STAT 135 Lab 2
Confidence Intervals, MLE and the Delta Method

Rebecca Barter

February 2, 2015

Confidence Intervals

Confidence intervals

What is a confidence interval?

- A confidence interval is calculated in such a way that the interval contains the true value of θ with some specified probability (**coverage probability**).

What kind of parameters can θ correspond to?

- $\theta = \mu$ from $N(\mu, \sigma^2)$
- $\theta = p$ from Binomial(n, p)

θ typically corresponds to a parameter from a distribution, F, from which we are sampling

$$X_i \sim F(\theta)$$

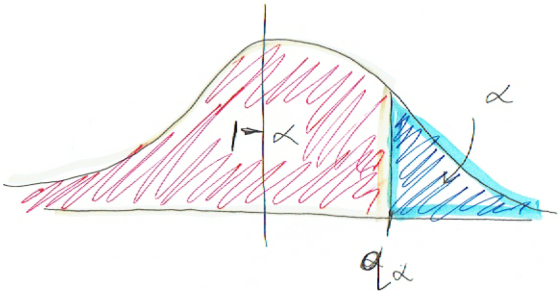
Confidence intervals

- We usually write the coverage probability in the form of $1-\alpha$
- If the coverage probability is 95%, then $\alpha = 0.05$.
- Let q_α be the number such that

$$P(Z < q_\alpha) = 1 - \alpha$$

  where $Z \sim N(0,1)$
- By symmetry of the normal distribution, we have also that

$$q_\alpha = -q_{(1-\alpha)}$$



Confidence intervals

q_α is the number such that

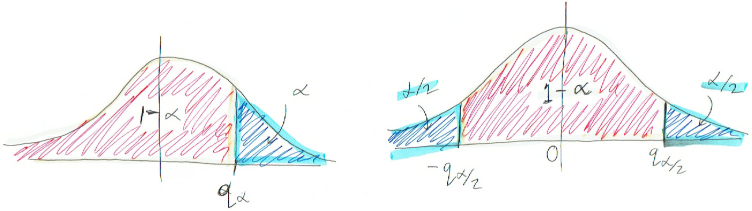
$$P(Z < q_\alpha) = 1 - \alpha$$

where $Z \sim N(0, 1)$.
Note also that by the **symmetry of the normal distribution**

$$1 - \alpha = P(Z < q_\alpha) = P(-q_{\alpha/2} < Z < q_{\alpha/2})$$

For a 95% CI, we have:

$$q_{0.05/2} = 1.96 \quad \text{because} \quad P(-1.96 < Z < 1.96) = 0.95$$



Confidence intervals

Suppose that our estimate, $\hat{\theta}_n$, of θ, asymptotically satisfies

$$\frac{\hat{\theta}_n - \theta}{\sigma_{\hat{\theta}_n}} \sim N(0, 1)$$

So in all of the equations in the previous slides, we can replace Z with $\frac{\hat{\theta}_n - \theta}{\sigma_{\hat{\theta}_n}}$ and rearrange so that θ is the subject.

Confidence intervals

Recall that

$$1 - \alpha = P(-q_{\alpha/2} < Z < q_{\alpha/2})$$

Given that $\frac{\hat{\theta}_n - \theta}{\sigma_{\hat{\theta}_n}} \sim N(0,1)$, we have also the result that

$$1 - \alpha = P\left(-q_{\alpha/2} \quad < \frac{\hat{\theta}_n - \theta}{\sigma_{\hat{\theta}_n}} < \quad q_{\alpha/2}\right)$$

rearranging to make θ the subject, we have

$$1 - \alpha = P\left(\hat{\theta}_n - q_{\alpha/2}\sigma_{\hat{\theta}_n} < \theta < \hat{\theta}_n + q_{\alpha/2}\sigma_{\hat{\theta}_n}\right)$$

Confidence intervals

We have that

$$1 - \alpha = P\left(\hat{\theta}_n - q_{\alpha/2}\sigma_{\hat{\theta}_n} < \theta < \hat{\theta}_n + q_{\alpha/2}\sigma_{\hat{\theta}_n}\right)$$

Recall that if we're looking for a 95% confidence interval (CI), then we are looking for an interval (a, b) such $P(a < \theta < b) = 0.95$.

