

Name : Joseph Janone Tiwouw

Informatika

“Random Variable”

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.

$$E(x) = \sum xf(x) \quad (2)$$

$$E(x) = \int xf(x)dx \quad (3)$$

The variance of a random variable, denoted by $\text{Var}(x)$ 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.

$$\text{Var}(x) = \sigma^2 = \sum (x - \mu)^2 f(x) \quad (4)$$

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

Special probability distributions

The binomial distribution

Two of the most widely used discrete probability distributions are the binomial and Poisson. The binomial probability mass function (equation 6) provides the probability that x successes will occur in n trials of a binomial experiment.

Equation.

A binomial experiment has four properties: (1) it consists of a sequence of n identical trials; (2) two outcomes, success or failure, are possible on each trial; (3) the probability of success on any trial, denoted p , does not change from trial to trial; and (4) the trials are independent. For instance, suppose that it is known that 10 percent of the owners of two-year old automobiles have had problems with their automobile's electrical system. To compute the probability of finding exactly 2 owners that have had electrical system problems out of a group of 10 owners, the binomial probability mass function can be used by setting $n = 10$, $x = 2$, and $p = 0.1$ in equation 6; for this case, the probability is 0.1937.

The Poisson distribution

The Poisson probability distribution is often used as a model of the number of arrivals at a facility within a given period of time. For instance, a random variable might be defined as the number of telephone calls coming into an airline reservation system during a period of 15 minutes. If the mean number of arrivals during a 15-minute interval is known, the Poisson probability mass function given by equation 7 can be used to compute the probability of x arrivals.

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \quad (7)$$

For example, suppose that the mean number of calls arriving in a 15-minute period is 10. To compute the probability that 5 calls come in within the next 15 minutes, $\mu = 10$ and $x = 5$ are substituted in equation 7, giving a probability of 0.0378.

The normal distribution

The most widely used continuous probability distribution in statistics is the normal probability distribution. The graph corresponding to a normal probability density function with a mean of $\mu = 50$ and a standard deviation of $\sigma = 5$ is shown in Figure 3. Like all normal distribution graphs, it is a bell-shaped curve. Probabilities for the normal probability distribution can be computed using statistical tables for the standard normal probability distribution, which is a normal probability distribution with a mean of zero and a standard deviation of one. A simple mathematical formula is used to convert any value from a normal probability distribution with mean μ and a standard deviation σ into a corresponding value for a standard normal distribution. The tables for the standard normal distribution are then used to compute the appropriate probabilities.

There are many other discrete and continuous probability distributions. Other widely used discrete distributions include the geometric, the hypergeometric, and the negative binomial; other commonly used continuous distributions include the uniform, exponential, gamma, chi-square, beta, t , and F .