COMPSCI 240: Reasoning Under Uncertainty

Andrew Lan and Nic Herndon

University of Massachusetts at Amherst

Spring 2019

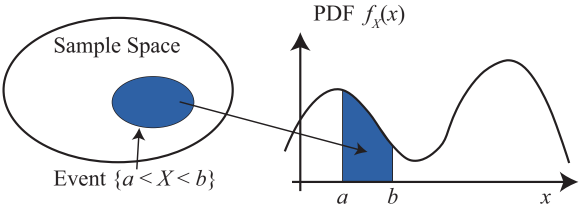
Lecture 14: Common Continuous Random Variables

Recap: Probability Density of Continuous RVs

- In the simplest case $A = [a, b]$ is a single interval and this definition reduces to a definite integral:

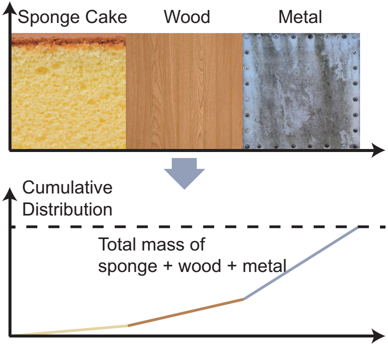
$$P(a < X < b) = \int_a^b f_X(x)dx$$

- Intuitively, the **probability mass** of an interval $[a, b]$ is $P(a < X < b)$.

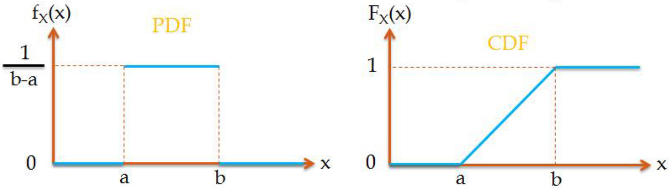


Cumulative Distribution Functions

- The cumulative distribution function (CDF) for a continuous random variable X is defined as
$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$
- Intuitively, the CDF accumulates probability upto the value of x.



CDF - Example



- The PDF of the above graph can be defined as

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise,} \end{cases}$$

- **Question**: What is its CDF?
- **Answer**:

$$F_X(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ 1, & x > b \end{cases}$$

Cumulative Distribution Functions

- CDF
  - is a continuous function of x, if X is a continuous RV.
  - is monotonically non-decreasing:

$$\text{if } x \leq y, \text{ then } F_X(x) \leq F_X(y).$$

  - approaches 0 as $x \to -\infty$, and 1 as $x \to \infty$.
- If CDF is known, its PDF can be similarly derived as

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

Cumulative Distribution Functions

- **Question:** Well, if we have PDF, why do we need CDF?
- **Answer**: If we have CDF, we do not need to integrate every time when we compute $P(a \leq X \leq b)$.

$$
\begin{aligned}
P(a < X < b) &= \int_a^b f_X(x)dx \\
&= \int_{-\infty}^b f_X(x)dx - \int_{-\infty}^a f_X(x)dx \\
&= F_X(b) - F_X(a)
\end{aligned}
$$

Common Continuous Random Variables

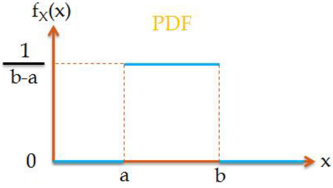
- There are some commonly used continuous RVs (PDFs)
  - The Uniform Random Variables
  - The Exponential Random Variables
  - The Normal (Gaussian) Random Variables
  - and many more...
- Let us explore some of these RVs

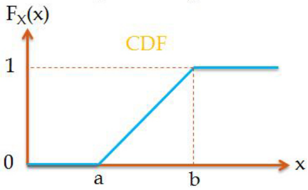
Uniform Continuous Random Variables

- Consider a RV that takes continuous values in an interval $[a, b]$.
- Uniform continuous RV has a uniform probability density in $[a, b]$.
- In other words, it has the same probability for two sub-intervals of the same length.
- Do not confuse with the discrete random variable!
- Its PDF can be defined as

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise,} \end{cases}$$

- We have looked at this distribution already.





Uniform Random Variables

- Its CDF can be defined as

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

- When $b = 2$ and $a = 0$, what is $P(0.5 < X < 1.5)$?
- **Answer**: $F_X(1.5) - F_X(0.5) = \frac{1.5}{2} - \frac{0.5}{2} = \frac{1}{2}$.