Thus, the 95% CI for θ can be found from

$$0.95 = P\left(\hat{\theta}_n - q_{0.025}\sigma_{\hat{\theta}_n} < \theta < \hat{\theta}_n + q_{0.025}\sigma_{\hat{\theta}_n}\right)$$

For a general $(1 - \alpha)$% CI, the interval

$$[\hat{\theta}_n - q_{(1-\alpha/2)}\sigma_{\hat{\theta}_n} \ , \ \hat{\theta}_n + q_{(1-\alpha/2)}\sigma_{\hat{\theta}_n}]$$

contains θ with probability $1 - \alpha$.

Exercise 1

Confidence intervals - exercise

CI exercises:

1. In R, generate 1000 random samples, $x_1, x_2, ..., x_{1000}$, from a (continuous) Uniform(5, 15) distribution
2. From the 1000 numbers you have just generated, draw 100 simple random samples (without replacement!), $X_1, ..., X_{100}$. Repeat this 1000 times, so that we have 1000 samples of size 100.
3. For each sample of size 100, compute the sample mean, and produce a histogram (preferably using `ggplot()`) of the 1000 sample means calculated above. What distribution does the sample mean (approximately) follow, and why?
4. For each sample, calculate the 95% confidence interval for the population mean.
5. Of the 1000 confidence intervals, what proportion of them cover the true mean $\mu = \frac{15+5}{2} = 10$?

Maximum likelihood estimation

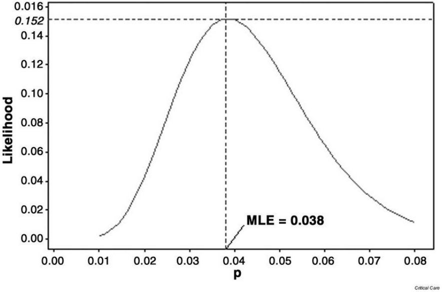
Maximum likelihood estimation

- **Confidence interval for θ**: calculate a **range of values** in which the true value of the parameter θ lies with some specified probability.
- **Maximum likelihood estimator for θ**: calculate a **single value** which estimates the true value of θ by maximizing the likelihood function with respect to θ
  - i.e. find the value of θ that maximizes the likelihood of observing the data given.

Maximum likelihood estimation

What is the likelihood function?

- The likelihood function, $lik(\theta)$, is a function of θ which corresponds to the probability of observing our sample for various value of θ.



How to find the value of θ that maximizes the likelihood function?

Maximum likelihood estimation

Assume that we have observed i.i.d. random variables $X_1, ..., X_n$ and that their distribution has density/frequency function f_θ. Suppose that the observed value of X_i is x_i for each $i = 1, 2, ..., n$

How do we write down the likelihood function? The (non-rigorous) idea:

$$
\begin{aligned}
lik(\theta) &= P(X_1 = x_1, ..., X_n = x_n) \\
&= P(X_1 = x_1)...P(X_n = x_n) \\
&= \prod_{i=1}^{n} f_\theta(X_i)
\end{aligned}
$$

(Note that this proof is not rigorous for continuous variables since they take on specific values with probability 0)

Maximum likelihood estimation

There are 4 main steps in calculating the MLE, $\hat{\theta}_{MLE}$, of θ.

1. Write down the likelihood function, $lik(\theta) = \prod_{i=1}^{n} f_\theta(X_i)$.
2. Calculate the log-likelihood function $\ell(\theta) = \log(lik(\theta))$ (Note: this is because it is often much easier to find the maximum of the log-likelihood function than the likelihood function)
3. Differentiate the log-likelihood function with respect to θ.
4. Set the derivative to 0, and solve for θ.

Maximum likelihood estimation - example

Example: Suppose $X_i \sim Bernoulli(p)$.

$$f_p(x) = p^x(1-p)^{1-x}$$

Step 1: Write down the likelihood function:

$$\begin{aligned} lik(p) &= \prod_{i=1}^{n} f_p(X_i) \\ &= \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} \\ &= p^{\sum_{i=1}^{n} X_i}(1-p)^{\sum_{i=1}^{n}(1-X_i)} \end{aligned}$$

Maximum likelihood estimation - example

Example: Suppose $X_i \sim Bernoulli(p)$.

