Chapter 4 Continuous Distributions

4.1. Probability Density

Probability Density Function

Let $F_X(x)$ be the *cumulative distribution function* (*cdf*) for a continuous random variable *X*, for $-\infty < x < \infty$. There is a function

$$f_X(x) = \frac{dF_X(x)}{dx} = F_X'(x)$$

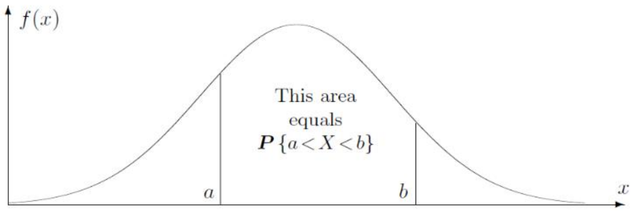
wherever the derivative exists, known as the ***probability density function*** (hereinafter abbreviated *pdf*) for the random variable *X*. It follows that $F_X(x)$ can be written as

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt.$$

The probability distribution curve of the random variable *X* is also recognized as *pdf*.

Properties of a probability density function.

1. $f_X(x) \ge 0$ for all $x \in \mathbb{R}$.
2. The total area between the curve of $f_X(x)$ and the *x*-axis is always 1, that is
$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$
3. $P(a < X < b) = \int_a^b f_X(x)\,dx = F(b) - F(a)$



Note. $P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$

The probability that a continuous random variable *X* assumes **a single value** is always **zero**. That is,

$$P(X = a) \equiv P(a \le X \le a) = \int_a^a f_X(x)\,dx = 0.$$

Analogy: pmf versus pdf

	Discrete	Continuous
Definitions	$P_X(x) = P(X = x)$, *pmf*	$f_X(x) = F_X'(x)$, *pdf*
Computing probabilities	$P(X \in A) = \sum_{x \in A} P_X(x)$	$P(X \in A) = \int_A f_X(x)\, dx$
Cumulative distribution function	$F_X(x) = P(X \le x) = \sum_{y \le x} P(y)$	$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\, dy$
Total probability	$\sum_x P_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Example 4.1.

Suppose that $F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}.$

Find the probability density function for *X*.

Solution:

Example 4.2.

Let *X* be a continuous random variable with probability density function given by

$$f_X(x) = \begin{cases} 4x^3, & 0 \le x \le 1 \\ 0, & elsewhere \end{cases}.$$

Find $F_X(x)$.

Solution:

Expectation and Variance

Discrete	Continuous
$E(X)=\sum_x x\cdot P_X(x)$	$E(X)=\int_{-\infty}^{\infty} x\cdot f_X(x)\,dx$
$Var(X)=E\left[(X-\mu)^2\right]$ $=\sum_x (x-\mu)^2\cdot P_X(x)$ $=\left[\sum_x x^2\cdot P_X(x)\right]-\mu^2$	$Var(X)=E\left[(X-\mu)^2\right]$ $=\int_{-\infty}^{\infty}(x-\mu)^2\cdot f_X(x)\,dx$ $=\left[\int_{-\infty}^{\infty}x^2\cdot f_X(x)\,dx\right]-\mu^2$
$Cov(X,Y)=E\left[(X-\mu_X)(Y-\mu_Y)\right]$ $=\sum_x\sum_y (x-\mu_X)(y-\mu_Y)P_{(X,Y)}(x,y)$ $=\sum_x\sum_y (xy)\cdot P_{(X,Y)}(x,y)-\mu_X\mu_Y$	$Cov(X,Y)=E\left[(X-\mu_X)(Y-\mu_Y)\right]$ $=\int\int (x-\mu_X)(y-\mu_Y)\,f_{(X,Y)}(x,y)\,dxdy$ $=\left[\int\int (xy)\cdot f_{(X,Y)}(x,y)\,dxdy\right]-\mu_X\mu_Y$

<u>*Example 4.3.*</u>

Given $f_X(x)=\begin{cases} kx\ , & 0\le x\le 4 \\ 0\ , & elsewhere \end{cases}$.

Find

(a) the value of k for which $f_X(x)$ is a valid density function,

(b) $P(1\le X\le 3)$ and $P(1< X\le 3)$,

(c) the mean and variance of X.

