Chapter 7

Confidence intervals

7.1 Concepts

We have introduced various estimators for the parameters. However, a point estimator is 100% sure incorrect. In stead of using one value to estimate the parameter, it is better to consider a range of possible values.

A $100(1-\alpha)\%$ confidence interval (CI) is an interval $(L(\boldsymbol{x}), U(\boldsymbol{x})$ that traps the parameter of interest, say θ, with $100(1-\alpha)\%$ "confidence". Thus, for all $\theta \in \Theta$,

$$P(L(\boldsymbol{x}) < \theta < U(\boldsymbol{x})) \geq 1 - \alpha.$$

Note that L and U are functions of data only. They are also random variables.

7.2 Methods to calculate CI

7.2.1 Pivot method

This approach finds a pivot which is a function of the random sample and the parameter whose distribution independent of the parameter of interest. One can get a confidence interval easily from the pivot. Invert the result and one can obtain a desired confidence interval.

Example 7.2.1 (GDP). *From the previous chapters, it is known that we can model the logarithm of the gdp per capita by a normal distribution. Let μ and σ^2 be the mean and variance of the logarithm of the gdp. Construct $1-\alpha$ confidence interval for μ.*

Interpretation of the coverage probability: from a frequentist's point of view, θ is fixed. We repeat the experiment 100 times, and calculate how many of those times do we expect the interval to contain θ. Suppose with $1-\alpha$ of the times, the interval contains θ. Then the CI is called a $1-\alpha$ CI.

Insert graph here.

Example 7.2.2 (GDP)**.** *Construct* $1-\alpha$ *CI for the variance* σ^2*.*

So far, the pivot is derived based on the normality assumption. In reality, the data are not normally distributed. But if the sample size is large, we can make confidence interval of this form even when the data are not normal.

Rationale. CLT. Suppose $Y_1, \cdots, Y_n$ are i.i.d. observations following a distribution with

$$EY_i = \mu, V(Y_i) = \sigma^2,$$

then as $n \to \infty$,

$$\frac{\sqrt{n}(\bar{Y} - \mu)}{\sigma} \to N(0, 1).$$

Example 7.2.3 (Poll)**.** *CNN/ORC Poll. Sept. 6-8, 2013. $N = 1,022$ adults nationwide. "Which of the following is the most important issue facing the country today? The economy. Health care. The situation in Syria. The federal budget deficit. The environment. Gun policy. Immigration."*

	%
The economy	41
Health Care	16
The situation in Syria	15
The federal budget deficit	13
The environment	5
Gun policy	5
Immigration	3
Other	1.

Let p be the true proportion of Americans who think the economy is the most important issue facing the country today. And also, we would like to derive a 95% confidence interval for p.

7.2.2 Variance Stabilization

Now let's revisit the polling example. The variance of the pivot $\hat{p}$ actually involves p. As the estimator changes, the variance will also change. It is easy to see that the variance of the pivot is related to the width of the confidence interval. Therefore, using the pivot method in the polling example might cause the confidence interval not stable enough (sometimes wide, sometimes narrow). How to solve this problem. One intuitive method is to do transformation, so that we construct a new pivot, $g(\hat{p})$, whose variance is irrelevant of p.

Assume that X has the mean μ and $V(X) = V(\mu)$. Let

$$g(x) = \int^x \frac{1}{\sqrt{V(\nu)}} d\nu,$$

where the arbitrary constant of integration can be chosen for convenience. Let $Y = g(X)$. Then the variance of Y is approximately constant.

Example 7.2.4 (Binomial)**.**

Delta Method

Using the variance stabilization transformation, we can make the variance approximately a constant. How to construct the confidence interval? We need to use the Delta method. For a sequence of random variable W_n, if

$$a_n(W_n - b) \xrightarrow{\mathcal{D}} N(0,1),$$

then

$$\frac{a_n(g(W_n) - g(b))}{|g'(b)|} \xrightarrow{\mathcal{D}} N(0,1).$$