CS 70 Discrete Mathematics and Probability Theory
Fall 2025 Course Notes **Note 23**

Continuous Probability Distributions

Up to now we have focused exclusively on *discrete* sample spaces Ω, where the number of sample points $\omega \in \Omega$ is either finite or countably infinite (such as the integers). As a consequence, we have only been able to talk about *discrete* random variables, which take on only a finite or countably infinite number of values.

But in real life many quantities that we wish to model probabilistically are *real-valued*; examples include the position of a particle in a box, the time at which a certain incident happens, or the direction of travel of a meteorite. In this note, we discuss how to extend the concepts we've seen in the discrete setting to this *continuous* setting. As we shall see, everything translates in a natural way once we have set up the right framework. The framework involves some elementary calculus but (at this level of discussion) nothing too daunting.

1 Continuous Uniform Probability Space

Suppose we spin a "wheel of fortune" and record the position of the pointer on the outer circumference of the wheel. Assuming that the circumference is of length ℓ and that the wheel is unbiased, the position is presumably equally likely to take on any value in the real interval $[0,\ell]$.[1] How do we model this experiment using a probability space?

Consider for a moment the analogous discrete setting, where the pointer can stop only at a finite number m of positions distributed evenly around the wheel. (If m is very large, then this is in some sense similar to the continuous setting, which we can think of as the limit $m \to \infty$.) Then we would model this situation using the discrete sample space $\Omega = \{0, \frac{\ell}{m}, \frac{2\ell}{m}, \dots, \frac{(m-1)\ell}{m}\}$, with uniform probabilities $\mathbb{P}[\omega] = \frac{1}{m}$ for each $\omega \in \Omega$.

In the continuous setting, however, we get into trouble if we try the same approach. If we let ω range over all real numbers in $\Omega = [0,\ell]$, what value should we assign to each $\mathbb{P}[\omega]$? By uniformity, this probability should be the same for all ω. But if we assign $\mathbb{P}[\omega]$ to be any positive value, then because there are infinitely many ω in Ω, the sum of all probabilities $\mathbb{P}[\omega]$ will be ∞! Thus, $\mathbb{P}[\omega]$ must be zero for all $\omega \in \Omega$. But if all of our sample points have probability zero, then we are unable to assign meaningful probabilities to any events!

To resolve this problem, consider instead any *interval* $[a,b] \subseteq [0,\ell]$, where $b > a$. Can we assign a non-zero probability value to this interval? Since the total probability assigned to $[0,\ell]$ must be 1, and since we want our probability to be uniform, the logical value for the probability of interval $[a,b]$ is

$$\frac{\text{length of } [a,b]}{\text{length of } [0,\ell]} = \frac{b-a}{\ell}.$$

In other words, the probability of an interval is proportional to its length.

Note that intervals are subsets of the sample space Ω and are therefore events. So in contrast to discrete probability, where we assigned probabilities to *points* in the sample space, in continuous probability we

[1] As we shall see shortly, it doesn't matter whether we consider the closed interval $[0,\ell]$ or the half-open interval $[0,\ell)$.

are assigning probabilities to certain basic *events* (in this case intervals). What about probabilities of other events? By specifying the probability of intervals, we have also specified the probability of any event E which can be written as the disjoint union of (a finite or countably infinite number of) intervals, $E = \cup_i E_i$. For then we can write $\mathbb{P}[E] = \sum_i \mathbb{P}[E_i]$, in analogous fashion to the discrete case. Thus for example the probability that the pointer ends up in the first or third quadrants of the wheel is $\frac{\ell/4}{\ell} + \frac{\ell/4}{\ell} = \frac{1}{2}$. For all practical purposes, such events are all we really need.[2]

2 Continuous Random Variables

Recall that in the discrete setting we typically work with *random variables* and their distributions, rather than directly with probability spaces and events. The simplest example of a continuous random variable is the position X of the pointer in the wheel of fortune, as discussed above. This random variable has the *uniform* distribution on $[0,\ell]$. How, precisely, should we define the distribution of a continuous random variable? In the discrete case the distribution of a r.v. X is described by specifying, for each possible value a, the probability $\mathbb{P}[X = a]$. But for the r.v. X corresponding to the position of the pointer, we have $\mathbb{P}[X = a] = 0$ for every a, so we run into the same problem as we encountered above in defining the probability space.