Solution:

Joint and Marginal Densities
For a vector of random variables, the ***joint cumulative distribution function*** is defined as

$$F_{(X,Y)}(x,y) = P(X \le x \cap Y \le y).$$

The ***joint density*** is the *mixed derivative* of the joint *cdf*,

$$f_{(X,Y)}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y).$$

Similarly to the discrete case, a marginal density of *X* or *Y* can be obtained by integrating out the other variable. Variables *X* and *Y* are *independent* if their joint density factors into the product of marginal densities. Probabilities about *X* and *Y* can be computed by integrating the joint density over the corresponding set of vector values $(x,y) \in \mathbb{R}^2$. This is also analogous to the discrete case.

	Discrete	Continuous
Marginal distributions	$P_X(x) = \sum_y P(x,y)$ $P_Y(y) = \sum_x P(x,y)$	$f_X(x) = \int f_{(X,Y)}(x,y)\,dy$ $f_Y(y) = \int f_{(X,Y)}(x,y)\,dx$
Independence	$P_{(X,Y)}(x,y) = P_X(x)P_Y(y)$	$f_{(X,Y)}(x,y) = f_X(x)\,f_Y(y)$
Computing probabilities	$P((X,Y)\in A)$ $= \sum_{(x,y)} \sum_{\in A} P_{(X,Y)}(x,y)$	$P((X,Y)\in A)$ $= \int\int_{(x,y)\in A} f_{(X,Y)}(x,y)\,dxdy$

<u>*Example 4.4.*</u>
Let *X* and *Y* have the joint *pdf*

$$f(x,y) = \frac{4}{3}(1-xy),\ 0 \le x \le 1,\ 0 \le y \le 1.$$

The marginal *pdfs* are

$$f_X(x) =$$

and

$$f_Y(y) =$$

The following probability is computed by a double integral

$$P\left(Y \le \frac{X}{2}\right) =$$

4.2. Families of Continuous Distributions – The Uniform Distribution

For *continuous random variables*, there is an infinite number of values in the sample space, but in some cases the values may appear to be equally likely. Continuous random variables that appear to have equally likely outcomes over their range of possible values possess a ***uniform probability distribution***.

Let the random variable X denote the outcome when a point is selected at random from an interval $[a,b]$, where $-\infty < a < b < \infty$. The interval may be open, or half-closed, or closed, as for a continuous probability distribution changing a value of density at one point does not change any probability.

Recall.

$P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$ & $P(X = a) = 0$.

The distribution function of X is

$$F_X(x) = \begin{cases} 0 & , \ x < a \\ \dfrac{x-a}{b-a} & , \ a \le x < b \\ 1 & , \ x \ge b \end{cases}.$$

Uniform Probability Distribution

The density function of X follows a uniform distribution on the interval $[a,b]$ is given by

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & , \ a \le x \le b \\ 0 & , \ elsewhere \end{cases}.$$

Mean and Variance

Suppose $X \sim U(a,b)$. Then

$$\mu = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = \frac{(b-a)^2}{12}.$$

Note.

The uniform probability distribution is also known as ***rectangular distribution***.

Example 4.5.

Suppose that a large conference room for a certain company can be reserved for no more than 4 hours. However, the use of the conference room is such that both long and short conferences occur quite often. In fact, it can be assumed that length X of a conference has a uniform distribution on the interval [0, 4].

(a) What is the probability density function of X ?

(b) What is the probability that any given conference lasts at least 3 hours?

Solution:

4.3. Families of Continuous Distributions – The Exponential Distribution

Exponential distribution is often used to model time: waiting time, interarrival time, hardware lifetime, failure time, time between telephone calls, etc.

Exponential distribution has density

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0.$$

With this density, the Exponential *cdf*, mean and variance are computed as

$$F(x) = 1 - e^{-\lambda x}$$

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}.$$

The quantity λ is a parameter of Exponential distribution, it is the frequency or the number of events per time unit. This λ has the same meaning as the parameter of Poisson distribution.

Example 4.6.
Jobs are sent to a printer at an average rate of 3 jobs per hour.
(a) What is the expected time between jobs?
(b) What is the probability that the next job is sent within 5 minutes?
Solution:

Memoryless Property
It is said that "Exponential variables lose memory." What does it mean?

Suppose that an Exponential variable *T* represents waiting time. Memoryless property means that the fact of having waited for *t* minutes gets "forgotten," and it does not affect the future waiting time. Regardless of the event $T > t$, when the total waiting time exceeds *t*, the remaining waiting time still has Exponential distribution with the same parameter. Mathematically,

$$P(T > t + x \mid T > t) = P(T > x) \text{ for } t, x > 0.$$

In this formula, t is the already elapsed portion of waiting time, and x is the additional, remaining time.

