

CEM1002

2015-10-05

“Parameter”, “Statistic”, “Estimator”

Parameter

- ▶ A parameter is a *constant* numerical property of a distribution (in our model, c.f. Bayesian statistics.)
- ▶ Conventional notation: small letter, often Greek, generically θ .

Statistic

- ▶ Recall: a *sample* is an i.i.d. sequence of random variables X_1, \dots, X_n
- ▶ A “statistic” is *any function of a sample*, e.g. “sample mean/average”: $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
- ▶ Statistics are *random variables*.

Estimator

- ▶ An estimator for a parameter, θ , is a statistic whose formula does not contain θ .

Mean, Sample Mean, Observed Sample Mean

- ▶ The “mean” is a distribution parameter.

$$E(X) = \int xf(x) dx$$

- ▶ Often denoted μ (Bad idea?)
- ▶ In practice it is *unknown*

Sample mean

- ▶ \bar{X} is a random variable.
- ▶ Often the best estimator for μ .

Observed sample mean

- ▶ \bar{x} is a number calculated from a variable in an actual dataset.
- ▶ Can be thought of as a realization of the random variable \bar{X} .
- ▶ Is used as the actual estimate of μ for a given variable.

Variance, Sample Variance, Observed Sample Variance

- ▶ The “variance” is a distribution parameter.

$$\text{Var}(X) = \int (x - E(X)) f(x) dx$$

- ▶ Often denoted σ^2
- ▶ In practice is *unknown*

Sample Variance

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

* Is a *random variable*. * Often used to estimate σ^2

Observed sample variance

- ▶ s^2 is a number calculated from a variable in an actual dataset.
- ▶ etc., etc.