(Discrete & Continuous probability distribution)

- **Bernoulli distribution:** A Bernoulli random variable is the simplest kind of random variable. It can take on two values, 1 and 0. It takes on a 1 if an experiment with probability p resulted in success and a 0 otherwise. Ex.: an item is defective or not. a valve is open or shut.

If X is a Bernoulli random variable, denoted $X \sim \text{Ber}(p)$:

Probability mass function:	$P(X = 1) = p$ $P(X = 0) = (1 - p)$
Expectation:	$E[X] = p$
Variance:	$\text{Var}(X) = p(1 - p)$

- **Binomial distribution:** A binomial random variable is random variable that represents the number of successes in n successive independent trials of a Bernoulli experiment.

The Binomial Distribution

Consider an experiment consisting of

- n Bernoulli trials
- that are independent and
- that each have a constant probability p of success.

Then the total number of successes X is a random variable that has a **binomial** distribution with parameters n and p, which is written

$$X \sim B(n, p)$$

The probability mass function of a $B(n, p)$ random variable is

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for $x = 0, 1, \ldots, n$, with

$$E(X) = np \qquad \text{and} \qquad \text{Var}(X) = np(1 - p)$$

- **Example:** Suppose a milk factory has 20 containers and there is a probability of 0.261 that a milk container is underweight. a) What is the distribution of the number of underweight containers in a box? b) Calculate expected number of underweight cartons in a box and also calculate its variance. c) Calculate the probability that a box contains exactly seven underweight containers. d) Calculate the probability that a box contains no more than three underweight containers.

Solution:

If the milk contents of two different milk containers are independent of each other, the number of underweight containers in a box has a binomial distribution with parameters $n = 20$ and $p = 0.261$. This is because each individual milk container represents a Bernoulli trial with a constant probability $p = 0.261$ of being underweight, as illustrated in Figure 3.5. The expected number of underweight cartons in a box is

$$E(X) = np = 20 \times 0.261 = 5.22$$

and the variance is

$$\mathrm{Var}(X) = np(1-p) = 20 \times 0.261 \times 0.739 = 3.86$$

so that the standard deviation is $\sigma = \sqrt{3.86} = 1.96$.

The probability mass function of the number of underweight containers is shown in Figure 3.6. The probability that a box contains exactly *seven* underweight containers, for example, is

$$P(X = 7) = \binom{20}{7} \times 0.261^7 \times 0.739^{13} = \frac{20!}{7!\,13!} \times 0.261^7 \times 0.739^{13} = 0.125$$

The probability that a box contains *no more than three* underweight containers is

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{20}{0} \times 0.261^0 \times 0.739^{20} + \binom{20}{1} \times 0.261^1 \times 0.739^{19} \\ &\quad + \binom{20}{2} \times 0.261^2 \times 0.739^{18} + \binom{20}{3} \times 0.261^3 \times 0.739^{17} \\ &= 0.0024 + 0.0167 + 0.0559 + 0.1185 = 0.1935 \end{aligned}$$

- **Homework**
  An archer hits a bull's-eye with a probability of 0.09, and the results of different attempts can be taken to be independent of each other. If the archer shoots nine arrows, calculate the probability that:
  (a) Exactly two arrows score bull's-eyes. (b) At least two arrows score bull's-eyes. (c) What is the expected number of bull's-eyes scored? (d) What is the variance and standard deviation of bull's-eyes scored?

- **Geometric distribution:** Consider a sequence of independent Bernoulli trials with a constant success probability p. Whereas the binomial distribution is the distribution of the number of successes occurring in a fixed number of trials n, it is sometimes of interest to count instead the number of trials performed until the first success occurs. Such a random variable is said to have a geometric distribution with parameter p.

- **Negative binomial distribution:** Consider a sequence of independent Bernoulli trials with a constant success probability p. Whereas the binomial distribution is the distribution of the number of successes occurring in a fixed number of trials n, it is sometimes of interest to count instead the number of trials performed until the rth success occurs. Such a random variable is said to have a negative binomial distribution distribution with parameter p.

- **Poisson distribution:** It is often useful to define a random variable that counts the number of 'events' that occur within certain specified boundaries. For example, the number of telephone calls received by an operator within a certain time limit. The Poisson distribution is often appropriate to model such situations.

  The Poisson Distribution

  A random variable X distributed as a **Poisson** random variable with parameter λ, which is written

  $$X \sim P(\lambda)$$

  has a probability mass function

  $$P(X = x) = \frac{e^{-\lambda}\,\lambda^x}{x!}$$

  Where x=0,1,2,...

