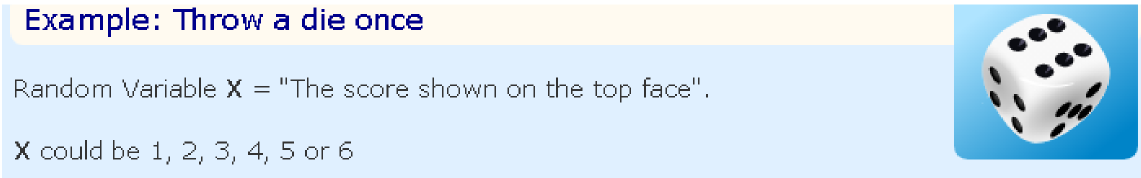
Lecturer 3 (Random variables)

Israt Jahan
M.S. (Hasselt University, Belgium), M.S. (SUST), B.Sc (SUST)

- **Random variable:** A random variable is obtained by assigning a numerical value to each outcome of a particular experiment.



  Random Variables can be either Discrete or Continuous:

  Discrete random variable can only take certain values (such as 1,2,3,4,5)
  Continuous random variable can take any value within a range (ex: A company manufactures metal cylinders that are used in the construction of a particular type of engine. The company discovers that the cylinders it manufactures can have a diameter anywhere between 49.5 and 50.5 mm. Suppose that the random variable X is the diameter of a randomly chosen cylinder manufactured by the company. Since this random variable can take any value between 49.5 and 50.5, it is a continuous random variable.)

- **Probability mass function:** The probability mass function (p.m.f.) of a random variable X is a set of probability values p_i assigned to each of the values x_i taken by the discrete random variable. These probability values must satisfy $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. The probability that the random variable takes the value x_i is said to be p_i , and this is written $P(X = x_i) = p_i$

- **Cumulative distribution function (cdf):** Cdf is the probability that X will take a value less than or equal to x. i.e $F(x) = P(X \leq x)$

  Example: A manager supervises the operation of three power plants, plant X, plant Y, and plant Z. At any given time, each of the three plants can be classified as either generating electricity (1) or being idle (0). With the notation (0, 1, 0) used to represent the situation where plant Y is generating electricity but plants X and Z are both idle.

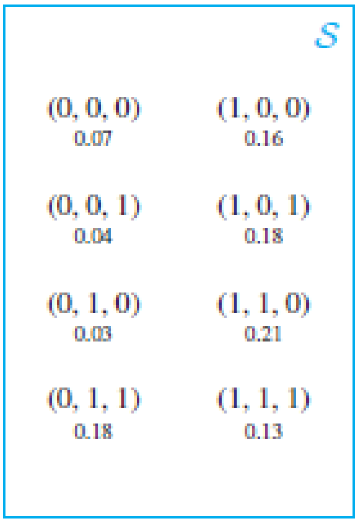


FIGURE 1.15

Probability values for power plant example

The probability that no plants are generating electricity (X=0) is 0.07. The probability that exactly one plant is generating electricity ($X = 1$) is the sum of the probabilities of the outcomes (1, 0, 0), (0, 1, 0), and (0, 0, 1), which is $0.04 + 0.03 + 0.16 = 0.23$.

x_i	0	1	2	3
p_i	0.07	0.23	0.57	0.13

FIGURE 2.7

Tabular presentation of the probability mass function for power plant example

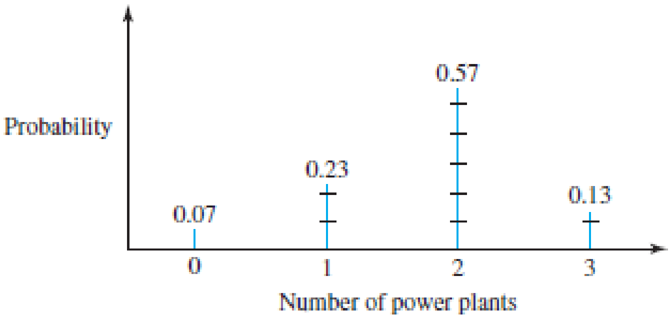
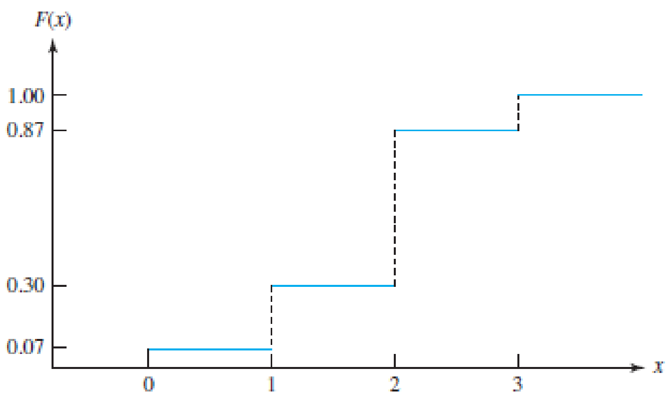


