Confidence Intervals

Confidence intervals can be computed using several different approaches. The usual method assumes that the sampling distribution of the estimator is approximately normal (MLEs are asymptotically normal). Thus, the common approach is

$$\hat{\theta} \pm t \cdot \widehat{\text{se}}(\hat{\theta}),$$

where t is the t distribution with the appropriate number of degrees for freedom and confidence level (usually 0.95).

A second approach assumes the sampling distribution is log-normal (long tail to the right), then the lower and upper confidence limits are computed as

$$\hat{\theta}_L = \hat{\theta}/C \qquad \text{and} \qquad \hat{\theta}_U = \hat{\theta} \cdot C,$$

where

$$C = \exp\left(z_{\alpha/2} \sqrt{\log_e\left(1 + [\text{cv}(\hat{\theta})]^2\right)} \right),$$

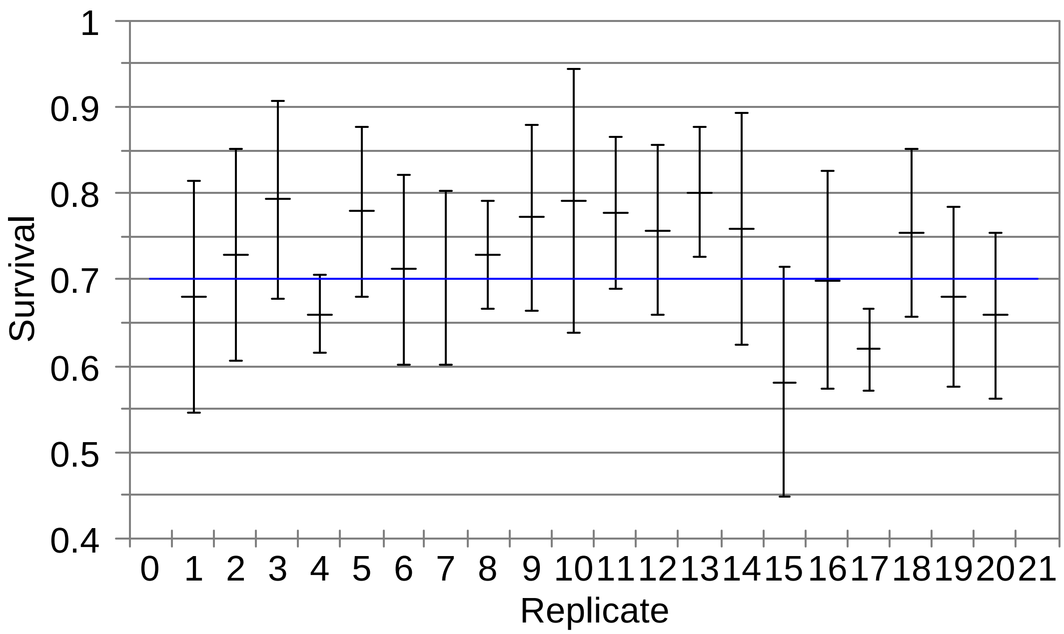
(see Burnham et al. 1987:212). This is the approach used in Program MARK for population estimates (N) and population rate of change (λ).

A third approach is often used in program *MARK* where a parameter (not the data) is transformed onto a new scale. For example, θ is mapped to $\log_e(\theta)$ or $\text{logit}(\theta)$ for analysis and then back-transformed. In these cases, intervals with good properties set lower and upper confidence limits on the transformed parameter and then back-transform the endpoints to the original scale. The idea here is that the estimator of the transformed parameter has a more normally sampling distribution than the estimator of the original parameter. Thus, set the interval endpoint, using the first approach (above) and then back-transform these values to the original parameter.