The resolution is the same: instead of specifying $\mathbb{P}[X = a]$, we specify $\mathbb{P}[a \le X \le b]$ for all intervals $[a,b]$.[3] To do this formally, we need to introduce the concept of a *probability density function* (sometimes referred to just as a “density", or a “p.d.f.").

Definition 23.1 (Probability Density Function)**.** *A* probability density function *(p.d.f.) for a real-valued random variable X is a function $f : \mathbb{R} \to \mathbb{R}$ satisfying:*

1. *f is non-negative: $f(x) \ge 0$ for all $x \in \mathbb{R}$.*
2. *The total integral of f is equal to 1: $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.*

Then the distribution of X is given by:

$$\mathbb{P}[a \le X \le b] = \int_a^b f(x)\,\mathrm{d}x \qquad \textit{for all } a < b.$$

Let us examine this definition. Note that the definite integral is just the area under the curve f between the values a and b. Thus f plays a similar role to the “histogram" we sometimes draw to picture the distribution of a discrete random variable. The first condition that f be non-negative ensures that the probability of every event is non-negative. The second condition that the total integral of f equal to 1 ensures that it defines a valid probability distribution, because the r.v. X must take on real values:

$$\mathbb{P}[X \in \mathbb{R}] = \mathbb{P}[-\infty < X < \infty] = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1. \tag{1}$$

For example, consider the wheel-of-fortune r.v. X, which has uniform distribution on the interval $[0,\ell]$. This means the density f of X vanishes outside this interval: $f(x) = 0$ for $x < 0$ and for $x > \ell$. Within the interval $[0,\ell]$ we want the distribution of X to be uniform, which means we should take f to be a constant $f(x) = c$ for $0 \le x \le \ell$. The value of c is determined by the requirement in (1) that the total area under f is 1:

$$1 = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^{\ell} c\,\mathrm{d}x = c\ell,$$

[2] A formal treatment of which events can be assigned a well-defined probability requires a discussion of *measure theory*, which is beyond the scope of this course.

[3] Note that it does not matter whether or not we include the endpoints a, b; since $\mathbb{P}[X = a] = \mathbb{P}[X = b] = 0$, we have $\mathbb{P}[a < X < b] = \mathbb{P}[a \le X \le b]$.

which gives us $c = \frac{1}{\ell}$. Therefore, the density of the uniform distribution on $[0,\ell]$ is given by

$$f(x) = \begin{cases} 0, & \text{for } x < 0, \\ 1/\ell, & \text{for } 0 \le x \le \ell, \\ 0, & \text{for } x > \ell. \end{cases}$$

Remark: Following the “histogram" analogy above, it is tempting to think of $f(x)$ as a “probability." However, $f(x)$ doesn’t itself correspond to the probability of anything! In particular, there is no requirement that $f(x)$ be bounded by 1. For example, the density of the uniform distribution on the interval $[0,\ell]$ with $\ell = \frac{1}{2}$ is equal to $f(x) = 1/(\frac{1}{2}) = 2$ for $0 \le x \le \frac{1}{2}$, which is greater than 1. To connect density $f(x)$ with probabilities, we need to look at a very small interval $[x, x+\mathrm{d}x]$ close to x; then we have

$$\mathbb{P}[x \le X \le x + \mathrm{d}x] = \int_x^{x+\mathrm{d}x} f(z)\,\mathrm{d}z \approx f(x)\,\mathrm{d}x. \tag{2}$$

Thus, we can interpret $f(x)$ as the “probability per unit length" in the vicinity of x.