FIGURE 2.8

Line graph of the probability mass function for power plant example

FIGURE 2.13

Cumulative distribution function for power plant example



The probability that no more than one plant is generating electricity is simply $F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = 0.07 + 0.23 = 0.30$

Homework

2.1.1

- **Probability density function:**

Probability Density Function

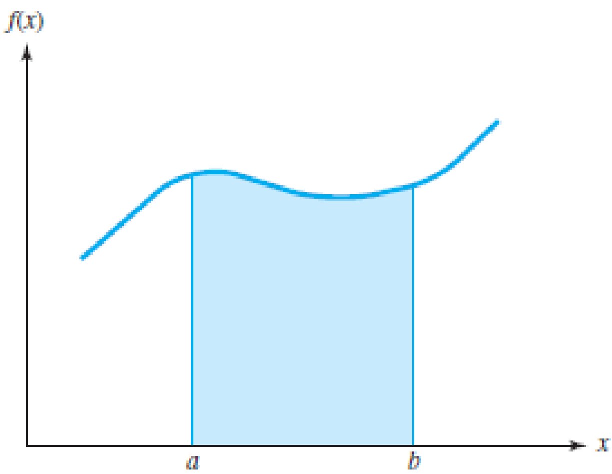
A **probability density function** $f(x)$ defines the probabilistic properties of a *continuous* random variable. It must satisfy $f(x) \geq 0$ and

$$\int_{\text{state space}} f(x)\, dx = 1$$

The probability that the random variable lies between two values is obtained by integrating the probability density function between the two values.

FIGURE 2.20

$P(a \leq x \leq b)$ is the area under the probability density function $f(x)$ between the points a and b



Example 14

Metal Cylinder Production

Suppose that the diameter of a metal cylinder has a probability density function

$$f(x) = 1.5 - 6(x - 50.0)^2$$

for $49.5 \leq x \leq 50.5$ and $f(x) = 0$ elsewhere, as shown in Figure 2.21. This is a valid probability density function because it is positive within the state space [49.5, 50.5] and because

$$\begin{aligned}
\int_{49.5}^{50.5} (1.5 - 6(x - 50.0)^2)\, dx &= [1.5x - 2(x - 50.0)^3]_{49.5}^{50.5} \\
&= [1.5 \times 50.5 - 2(50.5 - 50.0)^3] \\
&\quad - [1.5 \times 49.5 - 2(49.5 - 50.0)^3] \\
&= 75.5 - 74.5 = 1.0
\end{aligned}$$

The probability that a metal cylinder has a diameter between 49.8 and 50.1 mm can be calculated to be

$$\begin{aligned}
\int_{49.8}^{50.1} (1.5 - 6(x - 50.0)^2)\, dx &= [1.5x - 2(x - 50.0)^3]_{49.8}^{50.1} \\
&= [1.5 \times 50.1 - 2(50.1 - 50.0)^3] \\
&\quad - [1.5 \times 49.8 - 2(49.8 - 50.0)^3] \\
&= 75.148 - 74.716 = 0.432
\end{aligned}$$

about 43% of the cylinders will have diameter have diameters within these limits.

- **Cumulative distribution function**The cumulative distribution function of a continuous random variable X is defined in exactly the same way as for a discrete random variable, namely $F(x) = P(X \leq x)$

Example 14

Metal Cylinder Production

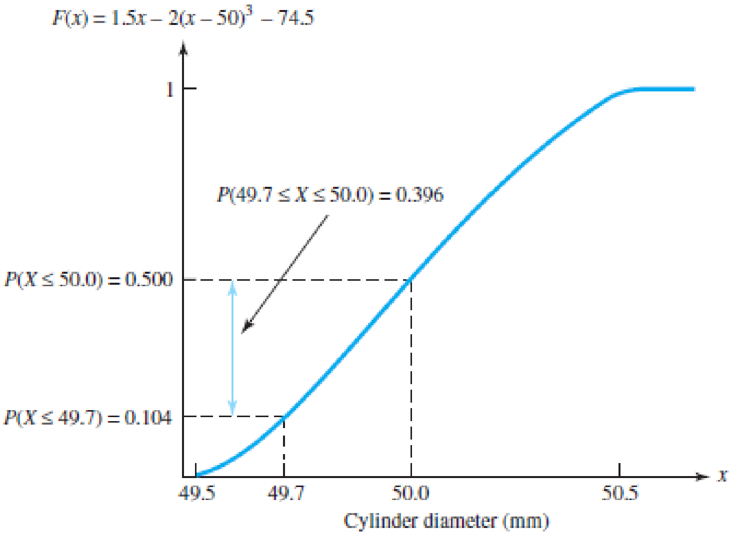
The cumulative distribution function of the metal cylinder diameters can be constructed from the probability density function as

