

CENG 222 - Chapter 4: Continuous Distributions

Burak Metehan Tuncel - April 2022

1 Probability Density

For all *continuous variables*, the probability mass function (pmf) is always equal to zero

$$P(x) = 0 \text{ for all } x$$

As a result, the pmf does not carry any information about a random variable. Rather, we can use the *cumulative distribution function* (cdf) $F(x)$. In the continuous case, it equals

$$F(x) = \mathbf{P}\{X \leq x\} = \mathbf{P}\{X < x\}$$

These two expression for $F(x)$ differ by $\mathbf{P}\{X = x\} = P(x) = 0$

In both continuous and discrete cases, the cdf $F(x)$ is a *non-decreasing* function that *ranges from 0 to 1*. In the discrete case, the graph of $F(x)$ has jumps of magnitude $P(x)$. For continuous distributions, $P(x) = 0$, which means no jumps. The cdf in this case is a continuous function.

Assume, additionally, that $F(x)$ has a derivative. This is the case for all commonly used continuous distributions, but in general, it is not guaranteed by continuity and monotonicity (the famous Cantor function is a counterexample).

Definition 1

Probability density function (pdf, density) is the derivative of the cdf, $f(x) = F'(x)$. The distribution is called **continuous** if it has a density.

Then, $F(x)$ is an antiderivative of a density. The integral of a density from a to b equals to the difference of antiderivatives, i.e.,

$$\int_a^b f(x)dx = F(b) - F(a) = \mathbf{P}\{a < X < b\}$$

where we notice again that the probability in the right-hand side also equals $\mathbf{P}\{a \leq X < b\}$, $\mathbf{P}\{a < X \leq b\}$, and $\mathbf{P}\{a \leq X \leq b\}$.

Probability density function

$$f(x) = F'(x)$$
$$\mathbf{P}\{a < X < b\} = \int_a^b f(x)dx$$

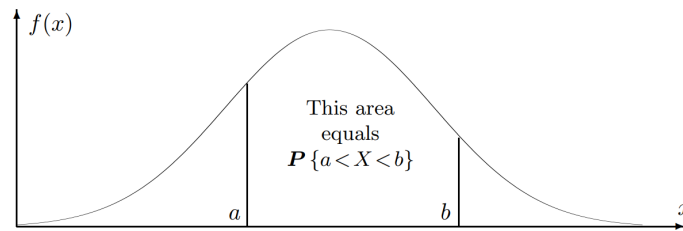


Figure 1

Thus, probabilities can be calculated by integrating a density over the given sets. Furthermore, the integral $\int_a^b f(x)dx$ equals the area below the density curve between the points a and b . Therefore, geometrically, probabilities are represented by areas (Figure 1). Substituting $a = -\infty$ and $b = +\infty$, we obtain

$$\int_{-\infty}^b f(x)dx = \mathbf{P}\{-\infty < X < b\} = F(b) \quad \text{and} \quad \int_{-\infty}^{+\infty} f(x)dx = \mathbf{P}\{-\infty < X < +\infty\} = 1$$

That is, the total area below the density curve equals 1.

Looking at Figure 1, we can see why $P(x) = 0$ for all continuous random variables. That is because

$$P(x) = \mathbf{P}\{x \leq X \leq x\} = \int_x^x f = 0$$

Geometrically, it is the area below the density curve, where two sides of the region collapse into one.

1.1 Analogy: pmf versus pdf

Distribution	Discrete	Continuous
Definition	$P(x) = \mathbf{P}\{X = x\}$	$f(x) = F'(x)$
Computing probabilities	$\mathbf{P}\{X \in A\} = \sum_{x \in A} P(x)$	$\mathbf{P}\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(x) = \mathbf{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = \mathbf{P}\{X \leq x\} = \int_{-\infty}^x f(y)dy$
Total probability	$\sum_x P(x) = 1$	$\int_{-\infty}^{+\infty} f(x)dx = 1$

Table 1: *Pmf $P(x)$ versus pdf $f(x)$*

The role of a density for continuous distributions is very similar to the role of the probability mass function for discrete distributions. Most vital concepts can be translated from the discrete case to the continuous case by replacing pmf $P(x)$ with pdf $f(x)$ and integrating instead of summing, as in Table 1.

1.2 Joint and Marginal Densities

Definition 2

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

$$F_{(X,Y)}(x,y) = \mathbf{P}\{X \leq x \cap Y \leq y\}$$

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; see Table 2.

Distribution	Discrete	Continuous
Marginal Distributions	$P(x) = \sum_y P(x,y)$ $P(y) = \sum_x P(x,y)$	$f(x) = \int f(x,y)dy$ $f(y) = \int f(x,y)dx$
Independence	$P(x,y) = P(x)P(y)$	$f(x,y) = f(x)f(y)$
Computing Probabilities	$\mathbf{P}\{(X,Y) \in A\} = \sum_{(x,y) \in A} P(x,y)$	$\mathbf{P}\{(X,Y) \in A\} = \int \int_{(x,y) \in A} f(x,y)dxdy$

Table 2: *Joint and marginal distributions in discrete and continuous cases.*

1.3 Expectation and Variance

Continuing our analogy with the discrete case, *expectation* of a continuous variable is also defined as a center of gravity,

$$\mu = \mathbf{E}(X) = \int xf(x)dx$$

This time, if the entire region below the density curve is cut from a piece of wood, then it will be balanced at a point with coordinate $\mathbf{E}(X)$, as shown in Figure 2.

