

Confidence Intervals for the Binomial Distribution

Frank Porter

July 29, 2004

The binomial distribution is:

$$P_N(n; \theta) = \binom{N}{n} \theta^n (1 - \theta)^{N-n}, \quad (1)$$

where N is the number of “trials”, n is the number of “successes”, and θ is the probability of success.

The maximum likelihood estimator for θ , $\hat{\theta}$, given a sampling of n successes, is:

$$\hat{\theta} = \frac{n}{N}. \quad (2)$$

A two-sided interval, (θ_ℓ, θ_u) , for θ at the α confidence level may be constructed according to:

$$(1 - \alpha)/2 = \sum_{k=n}^N P_N(k; \theta_\ell) \quad (3)$$

$$(1 - \alpha)/2 = \sum_{k=0}^n P_N(k; \theta_u). \quad (4)$$

For computational purposes, we may note that the first sum may be written:

$$(1 - \alpha)/2 = 1 - \sum_{k=0}^{n-1} P_N(k; \theta_\ell). \quad (5)$$

Note that if $n = 0$ then we take $\theta_\ell = 0$, and if $n = N$ then we take $\theta_u = 1$. Further it should be noted that the interval obtained by this method does not give exact coverage. This is due to the discrete nature of the binomial distribution (the observed result, n , has finite probability, and appears in the summation limits of both Eqn. 3 and Eqn. 4). The intervals are constructed such that they cover with at least the stated probability (α).