2.1 Cumulative Distribution Function

For a continuous random variable X, one often starts the discussion with the *cumulative distribution function (c.d.f.)*, which is the function $F(x) = \mathbb{P}[X \le x]$. It is closely related to the probability density function for X, $f(x)$, as

$$F(x) = \mathbb{P}[X \le x] = \int_{-\infty}^{x} f(z)\,\mathrm{d}z. \tag{3}$$

Thus, one can describe a random variable X by its c.d.f., denoted by $F(x)$, which then gives the probability density function, $f(x)$, as

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}.$$

To connect to discrete probability, one might think of approximating a continuous random variable, X, as the set of probabilities for X being in one of a countably infinite set of intervals of length $\mathrm{d}x$ on the real line. That is, the set of probabilities $\mathbb{P}[x_k < X \le x_k + \mathrm{d}x]$ where $x_k = k\,\mathrm{d}x$ for $k \in \mathbb{Z}$. In this view, $\mathbb{P}[X \le x_i] = \sum_{j \le i} \mathbb{P}[x_j < X \le x_j + \mathrm{d}x]$. Connecting this to the probability density function, $f(x)$, we have

$$\mathbb{P}[x_j < X \le x_j + \mathrm{d}x] \approx f(x_j)\,\mathrm{d}x$$

for “small” $\mathrm{d}x$, and

$$F(x_i) = \mathbb{P}[X \le x_i] \approx \sum_{j \le i} f(x_j)\,\mathrm{d}x.$$

Calculus comes in when we see the expression above is a Riemann sum. Taking the limit as $\mathrm{d}x$ goes to zero yields the integral we see in (3).

2.2 Expectation and Variance

As in the discrete case, we define the expectation of a continuous r.v. as follows:

Definition 23.2 (Expectation)**.** *The expectation of a continuous r.v. X with probability density function f is*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x.$$

Note that the integral plays the role of the summation in the discrete formula $\mathbb{E}[X] = \sum_a a\,\mathbb{P}[X = a]$. Similarly, we can define the variance as follows:

Definition 23.3 (Variance)**.** *The variance of a continuous r.v. X with probability density function f is*

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x - \left(\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x\right)^2.$$

Example

Let X be a uniform r.v. on the interval $[0,\ell]$. Then intuitively, its expected value should be in the middle, $\frac{\ell}{2}$. Indeed, we can use our definition above to compute

$$\mathbb{E}[X] = \int_0^{\ell} x\frac{1}{\ell}\,\mathrm{d}x = \frac{x^2}{2\ell}\bigg|_0^{\ell} = \frac{\ell}{2},$$

as claimed. We can also calculate its variance using the above definition and plugging in the value $\mathbb{E}[X] = \frac{\ell}{2}$ to get:

$$\mathrm{Var}(X) = \int_0^{\ell} x^2\frac{1}{\ell}\,\mathrm{d}x - \mathbb{E}[X]^2 = \frac{x^3}{3\ell}\bigg|_0^{\ell} - \left(\frac{\ell}{2}\right)^2 = \frac{\ell^2}{3} - \frac{\ell^2}{4} = \frac{\ell^2}{12}.$$

The factor of $\frac{1}{12}$ here is not particularly intuitive, but the fact that the variance is proportional to ℓ^2 should come as no surprise. Like its discrete counterpart, this distribution has large variance.

2.3 Joint Density

Recall that for discrete random variables X and Y, their joint distribution is specified by the probabilities $\mathbb{P}[X = a, Y = c]$ for all possible values a, c. Similarly, if X and Y are continuous random variables, then their joint distribution is specified by the probabilities $\mathbb{P}[a \le X \le b, c \le Y \le d]$ for all $a \le b$, $c \le d$. Moreover, just as the distribution of X can be characterized by its density function, the joint distribution of X and Y can be characterized by their joint density.

Definition 23.4 (Joint Density)**.** *A* joint density function *for two random variable X and Y is a function $f : \mathbb{R}^2 \to \mathbb{R}$ satisfying:*

1. *f is non-negative: $f(x,y) \ge 0$ for all $x, y \in \mathbb{R}$.*
2. *The total integral of f is equal to 1: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$.*

Then the joint distribution of X and Y is given by:

$$\mathbb{P}[a \le X \le b,\ c \le Y \le d] = \int_c^d \int_a^b f(x,y)\,\mathrm{d}x\,\mathrm{d}y \qquad \textit{for all } a \le b \textit{ and } c \le d.$$

In analogy with (2), we can connect the joint density $f(x,y)$ with probabilities by looking at a very small square $[x, x + \mathrm{d}x] \times [y, y + \mathrm{d}y]$ close to (x,y); then we have

$$\mathbb{P}[x \le X \le x + \mathrm{d}x,\ y \le Y \le y + \mathrm{d}y] = \int_y^{y+\mathrm{d}y} \int_x^{x+\mathrm{d}x} f(u,v)\,\mathrm{d}u\,\mathrm{d}v \approx f(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{4}$$

Thus we can interpret $f(x,y)$ as the “probability per unit area" in the vicinity of (x,y).

