Lecture in FYS4411, March 8, 2024

FYS 4411/9411, MARCH 8, 2024

Standard Bootstrap assumes that the stochastic variables x_i are i.i.d. ← distributed,

independent identically

The estimators are typically the mean value (sample mean) and the variance

$$\mathbb{E}[x] = \int_{x \in D} p(x)\, x \, dx$$

PDF

$p(x)dx$

or $\sum_{i \in D} x_i \, p(x_i)$

in our case we don't have $p(x)$

sample mean

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \neq \mathbb{E}[x]$$

when n is not infinitely large.

$$\int_{x \in D} p(x)\, dx = 1$$

$$var[x] = \int_{x \in D} p(x)\left(x - \mathbb{E}[x]\right)^2 dx$$

$$\sum_{i \in D} p(x_i)\left(x_i - \mathbb{E}[x]\right)^2$$

sample variance

$$\sigma^2 = \frac{1}{n} \sum_{i \in D} (x_i - \mu)^2$$

$$\left(p(x_i) = \frac{1}{n} \right)$$

Bootstrap

original sample $X = \{x_1, x_2, \ldots, x_n\}$
we assume we can produce
new samples X^* and
produce new estimates for
μ and σ^2

$$\Theta_n^{*(1)}, \Theta_n^{*(2)}, \ldots, \Theta_n^{*(B)}$$

$$\Theta_n^{*(i)} = \frac{1}{n} \sum_{j=1}^{n} x_j^{*(i)}$$

Example

$X = \{1, 2, 3, 4, 5\}$

$X^* = \{2, 3, 3, 5, 1\}$ shuffle randomly with replacement

Do B-times

$$\overline{\Theta}^*_B = \frac{1}{B} \sum_{i=1}^{B} \Theta_n^{*(i)}$$

$$\mathrm{var}\left[\overline{\Theta}^*_B\right] = \frac{1}{B-1} \sum_{i=1}^{B} \left(\Theta_n^{*(i)} - \overline{\Theta}^*_B\right)^2$$

Central limit theorem,

m-experiments, each of these have

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_j^{(i)}$$

iid

simplify notation

$\bar{X}_i \Rightarrow X_i$ — same PDF

Total mean value

$$z = \frac{X_1 + X_2 + \dots + X_m}{m}$$

what is the distribution of z? $\qquad P(z) = ?$

$$P(z) = \int_{x_1 \in D} dx_1 \, \underline{P}(x_1) \int_{x_2 \in D} dx_2 \, \underline{P}(x_2) \cdots \int_{x_m \in D} dx_m \, P(x_m)$$

$$\times \, \delta\left(z - \frac{(x_1 + x_2 + \cdots + x_m)}{m}\right)$$

$$\delta\left(z - \frac{(x_1 + x_2 + \dots x_m)}{m}\right)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} dq \exp\left[iq\left(z - \overbrace{\frac{(x_1 + x_2 + \dots + x_m)}{m}}^{\mu}\right)\right]$$

in $p(z)$ insert

$$1 = e^{i\mu q - i\mu q}$$

$$\mu = \int_{x \in D} p(x)\, x\, dx$$

$$p(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dq \, e^{iq(z-\mu)}$$

$$\times \left[\int_{x \in D} dx \, p(x) \exp\left[\frac{iq(\mu - x)}{m} \right] \right]^m$$

$$\int_{x \in D} p(x) \, dx = \underline{1}$$

$$\int_{x \in D} dx \, p(x) \, e^{iq(\mu - x)/m} =$$

$$= \int_{x \in D} p(x)\,dx \left[1 + \frac{iq(\mu - x)}{m} - \frac{q^2(\mu - x)^2}{2m^2} + \cdots \right]$$

$$\left(\mu = \int_{x \in D} x\, p(x)\,dx \qquad \int_{x \in D} p(x)\,dx = 1 \right)$$

$$= \left[1 + 0 - \frac{q^2 \sigma^2}{2m^2} + \cdots \right]$$

in the limit $m \to \infty$

$$P(z) = \frac{1}{\sqrt{2\pi\,\sigma^2/m}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2/m}\right]$$

Gaussian with mean value μ and variance

$$\sigma^2/m \to STD = \sigma/\sqrt{m}$$

$$\sigma^2 = \int_{x \in D} p(x)(x-\mu)^2 dx$$