STAT 200B 2019 Week02

Soyeon Ahn

1 Empirical distribution function

1.1 Dirac delta function in statistics

1.1.1 The properties of delta function

Dirac delta function (δ-function) is used to model the density of an idealized point mass or point charge as a function equal to zero everywhere except for zero and whose integral over the entire real line is equal to one.

Properties of the delta function

- $\delta(x) = 0$, if $x \neq 0$, and $\int_{-\infty}^{\infty} \delta(x)dx = 1$
- $x\delta(x) = 0$ for all x
- If $f(x)$ is any function which is continuous in a neighbourhood of the point x_0, then
$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0)dx = f(x_0)$$
, sifting, or sampling, property of the δ-function
- A closely related function to the δ-function is the Heaviside function $H(x)$ which is defined as the unit step function
$$H(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$
The generalized derivative of $H(x)$ is $\delta(x)$, that is
$$\delta(x) = \frac{dH(x)}{dx}$$
It follows that for any fixed x_0,
$$\delta(x - x_0) = \frac{dH(x - x_0)}{dx} = -\frac{dH(x_0 - x)}{dx}$$

1.1.2 The discrete random variable

X is a discrete random variable that assumes the values $a_1, \ldots, a_n$ with probability p_1. The probability mass function can be represented as a function of the form

$$p_X(x) = \sum_{i=1}^{n} p_i \delta(x - a_i)$$

, and the CDF for X can be written as

$$F_X(x) = \sum_{i=1}^{n} p_i H(x - a_i)$$

1.2 Empirical density function

If we could define an empirical density function by

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i)$$

, since we have

$$F(u) = \int_{-\infty}^{u} f(t)dt = \int_{\mathbb{R}} 1(t \leq u) f(t) dt$$

and

$$\int_{\mathbb{R}} 1(t \leq u) \hat{f}_n(x) dt = \hat{F}_n(u)$$

It does not seem to be a probability density since it takes an infinite value, although we could treat it as a random variable having zero variance. To produce continuous estimates of the density, the estimate of the density is given by a convolution of empirical density function and a kernel.

1.3 Kernel density function

Since a convolution of f, g is

$$f(x) * g(x) = \int_{\mathbb{R}} f(u) g(u - x) du$$

and a kernel is a function K such that

- $K(x) \geq 0$
- $\int K(x)dx = 1$
- $\int xK(x)dx = 0$
- $\int x^2 K(x) dx \equiv \sigma_K^2 > 0$

The kernel density estimator of f is a convolution of $\hat{f}_n(x)$ and $\frac{1}{h}K(\frac{u}{h})$,

$$\hat{f}_{n,kde}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$$

1.4 The plug-in estimates of statistical functionals

The empirical CDF $\hat{F}_n$ puts mass $1/n$ at each datapoint.

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

This can be represented as

$$\hat{F}_n = \frac{1}{n}\sum_{i=1}^{n} H(x - x_i)$$

when $x_1, \ldots, x_n$ be a sample of I.I.D. random variables $X_1, \ldots, Xn$, respectively.

The plug-in estimator of $T(F)$ is $T(\hat{F}_n)$.
For example, the mean

$$\theta = E(X) = \int x dF(x) = \begin{cases} \int x f(x) dx & \text{if } x \text{ is a continuous r.v.} \\ \sum_k x_k p(x_k) & \text{if } x \text{ is a discrete r.v.} \end{cases}$$

Since $\hat{F}_n$ is a discrete probability distribution with the mass $\frac{1}{n}$ at every sample point,

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

or, equivalently

$$\hat{\theta}_n = \int x\, d\hat{F}_n = \frac{1}{n}\sum_{i=1}^{n}\int x\, dH(x - x_i).$$

The derivative of the Heaviside function is the Dirac delta function $\delta(x)$,

$$\int x\, dH(x - x_i) = \int x\delta(x - x_i)\, dx = x_i.$$

Therefore, the plug-in estimate of the mean as the sample mean.

$$\hat{\theta}_n = T(\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

When T is a linear functional,

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

.

2 Properties of EDF

2.1 Mean and Variance of $\hat{F}_n$

Please remind that

$$\hat{F}_n \quad = \quad \frac{\sum_{i=1}^{n} I(X_i \leq x)}{n}$$

We can present EDF as

$$\hat{F}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$$

where $Y_i = I(X_i \leq x)$. In other words, EDF is the average of $I(X_i < x)$. Since $Y_i = I(X_i \leq x)$ is a function of a random variable X_i, we can calculate an expectation of and a variance Y_i .

For Y_i, the associated outcome is whether we observe $X_i \leq x$ or not.

$$Y_i = \begin{cases} 1 & \text{if } X_i \leq x \\ 0 & \text{if } X_i < x \end{cases}$$

Therefore

$$Y_i \sim Bernoulli(F(x))$$

since $P(Y_i = 1) = P(X_i \leq x) = F(x)$. This leads to us that

$$\begin{aligned} E(I(X_i \leq x)) &= E(Y_i) = F(x) \\ Var(I(X_i \leq x)) &= Var(Y_i) = F(x)(1 - (F(x))) \end{aligned}$$

for a given x.

The expectation and variance of $\hat{F}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$

$$\begin{aligned} E(\hat{F}_n) &= E(Y_i) = F(x) \\ Var(\hat{F}_n) &= \frac{Var(Y_i)}{n} = \frac{F(x)(1 - (F(x)))}{n} \end{aligned}$$

2.2 $\hat{F}_n$ a consistent estimator F

By Chebyshev's inequality,

$$P(|\hat{F}_n - F(x)| \geq \epsilon) \leq \frac{F(x)(1 - (F(x)))}{n\epsilon^2}$$

for any $\epsilon > 0$. This proves $\hat{F}_n(x)$ converges in probability to $F(x)$ as $n \to \infty$.