Sampling methods → From Slides

Monte Carlo
Bootstrap and Jacknife
Cross - Validation.

MC

we assume a model & then run a number of simulations

↳ Tell us average behaviour

Bootstrapping

→ Looks at a dataset

Monte Carlo Integration.

$$\theta = \int_a^b g(x)\,dx$$

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$$

Then if

His Notes

Integration ⟶ Simulation

we are interested in studying the statistical properties of some statistic $T(X)$

~~E.g~~ Given a sample of N values sampled at random from a population, we are interested in evaluating the "properties" or performance of $T(x)$

$$T(x) = \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$ (Sample mean)

$$T(x) = \left(\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \right)^{1/2}$$ (std dev)

$$T(x) = \hat{\beta} = (X^T X)^{-1} X^T y$$

A MC Sim consists of implementing a large number of repititions of the analysis

Ex

$$T(x) = \bar{x}$$

$m=1$: Generate a sample $\{x_1, \dots, x_N\}_1$ Calc $\bar{x}$ for the sample.

$\Rightarrow \bar{x}_1$

$m=2$: Generate another sample $\{x_1, \dots, x_N\}_2$

$\Rightarrow \bar{x}_2$

⋮

$m=M \longrightarrow \bar{x}_M$

$\Rightarrow$ Analyse statistical properties of $\{\bar{x}_1, \dots, \bar{x}_M\}$ in particular

Theoretic $E(\bar{x}) \simeq \hat{E}(\bar{x}) = \frac{1}{M}\sum_{m=1}^{M} \bar{x}_m$

(MC mean of sample means)

Theoretic $var(\bar{x}) \simeq \hat{var}(\bar{x}) = \frac{1}{M-1}\sum_{m=1}^{M}\left(\bar{x}_m - \hat{E}(\bar{x})\right)^2$

MC variance of the sample means

Many other characteristics may be considered

N.B

In theory, we know that when x_i are iid with $E(x) = \mu$, $var(x) = \sigma^2$ then

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{N}\right) \qquad \text{(CLT)}$$

$\Rightarrow$ The MC estimate should be s.t

$$\hat{E}(\bar{x}) \simeq \mu$$
$$\hat{var}(\bar{x}) \simeq \frac{\sigma^2}{N}$$

N.B

The std dev $\frac{\sigma}{\sqrt{N}}$ of $\bar{x}$ is called the std error of the estimator $\bar{x}$

Evaluating the std dev.

$$\hat{SD}(\bar{x}) = \left(\hat{var}(x)\right)^{1/2}$$
$$= \left[\frac{1}{M-1}\sum_{m=1}^{M}\left(x_m - \hat{E}(\bar{x})\right)^2\right]^{1/2}$$

gives us an estimate of the standard error.

$$var(x)^{1/2} = \frac{\sigma}{\sqrt{N}}$$

Likewise if we analyse a regression problem e.g $Y = \beta X + \varepsilon$ and consider some estimator $\hat{\beta}$ of β, then we can evaluate the statistical properties of $\hat{\beta}$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (LSE)$$

$$\Rightarrow \begin{cases} \hat{E}(\hat{\beta}) = \frac{1}{M} \sum_{m=1}^{M} \hat{\beta}_m & \text{Bias}(\hat{\beta}) \\ var(\hat{\beta}) = \frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{\beta}_m - E(\bar{\beta})\right)^2 & \rightarrow \text{Precision}(\hat{\beta}) \end{cases}$$

This is useful for benchmarking