Non-Parametric Inference

September 2, 2025

RECAP: Introduction to Inference

Different types of inference model/methods:

- Nonparametric inference
- Parametric inference: Frequentist inference; Bayesian inference

Different types of inferential problems:

- point estimation
- confidence sets
- hypothesis testing

Plug–in (or Substitution) Principle: a non-parametric estimation method

Let $X_1, ..., X_n \overset{iid}{\sim} F$, where F can be parametric or nonparametric. Assume that we are interested in estimating the quantities that are related to F, such as the mean, median, variance, quantiles, etc, by a nonparametric way. No matter F is parametric or non-parametric, we can write the quantities of interest as a function of F, $\theta(F)$.

The substitution (plug–in) method is to estimate $\theta(F)$ with $\theta(\hat{F}_n)$, where $\hat{F}_n$ is the empirical distribution of F.

Note: We also use F to denote the CDF of the distribution.

The Empirical CDF and Statistical Functionals

Let $X_1, \ldots, X_n \overset{iid}{\sim} F$. The empirical CDF $\hat{F}_n$ puts mass $1/n$ at each datapoint.

$$\begin{aligned}\hat{F}_n(x) &= \frac{\sum_{i=1}^n I(X_i \leq x)}{n} \\ &= \#\{X_i \leq x\}/n\end{aligned}$$

Note that $P_{\hat{F}_n}(X \leq x) = \hat{F}_n(x)$ is often different from $P_F(X \leq x) = F(x)$.

It's also helpful to note that $Y_i = I(X_i \leq x), i = 1, \ldots, n$ are iid Bernoulli r.v.'s, with

$$p = P(Y_i = 1) = P(X_i \leq x) = F(x)$$

Examples of Plug–in Estimators

- $\theta(F) = \mathrm{E}_F(X)$. Then plug–in estimator will be

$$\begin{aligned}
\theta(\hat{F}_n) &= \mathrm{E}_{\hat{F}_n}(X) = \sum_t t P_{\hat{F}_n}(X = t) \\
&= \sum_t t \cdot \frac{\sum_{i=1}^n I(X_i = t)}{n} \\
&= \frac{\sum_{i=1}^n X_i}{n} \\
&= \bar{X}_n
\end{aligned} \tag{1}$$

(This result is independent of the distribution of X.)

- $\theta(F) = var_F(X)$. Then the plug–in estimate will be

$$
\begin{aligned}
\theta(\hat{F}_n) &= var_{\hat{F}_n}(X) = \mathrm{E}_{\hat{F}_n}(X^2) - (\mathrm{E}_{\hat{F}_n}(X))^2 \\
&= \frac{\sum_{i=1}^n X_i^2}{n} - (\frac{\sum_{i=1}^n X_i}{n})^2 \\
&= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}
\end{aligned} \tag{2}
$$

- $\theta(F) = median(X) = \inf_t\{t|F(t) \geq \frac{1}{2}\}$. Then the plug–in estimate for the median is $\theta(\hat{F}_n) = \inf_t\{t|\hat{F}_n(t) \geq \frac{1}{2}\}$.

Note: In general, how plug-in estimator works depending on the properties of $\hat{F}_n$ and also the property of θ function.

Properties of the Empirical CDF

For any fixed x,

$$\begin{aligned} E[\hat{F}_n(x)] &= F(x) \\ V[\hat{F}_n(x)] &= \frac{F(x)[1-F(x)]}{n} \\ MSE[\hat{F}_n(x)] &= V[\hat{F}_n(x)] \to 0 \\ \hat{F}_n(x) &\xrightarrow{P} F(x) \end{aligned}$$

The Glivenko-Cantelli Theorem is even stronger, giving uniform convergence almost surely:
Let $X_1, \ldots X_n \overset{iid}{\sim} F$. Then

$$\sup_x |\hat{F}_n(x) - F(x)| \xrightarrow{as} 0$$

A theorem on Plug–in estimator

Theorem. Suppose the function $\theta(F)$ is continuous in the sup-norm:

$\forall \epsilon > 0, \exists \delta > 0$ such that "$||G - F||_\infty < \delta$ implies $|\theta(G) - \theta(F)| < \epsilon$".

[That is for any ϵ, if there is some G close enough to F, then $\theta(G)$ is close to $\theta(F)$.]

Then,

$$\theta(\hat{F}_n) \xrightarrow{P} \theta(F).$$

Linear function of F

A statistical functional $T(F)$ (or $\theta(F)$) is any function of F. Some examples are the mean $\int x dF(x)$, variance $\int x^2 dF(x) - \left(\int x dF(x)\right)^2$, and p^{th} quantile

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}$$

A linear functional can be written as $T(F) = \int r(x) dF(x)$. The mean is a linear functional, but the variance and quantile function are not.

The plug-in estimator of $T(F)$ is just $T(\hat{F}_n)$. When T is a linear functional,

$$T(\hat{F}_n) = \int r(x) d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} r(X_i)$$

.

Let $X_1, \ldots, X_n \overset{iid}{\sim} F$. Find the plug-in estimators for

- the expected value of X_1
- the expected value of $\exp(X_1)$
- the variance of X_1
- the median of F

Confidence Interval of the Empirical CDF

Dvoretzky-Kiefer-Wolfowitz Inequality: Let $X_1, \ldots, X_n \overset{iid}{\sim} F$. For any $\epsilon > 0$,

$$P\left(\sup_x |F(x) - \hat{F}_n(x)| > \epsilon\right) \leq 2e^{-2n\epsilon^2}$$

It follows that the functions

$$\begin{aligned} L(x) &= \max\{\hat{F}_n(x) - \epsilon_n, 0\} \\ U(x) &= \min\{\hat{F}_n(x) + \epsilon_n, 1\} \\ &\quad \text{for } \epsilon_n = \sqrt{\log(2/\alpha)/(2n)} \end{aligned}$$

form a global $1 - \alpha$ confidence band for F. That is,

$$P\left(L(x) \leq F(x) \leq U(x) \text{ for all } x\right) \geq 1 - \alpha$$

Often we have $T(\hat{F}_n) \approx N(T(F), \widehat{se}^2)$, which allows us to form an approximate $1 - \alpha$ confidence interval for $T(F)$ of

$$T(\hat{F}_n) \pm z_{\alpha/2}\widehat{se}$$

Example: Verify that the `R` expression

```
mean(x) + c(-2, 2) * sd(x)/sqrt(length(x))
```

produces an approximate 95% confidence interval for the mean waiting time for Old Faithful Geyser Data (built-in data in R).