = \phi(z) \text{ and } \frac{\partial \phi(z)}{\partial z} = (-z)\phi(z)$$

where ϕ and Φ = standard normal density and cumulative distribution functions, respectively.

Fisher Score

$$S_z = \frac{\partial l}{\partial z} = \frac{\partial l}{\partial p}\frac{\partial p}{\partial z} = \left[\frac{y-np}{p(1-p)}\right]\phi(z)$$

Observed information

$$I_z = \left(-\frac{\partial S_z}{\partial z`}\right) = \left[\frac{-\phi(z)}{p(1-p)}\right]\left\{-n\phi(z) + (y-np)\left[(-z) - \frac{\phi(z)}{p} + \frac{\phi(z)}{1-p}\right]\right\}$$

$$= \left[\frac{\phi(z)}{p(1-p)}\right]\left\{n\phi(z) + (y-np)\left[z + \frac{\phi(z)}{p(1-p)}\right]\right\}$$

Fisher Information

$$E(I_z) = \frac{n[\phi(z)]^2}{p(1-p)}$$

As with the logistic parameterization, when $y = 0$ or n, $\hat{p} = 0$ or 1, and the maximum likelihood estimate for the probit parameter is $\hat{z} = \mp\infty$. Again, the standard asymptotics apply with $\hat{z}$ having an approximate normal distribution with variance

$$\frac{p(1-p)}{n[\phi(z)]^2}$$

Poisson

$$p = e^{-\mu}\text{, so } \mu = [-\log(p)] \text{ and } \frac{\partial p}{\partial \mu} = (-p)$$

Fisher Score

$$S_\mu = \frac{\partial l}{\partial \mu} = \frac{\partial l}{\partial p}\frac{\partial p}{\partial \mu} = \left[\frac{y-np}{p(1-p)}\right](-p) = \left(\frac{np-y}{1-p}\right)$$

Observed information

$$I_\mu = \left(-\frac{\partial S_\mu}{\partial \mu}\right) = \left[-(1-p)^{-1}\right]\left\{-np + \left(\frac{np-y}{1-p}\right)p\right\} = \left(\frac{p}{1-p}\right)\left\{n + \left(\frac{y-np}{1-p}\right)\right\}$$

Fisher Information

$$E(I_\mu)=\left(\frac{np}{1-p}\right)$$

Similar to the logit and probit parameterizations, when $y = 0$, $\hat{\mu}=\infty$, which occurs with probability $(1-p)^n$. Nevertheless, the distribution of $\hat{\mu}$ still admits a normal approximation with variance parameter equal to the reciprocal of the Fisher information:

$$\frac{(1-p)}{np}$$

We know, however, for empirical studies that the distribution of a Poisson defect rate estimated from yield tends to be quite non-normal. We therefore would tend to avoid the use of this approximation. It is provided here largely for comparison with other transformations.

Complementary Log

$$p=e^{-\mu}\text{, with }\mu=e^{\gamma}\text{, so }\frac{\partial\mu}{\partial\gamma}=\mu\text{ and }\frac{\partial p}{\partial\gamma}=(-p\mu)$$

Fisher Score

$$S_\gamma=\frac{\partial l}{\partial\gamma}=\frac{\partial l}{\partial p}\frac{\partial p}{\partial\gamma}=\left[\frac{y-np}{p(1-p)}\right](-p\mu)=\left(\frac{np-y}{1-p}\right)\mu$$

Observed information

$$\begin{aligned}I_\gamma&=\left(-\frac{\partial S_\gamma}{\partial\gamma}\right)=\left(\frac{-\mu}{1-p}\right)\left\{-np\mu+(np-y)\left[\mu-\left(\frac{p\mu}{1-p}\right)\right]\right\}\\&=\left(\frac{\mu}{1-p}\right)\left\{np\mu-(np-y)\left[\mu-\left(\frac{p\mu}{1-p}\right)\right]\right\}\\&=\frac{\mu^2}{(1-p)^2}\{np(1-p)-(np-y)(1-2p)\}\end{aligned}$$

Fisher Information

$$E(I_\gamma) = \left(\frac{np\mu^2}{1-p}\right)$$

Similar to the logit and probit parameterizations, when y = 0 or 1, $\hat{\gamma} = \pm\infty$ with nonzero probability. As with other transformations with this property, the distribution of $\hat{\gamma}$ still admits a normal approximation with variance parameter equal to the reciprocal of the Fisher information:

$$\frac{(1-p)}{np\mu^2}$$

In this case, our empirical studies suggest that the distribution of a log(defect rate) seems to be more normal than either a defect rate or a yield.

Reference

Brown, Cai and DasGupta (2001) Statistical Science, 16: 101-133 and (2002) Annals of Statistics, 30: 160-2001