Math 502AB - Lecture 28

Dr. Jamshidian

December 6, 2017

1 Lecture - Part 1

1.1 Testing Hypotheses Using *Parametric* Bootstrap

Suppose you have a set of data $X_1, ..., X_n \sim gamma(\alpha, 1)$ and you would like to test the hypothesis

$$\begin{cases} H_0 : & \alpha = 2 \\ H_a : & \alpha \neq 2 \end{cases}$$

To do this, we would need to come up with a statistic to test this hypothesis. We know that α is the *expected value* for the gamma. Therefore, a reasonable rejection region would be

$$\mathcal{R} = \left\{ \overline{X} > c \right\}$$

Under normal circumstances, we would look at $\overline{X}$ and come up with a distribution for it. However, in a real example where you don't know what the distribution is, you would have to use *bootstrapping*.

Bootstrapping is concerned with finding p-values where,

$$p\text{-value} = P(\text{having observed a value}|\ H_0 \text{ is true})$$

In a parametric bootstrap, we assume that we *know* what the distribution of X is. To conduct the bootstrap, we would:

1. Generate $X_1, ..., X_{n_1}$ samples under the assumption that H_0 is true, and compute $\overline{x}_1$.
2. Repeat this b times, and store all of the $\overline{x}_i$'s
3. Calculate the proportion of bootstrapped $\overline{x}_i$ which lie above the *observed*, $\overline{x}_{obs}$

Note: Take note that you must be careful with *bootstrap*. If you arbitrarily increase sample size, you will make the test *too powerful*.

1.2 Testing Hypotheses with *Non-Parametric* Bootstrap

In the *non-parametric* bootstrap situation, all we have is the observed data:

$$\begin{array}{rcccc} & X_1, & X_2, & \ldots, & X_n \\ & \uparrow & \uparrow & & \uparrow \\ \text{Probability:} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{array}$$

In *non-parametric* bootstrapping, essentially the goal is to model generating a population characterized by the observed data.

Suppose we wanted to test μ by using the statistic $\overline{X} \sim N$. Then, we could come up with a confidence interval to test

$$\overline{X} \pm Z^* S.E.\left(\overline{X}\right)$$

where we estimate the "distribution" of $\overline{X}$ by sampling b times, calculating $\overline{x}_i$, and calculating the p-value as

$$p\text{-value} = P(\overline{X} > \overline{x}_{obs} |\ H_0 \text{ is true})$$

1.2.1 *Non-Parametric* Bootstrap with Two Populations

Suppose we have $X_1, ..., X_n$ and $Y_1, ..., Y_n$. Suppose we were interested in testing the hypothesis

$$\begin{cases} H_0: & \mu_X - \mu_Y = 0 \\ H_a: & \mu_X - \mu_Y \neq 0 \end{cases}$$

Recall, that if you had the statistic

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{sp^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

this is a t-distribution with n_1+n_2-2 degrees of freedom, **if** the two populations have the same variance.

If the populations have different variance, then you'd use the statistic

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

This is a great opportunity to use *bootstrapping*. Suppose that

$$\overline{W} = \frac{1}{n_1 + n_2}\left\{\sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i\right\}$$

Then consider

$$X_i^* = X_i - \overline{X} + \overline{W}$$
$$Y_i^* = Y_i - \overline{Y} + \overline{W}$$

Thus, the expected value of both X_i^* and Y_i^* are the same. So our process will be

1. Generate $X_1^*, ..., X_{n_1}^*$ and $Y_1^*, ..., Y_{n_2}^*$, and generate a statistic $\overline{W}$.
2. Repeat this process b times
3. Calculate the p-value, this time based on the statistic $\overline{W}$

2 Lecture - Part 2

2.1 Asymptotic Results

2.1.1 Definition

Let $X_1, ..., X_n$ be a sample from a distribution with *pdf* (or *pmf*) $f(x|\theta)$. Let $T_n = T(X_1, ..., X_n)$ denote a statistic. We say that T_n is a **consistent estimator** of θ if

$$T_n \xrightarrow{P} \theta$$

2.1.2 Theorem

If T_n is a sequence of estimators for a parameter θ that satisfy

1. $$\lim_{n\to\infty} E[T_n] = \theta \leftarrow \text{ asymptotically unbiased}$$
2. $$\lim_{n\to\infty} Var(T_n) = 0$$

then T_n is a **consistent estimator** of θ.

Example

1. Consider $T = \overline{X}$. Then,
$$Var(\overline{X}) = \frac{\sigma^2}{n}$$
Thus, $var \to 0$ as $n \to \infty$.

Proof:

Let $\epsilon > 0$. By the *Chebychev Inequality*, we have

$$P(|T_n - \theta| > \epsilon) \leq \frac{E[T_n - \theta]^2}{\epsilon^2}$$

$$
\begin{aligned}
E[T_n-\theta]^2 &= E[T_n - E(T_n) + E(T_n) - \theta]^2 \\
&= E[T_n - E(T_n)]^2 + E[E(T_n)-\theta]^2 + 2E[(T_n - E(T_n))(E(T_n)-\theta)] \\
&= Var(T_n) + (E[T_n]-\theta)^2 + 2[E(T_n)-\theta](E[T_n]-E[T_n]) \\
E[T_n-\theta]^2 &= Var(T_n) + [E(T_n)-\theta]^2 \\
&= 0+0 \text{ , by assumption 2 and 1, respectively}
\end{aligned}
$$

2.1.3 Theorem

If $\hat{\theta}_n$ is the *MLE* of θ based on a sample of size n, then, under certain regularity conditions,

$$\hat{\theta}_n \xrightarrow{P} \theta_0$$

where θ_0 is the true value of θ.

2.1.4 Definition

A sequence of estimators T_n is *asymptotically efficient* for a parameter θ, if

$$\sqrt{n}(T_n-\theta) \to N(0, V(\theta))$$

Where

$$V(\theta) = \frac{1}{E\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right]^2}$$

In other words, if the variance of the estimator satisfies the *Cramer-Rao Lower Bound*, then the estimator is *asymptotically efficient*.