This property is unique for Exponential distribution. No other continuous variable $X \in (0, \infty)$ is memoryless. Among discrete variables, such a property belongs to Geometric distribution.

4.4. <u>Families of Continuous Distributions – The Gamma Distribution</u>

When a certain procedure consists of α independent steps, and each step takes $Exponential(\lambda)$ amount of time, then the total time has Gamma distribution with parameters α and λ.

Thus, Gamma distribution can be widely used for the total time of a multistage scheme, for example, related to downloading or installing a number of files. In a process of rare events, with Exponential times between any two consecutive events, the time of the α^{th} event has Gamma distribution because it consists of α independent Exponential times.

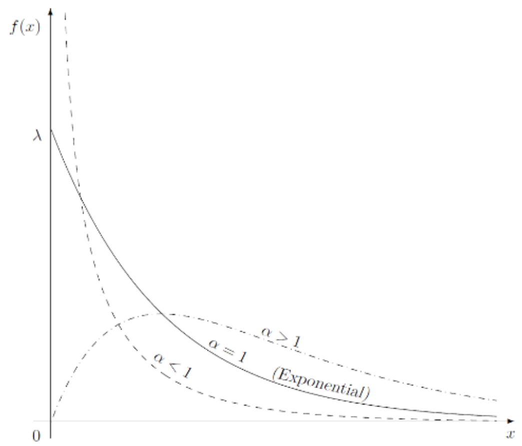
<u>Example.</u> Users visit a certain internet site at the average rate of 12 hits per minute. Every sixth visitor receives some promotion that comes in a form of a flashing banner. Then the time between consecutive promotions has Gamma distribution with parameters $\alpha = 6$ and $\lambda = 12$.

Having two parameters, Gamma distribution family offers a variety of models for positive random variables. Besides the case when a Gamma variable represents a sum of independent Exponential variables, Gamma distribution is often used for the amount of money being paid, amount of a commodity being used (gas, electricity, etc.), a loss incurred by some accident, etc.

Gamma distribution has a density

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \text{ for } x > 0.$$

In fact, α can take any positive value, not necessarily integer. With different α, the Gamma density takes different shapes. For this reason, α is called a *shape parameter*.



Notice two important special cases of a Gamma distribution. When $\alpha = 1$, the Gamma distribution becomes Exponential. Another special case with $\lambda = \frac{1}{2}$ and any $\alpha > 0$ results in a so-called ***Chi-square distribution*** with 2α degrees of freedom.

$$Gamma(1, \lambda) \quad = \quad Exponential(\lambda)$$

$$Gamma\left(\alpha, \frac{1}{2}\right) \quad = \quad Chi-square(2\alpha)$$

Expectation, Variance and Some Useful Integration Remarks

Gamma cdf has the form

$$F(t)=\int_0^t f(x)\,dx=\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^t x^{\alpha-1}e^{-\lambda x}\,dx\,.$$

This expression, related to a so-called incomplete Gamma function, does not simplify, and thus, computing probabilities is not always trivial. Let us offer several computational shortcuts.

Notice that $\int_0^\infty f(x)\,dx=1$ for Gamma and all the other densities. Hence, we obtain that

$$\int_0^\infty x^{\alpha-1}e^{-\lambda x}\,dx=\frac{\Gamma(\alpha)}{\lambda^\alpha}\text{ for any }\alpha>0\text{ and }\lambda>0\,.$$

With this, we can prove that

$$E(X)=\frac{\alpha}{\lambda}$$

and

$$Var(X)=\frac{\alpha}{\lambda^2}\,.$$

<u>*Example 4.7.*</u>

Compilation of a computer program consists of 3 blocks that are processed sequentially, one after another. Each block takes Exponential time with the mean of 5 minutes, independently of other blocks.

(a) Compute the expectation and variance of the total compilation time.

(b) Compute the probability for the entire program to be compiled in less than 12 minutes.

Solution:

Gamma – Poisson Formula

For a $\Gamma(\alpha,\lambda)$ variable T and a $Poi(\lambda t)$ variable X,

$$P(T>t)=P(X<\alpha)$$
$$P(T\le t)=P(X\ge\alpha).$$

<u>Recall.</u> $P(T>t)=P(T\ge t)$ and $P(T<t)=P(T\le t)$ for a Gamma variable T, because it is continuous.

