Unit 4: Probability

Probability is a subject that deals with uncertainty. In everyday terminology, probability can be thought of as a numerical measure of the likelihood that a particular event will occur. Probability values are assigned on a scale from 0 to 1, with values near 0 indicating that an event is unlikely to occur and those near 1 indicating that an event is likely to take place. A probability of 0.50 means that an event is equally likely to occur as not to occur.

Events and their probabilities

Oftentimes probabilities need to be computed for related events. For instance, advertisements are developed for the purpose of increasing sales of a product. If seeing the advertisement increases the probability of a person buying the product, the events “seeing the advertisement” and “buying the product” are said to be dependent. If two events are independent, the occurrence of one event does not affect the probability of the other event taking place. When two or more events are independent, the probability of their joint occurrence is the product of their individual probabilities. Two events are said to be mutually exclusive if the occurrence of one event means that the other event cannot occur; in this case, when one event takes place, the probability of the other event occurring is zero.

Random variables and probability distributions

A random variable is a numerical description of the outcome of a statistical experiment. A random variable that may assume only a finite number or an infinite sequence of values is said to be discrete; one that may assume any value in some interval on the real number line is said to be continuous. For instance, a random variable representing the number of automobiles sold at a particular dealership on one day would be discrete, while a random variable representing the weight of a person in kilograms (or pounds) would be continuous.

The probability distribution for a random variable describes how the probabilities are distributed over the values of the random variable. For a discrete random variable, x, the probability distribution is defined by a probability mass function, denoted by $f(x)$. This function provides the probability for each value of the random variable. In the development of the probability function for a discrete random variable, two conditions must be satisfied: (1) $f(x)$ must be nonnegative for each value of the random variable, and (2) the sum of the probabilities for each value of the random variable must equal one.

A continuous random variable may assume any value in an interval on the real number line or in a collection of intervals. Since there is an infinite number of values in any interval, it is not meaningful to talk about the probability that the random variable will take on a specific value; instead, the probability that a continuous random variable will lie within a given interval is considered.

In the continuous case, the counterpart of the probability mass function is the probability density function, also denoted by $f(x)$. For a continuous random variable, the probability density function provides the height or value of the function at any particular value of x; it does not directly give the probability of the random variable taking on a specific value. However, the area under the graph of $f(x)$ corresponding to some interval, obtained by computing the integral of $f(x)$ over that interval, provides the probability that the variable will take on a value within that interval. A probability density function must satisfy two requirements: (1) $f(x)$ must be nonnegative for each value of the random variable, and (2) the integral over all values of the random variable must equal one.

The expected value, or mean, of a random variable—denoted by $E(x)$ or μ—is a weighted average of the values the random variable may assume. In the discrete case the weights are given by the probability mass function, and in the continuous case the weights are given by the probability density function. The formulas for computing the expected values of discrete and continuous random variables are given by equations 2 and 3, respectively.

$$E(x) = \Sigma x f(x) \qquad (2)$$

$$E(x) = \int x f(x) dx \qquad (3)$$

The variance of a random variable, denoted by Var(x) or σ^2, is a weighted average of the squared deviations from the mean. In the discrete case the weights are given by the probability mass function, and in the continuous case the weights are given by the probability density function. The formulas for computing the variances of discrete and continuous random variables are given by equations 4 and 5, respectively. The standard deviation, denoted σ, is the positive square root of the variance. Since the standard deviation is measured in the same units as the random variable and the variance is measured in squared units, the standard deviation is often the preferred measure.

$$\mathrm{Var}(x) = \sigma^2 = \Sigma(x - \mu)^2 f(x) \qquad (4)$$

$$\mathrm{Var}(x) = \sigma^2 = \int (x - \mu)^2 f(x) dx \qquad (5)$$

Comprehension Exercises

Choose a, b, c or d which best completes each item.

1) If an event is equally likely to occur as not to occur, then the probability of its occurrence is _______.

a) 0 b) 1 c) 0.5 d) 0.25

2) A random variable representing the number of earthquakes occurred in a geographical area during the past year is a ______ random variable.

a) independent b) dependent c) continuous d) discrete

3) The area under the graph of a probability density function corresponding to some interval provides the ______ that the random variable will take on a value within that interval.

a) probability b) integral c) area d) expected value

4) For computing the expected value of a discrete random variable, the weights are given by the probability ______ function.

a) meant b) variance c) density d) mass

Words to Learn

Find the Persian equivalents of the following terms and expressions.

probability	likelihood	event	occur
uncertainty	likely	mutually exclusive	occurrence
dependent	unlikely	nonnegative	equally
independent	equally likely	real number line	finite
random variable	experiment	condition	infinite
discrete	outcome	requirement	respectively
continuous	integral	satisfy	assign
mass function	joint occurrence	affect	directly
density function	sequence	squared deviations	counterpart
expected value	variance	weighted average	meaningful