Variance, *standard deviation*, *covariance*, and *correlation* of continuous variables are defined similarly to the discrete case, see Table 3. All the properties in discrete cases extend to the continuous distributions. In calculations, don't forget to replace a pmf with a pdf, and a summation with an integral.

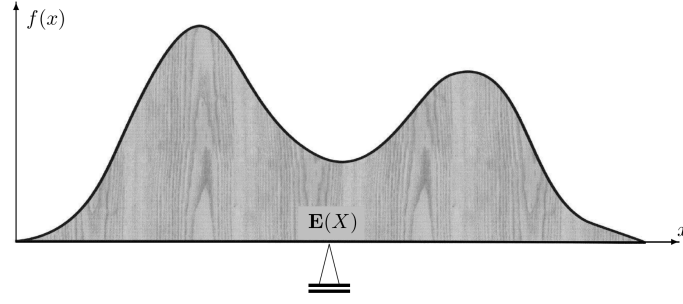


Figure 2: *Expectation of a continuous variable as a center of gravity.*

Discrete	Continuous
$\mathbf{E}(X) = \sum_x xP(x)$	$\mathbf{E}(X) = \int xf(x)dx$
$\text{Var}(X) = \mathbf{E}(X - \mu)^2$ $= \sum_x (x - \mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\text{Var}(X) = \mathbf{E}(X - \mu)^2$ $= \int (x - \mu)^2 f(x)dx$ $= \int x^2 f(x)dx - \mu^2$
$\text{Cov}(X, Y) = \mathbf{E}(X - \mu_X)(Y - \mu_Y)$ $= \sum_x \sum_y (x - \mu_X)(y - \mu_Y)P(x, y)$ $= \sum_x \sum_y (xy)P(x, y) - \mu_X \mu_Y$	$\text{Cov}(X, Y) = \mathbf{E}(X - \mu_X)(Y - \mu_Y)$ $= \int \int (x - \mu_X)(y - \mu_Y)f(x, y)dxdy$ $= \int \int (xy)f(x, y)dxdy - \mu_X \mu_Y$

Table 3: *Moments for discrete and continuous distributions.*

Example 1

The lifetime, in years, of some electronic component is a continuous random variable with the density

$$f(x) = \begin{cases} \frac{k}{x^3} & \text{for } x \geq 1 \\ 0 & \text{for } x < 1 \end{cases}$$

Find k , draw a graph of the cdf $F(x)$, and compute the probability for the lifetime to exceed 5 years. Then compute expectation and variance.

Find k from the condition $\int f(x)dx = 1$:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_1^{+\infty} \frac{k}{x^3}dx = -\frac{k}{2x^2} \Big|_{x=1}^{+\infty} = \frac{k}{2} = 1$$

Hence, $k = 2$. Integrating the density, we get the cdf,

$$F(x) = \int_{-\infty}^x f(y)dy = \int_1^x \frac{2}{y^3}dy - \frac{1}{y^2} \Big|_{y=1}^x = 1 - \frac{1}{x^2} \quad \text{for } x > 1$$

Next, compute the probability for the lifetime to exceed 5 years,

$$\mathbf{P}\{X > 5\} = 1 - F(5) = 1 - \left(1 - \frac{1}{5^2}\right) = 0.04$$

We can also obtain this probability by integrating the density,

$$\mathbf{P}\{X > 5\} = \int_5^{+\infty} f(x)dx = \int_5^{+\infty} \frac{2}{x^3}dx = -\frac{1}{x^2} \Big|_{x=5}^{+\infty} = \frac{1}{25} = 0.04$$

Its expectation equals

$$\mu = \mathbf{E}(X) = \int xf(x)dx = \int_1^{\infty} 2x^{-2}dx = -2x^{-1} \Big|_1^{\infty} = 2$$

Computing its variance, we run into a “surprise”: This variable does not have a finite variance!

$$\sigma^2 = \text{Var}(X) = \int x^2 f(x)dx - \mu^2 = \int_1^{\infty} 2x^{-1}dx - 4 = 2 \ln x \Big|_1^{\infty} - 4 = +\infty$$

2 Families of Continuous Distributions

As in the discrete case, varieties of phenomena can be described by relatively few families of continuous distributions. Here, we shall discuss **Uniform**, **Exponential**, **Gamma**, and **Normal** families, adding Student's t , Pearson's χ^2 , and Fisher's F distributions in later chapters.

2.1 Uniform Distribution

Uniform distribution plays a unique role in stochastic modeling. A random variable with any thinkable distribution can be generated from a Uniform random variable. Many computer languages and software are equipped with a random number generator that produces Uniform random variables. Users can convert them into variables with desired distributions and use for computer simulation of various events and processes.

