Probability Theory

Applications for Data Science
Module 4 Continuous Random Variables

Anne Dougherty

March 1, 2021

TABLE OF CONTENTS

Continuous Random Variables

Random Variables

At the end of this module, students should be able to

- Define a continuous random variable and give examples of a probability density function and a cumulative distribution function.
- Identify and discuss the properties of a uniform, **exponential**, and normal random variable.
- Calculate the expectation and variance of a continuous rv.

The Poisson random variable

The number of customers who arrive for service, and their waiting times, are described by the Poisson rv and the exponential rv, respectively.

A Poisson rv is a discrete rv that describes the total number of events that happen in a certain time period.

- # of vehicles crossing a bridge in one day
- # of gamma rays hitting a satellite per hour
- #of cookies sold at a bake sale in one hour
- # of customers arriving at a bank in a week

Definition: A discrete random variable X has **Poisson distribution** with parameter λ ($\lambda > 0$) if the probability mass function of X is

pmf: $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ for $k = 0, 1, 2, \ldots$

$$\sum_{k=0}^{\infty} P(X=k) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{e^{\lambda}} = e^{-\lambda} \cdot e^{\lambda} = 1 \checkmark$$

$$E(X) = \sum_{k=0}^{\infty} k\, P(X=k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} e^{-\lambda}$$

$$E(X) = \lambda$$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda(\lambda+1)$$

$$V(X) = E(X^2) - (E(X))^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$$

Notation: $X \sim \text{Poisson}(\lambda)$

Example: The number of mosquitoes captured in a trap during a given period of time can be modeled by a Poisson rv with $\lambda = 4.5$. What is the probability that the trap contains exactly 5 mosquitoes? 5 or fewer mosquitoes?

$$X \sim \text{Poisson}(\lambda = 4.5)$$

$$P(X=5) = \frac{(4.5)^5}{5!} e^{-4.5} \approx .1708\ldots$$

$$P(X \leq 5) = \sum_{k=0}^{5} P(X=k) = \sum_{k=0}^{5} \frac{(4.5)^k}{k!} e^{-4.5} \approx .7029\ldots$$

Example: A factory makes parts for a medical device company. On average, the rate of defective parts per day is 10. You are responsible for monitoring the number of defective parts on a particular day.

- Define an appropriate random variable for this experiment.
- Give the values that the random variable can take on.
- Find the probability that the random variable equals 2.
- What assumptions do you need to make?

Let X = # of defective parts (that day)

Model $X \sim \text{Poisson}(\lambda = 10)$, $X \in \{0, 1, 2, \ldots\}$

Assumption: X, as a Poisson, can take on an infinite number of values, but we didn't make an infinite #.

$$P(X = 2) = e^{-\lambda} \frac{\lambda^2}{2!} = e^{-10} \frac{(10^2)}{2!} \approx .0023$$

The exponential random variable

The family of exponential distributions provides probability models that are widely used in engineering and science disciplines to describe **time-to-event data**. An exponential rv is continuous.

- Time until birth
- Time until a light bulb fails
- Waiting time in a queue
- Length of service time
- Time between customer arrivals

Definition: A continuous random variable X has the exponential distribution with rate parameter λ ($\lambda > 0$) if the pdf of X is

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

Verify $\int_{-\infty}^{\infty} f(x)\,dx = 1 \implies \int_0^{\infty} \lambda e^{-\lambda x}\,dx = \lim_{t \to \infty} \int_0^t \lambda e^{-\lambda x}\,dx$

$$= \lim_{t \to \infty} \frac{\lambda e^{-\lambda x}}{-\lambda} \Big|_0^t = \lim_{t \to \infty} (-e^{-\lambda t} + 1)$$

$$= 1 \checkmark$$

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x}\,dx \underset{\text{IBP}}{=} \frac{1}{\lambda}$$

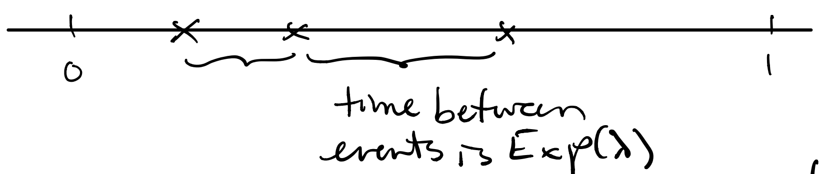
$$E(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx = \frac{2}{\lambda^2}$$

$$V(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Notation: $X \sim \exp(\lambda)$

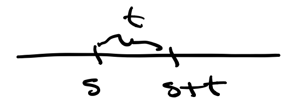
Two useful properties of the exponential: First, if the number of events occurring in a unit of time is a Poisson rv with parameter λ, the the time between events is exponential, also with parameter λ.

Example: Suppose the number of customers arriving for service is modeled by a Poisson rv with $\lambda = 5$. That is, an average of 5 customers arrive per hour. Then, the time between arrivals is exponential with $1/\lambda = 1/5$. That is, the expected time between arrivals is $1/5$ hour. This relationship is only between Poisson & exp'l r.v. It is a very important relationship that is used heavily in queueing theory & statistical analysis of customer service.



The second important property is the memoryless property of the exponential rv: If $X \sim Exp(\lambda)$, then
$P(X > s+t \,|X > s) = P(X > t)$ for all $s, t \geq 0$.

First: RHS
$$P(X>t) = 1 - P(X \le t) = 1 - \int_0^t \lambda e^{-\lambda x}\,dx = 1 - \left(\frac{\lambda}{-\lambda} e^{-\lambda x}\right)\Big|_0^t$$
$$= 1 + e^{-\lambda t} - 1 = e^{-\lambda t}$$



Then: LHS
$$P(X > s+t \,|\, X > s) = \frac{P(X > s+t \cap X > s)}{P(X > s)}$$
$$= \frac{P(X > s+t)}{P(X>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

Intuitively: If r.v. (say length of time of service) has the exp'l dist. then if you know that service has gone on for s time units, ~~then~~ the prob. that it will last to s+t is the same as if we had started at 0 & asked if it would lasting more than time t.

Now, many situations are not memoryless, like how long a piece of machinery works, but some, like time between customer arrivals, are.

Example: Suppose the service time at a bank with one teller is modeled by a rv X with $X \sim Exp(\lambda = 1/5)$. Then, $E(X) = 1/\lambda = 5$ minutes. If there is a customer in service when you enter the bank, find the probability that the customer is still in service 4 minutes later.

It's natural to ask: "When did the customer begin service?" or "how long has the customer been in service?" The answer is, as long as the service is modeled as an exp'l r.v. then it doesn't matter.

Let X = service time of customer (starting from your entrance into the bank)
= (your waiting time for service)

$$P(X \geq 4) = \int_4^\infty \lambda e^{-\lambda x}\,dx = e^{-4/5} \approx .449$$

Suppose we know customer started service 5 min ago? Then, what is the prob. they will need at least 4 more min?

$$P(X \geq 9 \mid X \geq 5) = \frac{P(X \geq 9 \cap X \geq 5)}{P(X \geq 5)} = \frac{P(X \geq 9)}{P(X \geq 5)} = \frac{e^{-9/5}}{e^{-5/5}} = e^{-4/5}$$

In summary: we have learned about the Poisson r.v. (which can model the number of arrivals in a certain time period) & we've looked at the exp'l r.v. & its two main properties

1. models time between events
2. memoryless

In the next video we'll study the Gaussian or normal r.v. See you then!