Chapter 4

Continuous Random Variables and Their Probability Distributions

The Probability Distribution for a Continuous Random Variable

Review

Recall :

1. A random variable Y is continuous provided the set of possible values of Y is an interval (possibly infinite)
2. For continuous random variables probability is now interpreted as area.

P($a \leq Y \leq b$) = area under the curve associated with the probability function of Y.

The Probability Dist$\underline{^n}$ for a Continuous Random Variable

Definition: Let Y denote any random variable. The *cumulative distribution function* (*cdf*) of Y, denoted by F(y), is given by

$$F(y) = P(Y \le y) \text{ for } -\infty < y < \infty .$$

Example

Y	1	2	3	4
P(Y=y)	.4	.3	.2	.1

Find F(y).

Example (con't)

Sol^n :

Definitions

<u>Definition</u>: Let Y denote a random variable with distribution function F(y). Y is said to be *continuous* if the distribution function F(y) is continuous, for

$-\infty < y < \infty$.

→Y must take on values in an interval.

Definition: Let F(*y*) be the distribution function for a continuous random variable Y. Then *f*(*y*), given by

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

wherever the derivative exists, is called the *probability density function* (*pdf*) for the random variable Y.

Properties of a Density Function

Theorem 4.2: If f(y) is a density function, then

1. $f(y) \geq 0$ for any value of y.

2. $\int_{-\infty}^{\infty} f(y)dy = 1.$

Example 1: Let Y be a continuous random variable with probability density function given by

$$f(y) = \begin{cases} 3y^2 , & 0 \le y \le 1 \\ 0 , & \text{otherwise} . \end{cases}$$

Find F(y).

Theorem 4.3: If the random variable Y has density function $f(y)$ and $a \le b$, then the probability that Y falls in the interval $[a,b]$ is

$$P(a \le Y \le b) = \int_a^b f(y)dy \quad .$$

Examples

Example1 : Given $f(y) = cy^2$, $0 \leq y \leq 2$, and $f(y) = 0$ elsewhere, find the value of c for which $f(y)$ is a valid density function.

Solution :

Example2 : Find $P(1 \leq Y \leq 2)$ for Example 1. Also find $P(1 < Y < 2)$.

Solution:

Example3 : Given $f(y) = ky(2-y)$, $0 < y < 2$, and $f(y) = 0$ elsewhere.

1. Find the value of k for which $f(y)$ is a valid density function.
2. Find $P(\frac{1}{2} \le Y \le 1)$.
3. Find the cumulative density function, F(Y).
4. Use F(Y) to find $P(Y \ge 1 \mid Y \le 2)$.

Sol^n :

1. Find the value of k for which $f(y)$ is a valid density function.

Sol^n :

2. $P(\frac{1}{2} \leq Y \leq 1) = ?$

Sol^n :

3. Find the cumulative density function, F(Y).

Sol^n :

4. $P(Y \geq 1 | Y \leq 2) =$

Expected Values for Continuous Random Variables

Expected Value

- As with discrete random variables we are also interested in the means, variances, and standard deviations of continuous random variables.

<u>Definition</u>: The expected value of a continuous random variable Y is

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

provided that the integral exists.

Analogy with the discrete case $E(Y) = \sum_{y} yp(y)$

Expected Value

<u>Theorem 4.4</u>: Let g(Y) be a function of Y; then the expected value of g(Y) is given by

$$E(g(Y)) = \int_{-\infty}^{\infty} g(y) f(y) dy$$

<u>Theorem 4.5</u>: Let c be a constant, and let $g(Y), g_1(Y), \ldots, g_k(Y)$ be functions of a continuous random variable Y. Then the following results hold:

1. E(c) = c
2. E[cg(Y)] = cE[g(Y)]
3. $E[g_1(Y)+ \ldots +g_k(Y) = E[g_1(Y)]+ \ldots +E[g_k(Y)]$

Examples

Example1 : Suppose Y is a continuous random variable with pdf

$$f(y) = \begin{cases} \dfrac{4}{\pi(1+y^2)} & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find E(Y).

Solution:

Example2 : Suppose Y is a continuous random variable with pdf

$$f(y) = \begin{cases} e^{-y} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $g(y)=e^{(3y/4)}$. Find E(g(Y))

Solution:

Variance

- Also as in the discrete case we define

$$V(Y) = E\left(\left(Y - \mu\right)^2\right) \quad , \quad \mu = E(Y)$$

$$= E(Y^2) - \left(E(Y)\right)^2$$

And

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y) dy$$

Examples

Example3: Suppose Y is a continuous random variable with pdf

$$f(y) = \begin{cases} \dfrac{4}{\pi(1+y^2)} & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find V(Y).