Also, Uniform distribution is used in any situation when a value is picked “at random” from a given interval; that is, without any preference to lower, higher, or medium values. For example, locations of errors in a program, birthdays throughout a year, and many continuous random variables modulo 1, modulo 0.1, 0.01, etc., are uniformly distributed over their corresponding intervals.

To give equal preference to all values, the Uniform distribution has a *constant density* (Figure 3). On the interval (a, b) , its density equals

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

because the rectangular area below the density graph must equal 1.

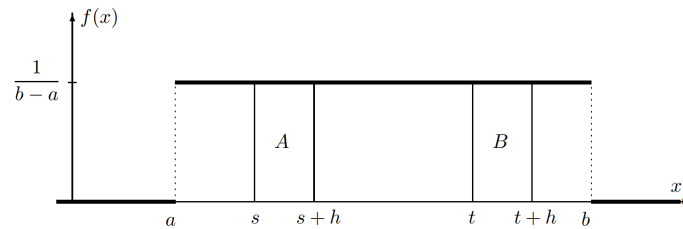


Figure 3: *The Uniform density and the Uniform property.*

For the same reason, $|b-a|$ has to be a finite number. There does not exist a Uniform distribution on the entire real line. In other words, if you are asked to choose a random number from $(-\infty, +\infty)$, you cannot do it uniformly.

2.1.1 The Uniform Property

For any $h > 0$ and $t \in [a, b-h]$, the probability

$$P\{t < X < t+h\} = \int_t^{t+h} \frac{1}{b-a} dx = \frac{h}{b-a}$$

is *independent* of t . This is the **Uniform property**: the probability is only determined by the length of the interval, but not by its location.

Example 2

In Figure 3, rectangles A and B have the same area, showing that

$$P\{s < X < s+h\} = P\{t < X < t+h\}$$

2.1.2 Standard Uniform Distribution

The Uniform distribution with $a = 0$ and $b = 1$ is called *Standard Uniform distribution*. The Standard Uniform density is $f(x) = 1$ for $0 < x < 1$. Most random number generators return a Standard Uniform random variable.

All the Uniform distributions are related in the following way. If X is a Uniform(a, b) random variable (*not standard uniform random variable*), then

$$Y = \frac{X-a}{b-a}$$

is Standard Uniform (If X is not inside $(0,1)$, “ $X-a$ ” slides X to the $(0,1)$). Likewise, if Y is Standard Uniform, then

$$X = a + (b-a)Y$$

is Uniform(a, b). Check that $X \in (a, b)$ if and only if $Y \in (0, 1)$.

A number of other families of distributions have a “standard” member. Typically, a simple transformation converts a standard random variable into a non-standard one, and vice versa.

2.1.3 Expectation and Variance

For a Standard Uniform variable Y ,

$$\mathbf{E}(Y) = \int_0^1 yf(y)dy = \int_0^1 ydy = \frac{1}{2}$$

and

$$\text{Var}(Y) = \mathbf{E}(Y^2) - \mathbf{E}^2(Y) = \int_0^1 y^2 dy - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Now, consider the general case (*Case of not being the standard uniform*). Let $X = a + (b - a)Y$ which has a $\text{Uniform}(a, b)$ distribution. By the properties of expectations and variances,

$$\mathbf{E}(X) = \mathbf{E}(a + (b - a)Y) = a + (b - a)\mathbf{E}(Y) = a + \frac{b - a}{2} = \frac{a + b}{2}$$

and

$$\text{Var}(X) = \text{Var}(a + (b - a)Y) = (b - a)^2 \text{Var}(Y) = \frac{(b - a)^2}{12}$$

The expectation is precisely the middle of the interval $[a, b]$. Giving no preference to left or right sides, this agrees with the Uniform property and with the physical meaning of $\mathbf{E}(X)$ as a center of gravity.

2.1.4 Uniform Distribution Functions and Variables

Uniform Distribution

(a, b) = range of values

$$f(x) = \frac{1}{b - a}, \quad a < x < b$$

$$\mathbf{E}(X) = \frac{a + b}{2}$$

$$\text{Var}(X) = \frac{(b - a)^2}{12}$$

2.2 Exponential Distribution

Exponential distribution is often used to model *time*: waiting time, interarrival time, hardware lifetime, failure time, time between telephone calls, etc. As we shall see below, in a sequence of rare events, *when the number of events is Poisson, the time between events is Exponential*.