$$\begin{aligned} F(x) = P(X \le x) &= \int_{49.5}^{x} (1.5 - 6(y - 50.0)^2)\, dy \\ &= [1.5y - 2(y - 50.0)^3]_{49.5}^{x} \\ &= [1.5x - 2(x - 50.0)^3] - [1.5 \times 49.5 - 2(49.5 - 50.0)^3] \\ &= 1.5x - 2(x - 50.0)^3 - 74.5 \end{aligned}$$

for $49.5 \le x \le 50.5$. As expected, the cumulative distribution function is an increasing function between the limits $x = 49.5$ and $x = 50.5$, as illustrated in Figure 2.27, with $F(49.5) = 0$ and $F(50.5) = 1$. Technically, in addition $F(x) = 0$ for $x < 49.5$ and $F(x) = 1$ for $x > 50.5$.

FIGURE 2.27

Cumulative distribution function for metal cylinder diameters illustrating $P(49.7 \le X \le 50.0)$



Homework

2.2.2, 2.2.6

- **Expectation of a Random Variable:** Although the probability mass function or the probability density function provides complete information about the probabilistic properties of a random variable, it is often useful to employ some summary measures of these properties. One of the most basic summary measures is the expectation or mean of a random variable, which is denoted by E(X) and represents an 'average' value of the random variable.

Expected Value of a Discrete Random Variable

The **expected value** or **expectation** of a discrete random variable with a probability mass function $P(X = x_i) = p_i$ is

$$E(X) = \sum_i p_i x_i$$

$E(X)$ provides a summary measure of the average value taken by the random variable and is also known as the **mean** of the random variable.

Example 4
Power Plant Operation

The expected number of power plants generating electricity is

$$E(X) = (0 \times 0.07) + (1 \times 0.23) + (2 \times 0.57) + (3 \times 0.13) = 1.76$$

which is illustrated in Figure 2.34. This expected value, 1.76, provides a summary measure of the average number of power plants generating electricity at particular points in time.

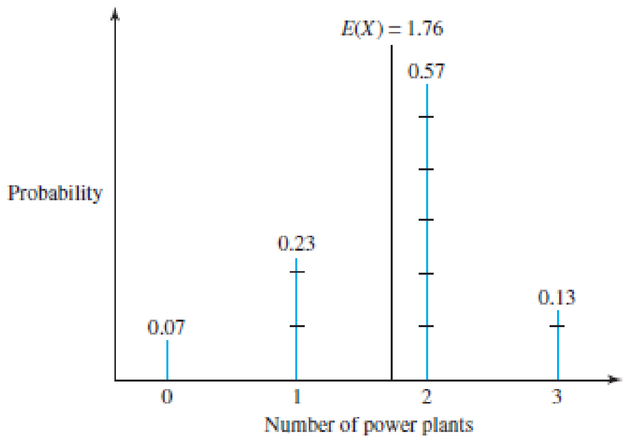


FIGURE 2.34

Expected value for power plant example

Expected Value of a Continuous Random Variable

The **expected value** or **expectation** of a continuous random variable with a probability density function $f(x)$ is

$$E(X) = \int_{\text{state space}} x f(x)\, dx$$

The expected value provides a summary measure of the average value taken by the random variable, and it is also known as the **mean** of the random variable.

Example 14

Metal Cylinder Production

The expected diameter of a metal cylinder is

$$E(X) = \int_{49.5}^{50.5} x(1.5 - 6(x - 50.0)^2)\,dx$$

The evaluation of this integral can be simplified using the transformation $y = x - 50.0$, so that

$$\begin{aligned} E(X) &= \int_{-0.5}^{0.5} (y + 50.0)(1.5 - 6y^2)\,dy \\ &= \int_{-0.5}^{0.5} (-6y^3 - 300y^2 + 1.5y + 75)\,dy \\ &= [-3y^4/2 - 100y^3 + 0.75y^2 + 75y]_{-0.5}^{0.5} \\ &= [25.09375] - [-24.90625] = 50.0 \end{aligned}$$

Consequently, the metal cylinders have an average diameter of 50.0 mm.