- **Difference between binomial distribution and Poisson distribution:**
  1) In binomial distribution, number of trials are fixed. In poisson distribution, number of trials are infinite.
  2) In binomial distribution, Variance $\leq$ Mean. In Poisson distribution, Mean $=$ Variance.
  3) Ex of binomial distribution: Coin tossing experiment. Ex of Poisson distribution: Printing mistakes/page of a large book.

- **Example:**
  Suppose that the number of errors in a piece of software has a parameter $\lambda = 3$. This parameter immediately implies that the expected number of errors is three and that the variance in the number of errors is also equal to three.
  a) What is distribution of the number of errors in a piece of software.
  b) Calculate the probability that a piece of software has no errors.
  c) Calculate the probability that there are three or more errors in a piece of software.
  Solution: The number of errors in a piece of software has a poisson distribution.
  The probability that a piece of software has no errors is

  $$P(X = 0) = \frac{e^{-3} \times 3^0}{0!} = e^{-3} = 0.050$$

  The probability that there are three or more errors in a piece of software is

  $$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \frac{e^{-3} \times 3^0}{0!} - \frac{e^{-3} \times 3^1}{1!} - \frac{e^{-3} \times 3^2}{2!} \\ &= 1 - e^{-3}\left(\frac{1}{1} + \frac{3}{1} + \frac{9}{2}\right) = 1 - 0.423 = 0.577 \end{aligned}$$

Homework

The number of cracks in a ceramic tile has a mean of $\lambda = 2.4$. a) What is the distribution of the number of cracks in a ceramic tile. b) What is the probability that a tile has no cracks? c) What is the probability that a tile has four or more cracks?

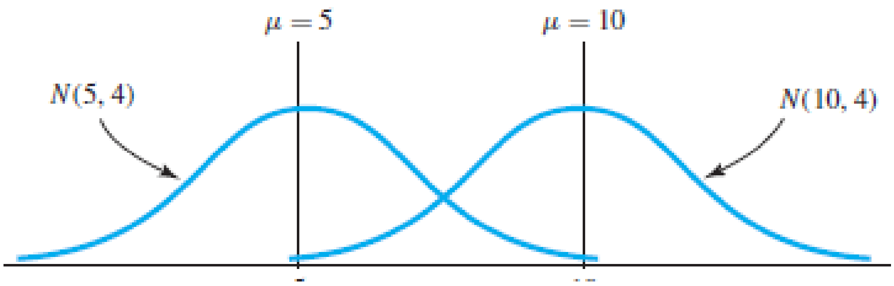
- **Normal distribution:**

The **normal** or **Gaussian distribution** has a probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}$$

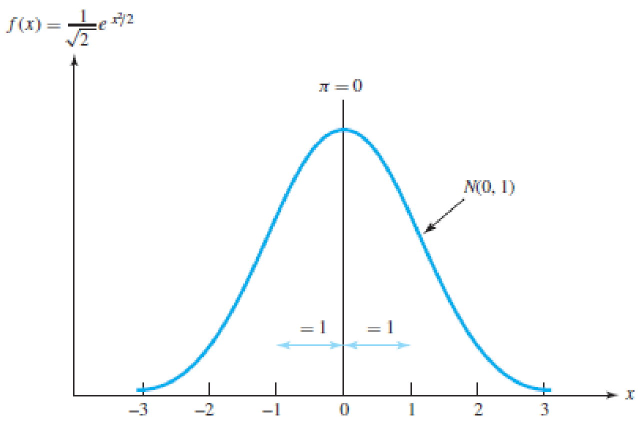
for $-\infty \leq x \leq \infty$, depending upon two parameters, the mean and the variance

$$E(X) = \mu \qquad \text{and} \qquad \mathrm{Var}(X) = \sigma^2$$



A normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is known as the **standard normal distribution**. Its probability density function has the notation $\phi(x)$ and is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$



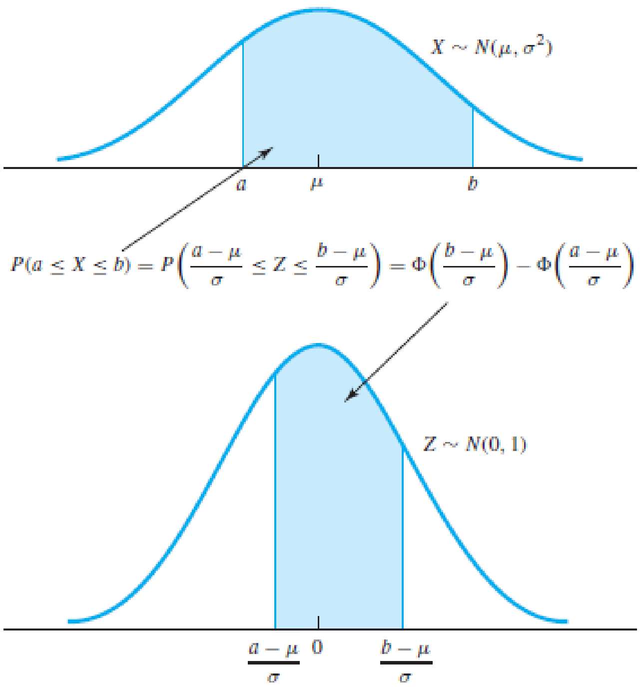
- **Probability Calculations for Normal Distributions:**