$$f_p(x) = p^x(1-p)^{1-x}$$

Step 1: $lik(p) = p^{\sum_{i=1}^n X_i}(1-p)^{\sum_{i=1}^n (1-X_i)}$
Step 2: Calculate the log-likelihood function:

$$\ell(p) = \log(lik(p)) = \sum_{i=1}^n X_i \log(p) + \sum_{i=1}^n (1-X_i)\log(1-p)$$

Maximum likelihood estimation - example

Example: Suppose $X_i \sim Bernoulli(p)$.

$$f_p(x) = p^x(1-p)^{1-x}$$

Step 2: $\ell(p) = \sum_{i=1}^n X_i \log(p) + \sum_{i=1}^n (1-X_i)\log(1-p)$
Step 3: Differentiate the log-likelihood function with respect to p**:**

$$\frac{d\ell(p)}{dp} = \frac{\sum_{i=1}^n X_i}{p} - \frac{\sum_{i=1}^n (1-X_i)}{1-p}$$

Maximum likelihood estimation - example

Example: Suppose $X_i \sim Bernoulli(p)$.

$$f_p(x) = p^x(1-p)^{1-x}$$

Step 3: $\frac{d\ell(p)}{dp} = \frac{\sum_{i=1}^n X_i}{p} - \frac{\sum_{i=1}^n (1-X_i)}{1-p}$

Step 4: Set the derivative to 0, and solve for p:

$$\frac{d\ell(p)}{dp} = 0 \implies \hat{p}_{MLE} = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}$$

So the MLE for p where $X_i \sim Bernoulli(p)$ is just equal to the sample mean.

Method of Moments (MOM)

Method of Moments

- **Confidence interval for** θ: calculate a **range of values** in which the true value of the parameter θ lies with some specified probability.
- **Maximum likelihood estimator for** θ: calculate a **single value** which estimates the true value of θ by maximizing the likelihood function with respect to θ.
- **Method of moments estimator for** θ: By equating the theoretical moments to the empirical (sample) moments, derive equations that relate the theoretical moments to θ. The equations are then solved for θ.

Suppose X follows some distribution. The kth **moment of the distribution** is defined to be

$$\mu_k = E[X^k] = g_k(\theta)$$

which will be some function of θ.

Method of Moments

MOM works by equating the theoretical moments (which will be a function of θ) to the empirical moments.

Moment	Theoretical Moment	Empirical Moment
first moment	$E[X]$	$\frac{\sum_{i=1}^{n} X_i}{n}$
second moment	$E[X^2]$	$\frac{\sum_{i=1}^{n} X_i^2}{n}$
third moment	$E[X^3]$	$\frac{\sum_{i=1}^{n} X_i^3}{n}$

Method of Moments

MOM is perhaps best described by example.
Suppose that $X \sim Bernoulli(p)$. Then the first moment is given by

$$E[X] = 0 \times P(X = 0) + 1 \times P(X = 1) = p$$

Moreover, we can estimate the E$[X]$ by taking a sample $X_1, ..., X_n$ and calculating the sample mean :

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

We approximate the first theoretical moment, $E[X]$, by the first empirical moment, $\overline{X}$, i.e.

$$\hat{p}_{MOM} = \overline{X}$$

which is the same as the MLE estimator! (note that this is not always the case...)

Exercise 2

Exercise – Question 43, Chapter 8 (page 320) from John Rice

The file `gamma-arrivals` contains a set of gamma-ray data consisting of the times between arrivals (interarrival times) of 3,935 photons (units are seconds)

1. Make a histogram of the interarrival times. Does it appear that a gamma distribution would be a plausible model?
2. Fit the parameters by the method of moments and by maximum likelihood. How do the estimates compare?
3. Plot the two fitted gamma densities on top of the histogram. Do the fits look reasonable?

Hint 1: the gamma distribution can be written as

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Hint 2: the MLE for α has no closed-form solution - use:

$$\hat{\alpha}_{MLE} = 1$$

The δ-method

The δ-method

Recall that the CLT says

$$\sqrt{n}(\overline{X}_n - \mu) \to N(0, \sigma^2)$$

What if we have some general function $g(\cdot)$?

$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \to ?$$

The δ-method

The δ-method tells us that

$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \to N(0, \sigma^2(g'(\mu))^2)$$

For a proof for the general case, see
http://en.wikipedia.org/wiki/Delta_method

This method can be used to find the variance of a function of our random variables!