Published in Towards Data Science

Aerin Kim

Aug 6, 2019 · 8 min read · Member-only · Listen

Exponential Distribution — Intuition, Derivation, and Applications

When to Use an Exponential Distribution

We always start with the "why" instead of going straight to the formulas. If you understand the why, it actually sticks with you and you'll be a lot more likely to apply it in your own line of work.

1. Why did we have to invent Exponential Distribution?

To predict the amount of waiting time until the next event (i.e., success, failure, arrival, etc.).

For example, we want to predict the following:

- **The amount of time until** the customer finishes browsing and actually purchases something in your store (success).
- **The amount of time until** the hardware on AWS EC2 fails (failure).
- **The amount of time** you need to **wait until** the bus arrives (arrival).

Then, my next question is this: **Why is λ * e^(−λt) the PDF of the time**

until the next event happens?

And the follow-up question would be: **What does X ~ Exp(0.25) mean?** Does the parameter **0.25** mean **0.25 minutes**, **hours**, or **days**, or is it **0.25 events**?

From this point on, I'll assume you know Poisson distribution inside and out. If you don't, this article will give you a clear idea.

X ~ Exp(λ) 👉 Is the exponential parameter λ the same as λ in Poisson?

One thing that would save you from the confusion later about **X ~ Exp(0.25)** is to remember that 0.25 is ***not* a time duration**, but it is **an event rate,** which is **the same as the**

parameter λ in a Poisson process.

For example, your blog has *500 visitors a day*. That is a **rate**. *The number of customers arriving at the store in an hour, the number of earthquakes per year, the number of car accidents in a week, the number of typos on a page, the number of hairs found in Chipotle, etc.,* are all **rates (λ) of the unit of time,** which is the parameter of the Poisson distribution.

However, when we model **the elapsed time between events**, we tend to speak in terms of **time** instead of rate, e.g., *the number of years a computer can power on without failure is **10 years** (instead of saying **0.1 failure/year,** which is a rate), a customer arrives every 10 minutes, major hurricanes come every 7 years,* etc. When you see **the terminology — "mean" of the exponential distribution — 1/λ is what it means.**

The confusion starts when you see the term “***decay parameter***”, or even worse, the term “***decay rate***”, which is frequently used in exponential distribution. The *decay parameter* is expressed in terms of **time** (e.g., every 10 mins, every 7 years, etc.), which is **a reciprocal (1/λ) of the rate (λ) in Poisson.** Think about it: If you get 3 customers per hour, it means you get one customer every 1/3 hour.

So, now you can answer the following:
What does it mean for “X ~ Exp(0.25)”?

It means **the Poisson rate will be 0.25.** During a **unit time** (either it’s a minute, hour or year), **the event occurs 0.25 times** on average. Converting this into *time terms*, it takes **4 hours** (a reciprocal of 0.25) until the event occurs, assuming your unit time is an hour.

```
* Confusion-proof :
Exponential's parameter λ is
the same as that of Poisson
process (λ).
```

2. Let's derive the PDF of Exponential from scratch!

Our first question was: **Why is λ * e^(−λt) the PDF of the time until the next event occurs?**

The definition of exponential distribution is *the probability distribution of* ***the time *between* the events*** *in a Poisson process.*

If you think about it, ***the amount of time until the event*** **occurs** means **during the waiting period, not a single event has happened.**

This is, in other words, **Poisson (X=0).**

$$\text{Poisson}\quad P(X=k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

$$\text{"}\quad P(X=0) = \frac{\lambda^0 \cdot e^{-\lambda}}{0!} = e^{-\lambda}$$

Poisson(X=0): the first step of the derivation of Exponential dist.

One thing to keep in mind about Poisson PDF is that **the time period in which Poisson events (X=k) occur is just one (1) unit time.**

If you want to model the probability distribution of "nothing happens during the ***time duration t***," *not just during* ***one unit time,*** how will you do that?

```
P(Nothing happens during t
time units)

= P(X=0 in the first time
unit)
  * P(X=0 in the second time
unit)
  * … * P (X=0 in the t-th
```

```
time unit)
= e^−λ ∗ e^−λ ∗ … ∗ e^−λ =
e^(−λt)
```

The Poisson distribution assumes that events occur independent of one another. Therefore, we can calculate the probability of zero success during **t** units of time by multiplying P(**X**=0 in a single unit of time) **t** times.

```
P(T > t) = P(X=0 in t time
units) = e^−λt

∗ T : the random variable of
our interest!
      the random variable for
the waiting time until the
first event
∗ X : the # of events in the
future which follows the
Poisson dist.

∗ P(T > t) : The probability
that the waiting time until
the first event is greater
than t time units
∗ P(X = 0 in t time units) :
The probability of zero
successes in t time units
```

A PDF is the derivative of the CDF.

Since we already have the CDF, **1 - P(T > t),** of exponential, we can get its PDF by differentiating it.

