Chapter 6 Continuous Probability Distributions

The observations generated by different statistical experiments have the same general type of behavior. The followings are the probability distributions that will be covered in this chapter:

- Continuous Uniform Distribution
- Normal Distribution
- Normal Approximation to Binomial Distribution
- Gamma and Exponential Distribution
- Chi-Squared Distribution

The Continuous Uniform Distribution

One of the simplest continuous distributions in all statistics is the continuous uniform distribution. This distribution is characterized as follows:

**Definition.** The density function of the continuous random variable $X$ on the interval $[A, B]$ is

$$f(x; A, B) = \begin{cases} 
\frac{1}{B - A}, & A \leq x \leq B \\
0, & \text{elsewhere.} 
\end{cases}$$

**Theorem.** The mean and variance of the continuous uniform distribution are

$$\mu = \frac{A + B}{2} \quad \text{and} \quad \sigma^2 = \frac{(B - A)^2}{12}.$$
The Normal Distribution

The most important continuous probability distribution in the entire field of statistics is the normal distribution. Normal distributions are a family of distributions that have the same general shape. They are symmetric with scores more concentrated in the middle than in the tails. Normal distributions are sometimes described as bell shaped which are shown below. Notice that they differ in how spread out they are. The area under each curve is the same. The height of a normal distribution can be specified mathematically in terms of two parameters: the mean ($\mu$) and the standard deviation ($\sigma$).

![Normal Distribution Graph]

**Definition.** The density function of the normal random variable $X$, with mean $\mu$ and variance $\sigma^2$, is

$$N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \quad -\infty < x < \infty$$

where $\pi = 3.14159\ldots$ and $e = 2.71828$.

**Standard normal distribution**

The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1. Normal distributions can be transformed to standard normal distributions by the formula:

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

where $X$ is a score from the original normal distribution, $\mu$ is the mean of the original normal distribution, and $\sigma$ is the standard deviation of original normal distribution. The standard normal distribution is sometimes called the $z$ distribution. A $z$ score always reflects the number of standard deviations above or below the mean a particular score is. For instance, if a person scored a 70 on a test with a mean of 50 and a standard deviation of 10, then they scored 2 standard deviations above the mean. Converting the test scores to $z$ scores, an $X$ of 70 would be:

$$z = \frac{70-50}{10} = 2$$
So, a z score of 2 means the original score was 2 standard deviations above the mean. Note that the z distribution will only be a normal distribution if the original distribution (X) is normal.

**Note:** The following figures give us the areas to the right of some $z$-value, to the left of some $z$-value and between two $z$-values.

Areas under portions of the standard normal distribution are shown to the right. About 0.68 (0.34 + 0.34) of the distribution is between -1 and 1 while about 0.96 of the distribution is between -2 and 2.
Example 1. Given a standard normal distribution, find the area under the curve that lies
(a) to the right of \( z = 1.84 \)
(b) between \( z = -1.97 \) and \( z = 0.86 \)

Solution. (a) \( P(Z > 1.84) = 1 - P(Z \leq 1.84) \) by table A3
\[ = 1 - 0.9671 = 0.0329 \]

(b) \( P(-1.97 < Z < 0.86) = P(Z < 0.86) - P(Z < -1.97) \) by table A3
\[ = 0.8051 - 0.0244 = 0.7807 \]

Example 2. Given a normal distribution with \( \mu = 50 \) and \( \sigma = 10 \), find the probability that
\( X \) assumes a value between 45 and 62.