Exponential distribution has density

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0 \quad (1)$$

With this density, we compute the Exponential cdf, mean, and variance as

$$\mathbf{P}\{X \leq x\} = F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad (x > 0) \quad (2)$$

$$\mathbf{E}(X) = \int_0^\infty t f(t)dt = \int_0^\infty t \lambda e^{-\lambda t} dt = \frac{1}{\lambda} \quad (\text{integral by parts}) \quad (3)$$

$$\begin{aligned} \text{Var}(X) &= \int_0^\infty t^2 f(t)dt - \mathbf{E}^2(X) \\ &= \int_0^\infty t^2 \lambda e^{-\lambda t} dt - \left(\frac{1}{\lambda}\right)^2 \quad (\text{by parts twice}) \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \end{aligned} \quad (4)$$

The quantity λ is a parameter of Exponential distribution, and its meaning is clear from $\mathbf{E}(X) = 1/\lambda$. If X is time, measured in minutes, then λ is a frequency, measured in min^{-1} . For example, if arrivals occur every half a minute, on the average, then $\mathbf{E}(X) = 0.5$ and $\lambda = 2$, saying that they occur with a frequency (arrival rate) of 2 arrivals per minute. This λ has the same meaning as the parameter of Poisson distribution.

Important Note

$$P\{X > x\} = 1 - P\{X \leq x\} = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

In here X is “time between two events” and x is “particular time”. So, if this particular time, x , is greater than or equal to time between two events, X , the eq. 2 is used. Otherwise, the equation inside this box is used.

“**Exponential distribution** is a continuous version of **Geometric distribution**. In the *Geometric distribution*, we analyze the how many trials are required before the first success, whereas, in the *Exponential distribution*, we analyze the time until the first success.”

2.2.1 Times between rare events are Exponential

What makes Exponential distribution a good model for interarrival times? Apparently, this is not only experimental, but also a mathematical fact.

As in Chapter 3, consider the sequence of rare events, where the number of occurrences during time t has Poisson distribution with a parameter proportional to t .

Event “the time T until the next event is greater than t ” can be rephrased as “zero events occur by the time t ”, and further, as “ $X = 0$ ”, where X is the number of events during the time interval $[0, t]$ (*It is related with the note in previous page*). This X has Poisson distribution with parameter λt . It equals 0 with probability

$$P_X(0) = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

Then we can compute the cdf of T as

$$F_T(t) = 1 - P\{T > t\} = 1 - P\{X = 0\} = 1 - e^{-\lambda t} \quad (5)$$

and here we recognize the Exponential cdf. Therefore, the time until the next arrival has Exponential distribution.

Example 3

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: Job arrivals represent rare events, thus the time T between them is Exponential with the given parameter $\lambda = 3 \text{ hrs}^{-1}$ (jobs per hour).

(a) $E(T) = 1/\lambda = 1/3$ hours or 20 minutes between jobs.

(b) Convert to the same measurement unit: 5 min = $(1/12)$ hrs. Then,

$$P\{T < 1/12 \text{ hrs}\} = F(1/12) = 1 - e^{-\lambda(1/12)} = 1 - e^{-1/4} = 0.2212$$

2.2.2 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,

$$P\{T > t + x \mid T > t\} = P\{T > x\} \quad \text{for } t, x > 0 \quad (6)$$

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

Proof. $P\{T > x\} = e^{-\lambda x}$. Also, by the formula for conditional probability,

$$P\{T > t + x \mid T > t\} = \frac{P\{T > t + x \cap T > t\}}{P\{T > t\}} = \frac{P\{T > t + x\}}{P\{T > t\}} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}$$

□

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.

2.2.3 Exponential Distribution Functions and Variables

Exponential Distribution

λ = frequency parameter, the number of events per time unit

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

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

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

2.3 Gamma Distributinon

When a certain procedure consists of α independent steps, and each step takes $\text{Exponential}(\lambda)$ 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.

Example 4: Internet Promotions

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

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0 \quad (7)$$

The denominator contains a Gamma function, later. With certain techniques, this density can be mathematically derived for integer α by representing a Gamma variable X as a sum of Exponential variables each having a density (eq. 1).

In fact, α can take *any positive value*, not necessarily integer. With different α , the Gamma density takes different shapes (Figure 4.5). For this reason, α is called a *shape parameter*.

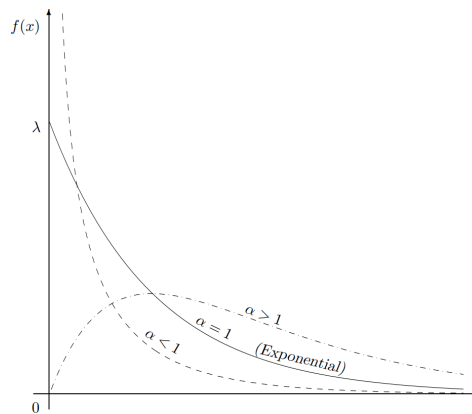


Figure 4: Gamma densities with different shape parameters α .

Notice two important special cases of a Gamma distribution. When $\alpha = 1$, the Gamma distribution becomes Exponential. This can be seen comparing (eq. 7) and (eq. 1) for $\alpha = 1$. Another special case with $\lambda = 1/2$ and any $\alpha > 0$ results in a so-called *Chi-square distribution* with (2α) degrees of freedom.

Special Cases

$$\text{Gamma}(1, \lambda) = \text{Exponential}(\lambda)$$

$$\text{Gamma}(\alpha, 1/2) = \text{Chi-square}(2\alpha)$$

2.3.1 Expectation, Variance, and some useful integration remarks

Gamma cdf has the form

$$F(t) = \int_0^t f(x)dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^t x^{\alpha-1} e^{-\lambda x} dx \quad (8)$$

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.