As an example, suppose $\widehat{\beta}$ the estimate of a survival rate, $\widehat{S}$, via the transformation $\widehat{S} = \exp(\widehat{\beta})/[1 + \exp(\widehat{\beta})]$. Then, $\widehat{\beta} \pm 2 \cdot \widehat{\text{se}}(\hat{\beta})$ provides a 95% confidence interval for β, and $\exp(\widehat{\beta} - 2 \cdot \widehat{\text{se}}(\hat{\beta}))/[1 + \exp(\widehat{\beta} - 2 \cdot \widehat{\text{se}}(\hat{\beta}))]$ to $\exp(\widehat{\beta} + 2 \cdot \widehat{\text{se}}(\hat{\beta}))/[1 + \exp(\widehat{\beta} + 2 \cdot \widehat{\text{se}}(\hat{\beta}))]$ provide a 95% confidence interval for S. This is the method used in program MARK for confidence intervals on parameters in the interval 0-1.

The fourth method, the profile likelihood approach, has advantages, but is sometime difficult to implement when there is more than one parameter involved. For now we will only briefly discuss this method. Consider a plot of the log-likelihood function and identify the exact point where this function is maximized. Now, draw a horizontal line down 1.92 (this corresponds to a 95% interval) units below the maximum point. The values of θ corresponding to the intersection of this horizontal line are taken as the lower and upper confidence limits. We will explore this idea is latter sessions of the course.

Example of coverage of an estimator:



Shape of the Likelihood and Log-likelihood Functions

As sample size increases, $\log_e(\mathcal{L}(p))$ becomes more concentrated around its maximum value (the MLE). The curvature at the peak becomes more pronounced.

In the limit, as sample size goes to infinity, the likelihood becomes a point; it will be located at the true p (if the model assumed in the data analysis is true).

In general as sample size increases, $\log_e(\mathcal{L}(p \mid \text{data}))$ concentrates at the MLE, and the MLE stabilizes at the true parameter.

The curvature at the maximum is inversely related to the sampling variance of the MLEs. As n increases, the $\log_e(\mathcal{L}(p))$ gets more curved, this curvature corresponds to more precision in the MLE (smaller $\widehat{\text{var}}(\hat{p})$).

Heuristically,

$$\widehat{\text{var}}(\hat{p}) \text{ is derived from } \quad \frac{1}{\text{curvature of } \log_e(\mathcal{L}(p)) \text{ at the MLE}}\ .$$

More aspects of likelihood theory:

As sample sizes get large, the MLE's have sampling distributions that are the *Normal distribution*. In the coin flipping example, for large n, $\hat{p}$ is distributed as

$$\text{Normal}(p, \ \frac{p(1-p)}{n})\ .$$

As sample size gets large any statistical bias in MLEs goes away; i.e, MLEs are unbiased in large samples if the model is correct.

These properties justify use of $\hat{p} \pm 1.96\widehat{\text{se}}(\hat{p})$ as a 95% confidence interval in large samples. Likelihood theory exist to provide better confidence intervals for not-so-large samples (likelihood and profile likelihood intervals).

Is there something better than maximum likelihood estimates of parameters (in a frequentist approach)? NO, at least not asymptotically as sample size goes to ∞. Thus for large samples, MLEs are the best you can do.

Likelihood theory also leads to optimal tests of hypotheses (again this is for large samples). These are called likelihood ratio tests (LRTs).

For example, test if the true $p = 0.5$. The LRT is based on the ratio

$$\text{LR} = \frac{\mathcal{L}(0.5)}{\mathcal{L}(\hat{p})}\ ,$$

where $\hat{p}$ is the MLE.

The actual test statistic is $-\mathbf{2log_{e}(LR)} \sim$ **chi-squared variable** with degrees of freedom 1 (here) is the number of estimated parameters in the more general model minus the number of estimated parameters in the simpler model (1-0 = 1 here).

For the actual "data,"

$$\mathrm{LR} = \frac{0.5^{11} \times 0.5^{5}}{0.6875^{11} \times 0.3125^{5}} = 0.3157,$$

$$-2\log_{e}(\mathrm{LR}) = 2.306, \qquad P = 0.1289$$

LRTs fundamentally involve using $\log_{e}(\mathcal{L})$. Moreover, other aspects of likelihood theory provide a deep motivation to work with log-likelihoods (not likelihoods directly).