Solution.
\[ P(45 < X < 62) = P\left(\frac{45 - 50}{10} < Z < \frac{62 - 50}{10}\right) = P(-0.5 < Z < 1.2) = \text{table}(1.2) - \text{table}(-0.5) \]
\[ = 0.8849 - 0.3088 = 0.5761 \]

Example 3. Given a standard normal distribution, find the value of \( k \) such that
(a) \( P(Z < k) = 0.0427 \)
(b) \( P(Z > k) = 0.2946 \)
(c) \( P(-0.93 < Z < k) = 0.7235 \)

Solution. (a) \( P(Z < k) = 0.0427 \Rightarrow \text{table}(k) = 0.0427 \Rightarrow k = -1.72 \)

(b) \[ P(Z > k) = 0.2946 \Rightarrow P(Z > k) = 1 - P(Z \leq k) = 0.2946 \]
\[ \Rightarrow P(Z \leq k) = 1 - 0.2946 = 0.7054 \]
\[ \Rightarrow \text{table}(k) = 0.7054 \Rightarrow k = 0.54 \]

(c) \[ P(-0.93 < Z < k) = 0.7235 \Rightarrow \text{table}(k) - \text{table}(-0.93) = 0.7235 \]
\[ \text{table}(k) = 0.7235 + 0.1762 = 0.8997 \]
\[ \Rightarrow k = 1.28 \]

Example 4. Given the normally distributed random variable \( X \) with \( \mu_x = 18 \) and \( \sigma_x^2 = 3 \)
(a) Compute \( P(X > 13.74) \)
(b) Compute \( x \) satisfying \( P(x < X < 18) = 0.4332 \).
Applications of the normal Distribution

Example 1. A certain machine makes electrical resistors having a mean resistance of 40 ohms and a standard deviation of 2 ohms. Assuming that the resistance follows a normal distribution and can be measured to any degree of accuracy, what percentage of resistors will have a resistance exceeding 43 ohms?

Solution. Let $X$ be the normal random variable, given $\mu = 40 \text{ohms}, \sigma = 2 \text{ohms}$,

$$P(X > 43) = P\left( \frac{X - \mu}{\sigma} > \frac{43 - 40}{2} \right) = P(Z > 1.5) = 1 - P(Z \leq 1.5)$$

$$= 1 - \text{table}(1.5) = 1 - 0.9332 = 0.0668 = 6.68\%$$

Example 2. The average grade for a exam is 74 and the standard deviation is 7. Assuming that the grades follow a normal distribution, what is the probability that a student will get a grade of at least 60?

Normal Approximation to Binomial Distribution

Theorem. If $X$ is a binomial random variable with mean $\mu = np$ and $\sigma^2 = npq$, then the limiting form of distribution of

$$Z = \frac{X_{bin} - np}{\sqrt{npq}} \text{ as } n \to \infty$$

is the standard normal distribution $N(0,1)$.

Note. We use the normal approximation to binomial distribution whenever $p$ is not close to 0 and 1. If both $np$ and $nq$ are greater than or equal to 5, the approximation will be good.

Example 1. The probability that a patient recovers from a blood disease is 0.4. If 100 people are known to have contracted this disease what is the probability that less than 30 survive?

Solution. Let $X$ be the number of surviving people from blood disease.

Given $n = 100$ and $p = 0.40$, $\mu = np = 100 \times 0.40 = 40$, $\sigma = \sqrt{100 \times 0.40 \times 0.60} = \sqrt{24}$,

$$P(X_{bin} < 30) \approx P(X_{nor} < 29.5) = P\left( \frac{X - np}{\sqrt{npq}} < \frac{29.5 - 40}{\sqrt{24}} \right) = P(Z < -2.14) = \text{table}(-2.14) = 0.0162$$
Example 2. A coin is tossed 400 times, use the normal approximation to binomial to find the probability of obtaining

(a) between 185 and 210 heads inclusive  
(b) exactly 205 heads  
(c) less than 176 or more than 227 heads

Example 3. A balanced die is rolled 180 times. Let be the number of cases when die shows the number 4 on its top face.

(a) Find $\mu_X$  
(b) Find $\sigma_X^2$  
(c) Use normal approximation to binomial to approximate $P(35 \leq X \leq 40)$.

Exercises

Exercise 1. If scores are normally distributed with a mean of 30 and a standard deviation of 5, what percent of the scores is: (a) greater than 30? (b) greater than 37? (c) between 28 and 34?