First, let us notice that $\int_0^\infty f(x)dx = 1$ for Gamma and all the other densities. Then, integrating (eq. 7) from 0 to ∞ , we obtain that

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha} \quad \text{for any } \alpha > 0 \text{ and } \lambda > 0 \quad (9)$$

Substituting $\alpha + 1$ and $\alpha + 2$ in place of α , we get for a Gamma variable X

$$\mathbf{E}(X) = \int x f(x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha + 1)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda} \quad (10)$$

(using the equality $\Gamma(t + 1) = t\Gamma(t)$ that holds for all $t \geq 0$),

$$\mathbf{E}(X^2) = \int x^2 f(x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha + 2)}{\lambda^{\alpha+2}} = \frac{(\alpha + 1)\alpha}{\lambda^2}$$

and therefore,

$$\text{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X) = \frac{(\alpha + 1)\alpha - \alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2} \quad (11)$$

For $\alpha = 1$, this agrees with (eq. 3) and (eq. 4). Moreover, for any integer α , (eq. 10) and (eq. 11) can be obtained directly from (eq. 3) and (eq. 4) by representing a Gamma variable X as a sum of independent $\text{Exponential}(\lambda)$ variables X_1, \dots, X_α ,

$$\mathbf{E}(X) = \mathbf{E}(X_1 + \dots + X_\alpha) = \mathbf{E}(X_1) + \dots + \mathbf{E}(X_\alpha) = \alpha \left(\frac{1}{\lambda} \right)$$

$$\text{Var}(X) = \text{Var}(X_1 + \dots + X_\alpha) = \text{Var}(X_1) + \dots + \text{Var}(X_\alpha) = \alpha \left(\frac{1}{\lambda^2} \right)$$

2.3.2 Gamma Distribution Functions and Variables

Gamma Distribution: Eq. 12

α = shape parameter

λ = frequency parameter

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

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

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

Example 5: Total compilation time

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: The total time T is a sum of three independent Exponential times, therefore, it has Gamma distribution with $\alpha = 3$. The frequency parameter $\lambda = (1/5) \text{ min}^{-1}$ because the Exponential compilation time of each block has expectation $1/\lambda = 5 \text{ min}$.

(a) For a Gamma random variable T with $\alpha = 3$ and $\lambda = 1/5$,

$$\mathbf{E}(T) = \frac{3}{1/5} = 15 \text{ (min)} \quad \text{and} \quad \text{Var}(T) = \frac{3}{(1/5)^2} = 75 \text{ (min}^2\text{)}$$

(b) A direct solution involves two rounds of integration by parts,

$$\begin{aligned} \mathbf{P}\{T < 12\} &= \int_0^{12} f(t)dt = \frac{(1/5)^3}{\Gamma(3)} \int_0^{12} t^2 e^{-t/5} dt \\ &= \frac{(1/5)^3}{2!} \left(-5t^2 e^{-t/5} \Big|_{t=0}^{t=12} + \int_0^{12} 10te^{-t/5} dt \right) \\ &= \frac{(1/125)}{2} \left(-5t^2 e^{-t/5} - 50te^{-t/5} \Big|_{t=0}^{t=12} + \int_0^{12} 50e^{-t/5} dt \right) \\ &= \frac{1}{250} \left(-5t^2 e^{-t/5} - 50te^{-t/5} - 250e^{-t/5} dt \right) \Big|_{t=0}^{t=12} \\ &= 1 - e^{-2.4} - 2.4e^{-2.4} - 2.88e^{-2.4} = 0.4303 \end{aligned} \tag{13}$$

A much shorter way is to apply the Gamma-Poisson formula below (example 6).

2.3.3 Gamma-Poisson Formula

Computation of Gamma probabilities can be significantly simplified by thinking of a Gamma variable as the time between some rare events. In particular, one can avoid lengthy integration by parts, as in Example 5, and use Poisson distribution instead.

Indeed, let T be a Gamma variable with an integer parameter α and some positive λ . This is a distribution of the time of the α -th rare event. Then, the event $\{T > t\}$ means that the α -th rare event occurs after the moment t , and therefore, *fewer than α rare events occur before the time t* . We see that

$$\{T > t\} = \{X < \alpha\}$$

where X is the number of events that occur before the time t . This number of rare events X has Poisson distribution with parameter (λt) ; therefore, the probability

$$\mathbf{P}\{T > t\} = \mathbf{P}\{X < \alpha\}$$

and the probability of a complement

$$\mathbf{P}\{T \leq t\} = \mathbf{P}\{X \geq \alpha\}$$

can both be computed using the Poisson distribution of X .