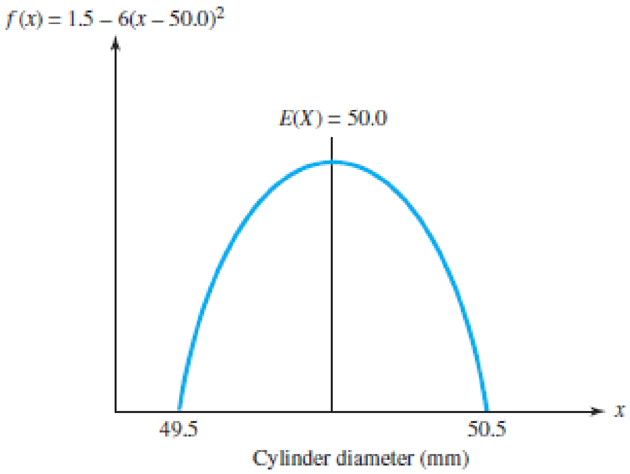


FIGURE 2.37

Expected value for metal cylinder diameters

Median

The **median** of a continuous random variable X with a cumulative distribution function $F(x)$ is the value x in the state space for which

$$F(x) = 0.5$$

The random variable is then equally likely to fall above or below the median value.

The median value of the metal cylinder diameters is the solution to

$$F(x) = 1.5x - 2(x - 50.0)^3 - 74.5 = 0.5$$

Homework

2.3.1, 2.3.12

Variance

The **variance** of a random variable X is defined to be

$$\text{Var}(X) = E((X - E(X))^2)$$

or equivalently

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

The variance is a positive quantity that measures the spread of the distribution of the random variable about its mean value. Larger values of the variance indicate that the distribution is more spread out.

$$\begin{aligned}
\text{Var}(X) &= E((X - E(X))^2) \\
&= E(X^2 - 2XE(X) + (E(X))^2) \\
&= E(X^2) - 2E(X)E(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}$$

Standard Deviation

The **standard deviation** of a random variable X is defined to be the positive square root of the variance. The symbol σ^2 is often used to denote the variance of a random variable, so that σ represents the standard deviation.

The expected squared breaking strength is

$$
\begin{aligned}
E(X^2) &= \int x^2 f(x)dx = \int_{120}^{150} x^2 0.128e^{-x/100}dx \\
&= [-12.8(20{,}000 + 200x + x^2)e^{-x/100}]_{120}^{150} \\
&= [-207{,}064.79] - [-225{,}148.70] = 18{,}083.91
\end{aligned}
$$

Since the mean breaking strength is $E(X) = 134.1$, the variance of the breaking strengths is

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 18{,}083.91 - 134.1^2 = 101.10$$

The standard deviation of the breaking strengths is thus $\sigma = \sqrt{101.10} = 10.05$, which is illustrated in Figure 2.49.

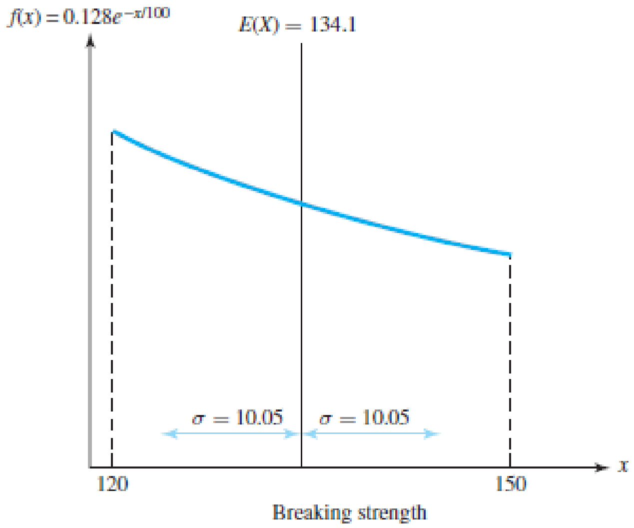


FIGURE 2.49

Mean and standard deviation of concrete breaking strengths

Quantiles

The pth **quantile** of a random variable X with a cumulative distribution function $F(x)$ is defined to be the value x for which

$$F(x) = p$$

Quartiles and Interquartile Range

The **upper quartile** of a distribution is defined to be the 75th percentile of the distribution, and the **lower quartile** of a distribution is defined to be the 25th percentile. The **interquartile range** is the distance between the two quartiles and like the variance provides an indication of the spread of the distribution.

The cumulative distribution function of the metal cylinder diameters is

$$F(x) = 1.5x - 2(x - 50.0)^3 - 74.5$$

for $49.5 \le x \le 50.5$. The *upper quartile* of the distribution is the value of x for which

$$F(x) = 0.75$$

which is 50.17 mm. The *lower quartile* satisfies

$$F(x) = 0.25$$

and is 49.83 mm. The interquartile range is therefore $50.17 - 49.83 = 0.34$ mm, and half of the cylinders will have diameters between 49.83 mm and 50.17 mm, as illustrated in Figure 2.56.

Homework

2.4.5