Lesson 3.5

The Gamma Family of Distributions

Learning Outcomes

At the end of the lesson, students must be able to

1. Describe the key properties of a random variable having a gamma distribution, such as the mean, variance, and moment generating function,
2. Describe the special types of gamma distributions, and
3. Compute probabilities associated with random variables having a gamma distribution.

Introduction

Some random variables are always non-negative and for various reasons yield distributions of data that are skewed (non-symmetric) to the right. That is, most of the area under the density function is located near the origin, and the density function drops gradually as y increases. The lengths of time between malfunctions for aircraft engines possess a skewed frequency distribution, as do the lengths of time between arrivals at a supermarket checkout queue (that is, the line at the checkout counter). Similarly, the lengths of time to complete a maintenance checkup for an automobile or aircraft engine possess a skewed frequency distribution. The populations associated with these random variables frequently possess density functions that are adequately modeled by a gamma density function.

Definition:

A random variable Y is said to have a gamma distribution with shape and scale parameters $0 < \alpha < 1$ and $0 < \beta < 1$, respectively, if its probability function is given by

$$f_Y(y) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} e^{-\frac{y}{\beta}},\ y \geq 0 \\ 0,\ \text{elsewhere} \end{cases}$$

where $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy$.

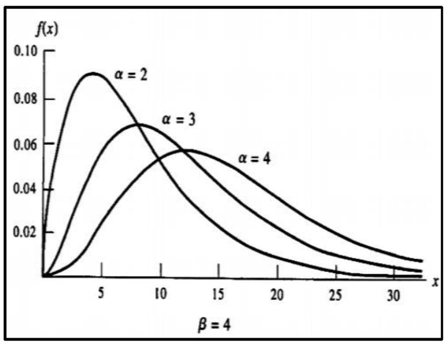
If a random variable has the above PDF then we write $Y \sim Gamma(\alpha, \beta)$.

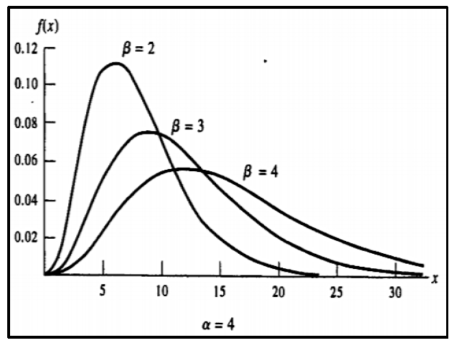
<u>Remarks</u>:

1. The gamma function satisfies certain properties:

   a. $\Gamma(1) = 1$
   b. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
   c. $\Gamma(n + 1) = (n)\Gamma(n) = n!$, for any integer n
   d. $\Gamma(\frac{1}{2}) = \pi$

2. The "flatness" and "peakedness" of the gamma density curve is determined by the parameters α and β, as shown in the following diagrams.





3. Upon closer inspection, we see that the nonzero part of the $Gamma(\alpha, \beta)$ PDF consists of two parts:

$$f_Y(y) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}}$$

a. the kernel of the PDF: $y^{\alpha-1}e^{-\frac{y}{\beta}}$

b. the constant out front: $\frac{1}{\Gamma(\alpha)\beta^\alpha}$

The kernel is the *guts* of the formula, while the constant out front is simply the *right quantity* that makes $f_Y(y)$ a valid pdf; i.e., the constant that makes $f_Y(y)$ integrate to 1. As such,

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}}\, dy = 1 \implies \int_0^\infty y^{\alpha-1} e^{-\frac{y}{\beta}}\, dy = \Gamma(\alpha)\beta^\alpha$$

This result is extremely useful and will be used repeatedly. You can use this result in evaluating integrals without going thru the tedious integration techniques, such as integration by parts, as illustrated below.

$$\int_0^{\infty} y^4 e^{-\frac{y}{3}}\, dy = \Gamma(5)3^5 = 4! \times 243 = 5832$$

Theorem:

If $Y \sim Gamma(\alpha, \beta)$, then

a. $E(Y) = \alpha\beta$

b. $V(Y) = \alpha\beta^2$

c. $m_Y(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha}$, $t < \frac{1}{\beta}$

PROOF: Left as classroom exercise!

<u>Example 1</u>:

The time to death (measured in days) for patients with a serious $\alpha = 2.7$ and $\beta = 100$.

a. What is the probability a patient with this type of cancer will live longer than one year?

b. What is the average time to death for this type of cancer patients?

SOLUTION:

Let Y be the time to death (measured in days), and as given $Y \sim Gamma(2.7, 100)$.

a. Thus,

$$\begin{aligned} P(Y > 365) &= 1 - P(Y \leq 365) \\ &= 1 - F_Y(365) \\ &= 1 - 07644654 \\ &\approx 0.236 \end{aligned}$$

$P(Y \leq 365)$ was determined using the R code **pgamma(365, 2.7, 1/100)**.

b. The mean time to death is $E(Y) = 2.7 \times 100 = 270$ days.