2.3.4 Gamma-Poisson Formula Functions

Gamma-Poisson Formula: Eq. 14

For a $\text{Gamma}(\alpha, \lambda)$ variable T and a $\text{Poisson}(\lambda t)$ variable X ,

$$\begin{aligned} \mathbf{P}\{T > t\} &= \mathbf{P}\{X < \alpha\} \\ \mathbf{P}\{T \leq t\} &= \mathbf{P}\{X \geq \alpha\} \end{aligned}$$

Remark: Recall that $\mathbf{P}\{T > t\} = \mathbf{P}\{T \geq t\}$ and $\mathbf{P}\{T < t\} = \mathbf{P}\{T \leq t\}$ for a Gamma variable T , because it is continuous. Hence, (eq. 14) can also be used for the computation of $\mathbf{P}\{T \geq t\}$ and $\mathbf{P}\{T < t\}$. Conversely, the probability of $\{X = \alpha\}$ cannot be neglected for the Poisson (discrete!) variable X , thus the signs in the right-hand sides of (eq. 14) cannot be altered.

Example 6: Total compilation time, continued

Here is an alternative solution to Example 5(b). According to the Gamma-Poisson formula with $\alpha = 3$, $\lambda = 1/5$, and $t = 12$,

$$\mathbf{P}\{T < 12\} = \mathbf{P}\{X \geq 3\} = 1 - F(2) = 1 - 0.5697 = 0.430$$

from Table A3 in book, for the Poisson distribution of X with parameter $\lambda t = 2.4$.

Furthermore, we notice that the four-term mathematical expression that we obtained in (eq. 13) after integrating by parts represents precisely $\mathbf{P}\{X \geq 3\} = 1 - P(0) - P(1) - P(2)$.

Example 7

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:

STEP 1, PARAMETERS. From the given data, compute parameters of this Gamma distribution. Using (eq. 12), obtain a system of two equations and solve them for α and λ ,

$$\begin{cases} \mu = \alpha/\lambda \\ \sigma^2 = \alpha/\lambda^2 \end{cases} \Rightarrow \begin{cases} \alpha = \mu^2/\sigma^2 = (12/4)^2 = 9 \\ \lambda = \mu/\sigma^2 = 15/4^2 = 0.75 \end{cases}$$

STEP 2, PROBABILITY. We can now compute the probability,

$$\mathbf{P}\{8 < T < 10\} = F_T(10) - F_T(8) \quad (15)$$

For each term in (eq. 15), we use the Gamma-Poisson formula with $\alpha = 9$, $\lambda = 0.75$, and $t = 8, 10$,

$$F_T(10) = \mathbf{P}\{T \leq 10\} = \mathbf{P}\{X \geq 9\} = 1 - F_X(8) = 1 - 0.662 = 0.338$$

from Table A3 in book, for a Poisson variable X with parameter $\lambda t = (0.75)(10) = 7.5$;

$$F_T(8) = \mathbf{P}\{T \leq 8\} = \mathbf{P}\{X \geq 9\} = 1 - F_X(8) = 1 - 0.847 = 0.153$$

from Table A3 in book, for a Poisson variable X with parameter $\lambda t = (0.75)(8) = 6$. then

$$\mathbf{P}\{8 < T < 10\} = 0.338 - 0.153 = 0.185$$

2.4 Normal Distribution

Normal distribution plays a vital role in Probability and Statistics, mostly because of the **Central Limit Theorem**, according to which sums and averages often have approximately Normal distribution. Due to this fact, various fluctuations and measurement errors that consist of accumulated number of small terms appear normally distributed.

Besides sums, averages, and errors, Normal distribution is often found to be a good model for *physical variables like weight, height, temperature, voltage, pollution level, and for instance, household incomes or student grades.*

Normal distribution has a density

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

where parameters μ and σ have a simple meaning of the expectation $\mathbf{E}(X)$ and the standard deviation $\text{Std}(X)$. This density is known as the bell-shaped curve, symmetric and centered at μ , its spread being controlled by σ . As seen in Figure 5, changing μ shifts the curve to the left or to the right without affecting its shape, while changing σ makes it more concentrated or more flat. Often μ and σ are called *location* and *scale* parameters.

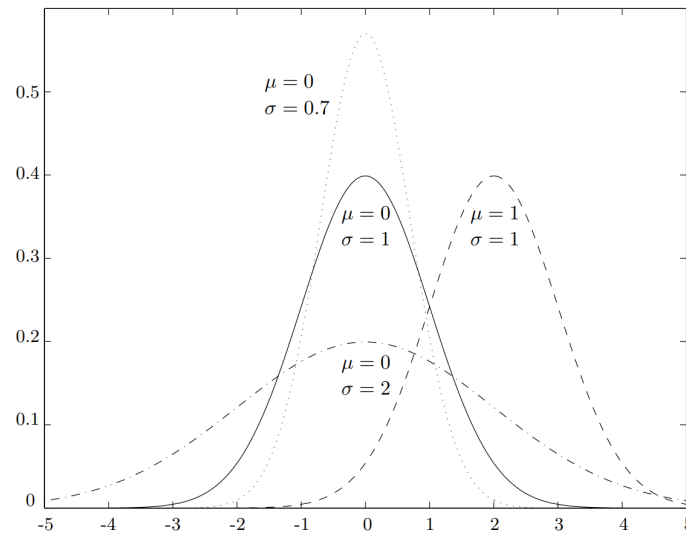


Figure 5: Normal densities with different location and scale parameters.