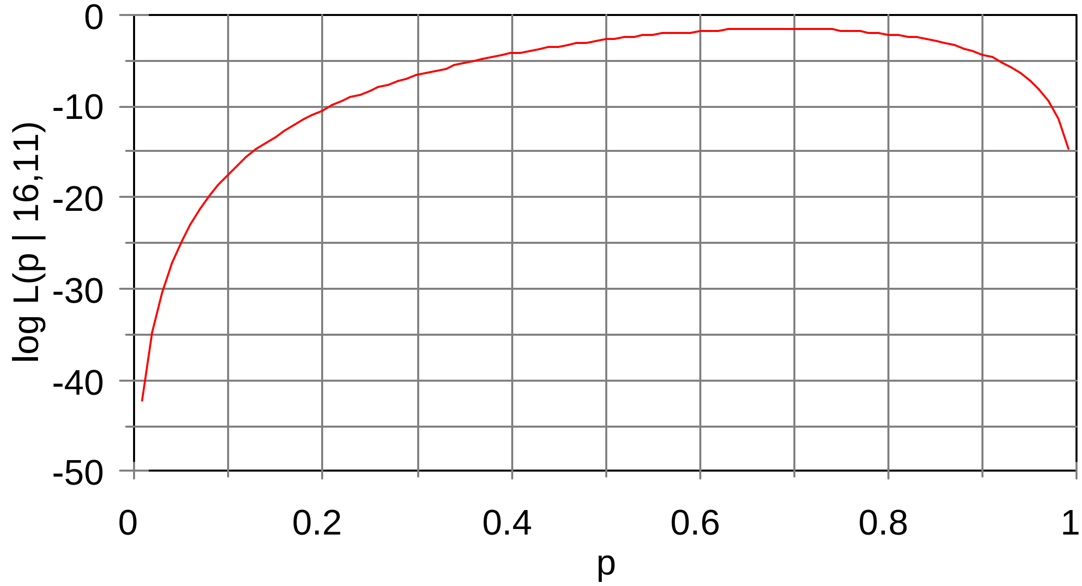
The examples:

$$\log_{e}(\mathcal{L}(p)) = 11\log_{e}(p) + 5\log_{e}(1-p) \quad (n = 16 \text{ case})$$

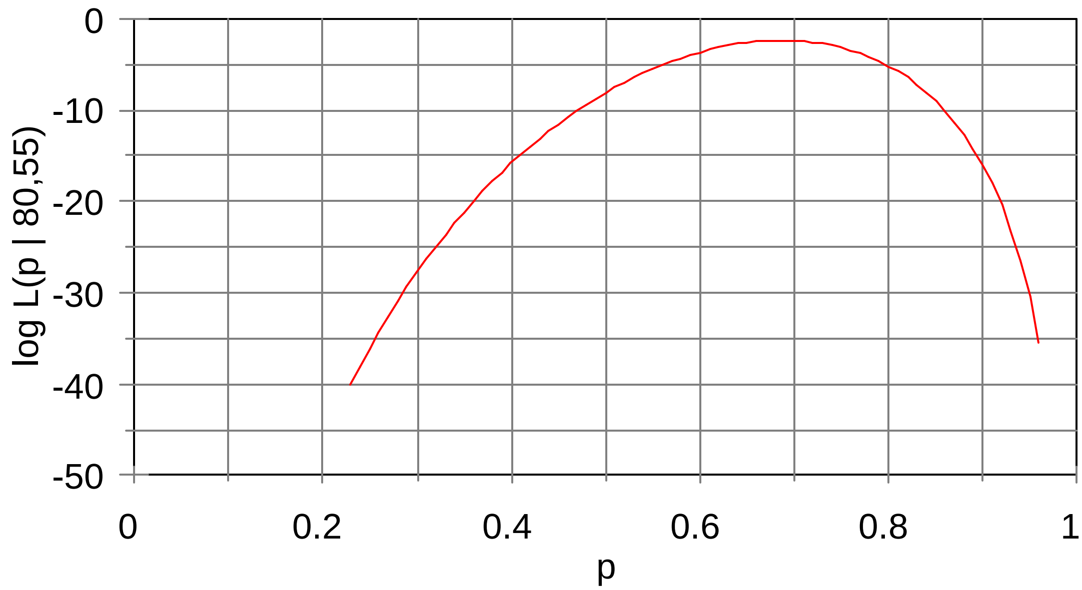
$$\log_{e}(\mathcal{L}(p)) = 55\log_{e}(p) + 25\log_{e}(1-p) \quad (n = 80 \text{ case})$$