Solution:

Example 4 : Daily total solar radiation for a specified location in Florida in October has probability density function given by

$$f(y) = \begin{cases} (3/32)(y-2)(6-y) & , \quad 2 \le y \le 6 \\ 0 & , \quad \text{otherwise} \end{cases}$$

with measurements in hundreds of calories. Find the expected daily solar radiation for October.

Solution:

- Let Y = # of daily solar radiation in October
- Want to find E(Y).

Example 5: Weekly CPU time used by an accounting firm has probability density function (measured in hours) given by

$$f(y) = \begin{cases} (3/64)y^2(4-y) & , \quad 0 \le y \le 4 \\ 0 & , \quad \text{otherwise} \end{cases}$$

(a) Find the expected value and variance of weekly CPU time.

(b) The CPU time costs the firm \$200 per hour. Find the expected value and variance of the weekly cost for CPU time.

(c) Would you expect the weekly cost to exceed \$600 very often?

(a) Find the expected value and variance of weekly CPU time.

Solution :

Let Y = weekly CPU time

(b) The CPU time costs the firm \$200 per hour. Find the expected value and variance of the weekly cost for CPU time.

Solution:

Let Y = weekly CPU time

C = weekly cost CPU time = $200Y$

(c) Would you expect the weekly cost to exceed $600 very often?

$P(C > 600) = ?$

Solution:

Example 6: Given

$$F(y) = \begin{cases} 0, & y \le 0 \\ y/8, & 0 < y < 2 \\ y^2/16, & 2 \le y < 4 \\ 1, & y \ge 4 \end{cases}$$

Find the mean and variance of Y.

Solution:

Example 7 : Refer to Example 6.

(a) Find $P(1 \leq Y \leq 3)$.

(b) Find $P(Y \geq 1.5)$.

(c) Find $P(Y \geq 1 \mid Y \leq 3)$.

Solution:

(a) $P(1 \leq Y \leq 3)$

(b) $P(Y \geq 1.5)$

(c) $P(Y \geq 1 \mid Y \leq 3)$

The Normal Probability Distribution

The Normal Distribution

The *normal distribution* is continuous, symmetric, bell-shaped distribution of a variable.

The normal distribution or bell-shaped distribution is the most important probability distribution in statistics. It is used for inferential statistics and also for advanced probability.

Normal Distribution

Definition : A random variable Y is said to have a normal probability distribution if and only if , for $\sigma > 0$ and $-\infty < \mu < \infty$, the density function of Y is

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \quad , \quad -\infty < y < \infty$$

Normal Distribution

<u>Theorem 4.7</u> : If Y is a normally distributed random variable with parameters μ and σ, then

$$E(Y) = \mu \qquad \text{and} \qquad V(Y) = \sigma^2$$

Normal Distribution

Remarks :

1. $$P(a \le Y \le b) = \int_a^b f(y)dy = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

   A closed-form expression for this integral does not exist; hence its evaluation requires the use of numerical integration techniques.

2. Always transform a normal random variable Y with parameters μ and σ to a standard normal random variable Z (μ=0 and σ=1) by using the relationship

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

The Normal Distribution (Cont.)

There are several cases for normal curve.

(1) Same means but different standard deviations.

(2) Different means but same standard deviations.

(3) Different means and different standard deviations.

The Normal Distribution (Cont.)

For any normal distribution,

1. The mode, which is the point on the horizontal axis where the curve is at max, occurs at $X=\mu$. (Note: *X* is a normal random variable.)
2. The curve is symmetric about μ.
3. The probability that *X* lies in an interval, say between *a* and *b* is equal to the area under the curve between two point *a* and *b*.
4. The total area under the curve is 1.
5. $P(X = a) = 0$.
6. $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

The Standard Normal Distribution

The Standard Normal Distribution

The *standard normal distribution* is a normal distribution with a mean 0 and standard deviation 1.

All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

$$z = \frac{\text{value - mean}}{\text{standard deviation}} \quad \text{OR} \quad z = \frac{X - \mu}{\sigma}$$

Note: the *z value* is actually the number of standard deviation that a particular *X* value is away from the mean.

Finding Areas Under the Standard Normal Distribution Curve

Example 1: Find the area under the curve that lies:

a) Between z = 0 and z = 1.64.
b) Between z = 0 and z = -1.64.
c) To the right of z = 1.73.
d) To the left of z = -1.22.
e) To the right of z = -0.5
f) To the left of z = 1.09.
g) Between z = -2.04 and z = 0.66.
h) Between z = 0.66 and z = 1.09.
i) Between z = -2.04 and z = -0.5.

The Normal Distribution Curve as a Probability Distribution Curve

The normal distribution curve can be used as a probability distribution curve for normally distributed variables. i.e. *Area and probability are the same.*

The Normal Distribution Curve as a Probability Distribution Curve (Cont.)

Example 2: Find the probability for each.

a. $P(0 < z < 1.64)$
b. $P(- 2.04 < z < 0)$
c. $P(z > 2.14)$
d. $P(z < 1.09)$

The Normal Distribution Curve as a Probability Distribution Curve (Cont.)

Example 3: Find the z value such that

a) the area under the normal distribution curve between 0 and the z value is .3888.

b) the area under the normal distribution curve lies to the left of it is .6915.

The Normal Distribution Curve as a Probability Distribution Curve (Cont.)

Example 4: Given a standard normal distribution, find value of k such that

a) $P(z > k) = .3228$

b) $P(z < k) = .1335$

c) $P(z < k) = .5199$

d) $P(-2.2 < z < k) = .8590$

e) $P(.5 < z < k) = .2667$

Applications of the Normal Distribution

Applications of the Normal Distribution

The standard normal distribution curve can be used to solve a variety of practical problems. The only requirement is that the variable be normally or approximately normally distributed.

To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable by using the formula

$$z = \frac{\text{value - mean}}{\text{standard deviation}} \quad \text{OR} \quad z = \frac{X - \mu}{\sigma}$$

Applications of the Normal Distribution (Cont.)

Example 5: Given a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 60.

Applications of the Normal Distribution (Cont.)

Example 6: Given a normal distribution with $\mu = 100$ and $\sigma = 120$, find the probability that X assumes a value less than 50.

Applications of the Normal Distribution (Cont.)

Example 7: Given a normal distribution with $\mu = 40$ and $\sigma = 5$, find the value of X that has

(a) 35% of the area to the left.

(b) 5% of the area to the right.

Applications of the Normal Distribution (Cont.)

Example 8 Suppose that the heights of 16-month-old oak seedling are normally distributed with a mean of 31.5 cm and standard deviation of 10 cm.

a. What is the probability that the height of a randomly selected seedling will be between 24 and 38 cm?

b. What is the height which 80% of the seedling will grow above?

c. Above what value will 25% of the seedling grow be?

d. Between which 2 values (assume symmetric) will 95% of the seeding grow fall?

The Normal Approximation to the Binomial Distribution

The Binomial Distribution

Recall : A *binomial distribution* is a probability experiment that satisfies the following four requirements.

(1) Fixed number (n) of trials.

(2) Each trial has only 2 possible outcomes.

Success (S)

Failure (F)

(3) (3) the trials are independent.

(4) Probability of success in each trial is the same.

Probability of success = $P(S) = p$

Probability of failure = $P(F) = 1 - p = q$.

The Normal Approximation to the Binomial Distribution

The normal distribution is often used to solve problems that involve the binomial distribution since, when n is large (say,100), the calculations are too difficult to do by hand using the binomial distribution.

Statisticians generally agree that the normal approximation should be used only when $n \cdot p$ and $n \cdot q$ are both greater than or equal to 5.

The Normal Approximation to the Binomial Distribution (Cont.)

A *correction for continuity* is a correction employed when a continuous distribution is used to approximate a discrete distribution.

For example, for $P(X < 4)$, the correction is $P(X < 3.5)$.
For $P(X \leq 4)$, the correction is $P(X < 4.5)$.

The Normal Approximation to the Binomial Distribution (Cont.)

Summary of the Normal Approximation to the Binomial Distribution:

Binomial	Normal
When finding:	Use:
1. P(X = a)	P(a – 0.5 < X < a + 0.5)
2. P(X ≥ a)	P(X > a – 0.5)
3. P(X > a)	P(X > a + 0.5)
4. P(X ≤ a)	P(X < a + 0.5)
5. P(X < a)	P(X < a – 0.5)

For all cases, $\mu = n \cdot p$, $\sigma = \sqrt{n \cdot p \cdot q}$, $n \cdot p \geq 5$ and $n \cdot q \geq 5$.

The Normal Approximation to the Binomial Distribution (Cont.)

Example 9: Vaccinate children for a disease , and the lifetime immunity is found to be 70%. If a random sample of 100 children are vaccinated, find the probability that

(a) Exactly 75 children immunized.

(b) Fewer than 80 children immunized.

(c) More than 65 but less than 79 children immunized.

The Normal Approximation to the Binomial Distribution (Cont.)

Example 10: The probability that a newborn child will be female is 0.5.

(a) Find the probability that in 100 births, 45 or fewer will be female.

(b) Find the probability that in 100 births, greater than 45 are male.