2.4. The Mean and Variance of a Sample

1. The word "sample" that will be defined in the next item is more abstract and is different from what we mean by the word in Chapter 1. Usually the meaning of the word will be clear from context.

2. *Definition.* (sample)

   Let X be a random variable with pmf $f(x)$.

   If $X_1, X_2, \cdots, X_n$ are random variables such that

   (a) $X_1, X_2, \cdots, X_n$ are mutually independent;

   (b) Each X_i has the pmf $f(x)$,

   then we say $X_1, X_2, \cdots, X_n$ is a random sample of size n from the population $f(x)$, or a random sample of size n from the population X.

3. *Definition.* If $X_1, X_2, \cdots, X_n$ is a random sample from a population, then we define the sample mean as follows:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

We define sample variance as follows:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right).$$

The sample standard deviation is:

$$S = \sqrt{S^2}.$$

4. Note that the sample mean $\bar{X}$, the sample variance S^2, and the sample standard deviation S are random variables, because they are functions of the random variables $X_1, X_2, \cdots, X_n$.

5. In probability and statistics, an observed value of a random variable X is the value that is actually observed after the experiment is performed. Conventionally, to avoid confusion, upper case letters denote random variables; the corresponding lower case letters denote their realizations.

   For example, consider the experiment of tossing a die, and let X be the outcome of the die. Before the experiment, X is a random variable with a distribution. After the experiment is done, we observe a number on the die, say $x = 2$, which is called an observed value. The observed value differs from experiment to experiment.

6. *Notation.* We use $x_1, x_2, \cdots, x_n$ to denote the observed values of $X_1, X_2, \cdots, X_n$, respectively.

   We denote by $\bar{x}$ the observed value of $\bar{X}$.

   We denote by s^2 the observed value of S^2.

   We denote by s the observed value of S.

7. It is clear that

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

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).$$

$$s = \sqrt{s^2}.$$

8. *Example.* Let the experiment consist of tossing seven dice in a row.

   Let X_i be the outcome of the i-th die, $i = 1, 2, 3, 4, 5, 6, 7$.

   Then, $X_1, X_2, \cdots, X_7$ form a random sample of size 7 from the population X, where $X \sim$ discrete uniform distribution over the numbers $1, 2, 3, 4, 5, 6$.

   Suppose $x_1 = 1$, $x_2 = 5$, $x_3 = 5$, $x_4 = 3$, $x_5 = 4$, $x_6 = 2$, $x_7 = 3$ are the observed values of $X_1, X_2, \cdots, X_n$. Find $\bar{x}$ and s^2.

   — *Solution.* We have

$$\bar{x} = \frac{1}{7}(1 + 5 + 5 + 3 + 4 + 2 + 3) = \frac{23}{7},$$

$$s^2 = \frac{1}{6} \sum_{i=1}^{7} (x_i - \bar{x})^2 = \frac{47}{21}.$$