2.4.1 Normal Distribution Functions and Variables

μ = expectation, *location* parameter

σ = standard deviation, *scale* parameter

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

$$\mathbf{E}(X) = \mu$$

$$\text{Var}(X) = \sigma^2$$

2.4.2 Standard Normal Distribution

Definition 3

Normal distribution with “standard parameters” $\mu = 0$ and $\sigma = 1$ is called **Standard Normal distribution**.

NOTATION

Z = Standard Normal random variable

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{ Standard Normal pdf}$$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \text{ Standard Normal cdf}$$

A Standard Normal variable, usually denoted by Z , can be obtained from a non-standard Normal(μ, σ) random variable X by *standardizing*, that is, subtracting the mean and dividing by the standard deviation,

$$Z = \frac{X - \mu}{\sigma} \quad (16)$$

Unstandardizing Z , we can reconstruct the initial variable X ,

$$X = \mu + \sigma Z \quad (17)$$

Using these transformations, any Normal random variable can be obtained from a Standard Normal variable Z ; therefore, we need a table of Standard Normal Distribution only (Table A4 in book).

Example 8: Computing non-standard Normal probabilities

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: Standardize and use Table A4. For a Normal($\mu = 900, \sigma = 200$) variable X ,

$$\begin{aligned} P\{600 < X < 1200\} &= P\left\{ \frac{600 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{1200 - \mu}{\sigma} \right\} \\ &= P\left\{ \frac{600 - 900}{200} < Z < \frac{1200 - 900}{200} \right\} = P\{-1.5 < Z < 1.5\} \\ &= \Phi(1.5) - \Phi(-1.5) = 0.9332 - 0.0668 = 0.8664 \end{aligned}$$

So far, we were computing probabilities of clearly defined events. These are direct problems. A number of applications require solution of an inverse problem, that is, finding a value of x given the corresponding probability.

Example 9: Inverse problem

The government of the country in Example 8 decides to issue food stamps to the poorest 3% of households. Below what income will families receive food stamps?

Solution: We need to find such income x that $P\{X < x\} = 3\% = 0.03$. This is an equation that can be solved in terms of x . Again, we standardize first, then use the table:

$$P\{X < x\} = P\left\{ Z < \frac{x - \mu}{\sigma} \right\} = \Phi\left(\frac{x - \mu}{\sigma} \right) = 0.03,$$

from where

$$x = \mu + \sigma \Phi^{-1}(0.03)$$

In Table A4, we have to find the probability, the table entry of 0.03. We see that $\Phi(-1.88) \approx 0.03$. Therefore, $\Phi^{-1}(0.03) = -1.88$, and

$$x = \mu + \sigma(-1.88) = 900 + (200)(-1.88) = 524 \text{ (coins)}$$

is the answer. In the literature, the value $\Phi^{-1}(\alpha)$ is often denoted by $z_{1-\alpha}$.

As seen in this example, in order to solve an inverse problem, we use the table first, then unstandardize, as in (eq. 17), and find the required value of x .

3 Central Limit Theorem

We now turn our attention to sums of random variables,

$$S_n = X_1 + \dots + X_n,$$

that appear in many applications. Let $\mu = \mathbf{E}(X_i)$ and $\sigma = \text{Std}(X_i)$ for all $i = 1, \dots, n$. How does S_n behave for large n ?

- The pure sum S_n diverges. In fact, this should be anticipated because

$$\text{Var}(S_n) = n\sigma^2 \rightarrow \infty$$

so that variability of S_n grows unboundedly as n goes to infinity.

- The average $\frac{S_n}{n}$ converges. Indeed, in this case, we have

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{\text{Var}(S_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \rightarrow 0$$

so that variability of $\frac{S_n}{n}$ vanishes as $n \rightarrow \infty$.

- An interesting normalization factor is $\frac{1}{\sqrt{n}}$. For $\mu = 0$, $\frac{S_n}{\sqrt{n}}$ *neither diverges nor converges!* It does not tend to leave 0, but it does not converge to 0 either. Rather, it behaves like some random variable. The following theorem states that this variable has approximately Normal distribution for large n .

Theorem 1: Central Limit Theorem

Let X_1, X_2, \dots be independent random variables with the same expectation $\mu = \mathbf{E}(X_i)$ and the same standard deviation $\sigma = \text{Std}(X_i)$, and let

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

As $n \rightarrow \infty$, the standardized sum

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

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

$$F_{Z_n}(z) = \mathbf{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right\} \rightarrow \Phi(z) \quad (18)$$

for all z .

This theorem is very powerful because it can be applied to random variables X_1, X_2, \dots having virtually any thinkable distribution with finite expectation and variance. As long as n is large (the rule of thumb is $n > 30$), one can use Normal distribution to compute probabilities about S_n .

Theorem 1 is only one basic version of the **Central Limit Theorem**. Over the last two centuries, it has been extended to large classes of dependent variables and vectors, stochastic processes, and so on.