These are easily plotted.

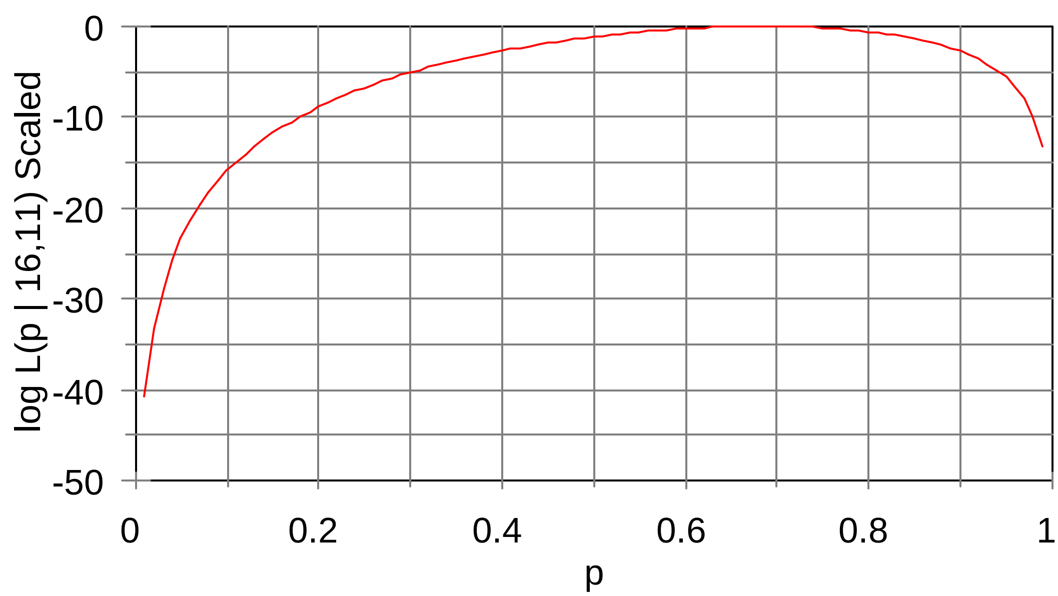
LOG_LIKELIHOOD, n = 16.