<u>Example 4.8.</u>

Reconsider *Example 4.7.*, use the Gamma – Poisson formula to find an alternative solution to *part (b).*

Solution:

<u>Example 4.9.</u>

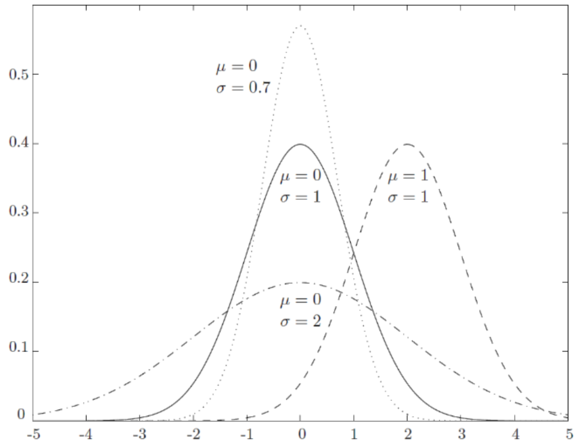
Lifetimes of computer memory chips have Gamma distribution with expectation $\mu=12$ years and standard deviation $\sigma=4$ years. What is the probability that such a chip has a lifetime between 8 and 10 years?

Solution:

4.5. The Normal Distribution

The *normal distribution* is one of the many probability distributions that a continuous random variable can possess. The normal distribution is the most important and most widely used of all probability distributions.

The *normal probability distribution* or the *normal curve* is a bell-shaped (symmetric) curve, centered at μ, its spread being controlled by σ.



As we can see from the graph, changing μ shifts the curve to the left or to the right without changing its shape, while changing σ makes it more concentrated or more flat. Often μ is called a *location parameter* and σ is called a *scale parameter*.

Normal Probability Distribution

The density function of *X* follows a normal distribution is given by

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

where $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ respectively.

Mean and Variance

Suppose $X \sim N\left(\mu, \sigma^2\right)$. Then

$$\mu_X = E(X) = \mu \quad \text{and} \quad \sigma_X^2 = Var(X) = \sigma^2.$$

4.6. The Standard Normal Distribution

A normally distributed variable having **mean 0** and **variance 1** is said to have the ***standard normal distribution***.

The standardized version of a normally distributed variable X, $Z = \dfrac{X-\mu}{\sigma}$, has the standard normal distribution. This is also called the Z *values* or Z *scores*.

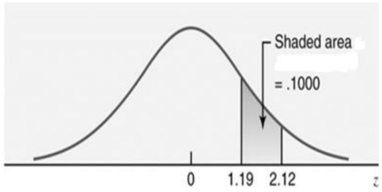
$$\therefore \text{ If } \quad X \sim N(\mu, \sigma^2)$$
$$\text{then } Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Example 4.10.