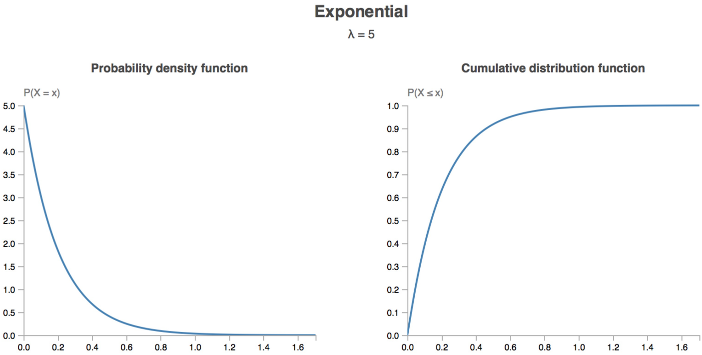
Exponential Random Variables

- An **exponential random variable** X is a continuous random variable with PDF:

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

  where λ must be strictly greater than 0.
- Exponential random variables are often used to model waiting times (eg: the length of time between calls at a call center, the length of time between people entering a store, the length of time between hits on a website, etc...).
- Closely connected to the geometric (discrete) random variable, which also relates to the discrete time that will elapse until an incident of interest occurs.

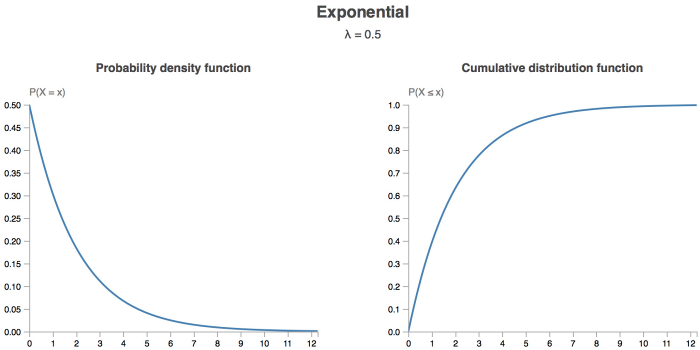
Exponential Random Variables: $\lambda = 5$



- The **probability density** can be greater than 1 at some points.

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

Exponential Random Variables: $\lambda = 0.5$



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

CDF of Exponential RV

- Its PDF is:

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

- We know that

$$\int_{-\infty}^{\infty} e^{ax} = \frac{1}{a} e^{ax}.$$

or

$$\frac{d}{dx} e^{ax} = a e^{ax}.$$

- By definition of CDF,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t) dt = \int_{0}^{x} \lambda e^{-\lambda t} dt \\ &= -e^{-\lambda x} \Big|_{0}^{x} \\ &= 1 - e^{-\lambda x}, \text{ if } x \geq 0 \end{aligned}$$

- Thus,

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Probability Mass

- Using these substitutions we can find the value of the probability mass for an interval $[a, b]$ as follows:

$$\begin{aligned} P(a < X < b) &= \int_a^b \lambda e^{-\lambda x} dx \\ &= \lambda \int_a^b e^{-\lambda x} dx \\ &= -e^{-\lambda x}\Big|_a^b \\ &= -(e^{-\lambda b}) - (-e^{-\lambda a}) \\ &= e^{-\lambda a} - e^{-\lambda b}. \end{aligned}$$

- Or Similarly

$$\begin{aligned} P(a < X < b) &= F_X(b) - F_X(a) \\ &= \left(1 - e^{-\lambda b}\right) - \left(1 - e^{-\lambda a}\right) \\ &= e^{-\lambda a} - e^{-\lambda b}. \end{aligned}$$

Normalization

- Normalization says that $P(0 < X < \infty)$ should be equal to 1. We can use the last result to verify normalization:

$$
\begin{aligned}
P(0 < X < \infty) &= \lim_{b \to \infty} F_X(b) - \lim_{a \to 0} F_X(a) \\
&= \lim_{b \to \infty} \left(1 - e^{-\lambda b}\right) - \lim_{a \to 0} \left(1 - e^{-\lambda a}\right) \\
&= (1) - (0) \\
&= 1
\end{aligned}
$$

Mean and Variance of Exponential RV

- The mean and the variance can be calculated as

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

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

- Show this by using the following:
  - Integration by parts: $\int udv = uv - \int vdu$
  - $\int_{-\infty}^{\infty} e^{ax} dx = \frac{1}{a} e^{ax}$ and/or $\frac{d}{dx} e^{ax} = ae^{ax}$.

Example

- **Question**: Let the number of miles traveled by a car before its engine fails to function be governed by the exponential distribution with a mean of $100,000$ miles. What is the probability that a car's engine will fail during its first $50,000$ miles of operation?
- **Solution**: Since $E(X) = \frac{1}{\lambda}$ for an exponential random variable X. Thus $\lambda = 1/100000$. Then,

$$
\begin{aligned}
P(X < 50,000) = F_X(50,000) &= 1 - e^{-\lambda 50{,}000} \\
&= 1 - e^{-\frac{50{,}000}{100{,}000}} \\
&= 1 - e^{-\frac{1}{2}} \\
&= 0.3934
\end{aligned}
$$