As with joint distributions for discrete random variables, we can define marginal and conditional density functions for continuous random variables as follows:

Definition 23.5 (Marginal and Conditional Density)**.** *For a joint distribution on X and Y with joint density function $f_{X,Y}(x,y)$,*

1. *The* marginal density function on X *is*

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y,$$

and similarly the marginal density function on Y *is*

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x.$$

2. *The* conditional density function $f_{Y|X}(x,y)$ *for two random variables X and Y with joint distribution $f_{X,Y}(x,y)$ is defined as:*

$$f_{Y|X}(x,y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

The integration above for the marginal density function converts the joint density function on X and Y to the density function for only X (or Y).

The conditional density function $f_{Y|X}(x,y)$ is the density function for Y conditioned on the event $X = x$. Recall that the probability of event B given A is computed as $\mathbb{P}[B \mid A] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[A]}$. For continuous random variables X and Y, the event B corresponds to the event $Y \in [y, y+\mathrm{d}y]$ and the event A is $X \in [x, x+\mathrm{d}x]$. Using (4), we can plug in $\mathbb{P}[A \cap B] \approx f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$ and $\mathbb{P}[B] \approx f_X(x)\,\mathrm{d}x$, which yields

$$\mathbb{P}[Y \in [y, y+\mathrm{d}y] \mid X \in [x, x+\mathrm{d}x]] \approx \frac{f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y}{f_X(x)\,\mathrm{d}x} \approx \frac{f_{X,Y}(x,y)}{f_X(x)}\,\mathrm{d}y.$$

Dividing by $\mathrm{d}y$ produces conditional density of Y conditioned on X. This last bit can be made formal using limits.

2.4 Independence

Recall that two discrete random variables X and Y are said to be independent if the events $X = a$ and $Y = c$ are independent for every possible values a, c. We have a similar definition for continuous random variables:

Definition 23.6 (Independence for Continuous R.V.'s)**.** *Two continuous r.v.'s X, Y are* independent *if the events $a \le X \le b$ and $c \le Y \le d$ are independent for all $a \le b$ and $c \le d$:*

$$\mathbb{P}[a \le X \le b,\ c \le Y \le d] = \mathbb{P}[a \le X \le b] \cdot \mathbb{P}[c \le Y \le d].$$

What does this definition say about the joint density of independent r.v.'s X and Y? Applying (4) to connect the joint density with probabilities, we get, for small $\mathrm{d}x$ and $\mathrm{d}y$:

$$\begin{aligned} f(x,y)\,\mathrm{d}x\,\mathrm{d}y &\approx \mathbb{P}[x \le X \le x+\mathrm{d}x,\ y \le Y \le y+\mathrm{d}y] \\ &= \mathbb{P}[x \le X \le x+\mathrm{d}x] \cdot \mathbb{P}[y \le Y \le y+\mathrm{d}y] && \text{(by independence)} \\ &\approx f_X(x)\,\mathrm{d}x \times f_Y(y)\,\mathrm{d}y \\ &= f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y, \end{aligned}$$

where f_X and f_Y are the (marginal) densities of X and Y respectively. So we get the following result:

Theorem 23.1. *The joint density of independent r.v.'s X and Y is the product of the marginal densities:*

$$f(x,y) = f_X(x) f_Y(y) \quad \textit{for all } \; x, y \in \mathbb{R}.$$

Exercise. Show that for independent r.v.'s X and Y that $f_{Y|X}(x,y) = f_Y(y)$.

2.5 Total probability

Equipped with the knowledge of conditional densities, we can define the *total probability rule* similarly to discrete probability.