$$X \sim N(\mu, \sigma^2)$$

then the transformed random variable

$$Z = \frac{X - \mu}{\sigma}$$

$$\begin{aligned}
P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\
&= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)
\end{aligned}$$



- **Property of a normal distribution:**
  1) It is symmetric.
  2) Mean=Mode=Median.
  3) It is unimodal.
  4) The total area under the curve is equal to one.
  5) The normal curve approaches, but never touches, the x-axis

Example 37 Concrete Block Weights

A company manufactures concrete blocks that are used for construction purposes. Suppose that the weights of the individual concrete blocks are normally distributed with a mean value of $\mu = 11.0$ kg and a standard deviation of $\sigma = 0.3$ kg.

The probability that a concrete block weighs less than 10.5 kg is

$$P(X \le 10.5) = P(-\infty \le X \le 10.5) = \Phi\left(\frac{10.5-\mu}{\sigma}\right) - \Phi\left(\frac{-\infty-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{10.5-11.0}{0.3}\right) - \Phi\left(\frac{-\infty-11.0}{0.3}\right)$$
$$= \Phi(-1.67) - \Phi(-\infty) = 0.0475 - 0 = 0.0475$$

$$P(X \le 10.5) = P(-\infty \le X \le 10.5) = \Phi\left(\frac{10.5-\mu}{\sigma}\right) - \Phi\left(\frac{-\infty-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{10.5-11.0}{0.3}\right) - \Phi\left(\frac{-\infty-11.0}{0.3}\right)$$
$$= \Phi(-1.67) - \Phi(-\infty) = 0.0475 - 0 = 0.0475$$

That is only about 1 in 20 concrete blocks weighs less than 10.5 kg.

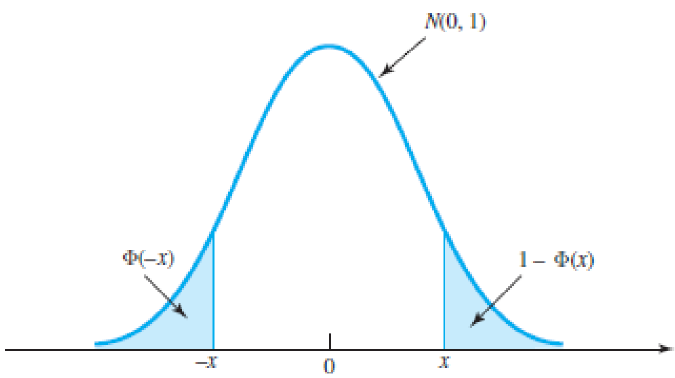
$$P(Z \le 0.31) = \Phi(0.31) = 0.6217$$

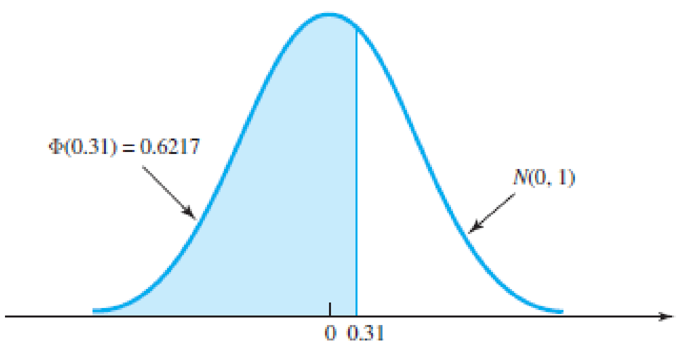
$$P(Z \ge 1.05) = 1 - \Phi(1.05) = 1 - 0.8531 = 0.1469$$

