5.1 Basic Concepts of Random Samples

Definition 5.1.1

The random variables $X_1, \ldots, X_n$ are called a random sample of size n from the population $f(x)$ if $X_1, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$. Alternatively, $X_1, \ldots, X_n$ are called *independent and identically distributed* (iid) random variables with pdf or pmf $f(x)$. This is commonly abbreviated to iid random variables.

If the population pdf or pmf is a member of a parametric family with pdf or pmf given by $f(x|\theta)$, then the joint pdf or pmf is

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta),$$

where the same parameter value θ is used in each of the terms in the product.

Example Let $X_1, \ldots, X_n$ be a random sample from an exponential(β) population. Specifically, $X_1, \ldots, X_n$ might correspond to the times (measured in years) until failure for n identical circuit boards that are put on test and used until they fail. The joint pdf of the sample is

$$f(x_1, \ldots, X_n|\beta) = \prod_{i=1}^{n} f(x_i|\beta) = \frac{1}{\beta^n} e^{-\sum_{i=1}^{n} x_i/\beta}.$$

This pdf can be used to answer questions about the sample. For example, what is the probability that all the boards last more than 2 years?

$$P(X_1 > 2, \ldots, X_n > 2) = P(X_1 > 2) \cdots P(X_n > 2)$$
$$= [P(X_1 > 2)]^n = (e^{-2/\beta})^n = e^{-2n/\beta}.$$

Random sampling models

(a) Sampling from an infinite population. The samples are iid.

(b) Sampling with replacement from a finite population. The samples are iid.

(c) Sampling without replacement from a finite population. This sampling is sometimes called simple random sampling. The samples are not iid exactly. However, if the population size N is large compared to the sample size n, the samples will be approximately iid.

<u>Example 5.1.3 (Finite population model)</u>

Suppose $\{1, \ldots, 1000\}$ is the finite population, so $N = 1000$. A sample of size $n = 10$ is drawn without replacement. What is the probability that all ten sample values are greater than 200? If $X_1, \ldots, X_{10}$ were mutually independent we would have

$$P(X_1 > 200, \ldots, X_{10} > 200) = (\frac{800}{1000})^{10} = .107374.$$

Without the independent assumption, we can calculate as follows.

$$P(X_1 > 200, \ldots, X_{10} > 200) = \frac{\binom{800}{10}\binom{200}{0}}{\binom{1000}{10}} = .106164.$$

Thus, the independence assumption is approximately correct.