If $Z \sim N(0,\ 1)$, find

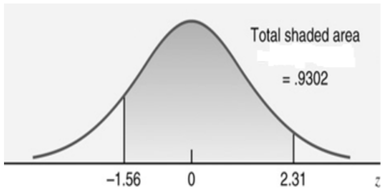
(a) $P(1.19 < Z < 2.12)$

P (1.19 < Z < 2.12) = P (Z > 1.19) – P (Z > 2.12)
= 0.1170 – 0.0170
= 0.1000



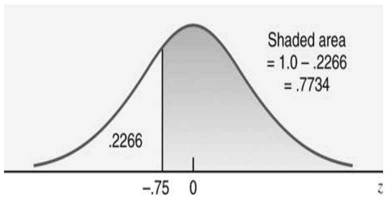
(b) $P(-1.56 < Z < 2.31)$

P (-1.56 < Z < 2.31) = 1 – P(Z < -1.56) – P(Z > 2.31)
= 1 – P(Z > 1.56) – P(Z > 2.31)
= 1 – 0.0594 – 0.01044
= 0.9302



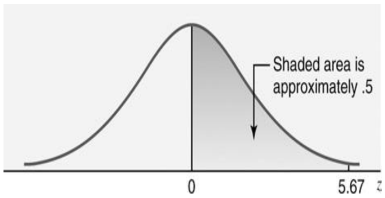
(c) $P(Z > -0.75)$

P(Z > -0.75) = 1 – P(Z < -0.75)
= 1 – P(Z > 0.75)
= 1 – 0.2266
= 0.7734



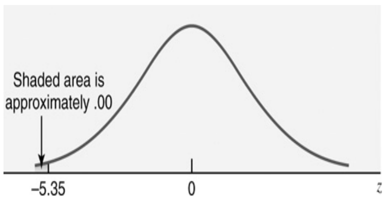
(d) $P(0 < Z < 5.67)$

P(0 < Z < 5.67) = P(Z > 0) – P(Z > 5.67)
= 0.5 – 0
= 0.5



(e) $P(Z < -5.35)$

P(Z < -5.35) = P(Z > 5.35)
= 0



(f) $P(-1.7 < Z < -0.88)$

P(-1.7 < Z < -0.88) = P(0.88 < Z < 1.7)
= P(Z > 0.88) – P(Z > 1.7)
= 0.1894 – 0.0446
= 0.1448

Example 4.11.
Suppose that the average household income in some country is 900 coins, and the standard deviation is 200 coins. Assuming the Normal distribution of incomes, compute the proportion of "the middle class," whose income is between 600 and 1200 coins.
Solution:

Example 4.12.
The government of the country in *Example 4.11.* decides to issue food stamps to the poorest 3% of households. Below what income will families receive food stamps?
Solution:

Example 4.13.
A study shown that 20% of all college textbooks have a price of RM90 or higher. It is known that the standard deviation of the prices of all college textbooks is RM9.50. Suppose the prices of all college textbooks have a normal distribution. What is the mean price of all college textbooks?
Solution:

4.7. Central Limit Theorem

Let $X_1, X_2, \ldots$ be *independent* random variables with the same expectation $\mu = E(X_i)$ and the same standard deviation $\sigma = Std(X_i)$, and let

$$S_n = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n .$$

As $n \to \infty$, the *standardized sum*

$$Z_n = \frac{S_n - E(S_n)}{Std(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a *Standard Normal* random variable, that is,

$$F_{Z_n}(z) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right) \to \Phi(z) \text{ for all } z .$$

As long as n is sufficiently large i.e. $n > 30$, one can use Normal distribution to compute probabilities about S_n.

Example 4.14.
A disk has free space of 330 megabytes. Is it likely to be sufficient for 300 independent images, if each image has expected size of 1 megabyte with a standard deviation of 0.5 megabytes?
Solution:

Example 4.15.
You wait for an elevator, whose capacity is 2000 pounds. The elevator comes with ten adult passengers. Suppose your own weight is 150 lbs, and you heard that human weights are normally distributed with the mean of 165 lbs and the standard deviation of 20 lbs. Would you board this elevator or wait for the next one?
Solution:

4.8. The Normal Approximation to the Binomial Distribution

The normal distribution can be used as an approximation to the binomial distribution when *np* and *nq* are both greater than 5.

Suppose $X \sim Bin(n, p)$. If $np > 5$ and $nq > 5$, then $X \approx N\left(\mu = np,\ \sigma^2 = npq\right)$.

An addition of 0.5 and/or subtraction of 0.5 from the value(s) of *X* when the normal distribution is used as an approximation to the binomial distribution is called the *continuity correction factor*.

Continuity Correction Factor
In general, if *a* is an integer,
(a) $P(X \geq a) = P(X > a - 0.5)$
(b) $P(X \leq a) = P(X < a + 0.5)$
(c) $P(X = a) = P(a - 0.5 < X < a + 0.5)$

Example 4.16.
A new computer virus attacks a folder consisting of 200 files. Each file gets damaged with probability 0.2 independently of other files. What is the probability that fewer than 50 files get damaged?
Solution:

4.9. The Normal Approximation to the Poisson Distribution

The normal distribution can be used as an approximation to the Poisson distribution when its mean, λ is relatively large, that is if $\lambda > 25$.

Suppose $X \sim Poi(\lambda)$. If $\lambda > 25$, then $X \approx N\left(\mu = \lambda,\ \sigma^2 = \lambda\right)$.

An addition of 0.5 and/or subtraction of 0.5 from the value(s) of X when the normal distribution is used as an approximation to the Poisson distribution is called the *continuity correction factor*.

Example 4.17.

The average number of e-mails received per week by a computer user is 26. Find the probability that there will be 26 to 28 (inclusive) e-mails received per week, using

(a) Poisson distribution,

(b) normal approximation to Poisson distribution.

Solution: