Chapter 4. Continuous Random Variables and Their Probability Distributions

4.1 Introduction

4.2 The Probability distribution for a continuous random variable

4.3 Expected value for continuous random variables

4.4-4.6 Well-known discrete probability distributions

   The Uniform probability distribution
   The Normal probability distribution
   The Gamma probability distribution

4.10 Tchebysheff’s theorem
4.1 Introduction

Recall what “a r.v. \( Y \) is discrete” means

- the support of a discrete \( Y \) is a countable set (i.e., a finite or countably infinite set)

- the probability distribution (probability mass function) for a discrete \( Y \) can always be given by assigning a positive probability to each of the possible values that the variable may assume.

- the sum of probabilities that we assign must be 1.

Are all r.v.s of interest discrete? No!. There also exist “continous” random variables whose set of possible values are uncountable.

(Example) Let \( T \) be the r.v. denoting the time until the first radioactive particle decays. Suppose the time could take on any value between 0 and 10 seconds. Then each of the uncountably infinite number of points in the interval (0,10) represents a distinct value possible value of the time.
(Example continued)
But the probability is zero that the first decay occurs exactly at any specific time, say, \( T = 2.0 \) seconds. Only the probability is positive that the first decay occurs in an interval.

A continuous r.v. takes on any value in an interval.

- Unfortunately, the probability distribution for a continuous r.v. cannot be specified in the same way as that of a discrete r.v.

- It is mathematically impossible to assign nonzero probabilities to all the points on a line interval, and at the same time satisfy the requirement that the probabilities of the distinct possible values sum to 1.

We need different method to describe the probability distribution for a continuous r.v.
4.2.1 The cumulative distribution function for a random variable

(Def 4.1) Let $Y$ denote a (discrete/continuous) random variable. The cumulative distribution function (C.D.F.) of $Y$, denoted by $F(y)$, is given by

$$F(y) \equiv P(Y \leq y)$$

for $-\infty < y < \infty$.

The nature of the C.D.F. associated with $y$ determines whether the variable is continuous or discrete.!!!

[Example for the properties of a CDF]
Suppose that $Y \sim b(n = 2, p = .5)$. Find and draw $F(y)$ for $Y$ (we learned the CDF of a discrete $Y$ in Lecture note 3.2). What we can see from this graph?
(Example continued)
The CDF $F(y)$ in this example
1. goes to 0 as $y$ goes to $-\infty$.
2. goes to 1 as $y$ goes to $\infty$.
3. is monotonic and nondecreasing.
4. is a step function (not continuous): CDFs for discrete random variables are always step functions because they increase only a countable number of points.

(Theorem 4.1)
If $F(y)$ for a (discrete/continuous) r.v. $Y$ is a C.D.F., then

1. $F(-\infty) \equiv \lim_{y \to -\infty} F(y) = 0$.
2. $F(\infty) \equiv \lim_{y \to \infty} F(y) = 1$.
3. $F(y)$ is a nondecreasing function of $y$ (i.e., $F(y_1) \leq F(y_2)$ for $y_1 < y_2$).
4.2.2 The Probability distribution for a continuous r.v.

(Def 4.2) Let $Y$ denote a r.v. with CDF $F(Y)$ satisfying the properties in (Theorem 4.1). $Y$ is said to be \textit{continuous} if $F(y)$ is continuous for $-\infty < y < \infty$.

- Figure for a CDF for a continuous r.v.
- For a continuous r.v. $Y$, $P(Y = y) = 0$ for any real number $y$
- If $P(Y = y_0) = p_0 > 0$, then $F(y)$ would have a discontinuity (jump) of size $p_0$ at the point $y_0$.
- If $F(y)$ is a step-function for $-\infty < y < \infty$, then $Y$ is a discrete r.v..

(Def 4.3) Let $F(Y)$ be the CDF for a continuous r.v. $Y$. Then $f(y)$, given by

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

wherever the derivative exists, is called the \textit{probability density function} (p.d.f.) for the r.v. $Y$. 
(Def 4.2) and (Def 4.3) give us

\[ F(y) = \int_{-\infty}^{y} f(t)dt. \]

[iii] probability distribution for a discrete \( Y \) : p.m.f. \( p(y) \)

[ii) probability distribution for a continuous \( Y \) : p.d.f., \( f(y) \)

(Theorem 4.2) **Properties of a p.d.f.** If \( f(y) \) is a p.d.f. for a continuous r.v. \( Y \), then

1. \( f(y) \geq 0 \) for any value of \( y \).
2. \( \int_{-\infty}^{\infty} f(y)dy = 1. \)
3. \( P(Y \in B) = \int_{B} f(y)dy. \)

[Recall] Properties of the probability distribution for a discrete \( Y \)? (See Theorem 3.1)

(Example 4.2)

(Example 4.3)
Note that $F(y_0)$ in (Def 4.1) gives the probability that $Y \leq y_0$, $P(Y \leq y_0)$. How about the probability that $Y$ falls in a specific interval, $P(a \leq Y \leq b)$?