Example 10: Allocation of disk space

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: We have $n = 300$, $\mu = 1$, $\sigma = 0.5$. The number of images n is large, so the Central Limit Theorem applies to their total size S_n . Then,

$$\begin{aligned} P\{\text{sufficient space}\} &= P\{S_n \leq 330\} = P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{330 - (300)(1)}{0.5\sqrt{300}}\right\} \\ &\approx \Phi(3.46) = 0.9997 \end{aligned}$$

This probability is very high, hence, the available disk space is very likely to be sufficient.

In the special case of Normal variables X_1, X_2, \dots , the distribution of S_n is always Normal, and (eq. 18) becomes exact equality for arbitrary, even small n .

Example 11: Elevator

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: In other words, is overload likely? The probability of an overload equals

$$\begin{aligned} P\{S_{10} + 150 > 2000\} &= P\left\{\frac{S_{10} - (10)(165)}{20\sqrt{10}} > \frac{2000 - 150 - (10)(165)}{20\sqrt{10}}\right\} \\ &= 1 - \Phi(3.16) = 0.0008 \end{aligned}$$

So, with probability 0.9992 it is safe to take this elevator. It is now for you to decide.

Among the random variables discussed in Chapters 3 and 4, at least three have a form of S_n :

Binomial variable	=	sum of independent Bernoulli variables
Negative Binomial variable	=	sum of independent Geometric variables
Gamma variable	=	sum of independent Exponential variables

Hence, the Central Limit Theorem applies to all these distributions with sufficiently large n in the case of Binomial, k for Negative Binomial, and α for Gamma variables.

3.1 Normal Approximation to Binomial Distribution

Binomial variables represent a special case of $S_n = X_1 + \dots + X_n$, where all X_i have Bernoulli distribution with some parameter p . We know that small p allows to approximate Binomial distribution with Poisson, and large p allows such an approximation for the number of failures. For the moderate values of p (say, $0.05 \leq p \leq 0.95$) and for large n , we can use Theorem 1:

$$\text{Binomial}(n, p) \approx \text{Normal}\left(\mu = np, \sigma = \sqrt{np(1-p)}\right) \quad (19)$$

3.2 Continuity Correction

This correction is needed when we approximate a discrete distribution (Binomial in this case) by a continuous distribution (Normal). Recall that the probability $P\{X = x\}$ may be positive if X is discrete, whereas it is always 0 for continuous X . Thus, a direct use of (eq. 19) will always approximate this probability by 0. It is obviously a poor approximation.

This is resolved by introducing a **continuity correction**. Expand the interval by 0.5 units in each direction, then use the Normal approximation. Notice that

$$P_X(x) = P\{X = x\} = P\{x - 0.5 < X < x + 0.5\}$$

is true for a Binomial variable X ; therefore, the continuity correction does not change the event and preserves its probability. It makes a difference for the Normal distribution, so every time when we approximate some discrete distribution with some continuous distribution, we should be using a continuity correction. Now it is the probability of an interval instead of one number, and it is not zero.

Example 12

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: The number X of damaged files has Binomial distribution with $n = 200$, $p = 0.2$, $\mu = np = 40$, and $\sigma = \sqrt{np(1-p)} = 5.657$. Applying the Central Limit Theorem with the continuity correction,

$$\begin{aligned} P\{X < 50\} &= P\{X < 49.5\} = P\left\{\frac{X - 40}{5.657} < \frac{49.5 - 40}{5.657}\right\} \\ &= \Phi(1.68) = 0.9535 \end{aligned}$$

Notice that the properly applied continuity correction replaces 50 with 49.5, not 50.5. Indeed, we are interested in the event that X is strictly less than 50. This includes all values up to 49 and corresponds to the interval $[0, 49]$ that we expand to $[0, 49.5]$. In other words, events $\{X < 50\}$ and $\{X < 49.5\}$ are the same; they include the same possible values of X . Events $\{X < 50\}$ and $\{X < 50.5\}$ are different because the former includes $X = 50$, and the latter does not. Replacing $\{X < 50\}$ with $\{X < 50.5\}$ would have changed its probability and would have given a wrong answer.

When a continuous distribution (say, Gamma) is approximated by another continuous distribution (Normal), the continuity correction is not needed. In fact, it would be an error to use it in this case because it would no longer preserve the probability.

4 Summary and Conclusion

- Continuous distributions are used to model various times, sizes, measurements, and all other random variables that assume an entire interval of possible values.
- Continuous distributions are described by their densities that play a role analogous to probability mass functions of discrete variables. Computing probabilities essentially reduces to integrating a density over the given set. Expectations and variances are defined similarly to the discrete case, replacing a probability mass function by a density and summation by integration.
- In different situations, one uses Uniform, Exponential, Gamma, or Normal distributions.
- The Central Limit Theorem states that a standardized sum of a large number of independent random variables is approximately Normal, thus Table A4 can be used to compute related probabilities. A continuity correction should be used when a discrete distribution is approximated by a continuous distribution.