Lecture - Confidence Interval

Dr. Arabin Kumar Dey

Assistant Professor
Department of Mathematics
Indian Institute of Technology Guwahati

August 27 - 28, 2013

Outline

Likelihood and Confidence Interval

- The likelihood of a sample is the joint PDF

$$L(\theta) = f(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

- The maximum likelihood estimate (MLE) $\hat{\theta}_{MLE}$ maximizes $L(\theta) : L(\hat{\theta}) \geq L(\theta), \quad \forall \theta$

- MLE is consistent, $\hat{\theta} \to \theta$ in probability, as $n \to \infty$
- MLE is efficient, has small $SE(\hat{\theta})$ as $n \to \infty$
- asymptotically normal, $\frac{\hat{\theta}-\theta}{SE(\hat{\theta})} \approx N(0,1)$

Asymptotic Confidence Interval

- Likelihood forms the basis of many approximate confidence interval.
- It depends on sample size. As n gets larger, we expect the following changes in the log-likelihood function :
- First, $l(p; x)$ is becoming more sharply peaked around $\hat{p}$. Let's assume p is parameter of interest and parameter of bernoulli distribution.
- Second, $l(p; x)$ is becoming more symmetric about $\hat{p}$.
- As sample size grows, likelihood function approaches a quadratic function centered at the MLE. The parabola is significant because that is the shape of the likelihood from the normal distribution.

- Interpretation of CI : If we took many samples, most of our intervals would capture true parameter (e.g. 95% of our intervals will contain the true parameter.)
- From asymptotic normality property of MLE we construct the 95% confidence interval for a parameter θ as

$$\hat{\theta} \pm 1.96 \frac{1}{\sqrt{-l''(\hat{\theta}; x)}} \tag{1}$$

where $l''(\hat{\theta}; x)$ is the second derivative of the log-likelihood function with respect to θ, evaluated at $\theta = \hat{\theta}$.

Observed and expected information

- The quantity $-l''(\hat{\theta}; x)$ is called the "observed information", and $1/\sqrt{-l''(\hat{\theta}; x)}$ is an approximate standard error for $\hat{\theta}$.
- As the likelihood becomes more sharply peaked about the MLE, the second derivative drops and the standard error goes down.
- When calculating asymptotic confidence intervals, statisticians often replace the second derivative of the log-likelihood by its expectation; i.e. replace $-l''(\theta; x)$ by the function

$$I(\theta) = -E(l''(\theta; x)),$$

which is called the expected information or the Fisher information. In that case, the 95% CI would become

$$\hat{\theta} \pm 1.96 \frac{1}{\sqrt{I(\hat{\theta})}}.$$

If X is bernoulli with success probability p, the likelihood is

$$l(p;x) = xlog(p) + (1-x)log(1-p)$$

the first derivative is

$$l^{'}(p;x) = \frac{x-p}{p(1-p)}$$

and the second derivative is

$$l^{''}(p;x) = \frac{-(x-p)^2}{p^2(1-p)^2}$$

$$E((x-p)^2) = V(x) = p(1-p)$$

the Fisher information is

$$I(p) = \frac{1}{p(1-p)}$$

A single bernoulli trial does provide enough information to get a reasonable confidence interval for p. Let's see what happens when we have multiple trials.

If $X \sim Bin(n, p)$, then the log-likelihood is

$$l(p; x) = x log(p) + (n - x) log(1 - p),$$

the second derivative is

$$l''(p; x) = -\frac{x - 2xp + np^2}{p^2(1-p)^2}$$

and the Fisher information is

$$I(p) = \frac{n}{p(1-p)}$$

Thus an approximate 95% confidence interval for p based on the Fisher information is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

where $\hat{p} = \frac{x}{n}$ is the MLE.

What happens if we use the observed information rather than the expected information ? Evaluating the second derivative $l^{''}(p; x)$ at the MLE $\hat{p} = \frac{x}{n}$ gives

$$l^{''}(\hat{p}; x) = -\frac{n}{\hat{p}(1-\hat{p})}.$$

so the 95% interval based on the observed information is identical to above confidence interval i.e.

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

Problems with the above procedures

Suppose $X = 2$ from a bin(20, p). The MLE is $\hat{p} = 2/20 = 0.1$ and log-likelihood is not very symmetric. The usual 95% confidence interval is $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.1 \pm 0.131$. or $(-0.031, 0.231)$, which stays outside the parameter space.

Remedy for the above problem

The "logit" transformation is defined as $\phi = \log(\frac{p}{(1-p)})$.
The logit is called the "log-odds", since $p/(1-p)$ is the odds associated with p.
p lies between 0 and 1, ϕ may take any value from $-\infty$ and ∞ and p produces back-transformation

$$p = \frac{e^{\phi}}{1 + e^{\phi}}.$$

Let's rewrite the binomial log-likelihood in terms of ϕ :

$$\begin{aligned} l(\phi; x) &= xlog(p) + (n-x)log(1-p) \\ &= x\phi + n\log(\frac{1}{(1+e^{\phi})}). \end{aligned}$$

An approximate 95% CI for ϕ is

$$\hat{\phi} \pm 1.96\frac{1}{\sqrt{I(\hat{\phi})}}.$$

where $\hat{\phi}$ is the MLE of ϕ. $I(\phi)$ is the Fisher information for ϕ.

General method for Reparameterization

First we choose a transformation $\phi = \phi(\theta)$ for which we think the log-likelihood will be symmetric.

We calculate $\hat{\theta}$, the MLE for θ, and transform it to the ϕ scale,

$$\hat{\phi} = \phi(\hat{\theta})$$

Next we need to calculate $I(\hat{\phi})$, the Fisher information for ϕ. It turns out that this is given by

$$I(\hat{\phi}) = \frac{I(\hat{\theta})}{[\phi'(\hat{\theta})]^2},$$

where $\phi^{'}(\theta)$ is the derivative of ϕ with respect to θ. Then the endpoints of a 95% confidence interval for ϕ are :

$$\phi_{low} = \hat{\phi} - 1.96 \times \frac{1}{\sqrt{I(\hat{\phi})}},$$

$$\phi_{high} = \hat{\phi} + 1.96 \times \frac{1}{\sqrt{I(\hat{\phi})}}.$$

- The approximate 95% confidence interval for ϕ is $[\phi_{low}, \phi_{high}]$.
- The corresponding confidence interval for θ is obtained by transforming ϕ_{low} and ϕ_{high} back to the original θ scale.

Intervals based on the likelihood ratio

Another way to form a confidence interval for a single parameter is to find all values of θ for which the loglikelihood $l(\theta; x)$ is within a given tolerance of the maximum value $l(\theta; x)$. Statistical theory tells us that, if θ_0 is the true value of the parameter, then the likelihood-ratio statistic

$$2\log\left(\frac{L(\hat{\theta}; x)}{L(\theta_0; x)}\right) = 2[l(\hat{\theta}; x) - l(\theta_0; x)]$$

is approximately distributed as χ_1^2 when the sample size n is large.

Likelihood Ratio Test

In LR test we consider the null hypothesis

$$H_0 : \theta = \theta_0$$

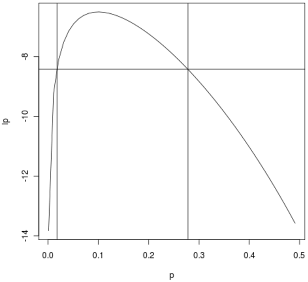
versus the two-sided alternative

$$H_1 : \theta \neq \theta_0$$

we would reject H_0 at the α-level if the LR statistic exceeds the 100(1 - α)-th percentile of the χ^2_1 distribution, i.e. for an $\alpha = 0.05$-level test, we would reject H_0 if the LR statistic is greater than 3.84.

The LR testing principle can also be used to construct confidence intervals. An approximate $100(1-\alpha)\%$ confidence interval for θ consists of all the possible θ_0's for which the null hypothesis $H_0 : \theta = \theta_0$ would not be rejected at the α-level. For a 95% interval, the interval would consist of all the values of θ for which $2[l(\hat{\theta}; x) - l(\theta; x)] \leq 3.84$ or $l(\theta; x) \geq l(\hat{\theta}; x) - 1.92$.
In other words, the 95% interval includes all values of θ for which the loglikelihood function drops off by no more than 1.92 units.

If we consider the previous binomial example, we observe that $X = 2$ from binomial distribution with $n = 20$ and p unknown. MLE is $\hat{p} = 0.1$, and the maximized likelihood is
$l(\hat{p}; x) = 2 \times \log(0.1) + 18 \times \log(0.9) = -6.50.$
Therefore we need the collection of p such that
$l(p; x) \geq -6.5 - 1.92 = -8.42$



Therefore, the LR confidence interval for p is $(0.018, 0.278)$

Interesting Observation

- The above example shows us how to get confidence interval inverting a known test.
- The above method works as long as likelihood function is unimodal.
- If we find the LR interval for a transformed version of the parameter such as $\phi = \log p/(1-p)$ and then transform the endpoints back to the p-scale, we get exactly the same answer as if we apply the LR method directly on the p-scale.
- For that reason, statisticians tend to like the LR method better.