[Recall] For a discrete r.v. $Y$ with a p.m.f $p(y)$, $P(a \leq Y \leq b)$? (see Lecture note 3.2)

(Theorem 4.3) If a continuous $Y$ has p.d.f. $f(y)$ and $a \leq b$, then the probability that $Y$ falls in the interval $[a, b]$ is

$$P(a \leq Y \leq b) = P(a < Y < b) = P(a < Y \leq b) = P(a \leq Y < b) = \int_{a}^{b} f(y)dy.$$  

Why?  

i) $P(Y = a) = 0$ and $P(Y = b) = 0$,  

ii) $P(a \leq Y \leq b) = P(Y \leq b) - P(Y < a) = P(Y \leq b) - P(Y \leq a)$  

$$= \int_{-\infty}^{b} f(y)dy - \int_{-\infty}^{a} f(y)dy = \int_{a}^{b} f(y)dy$$

[note] How about for a discrete $Y$?
(Example 4.4)

(Example 4.5)

(Exercise 4.11)

[ Summary of discrete and continuous r.v. ]
4.3 Expected value for continuous r.v.

(Def 4.4) The expected value of a continuous random variable $Y$ is

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

provided that the integral exists.

[Note] For a discrete $Y$, $E(Y) = \sum_y yp(y)$: the quantity $f(y)dy$ corresponds to $p(y)$ for the discrete case, and integration is analogous to summation.

(Theorem 4.4) Let $g(Y)$ be a function of $Y$. Then the expected value of $g(Y)$ is given by

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

provided that the integral exists.

[Note] If $g(Y) = (Y - \mu)^2$, the variance of $Y$ is $V(Y) = E(Y - \mu)^2 = E(Y^2) - \mu^2$. 

10
(Theorem 4.5) Let $c$ be a constant, and let $g(Y), g_1(Y), \ldots, g_k(Y)$ be functions of a continuous r.v. $Y$. Then the following results hold:

1. $E(c) = c$.

2. $E[cg(Y)] = cE[g(Y)]$.

3. $E[g_1(Y) + g_2(Y) + \ldots + g_k(Y)]$
   \[ = E[g_1(Y)] + E[g_2(Y)] + \ldots + E[g_k(Y)]. \]

Theorem 3.2 - 3.6 for discrete r.v.
Theorem 4.4 - 4.5 for continuous r.v.

(Example 4.6)

(Exercise 4.21)

(Exercise 4.26)
4.4 (Continuous) Uniform random variable

All the values of a uniform r.v. are as likely to occur.

(Def 4.5) A r.v. \( Y \) is said to have a *continuous uniform probability distribution* with the parameters, \( \theta_1 \) and \( \theta_2 \) where \( \theta_1 < \theta_2 \) (i.e., \( Y \sim U(\theta_1, \theta_2) \)) if and only if the p.d.f. of \( Y \) is

\[
f(y) = \begin{cases} 
\frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2, \\
0, & \text{elsewhere.}
\end{cases}
\]

(Question) Does \( f(y) \) in (Def 4.5) satisfy the necessary properties in (Theorem 4.2)?

(Question) What is \( F(y) \) if \( Y \sim U(\theta_1, \theta_2) \)?
(Theorem 4.6(p.168))
If $\theta_1 < \theta_2$ and $Y$ is a r.v. uniformly distributed on the interval $(\theta_1, \theta_2)$. Then

$$\mu = E(Y) = \frac{\theta_1 + \theta_2}{2} \quad \text{and} \quad \sigma^2 = V(Y) = \frac{\left(\theta_2 - \theta_1\right)^2}{12}$$

(Proof)

(Exercise 4.51) The cycle time for trucks hauling concrete to a highway construction site is uniformly distributed over the interval 50 to 70 minutes. What is the probability that the cycle time exceeds 65 minutes if it is known that the cycle time exceeds 55 minutes?
4.5 Normal random variable

The most widely used continuous probability distribution is the normal distribution with the familiar ‘bell’ shape (the empirical rule (p.10)).

(Def 4.7) A r.v. $Y$ is said to have a normal probability distribution with two parameters, mean $\mu$ and variance $\sigma^2$ (i.e., $Y \sim N(\mu, \sigma^2)$) if and only if, for $\sigma > 0$ and $-\infty < \mu < \infty$, the p.d.f. of $Y$ is

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

(Question) Does $f(y)$ in (Def 4.7) satisfy the necessary properties in (Theorem 4.2)?

(Theorem 4.7) If $Y$ is a normally distributed r.v. with parameters $\mu$ and $\sigma$, then

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

[Note] $\mu$ (location parameter) locates the center of the distribution and $\sigma$ (scale parameter) measures its spread.
[Properties of $Y \sim N(\mu, \sigma^2)$]

- $f(y)$ is symmetric at $\mu$
  
  i) $F(\mu) = P(Y \leq \mu) = P(Y \geq \mu) = 1 - F(\mu) = 0.5$,

  ii) For a such that $a \geq 0$,

  $\cdot F(\mu - a) = P(Y \leq \mu - a) = P(Y \geq \mu + a) = 1 - F(\mu + a)$,
  
  $\cdot F(\mu + a) - F(\mu - a) = P(\mu - a < Y < \mu + a)$
  
  $= 2P(\mu < Y < \mu + a) = 2P(\mu - a < Y < \mu)$

  So, it is enough to know the areas on only one side of the mean $\mu$.

(Question) What is $F(a) = P(Y \leq a)$ if $Y \sim N(\mu, \sigma^2)$?

The calculation of $P(Y \leq a)$ requires evaluation of the integral

$$\int_{-\infty}^{a} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

Unfortunately, a closed-form expression for this integral does not exist. Hence we need the use of numerical integration techniques.

But there is an easy way to do this job: Use Table 4, Appendix III. For the use of Table 4, we need to know the three following steps:
1) Standardize $Y$ by using $\mu$ and $\sigma^2$

- $Z \equiv \frac{Y - \mu}{\sigma} \sim N(0, 1)$

with $f(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}, \quad -\infty < z < \infty$

- $f(z)$ is symmetric at 0: For $a > 0$,
  
  $P(Z \leq -a) = P(Z \geq a),$
  
  $P(-a < Z < a) = 2P(0 < Z < a) = 2P(-a < Z < 0)$

- The tabulated areas in Table 4 are to the right of points $z$, $P(Z \geq a)$.

2) Apply $Z \equiv \frac{Y - \mu}{\sigma}$ to $P(a \leq Y \leq b)$

3) Read Table 4 and use the properties of $Z$ in 1) in order to find $P\left(\left.\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)\right.$

(Example) Find

- $P(Z > 2),\ P(-2 \leq Z \leq 2)$
- $P(0 \leq Z \leq 1.73),\ P(-2.37 \leq Z \leq 1.24)$

(Example) Find $a$ and $b$ such that

- $P(Z \leq a) = 0.9147$,

- $P(a \leq Z \leq b) = 0.0603$

- Percentile of $Z$: $z_\alpha = \text{the } 100(1-\alpha)\% \text{ percentile of } Z \text{ such that } P(Z \leq z_\alpha) = 1 - \alpha$.

- $P(Z \geq z_\alpha) = \alpha,\ P(Z \leq z_{1-\alpha}) = \alpha$. So $z_{1-\alpha} = -z_\alpha$.  

16
Let $Y \sim N(3, 16)$.
(a) Find $P(Y \leq 5)$
(b) Find $P(Y \geq 4)$
(c) Find $P(4 \leq Y \leq 8)$
(d) Find the value of $c$ such that $P(|Y - 3| \leq c) = 0.9544$.

A candy maker produces mints that have a label weight of 20.4 grams. Assuming that the distribution of the weights of these mints is $N(21.37, 0.16)$. Let $Y$ denote the weight of a single mint selected at random from the production line. Find $P(Y \geq 22.07)$. 

(Example) Let $Y \sim N(3, 16)$.
(Exercise 4.73)
4.6 Gamma random variable

[1] The Gamma probability distribution is widely used in engineering, science, and business, to model continuous variables that are always positive and have skewed distributions.

The lengths of time between malfunctions for aircraft engines possess a skewed distribution as do the lengths of time between arrivals at a supermarket checkout queue. The populations associated with theses random variables frequently possess distributions that are adequately modelled by a gamma density function.

[2] The Gamma probability function provides two important probability distributions:

- **Chi-squared distribution**
- **Exponential distribution**
(Def 4.8) A r.v. $Y$ is said to have a *gamma probability distribution with parameters* $\alpha > 0$ and $\beta > 0$ (i.e., $Y \sim \text{gamma}(\alpha, \beta)$) if and only if the p.d.f. of $Y$ is

$$f(y) = \begin{cases} \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)} & 0 \leq y < \infty, \\ 0 & \text{elsewhere.} \end{cases}$$

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

Note that

- $\Gamma(\alpha)$: the *gamma function*,
  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ where $\alpha > 0$,
  $\Gamma(1) = 1$, $\Gamma(n + 1) = n!$ for any integer $n > 0$.

- $\alpha$ is the shape parameter and $\beta$ is the scale parameter

(Question) Does $f(y)$ in (Def 4.8) satisfy the necessary properties in (Theorem 4.2)?

(Theorem 4.8) If $Y$ is a gamma distribution with parameters $\alpha$ and $\beta$, then

$$\mu = E(Y) = \alpha\beta \quad \text{and} \quad \sigma^2 = V(Y) = \alpha\beta^2.$$
Two special cases of gamma r.v.:

1) **Chi-squared distribution** (Def 4.9)
2) **Exponential distribution** (Def 4.10).

(Def 4.9) **Chi-squared distribution**
Let $\nu$ be a positive integer. A r.v. $Y$ is said to have a *chi-square distribution with $\nu$ degrees of freedom* (i.e., $Y \sim \chi^2(\nu)$) if and only if $Y$ is a gamma r.v. with parameters $\alpha = \nu/2$ and $\beta = 2$:

$$f(y) = \begin{cases} 
\frac{y^{\nu/2-1}e^{-y/2}}{2^{\nu/2}\Gamma(\nu/2)} & 0 \leq y < \infty, \\
0 & \text{elsewhere.}
\end{cases}$$

(Theorem 4.9) If $Y$ is a chi-square distribution with $\nu$ degrees of freedom, then

$$\mu = E(Y) = \nu \quad \text{and} \quad \sigma^2 = V(Y) = 2\nu.$$
Note that

- **Table 6, Appendix III** gives probabilities associated with $\chi^2$ distributions. This table gives percentage points associated with $\chi^2$ distributions for many choices of $\nu$.

- Denote $\chi^2_\alpha(\nu) = \text{the } 100(1-\alpha) \text{ percentile of } \chi^2(\nu) \text{ for } Y \sim \chi^2(\nu): P(Y \leq \chi^2_\alpha(\nu)) = 1 - \alpha$.

(Example) Find $\chi^2_{0.10}(5)$, $\chi^2_{0.90}(5)$, $\chi^2_{0.95}(17)$ and $\chi^2_{0.025}(17)$.

(Example)
(a) If $Y \sim \chi^2(5)$, find $P(1.145 \leq Y \leq 12.83)$.
(b) If $Y \sim \chi^2(7)$, find $P(1.690 \leq Y \leq 16.01)$.
(c) If $Y \sim \chi^2(17)$, find $P(Y < 7.564)$ and $P(Y > 27.59)$.
(d) If $Y \sim \chi^2(17)$, find $P(6.408 < Y < 27.59)$. 

21
A r.v. $Y$ is said to have an exponential distribution with parameter $\beta > 0$ (i.e., $Y \sim \exp(\beta)$) if and only if the density function of $Y$ is

$$f(y) = \begin{cases} \frac{1}{\beta}e^{-y/\beta} & 0 \leq y < \infty, \\ 0 & \text{elsewhere}. \end{cases}$$

[Note] The gamma density function in which $\alpha = 1$ is called the exponential density function.

(Theorem 4.10) If $Y$ is an exponential random variable with parameter $\beta$, then

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

Note that

- The exponential density function is useful for modelling the length of life of electronic components.
- The memoryless property of the exponential distribution:

$$P(Y > a + b \mid Y > a) = P(Y > b) \text{ for } a, b > 0$$
(Example) Let $Y$ have an exponential distribution with mean $= 40$. Find 
(a) probability density function of $Y$. 
(b) $P(Y < 36)$
(c) $P(Y > 36 \mid Y > 30)$

(Exercise 4.93) Times between accidents for all fatal accidents on scheduled American domestic passenger flights during the years 1948 through 1961 were found to have an approximately exponential distribution with mean 44 days.

a. If one of the accidents occurred on July 1 of a randomly selected year in the study period, what is the probability that another accident occurred that same month?

b. What is the variance of the times between accidents for the years just indicated?

(Exercise 4.110)
4.10 Tchebyseff’s Theorem (See 3.11)

(Theorem 4.13) Let $Y$ be a r.v. with finite mean $\mu$ and variance $\sigma^2$. Then, for any $k > 0$,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{or} \quad P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

(Example 4.17)

(Exercise 4.147) A machine used to fill cereal boxes dispenses, on the average, $\mu$ ounces per box. The manufacturer wants the actual ounces dispensed $Y$ to be within 1 ounce of $\mu$ at least 75% of the time. What is the largest value of $\sigma$, the standard deviation of $Y$, that can be tolerated if the manufacturer’s objectives are to be met?