$\boxed{\text{CDF}}$

$$P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t}$$

$$\Downarrow$$

$\boxed{\text{PDF}}$

$$\frac{d}{dt}(\text{CDF}) = \frac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t}$$

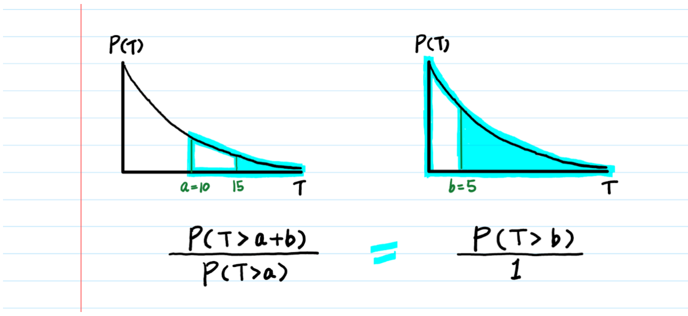
The probability density function is the derivative of the cumulative density function.

3. Memoryless Property

```
Defintion:

P(T > a + b | T > a) = P(T >
b)
```

This means…



Show me the proof?

① We already know $P(T>t) = e^{-\lambda t}$

② $P(T>a+b \mid T>a) = \dfrac{P(T>a+b)}{P(T>a)}$

$= \dfrac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b}$

③ $P(T>b) = e^{-\lambda b}$　　② = ③ ▨

> ***Is memoryless a "useful" property?***

Is it reasonable to model *the longevity of a mechanical device* using exponential distribution?

For example, if the device has lasted nine years already, then memoryless means the probability that it will last another three years (so, a total of 12 years) is exactly the same as that of a brand-new machine lasting for the next three years.

```
P(T > 12|T > 9) = P(T > 3)
```

Does this equation look reasonable to you?

For me, it doesn't. Based on my experience, the older the device is, the more likely it is to break down. To model this property— **increasing hazard rate —** we can use, for example, a Weibull distribution.

> *Then, when is it appropriate to use exponential distribution?*

Car accidents. It doesn't increase or decrease your chance of a car accident if no one has hit you in the past five hours. This is why **λ is often called a hazard rate.**

> ***Who else has memoryless***

> ***property?***

The exponential distribution is the **only** continuous distribution that is memoryless (or with a constant failure rate). Geometric distribution, its discrete counterpart, is the only discrete distribution that is memoryless.

4. Applications IRL 🔥

a) Waiting time modeling

Values for an exponential random variable have more small values and fewer large values. The bus that you are waiting for will probably come within the next 10 minutes rather than the next 60 minutes.

> *Using exponential distribution, we can answer the questions below.*
>
> *1. The bus comes in every 15 minutes*

> *on average. (Assume that the time that elapses from one bus to the next has exponential distribution, which means the total number of buses to arrive during an hour has Poisson distribution.) And I just missed the bus! The driver was unkind. The moment I arrived, the driver closed the door and left. If the next bus doesn't arrive within the next ten minutes, I have to call Uber or else I'll be late. What's the probability that it takes less than ten minute for the next bus to arrive?*
>
> *2. Ninety percent of the buses arrive within how many minutes of the previous bus?*
>
> *3. How long on average does it take for two buses to arrive?*
>
> ** Post your answers in the comment, if you want to see if your answer is correct.*

b) Reliability (failure) modeling

Since we can model the successful event (the arrival of the bus), why not the failure modeling — the amount of time a product lasts?

> *The number of hours that AWS hardware can run before it needs a restart is exponentially distributed with an average of 8,000 hours (about a year).*
>
> *1. You don't have a backup server and you need an uninterrupted 10,000-hour run. What is the probability that you will be able to complete the run without having to restart the server?*
>
> *2.What is the probability that the server doesn't require a restart between 12 months and 18 months?*

Note that sometimes, the exponential distribution might not be appropriate

— when the failure rate changes throughout the lifetime. However, it will be *the only distributio*n that has this unique property-- constant hazard rate.

c) Service time modeling (Queuing Theory)

The service times of agents (e.g., how long it takes for a Chipotle employee to make me a burrito) can also be modeled as exponentially distributed variables.

The total length of a process — a sequence of several independent tasks — follows the Erlang distribution: the distribution of the sum of several independent exponentially distributed variables.

5. Recap: Relationship between a Poisson and an Exponential

distribution

If the number of events per unit time follows a Poisson distribution, then the amount of time between events follows the exponential distribution.

Assuming that the time between events is not affected by the times between previous events (i.e., they are independent), then the number of events per unit time follows a Poisson distribution with the rate $\boldsymbol{\lambda = 1/\mu}$.

6. Exercise

I've found that most of my understanding of math topics comes from doing problems. So, I encourage you to do the same. Try to complete the exercises below, even if they take some time.

1. Let U be a uniform random variable between 0 and 1. Then an exponential random variable **X**

can be generated as

```
X = −1/λ ∗ ln(U)
```

Prove why.

2. The maximum value on the *y*-axis of PDF is λ. Why is it so?

Probability Density Function of Exponential Distribution

3. **X1** and **X2** are independent exponential random variables with the rate λ.

X1 ~ Exp(λ)

X2 ~ Exp(λ)

Let **Y=X1+X2**.

What is the PDF of **Y**?

Where can this distribution be used?

The answer is here.

Other intuitive articles that you might like:

Poisson Distribution…

…Why did Poisson invente…

Beta Distribution —…

…The difference between the…