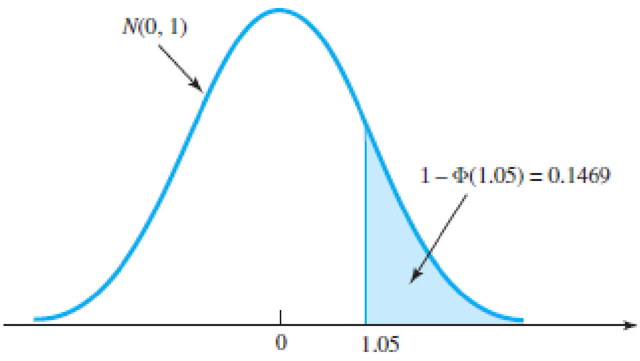
as illustrated in Figure 5.8, and that

$$P(-1.50 \le Z \le 1.18) = \Phi(1.18) - \Phi(-1.50) = 0.8810 - 0.0668 = 0.8142$$

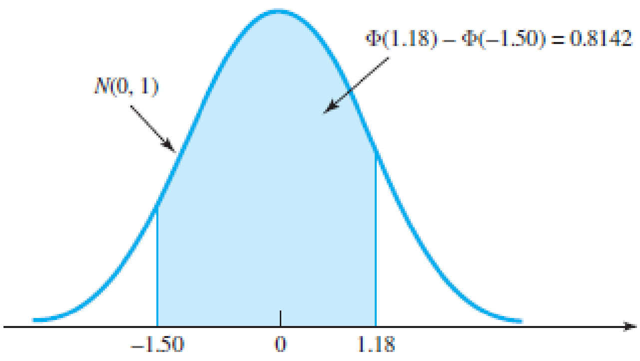
as illustrated in Figure 5.9.



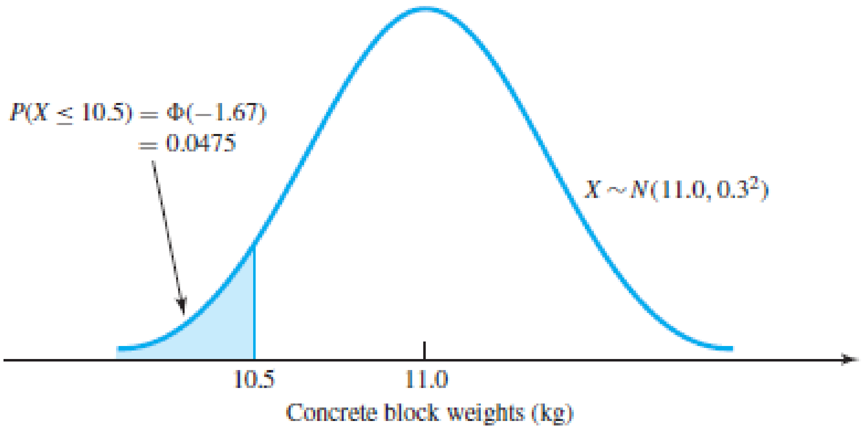




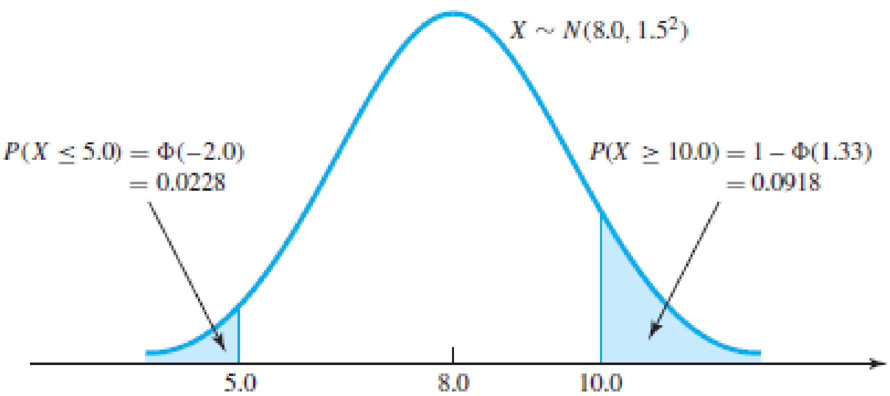
$N(0, 1)$
$1 - \Phi(1.05) = 0.1469$
0
1.05



$\Phi(1.18) - \Phi(-1.50) = 0.8142$
$N(0, 1)$
−1.50
0
1.18



$P(X \leq 10.5) = \Phi(-1.67)$
$= 0.0475$
$X \sim N(11.0, 0.3^2)$
10.5
11.0
Concrete block weights (kg)



$X \sim N(8.0, 1.5^2)$
$P(X \leq 5.0) = \Phi(-2.0)$
$= 0.0228$
$P(X \geq 10.0) = 1 - \Phi(1.33)$
$= 0.0918$
5.0
8.0
10.0