Recall that in discrete probability, we defined the total probability rule for an event A as

$$\mathbb{P}[A] = \sum_x \mathbb{P}[A \mid X = x]\, \mathbb{P}[X = x].$$

Working with a continuous random variable instead, the total probability rule is defined as

$$\mathbb{P}[A] = \int_{-\infty}^{\infty} \mathbb{P}[A \mid X = x] f_X(x)\, \mathrm{d}x.$$

Suppose now that the event A is defined as $y \le Y \le y +\, \mathrm{d}y$, where Y is another continuous random variable. Recall that the probability $\mathbb{P}[y \le Y \le y + \mathrm{d}y]$ can be approximated as $f_Y(y)\, \mathrm{d}y$; the conditional probability $\mathbb{P}[y \le Y \le y + \mathrm{d}y \mid X = x]$ can similarly be approximated as $f_{Y|X}(x,y)\, \mathrm{d}y$.

This means that we'd have

$$\begin{aligned}
f_Y(y)\, \mathrm{d}y &\approx \mathbb{P}[y \le Y \le y + \mathrm{d}y] \\
&= \int_{-\infty}^{\infty} \mathbb{P}[y \le Y \le y + \mathrm{d}y \mid X = x] f_X(x)\, \mathrm{d}x \\
&\approx \int_{-\infty}^{\infty} \big(f_{Y|X}(x,y)\, \mathrm{d}y\big) f_X(x)\, \mathrm{d}x \\
f_Y(y)\, \mathrm{d}y &\approx \left(\int_{-\infty}^{\infty} f_{Y|X}(x,y) f_X(x)\, \mathrm{d}x\right) \mathrm{d}y \\
f_Y(y) &= \int_{-\infty}^{\infty} f_{Y|X}(x,y) f_X(x)\, \mathrm{d}x
\end{aligned}$$

This is the continuous analog to the total probability rule for discrete random variables. Note that this can also be viewed as computing the marginal density by utilizing the definition of a conditional density:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, \mathrm{d}x = \int_{-\infty}^{\infty} f_{Y|X}(x,y) f_X(x)\, \mathrm{d}x.$$

3 Exponential Distribution

We have already seen one important continuous distribution, namely the uniform distribution. In this and the next sections we will see two more: the *exponential* distribution and the *normal* (or *Gaussian*) distribution. These three distributions cover the vast majority of continuous random variables arising in applications.

The exponential distribution is a continuous version of the geometric distribution, which we have already seen. Recall that the geometric distribution describes the number of tosses of a coin until the first Head

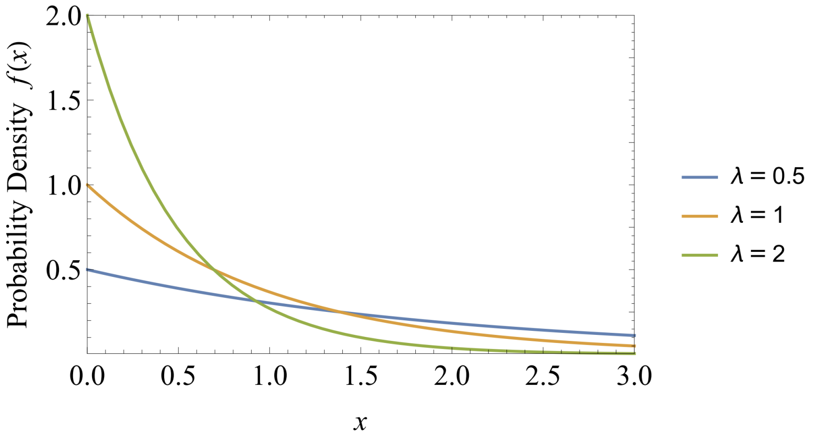

Figure 1: The probability density function $f(x)$ for the exponential distribution with $\lambda = 0.5, 1, 2$.

appears; the distribution has a single parameter p, which is the bias (Heads probability) of the coin. Of course, in real life applications we are usually not waiting for a coin to come up Heads but rather waiting for a system to fail, a clock to ring, an experiment to succeed, etc.

