Lecture Notes 19
36-705

So far we have focused on parametric models. Now we will turn our attention to nonparametric inference. In particular, we will discuss estimating quantities known as *statistical functionals.*

1 Statistical Functional

A statistical functional is a map ψ that maps a distribution P to a real number (or vector). Examples include:

the mean: $\psi(P) = \int x p(x) dx$

the variance $\psi(P) = \int x^2 p(x) dx - \left(\int x p(x) dx\right)^2$

the median $\psi(P) = F^{-1}(1/2)$ where F is the cdf

Sometimes people refer to an unknown statistical functional as a parameter. This should not be confused with the idea of a parameter in a parametric model.

At this point, let me remind you of some notation. If g is any function then we write

$$\int g(x) dP(x) = \begin{cases} \int g(x)p(x)dx & \text{if } X \text{ is continuous} \\ \sum_j g(x_j)p(x_j) & \text{if } X \text{ is discrete.} \end{cases}$$

We also write this as $\int g(x) dF(x)$ where F is the cdf.

2 Plug-In Estimators

Let $X_1, \ldots, X_n \sim P$. Recall that the empirical distribution P_n is the distribution that puts mass $1/n$ at each data point. Thus

$$P_n(A) = \frac{1}{n} \sum_i I(X_i \in A).$$

The corresponding cdf — the empirical cdf — is

$$F_n(t) = \frac{1}{n} \sum_i I(X_i \le t).$$

If g is any function then

$$\int g(x)dP_n(x) = \int g(x)dF_n(x) = \frac{1}{n}\sum_i g(X_i).$$

If $\psi(P)$ is a statistical functional, the *plug-in estimator* is

$$\widehat{\psi}_n = \psi(P_n).$$

For example, if $\psi(P) = \int x dP(x)$ is the mean then the plug-in estimator is

$$\widehat{\psi}_n = \psi(P_n) = \int x dP_n(x) = \frac{1}{n}\sum_i X_i.$$

If

$$\psi(P) = \int (x-\mu)^2 dP(x) = \int x^2 dP(x) - \left(\int x dP(x)\right)^2$$

is the variance then plug-in estimator is

$$\begin{aligned}\widehat{\psi}_n = \psi(P_n) &= \int x^2 dP_n(x) - \left(\int x dP_n(x)\right)^2 = \frac{1}{n}\sum_i X_i^2 - \left(\frac{1}{n}\sum_i X_i\right)^2 \\ &= \frac{1}{n}\sum_i (X_i - \overline{X}_n)^2.\end{aligned}$$

Let's consider a bivariate example. Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n) \sim P$. The covariance is

$$\psi(P) = \mathbb{E}[XY] - \mathbb{E}[X]\ \mathbb{E}[Y] = \int xy dP(x,y) - \int x dP(x) \int y dP(y)$$

and the plug-in estimator is

$$\widehat{\psi}_n = \frac{1}{n}\sum_i X_i Y_i - \overline{X}_n \overline{Y}_n = \sum_i (X_i - \overline{X}_n)(Y_i - \overline{Y}_n).$$

Is plug-in estimation a good idea? It depends. If the functional ψ satisfies some weak regularity conditions and P is well-behaved (for example, has some moments) that $\widehat{\psi}_n$ can be a good estimator. We won't go into details on this here.

The next question is: how do we do inference for a statistical functional? We'll dicuss two approaches: influence functions and the bootstrap.

3 Influence Functions

Let δ_x denote a point mass at x. The influence function for a statistical functional ψ is defined by

$$\varphi(x) = \lim_{\epsilon \to 0} \frac{\psi((1-\epsilon)P + \epsilon\delta_x) - \psi(P)}{\epsilon}.$$

For example, if $\psi(P)$ is the mean of P then

$$\frac{\psi((1-\epsilon)P + \epsilon\delta_x) - \psi(P)}{\epsilon} = \frac{(1-\epsilon)\psi(P) + \epsilon x) - \psi(P)}{\epsilon} = x - \psi(P).$$

Hence $\varphi(x) = x - \psi(P)$.

Let's consider another example. Suppose that $\psi(P)$ is the variance σ^2, that is, $\psi(P) = \int x^2 dP(x) - (\int x dP(x))^2$. Let μ denote the mean. Then

$$\psi((1-\epsilon)P + \epsilon\delta_x) - \psi(P) = (1-\epsilon)\int x^2 dP(x) + \epsilon x^2 - [(1-\epsilon)\mu + \epsilon x]^2 - (\int x^2 dP(X) - \mu^2)$$

and so

$$\varphi(x) = \lim_{\epsilon \to 0} \frac{\psi((1-\epsilon)P + \epsilon\delta_x) - \psi(P)}{\epsilon} = x^2 - \int x^2 dP(x) - 2\mu x.$$

Notice that the influence function φ is itself a statistical functional: it depends on P. For example, if ψ is the mean then $\varphi(x) = x - \psi = x - \psi(P)$. So we can write

$$\varphi(x) = \varphi(x, P).$$

The empirical influence function is an estimate of the influence function obtained by replacing P by P_n, that is,

$$\widehat{\varphi}(x) = \varphi(x, P_n).$$

So, for example, when $\varphi(x) = x - \psi = x - \psi(P)$ for the mean, the empirical influence function is

$$\widehat{\varphi}(x) = x - \psi(P_n) = x - \overline{X}_n.$$

Now we can use the following result.

Theorem 1 *If ψ satisfies some regularity conditions then*

$$\sqrt{n}(\psi(P_n) - \psi(x)) \rightsquigarrow N(0, \tau^2)$$

where

$$\tau^2 = \int \varphi^2(x) dP(x).$$

A consistent estimate of τ^2 *is*

$$\widehat{\tau}^2 = \frac{1}{n}\sum_i \widehat{\varphi}(X_i).$$

Hence, an asymptotic $1-\alpha$ *confidence interval for* $\psi(P)$ *is*

$$\widehat{\psi}_n \pm \frac{z_{\alpha/2}\widehat{\tau}}{\sqrt{n}}.$$

In the case where $\psi(P)$ is the mean we see that

$$\widehat{\tau}^2 = \frac{1}{n}\sum_i \widehat{\varphi}^2(X_i) = \frac{1}{n}\sum_i (X_i - \overline{X}_n)^2 = S_n^2$$

and the confidence interval is

$$\overline{X}_n \pm \frac{z_{\alpha/2}S_n}{\sqrt{n}}.$$

Of course, we did not need all this machinery to arrive at this confidence interval, but in more complicated cases these methods can be very useful. Let's consider another example.

Let $\psi(P)$ be the r^{th} quantile for $0 < r < 1$. Assume that the cdf is strictly increasing so that

$$\psi(P) = F^{-1}(r).$$

The plug-in estimator $\widehat{\psi}_n$ is the r^{th} sample quantile

$$\widehat{\psi}_n = \inf\{x : F_n(x) \geq r\}.$$

The influence function is

$$\varphi(x) = \begin{cases} \frac{r-1}{p(\psi)} & x \leq \psi \\ \frac{r}{p(\psi)} & x > \psi \end{cases}$$

where p is the density function. Hence,

$$\tau^2 = \int \varphi^2(x)dP(x) = \frac{r(1-r)}{p^2(\psi)}.$$

To estimate τ^2 we would need to estimate the density p. We'll discuss how to do that later in the course. However, there are simpler ways to get confidence intervals for quantiles.

There are many subtle technicalities associated with influence functions. These are beyond the scope of the course but if you are interested, search for *semiparametric inference*.