The Sample Mean Estimator

Dan Saunders

Definition

Suppose \( x_i \sim i.i.d. (\mu, \sigma^2) \) is a set of independent and identically distributed data, with mean \( \mu \) and variance \( \sigma^2 \). Then the sample mean estimator is defined as:

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\]

Since the sample mean is a function of each observation, \( \bar{x}(x_1, x_2, ..., x_n) \), it is also a random variable, whose distribution specifically depends upon the underlying distribution of each \( x_i \) in the following way:

\[
\bar{x} \sim \left( \mu, \frac{\sigma^2}{n} \right)
\]

Finite Sample Properties

Unbiased:

\[
E(\bar{x}) = E \left( \frac{1}{n} \sum_{i=1}^{n} x_i \right) = \frac{1}{n} E \left( \sum_{i=1}^{n} x_i \right) = \frac{1}{n} \sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\mu) = \mu.
\]

Variance:

\[
Var(\bar{x}) = Var \left( \frac{1}{n} \sum_{i=1}^{n} x_i \right) = \left( \frac{1}{n} \right)^2 Var \left( \sum_{i=1}^{n} x_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.
\]

Gaussian (Normality): If each observation is an i.i.d. sample from the same Normal distribution,

\[
x_i \sim N(\mu, \sigma^2) \Rightarrow z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0, 1)
\]

If we replace the variance, \( \sigma^2 \), with our sample variance estimator, \( s^2 \), then we have Student’s distribution (i.e. t-distribution).

\[
t = \frac{z}{\sqrt{\frac{s^2}{\sigma^2}}} = \frac{\sqrt{n}(\bar{x} - \mu)}{s} \sim t_{n-1} \text{ where } s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \sim \chi^2_{n-1}.
\]

Large Sample Properties

For any i.i.d. distribution of \( x_i \), the Law of Large Numbers says:

\[
\bar{x} \xrightarrow{p} \mu
\]

For any i.i.d. distribution of \( x_i \), the Central Limit Theorem says:

\[
\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)
\]