In such applications we are frequently not dealing with discrete events or discrete time, but rather with *continuous* time: for example, if we are waiting for an apple to fall off a tree, it can do so at any time at all, not necessarily on the tick of a discrete clock. This situation is naturally modeled by the exponential distribution, defined as follows.

Definition 23.7 (Exponential Distribution)**.** *For $\lambda > 0$, a continuous random variable X with p.d.f.*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \textit{if } x \geq 0, \\ 0, & \textit{otherwise}, \end{cases}$$

is called an exponential random variable with parameter λ, and we write $X \sim \mathrm{Exp}(\lambda)$.

Note that by definition $f(x)$ is non-negative. Moreover, we can check that it satisfies (1):

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^{\infty} \lambda e^{-\lambda x}\,\mathrm{d}x = -e^{-\lambda x}\Big|_0^{\infty} = 1,$$

so $f(x)$ is indeed a valid probability density function. Figure 1 shows the probability density function for the exponential distribution with a few different values of λ.

3.1 Mean and Variance of an Exponential Random Variable

Let us now compute the expectation and variance of $X \sim \mathrm{Exp}(\lambda)$.

Theorem 23.2. *Let X be an exponential random variable with parameter $\lambda > 0$. Then*

$$\mathbb{E}[X] = \frac{1}{\lambda} \qquad \textit{and} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Proof. We can calculate the expected value using integration by parts:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_0^{\infty} \lambda x e^{-\lambda x}\,\mathrm{d}x = -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,\mathrm{d}x = 0 + \left(-\frac{e^{-\lambda x}}{\lambda}\right)\Big|_0^{\infty} = \frac{1}{\lambda}.$$

To compute the variance, we first evaluate $\mathbb{E}\left[X^2\right]$, again using integration by parts:

$$\mathbb{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_0^{\infty} \lambda x^2 e^{-\lambda x}\,\mathrm{d}x = -x^2 e^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} 2x e^{-\lambda x}\,\mathrm{d}x = 0 + \frac{2}{\lambda}\mathbb{E}[X] = \frac{2}{\lambda^2}.$$

The variance is therefore

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

as claimed. □

3.2 As Continuous Time Analog of Geometric Distribution

Like the geometric distribution, the exponential distribution has a single parameter λ, which characterizes the *rate* at which events happen. Note that the exponential distribution satisfies, for any $t \geq 0$,

$$\mathbb{P}[X > t] = \int_t^{\infty} \lambda e^{-\lambda x}\,\mathrm{d}x = -e^{-\lambda x}\Big|_t^{\infty} = e^{-\lambda t}. \tag{5}$$

In other words, the probability that we have to wait more than time t for our event to happen is $e^{-\lambda t}$, which is an exponential decay with rate λ.

Now consider a discrete-time setting in which we perform one trial every δ seconds (where δ is very small — in fact, we will take $\delta \to 0$ to make time "continuous"), and where our success probability is $p = \lambda\delta$. Making the success probability proportional to δ makes sense, as it corresponds to the natural assumption that there is a fixed *rate* of success *per unit time*, which we denote by $\lambda = p/\delta$. In this discrete setting, the number of trials until we get a success has the geometric distribution with parameter p, so if we let the r.v. Y denote the time (in seconds) until we get a success, we have

$$\mathbb{P}[Y > k\delta] = (1-p)^k = (1-\lambda\delta)^k, \qquad \text{for any integer } k \geq 0.$$

Hence, for any $t > 0$, we have

$$\mathbb{P}[Y > t] = \mathbb{P}[Y > (\tfrac{t}{\delta})\delta] = (1-\lambda\delta)^{t/\delta} \approx e^{-\lambda t}, \tag{6}$$

where this final approximation holds in the limit as $\delta \to 0$ with λ and t fixed. (We are ignoring the detail of rounding $\frac{t}{\delta}$ to an integer since we are taking an approximation anyway.)

Comparing (6) with (5), we see that this distribution has the same form as the exponential distribution with parameter λ. Thus we may view the exponential distribution as a continuous time analog of the geometric distribution, where the parameter p (success probability per trial) is replaced by the parameter λ (success probability per unit time). Note, though, that λ is not constrained to be ≤ 1.