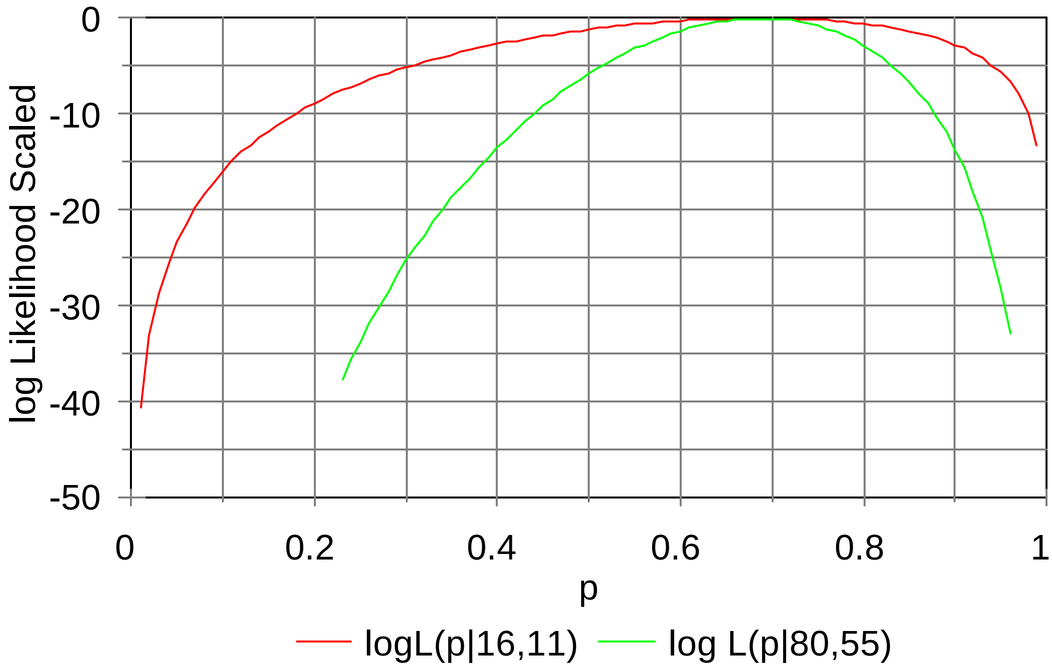
LOG_LIKELIHOOD, n = 80



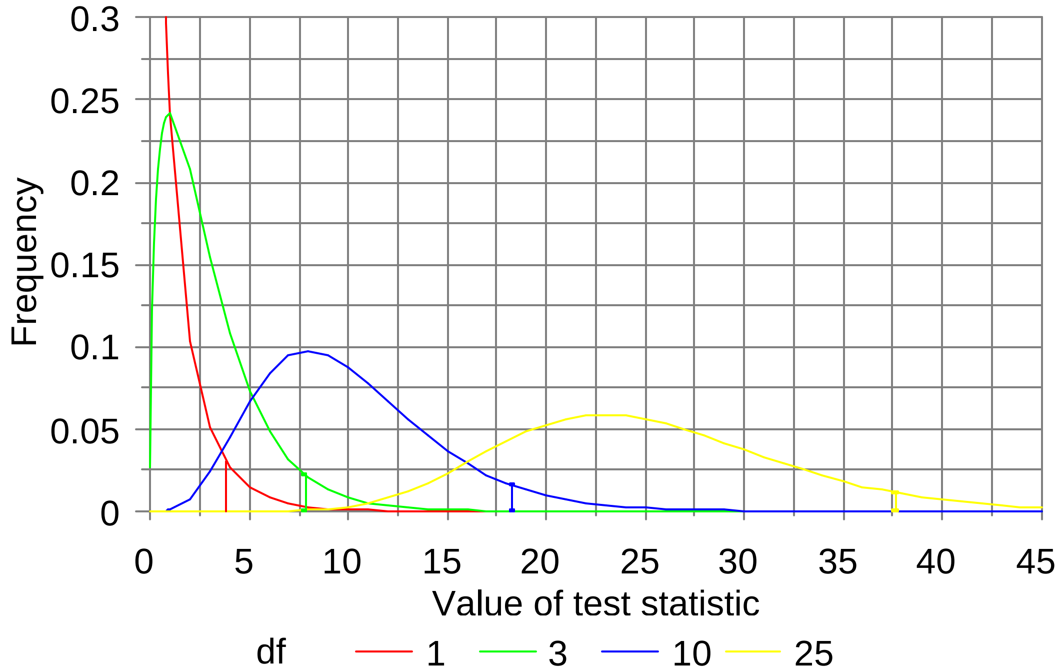
LOG_LIKELIHOOD n = 16 case scaled so that log-likelihood value at the MLE is 0.



LOG_LIKELIHOOD for both coin flipping cases scaled to a maximum of zero and then overlayed (with a reduced range of p for the plot).



Shape of the chi-squared distribution for various degrees of freedom:



The vertical lines designate the critical value of the distribution for $\alpha = 0.05$. That is, the area under the curve to the right of the vertical line is 0.05, whereas the area to the left of the vertical line is 0.95.