Exercise 2. Assume a normal distribution with a mean of 90 and a standard deviation of 7. What limits would include the middle 65% of the cases?

Exercise 3. If is the standard normal random variable, 
(a) Calculate $P(−1.3 < Z < 1.37)$  
(b) If $P(a < Z < 1.12) = 0.6845$, find the value of $a$.

Exercise 4. A research scientist reports that mice will live an average of 40 months when their diets are sharply restricted and enriched with vitamins and proteins. Assuming that the lifetimes of such mice are normally distributed with a standard deviation of 6.3 months, find the probability that a given mouse will live  
(a) more than 32 months  
(b) less than 28 months  
(c) between 37 and 49 months.
Exponential and Gamma Distributions

Another continuous distribution that has many useful applications is the exponential distribution, which has density

\[
f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}
\]

Because the sample space for this distribution consists of the positive real numbers, this distribution is sometimes used to model **time to failure or survival time of a system**. The distribution function is,

\[
F(x) = \int_{0}^{x} \frac{1}{\beta} e^{-t/\beta} dt = 1 - e^{-x/\beta}, \quad x > 0
\]

The exponential distribution has a special property that is unique to this distribution. Suppose that \( T \) is the time to failure of a randomly selected new component, and suppose that this r.v. has an exponential distribution with parameter \( \beta \). The probability that this new component survives to time \( t \) is,

\[
P(T > t) = 1 - F(t) = e^{-t/\beta}
\]

A generalization of the exponential distribution that can provide a much wider range of models is based on the gamma integral. Define a function \( f(t; \alpha, \beta) \) by

\[
f(x; \alpha, \beta) = \begin{cases} \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}
\]

where \( \alpha, \beta \) are positive constants. Note that in this parameterization, the parameter \( \beta \) is in the denominator of the exponential component. The reason for this modification will be shown below. Recall that

\[
\int_{0}^{\infty} t^{\alpha-1} e^{-t/\beta} dt = \beta^\alpha \Gamma(\alpha)
\]

which implies that

\[
\int_{0}^{\infty} f(t; \alpha, \beta) dt = 1
\]

and hence that \( f(x; \alpha, \beta) \) is a density function. This density function defines a distribution on the positive real numbers and is referred to as the **gamma distribution**. Note that the exponential distribution is a special case of the gamma distribution in which \( \alpha = 1 \). The parameter \( \alpha \) is referred to as the shape parameter and \( \beta \) is referred to as the scale parameter of the gamma distribution.
Theorem. The Mean and Variance of the Gamma Distribution are

\[ \mu = \alpha \beta \quad \text{and} \quad \sigma^2 = \alpha \beta^2 \]

Theorem. The Mean and Variance of the Exponential Distribution are

\[ \mu = \beta \quad \text{and} \quad \sigma^2 = \beta^2 \]

The following is the plot of the gamma probability density function.

Example 1. In a certain city the daily consumption of water (in millions of liters) follows a gamma distribution with \( \alpha = 2 \) and \( \beta = 3 \). If the daily capacity of that city is 9 million liters of water, what is the probability that on a given day the water supply will be inadequate?

Solution. Let \( X \) be the water supply in millions of liters of water. Given \( \alpha = 2 \) and \( \beta = 3 \),

\[ P(X > 9) = \int_9^\infty f(x) \, dx \quad \text{where} \quad f(x) = \begin{cases} \frac{1}{\Gamma(2)3^2} x e^{-x/3}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \]

Thus, \( P(X > 9) = \int_9^\infty \frac{1}{9} x e^{-x/3} \, dx = \frac{1}{e^3} \).