Two special cases of gamma-distributed random variables merit particular consideration. The gamma density function in which $\alpha = 1$ is called the exponential density function.

Definition:

A random variable Y is said to have an exponential distribution with parameter $\beta > 0$ if and only if the density function of Y is

$$f_Y(y) = \begin{cases} \frac{1}{\beta} e^{-\frac{y}{\beta}}, \ y \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

We write $Y \sim Exp(\beta)$.

Theorem:

If Y is distributed as $Exp(\beta)$, then

a. $E(Y) = \beta$

b. $V(Y) = \beta^2$

c. $m_Y(t) = \frac{1}{1-\beta t}$, $t < \frac{1}{\beta}$,

Remarks:

1. Note that in some textbooks, the exponential PDF is written as

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y}, \ y \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

where $\lambda = \frac{1}{\beta}$.

2. There is a unique connection between the Poisson and exponential distributions. The Poisson distribution is used to model the number of events occurring in an interval, while, the exponential distribution models the time in between occurrences of successive events.

An exponential random variable possesses a unique property which we call the "memoryless" property. Suppose that the length of time a component already has operated does not affect its chance of operating for at least b additional time units. That is, the probability that the component will operate for more than $a + b$ time units, given that it has already operated for at least a time units, is the same as the probability that a new component will operate for at least b time units if the new component is put into service at time 0. In symbols, this means that

$$P(Y > a + b | Y > a) = P(Y > b)$$

It can be easily demonstrated using the definition of conditional probability.

$$P(Y > a + b|Y > a) = \frac{P(\{Y > a + b\} \cap \{Y > a\})}{P(Y > a)} = \frac{P(Y > a + b)}{P(Y > a)}$$

Now,

$$P(Y > a + b) = \int_{a+b}^{\infty} \frac{1}{\beta} e^{-\frac{y}{\beta}}\, dy = -e^{-\frac{y}{\beta}} \Big|_{a+b}^{\infty} = e^{\frac{a+b}{\beta}}$$

Similarly,

$$P(Y > a) = \int_{a}^{\infty} \frac{1}{\beta} e^{-\frac{y}{\beta}}\, dy = -e^{-\frac{y}{\beta}} \Big|_{a}^{\infty} = e^{\frac{a}{\beta}}$$

Thus,

$$P(Y > a + b|Y > a) = \frac{e^{\frac{a+b}{\beta}}}{e^{\frac{a}{\beta}}} = e^{\frac{b}{\beta}} = P(Y > b)$$

Example 2:

The magnitude of earthquakes recorded in a certain region can be modeled as having an exponential distribution with mean 2.4, as measured on the Richter scale. Find the probability that an earthquake striking this region will

a) exceed 3.0 on the Richter scale.

b) fall between 2.0 and 3.0 on the Richter scale.

SOLUTION:

Let Y be the magnitude of an earthquake in the region, $Y \sim Exp(2.4)$.

a. Then

$$P(Y > 3.0) = \int_{3}^{\infty} \frac{1}{2.4} e^{-\frac{y}{2.4}}\, dy = -e^{-\frac{3}{2.4}} \approx 0.287$$

This probability can be obtained using the R command **1 - pgamma(3,1,1/ 2.4)** or **1 - pexp(3,1/2.4)**.

b. $P(2.0 < Y < 3.0) = \int_2^3 \frac{1}{2.4} e^{-\frac{y}{2.4}}\, dy \approx 0.148$

The second special case of a gamma distribution is the Chi-square distribution.

Definition:

A random variable Y is said to have a Chi-square distribution (χ^2) with ν degrees of freedom if its probability density function is given by

$$f_Y(y) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{\frac{\nu}{2}-1} e^{-y/2}, \ y \geq 0 \\ 0, \text{elsewhere} \end{cases}$$

We write $Y \sim \chi^2(\nu)$.

Notice that the above PDF is a Gamma PDF with $\alpha = \frac{\nu}{2}$ and $\beta = 2$.

Theorem:

If $Y \sim \chi^2(\nu)$, then

a. $E(Y) = \nu$

b. $V(Y) = 2\nu$

c. $m_Y(t) = \left(\frac{1}{1-2t}\right)^{\nu/2}, \ t < \frac{1}{2}$

The proofs of these results follow directly from the proof for $Y \sim Gamma(\alpha, \beta)$.

Example 3:

Customers arrive in a certain shop according to an approximate Poisson process with a mean rate of 20 per hour. What is the probability that the shopkeeper will have to wait more than 5 minutes for the arrival of the first customer?

SOLUTION: Left as a classroom exercise!

Example 4:

The weekly amount of downtime Y (in hours) for an industrial machine has an approximate gamma distribution with $\alpha = 3$ and $\beta = 2$. The loss L (in dollars) to the industrial operation as a result of this downtime is given by $L = 30Y + 2Y^2$. Find the expected loss.

SOLUTION: Left as a classroom exercise!