Note. There is a relationship between the Exponential and Poisson distributions. Suppose events are occurring in time according to Poisson distribution with a rate of \( \lambda \) events per hour. Thus in \( t \) hours, the number of events say \( Y \), will have a Poisson distribution with mean value \( \lambda t \).
Suppose we start at time zero and ask the question ‘How long do I have to wait to see the first event occur?’. Let \( X \) denote the length of time until the first event. Then

\[
P(X > t) = P(Y = 0 \text{ on the interval } (0, t)) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}
\]

and

\[
P(X \leq t) = 1 - P(X > t) = 1 - e^{-\lambda t}.
\]

Thus,

\[
P(X \leq t) = F(t), \text{ the distribution function for } X \text{ has the form of an exponential distribution with } \lambda = \frac{1}{\beta} \text{ (the failure rate)}. \text{ Upon differentiating, we see that}
\]

\[
f(t) = \frac{d(1-e^{-\lambda t})}{dt} = \frac{1}{\beta} e^{-\beta t}, \quad t > 0.
\]

**Example 2.** The life of a certain type of device has an advertised rate of 0.01 per hour. The failure rate is constant and the exponential distribution applies.

(a) What is the probability that 200-hours will pass before a failure is observed?

**Solution.** Given \( f(x) = \begin{cases} 0.01e^{-0.01x}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \),

\[
P(X > 200) = \int_0^{200} 0.01e^{-0.01x} \, dx = -e^{-0.01x} \bigg|_0^{200} = e^{-2}.
\]

(b) What is the mean time to failure?

**Solution.** Since the failure rate \( \frac{1}{\beta} = 0.01, \mu = \beta = 100 \). Therefore the mean failure time is 100-hours.

**Example 3.** The length of time for one individual to be served at a cafeteria is a random variable having an exponential distribution with a mean of 4-minutes. What is the probability that a person is served in less than 3-minutes on at least 4 of the next 6 days?

**Example 4.** The exponential distribution is frequently applied to the waiting times between successes in a Poisson process. If the number of calls received per hour by a telephone answering service is a Poisson random variable with parameter \( \lambda = 6 \), we know that the time, in hours, between successive calls has an exponential distribution with parameter \( \beta = \frac{1}{6} \). What is the probability of waiting more than 15 minutes between any two successive calls?

**Example 5.** In a certain city, the daily consumption of electric power in millions of kw-hours is a random variable \( X \) having a Gamma distribution with \( \mu = 6 \) and \( \sigma^2 = 12 \).
(a) Find the values of $\alpha$ and $\beta$.

\[
\begin{align*}
\mu &= \alpha \beta = 6 \\
\sigma^2 &= \alpha \beta^2 = 12 = 6.2
\end{align*}
\]

$\Rightarrow \beta = 2, \ \alpha = 3.$

(b) Find the probability that on any given day the daily power consumption will exceed 12 million kw-hours.

\[
P(X > 12) = \int_{12}^{\infty} \frac{1}{2^\frac{3}{2} \Gamma\left(\frac{3}{2}\right)} e^{-x/2} x^2 dx = \frac{1}{16} \int_{12}^{\infty} e^{-x/2} x^2 dx
\]

**Chi-Squared Distribution**

Another very important special case of the Gamma distribution is obtained by letting $\alpha = \nu / 2$ and $\beta = 2$, where $\nu$ is a positive integer. The result is called the **Chi-squared distribution**. The distribution has a single parameter, $\nu$, called the **degrees of freedom**.

**Definition.** The continuous random variable $X$ has a Chi-squared distribution, with $\nu$ degrees of freedom, if its density function is given by

\[
f(x) = \begin{cases} 
\frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\nu/2 - 1} e^{-x/2}, & x > 0 \\
0, & \text{elsewhere}
\end{cases}
\]

where $\nu$ is a positive integer.

The Chi-squared distribution plays a vital role in statistical inference that will be studied in the next chapter. Topics dealing with sampling distributions, analysis of variance, and non-parametric statistics involve extensive use of the Chi-squared distribution.

**Theorem.** The mean and variance of the Chi-squared distribution are

$\mu = \nu$ and $\sigma^2 = 2\nu$. 