Lesson 8. Review – Poisson, Exponential, and Erlang Random Variables

1 Discrete random variables

- A **random variable** is a variable that takes on its values by chance
- A random variable is **discrete** if it can take on only a finite or countably infinite number of values
- Let X be a discrete random variable that takes on values $a_1, a_2, \ldots$:

	notation	definition
probability mass function (pmf) probability X is equal to a	$p_X(a)$	=
cumulative distribution function (cdf) probability X is less than or equal to a	$F_X(a)$	=
expected value or **mean** weighted average of possible values of X	$E[X]$	=
variance spread of X around its mean	$\mathrm{Var}(X)$	=

- We can use the cdf to compute the probability that a random variable is between two values:

2 The Poisson random variable

- X is a **Poisson random variable with parameter** μ if its pmf is

$$p_X(a) = \begin{cases} \dfrac{e^{-\mu}\mu^a}{a!} & \text{if } a = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

  - Shorthand: $X \sim \text{Poisson}(\mu)$

- Using the definitions above, we can show that for $X \sim \text{Poisson}(\mu)$:

$$F_X(a) = \sum_{k=0}^{\lfloor a \rfloor} \frac{e^{-\mu}\mu^k}{k!} \qquad E[X] = \mu \qquad \mathrm{Var}(X) = \mu$$

Example 1. As a data analyst for the Markov Theater, you've been tasked with studying the customer arrival process. Based on historical data, you have determined that the number of customers that arrive on a Saturday night between 6pm and 8pm can be modeled as a Poisson random variable X with parameter $\mu = 40$.

a. Why is X a discrete random variable?
b. What is the probability that exactly 35 customers arrive during the given time period?
c. What is the probability that 30 or fewer customers arrive during the given time period?
d. What is the average number of customers that arrive during the given time period?

3 Continuous random variables

- A random variable is **continuous** if it can take on a continuum of values
- Let X be a continuous random variable:

	notation	definition
probability density function (pdf) relative likelihood X being near a	$f_X(a)$	$=$
cumulative distribution function (cdf) probability X is less than or equal a	$F_X(a)$	$=$
expected value or **mean** weighted average of possible values of X	$E[X]$	$=$
variance spread of X around its mean	$\mathrm{Var}(X)$	$=$

- Also for continuous random variables, we can use the cdf to compute the probability that a random variable is between two values:

$$\Pr\{a \leq X \leq b\} = F_X(b) - F_X(a)$$

4 Exponential random variables

- X is an **exponential random variable with parameter** λ if its pdf

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

  - Shorthand: $X \sim \text{Exponential}(\lambda)$

- Using the definitions above, we can show that for $X \sim \text{Exponential}(\lambda)$:

$$F_X(a) = \begin{cases} 1 - e^{-\lambda a} & \text{if } a \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad E[X] = \frac{1}{\lambda} \qquad \text{Var}(X) = \frac{1}{\lambda^2}$$

Example 2. The Town of Simplexville is studying the bus service at the Main Street stop. Based on historical data, it has determined that the time between bus arrivals (in hours) can be modeled as an exponential random variable X with parameter $\lambda = 2$.

a. Why is X a continuous random variable?
b. Plot the pdf of X. Which values are more or less likely?
c. What is the average time between bus arrivals?
d. What is the probability that the time between bus arrivals is between 20 and 40 minutes (1/3 and 2/3 hours)?
e. What is the probability that the time between bus arrivals is exactly 1 hour?

5 Erlang random variables

- Let $G_1, G_2, \dots, G_n$ be independent and identically distributed (iid) exponential random variables with common parameter λ
- Then $X = G_1 + G_2 + \cdots + G_n$ is an X is an **Erlang random variable with n phases and parameter λ**
  - Shorthand: $X \sim \text{Erlang}(n, \lambda)$
- We can show that for $X \sim \text{Erlang}(n, \lambda)$:

$$f_X(a) = \begin{cases} \dfrac{\lambda(\lambda a)^{n-1} e^{-\lambda a}}{(n-1)!} & \text{if } a \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad F_X(a) = \begin{cases} 1 - \displaystyle\sum_{k=0}^{n-1} \frac{e^{-\lambda a}(\lambda a)^k}{k!} & \text{if } a \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \frac{n}{\lambda} \qquad \text{Var}(X) = \frac{n}{\lambda^2}$$

Example 3. Back to Example 2...

Let X be a random variable that represents the time between the 2nd and 5th bus arrivals at the Main Street stop. Assume that the times between consecutive bus arrivals all have the same exponential distribution and are independent.

a. What kind of random variable is X?
b. What is the variance of X?
c. What is the probability that X between 1 and 2 hours?

6 Summary

	$X \sim \text{Poisson}(\mu)$	$X \sim \text{Exponential}(\lambda)$	$X \sim \text{Erlang}(n, \lambda)$
pmf / pdf	$p_X(a) = \begin{cases} \frac{e^{-\mu}\mu^a}{a!} & \text{if } a = 0, 1, 2, \dots \\ 0 & \text{o/w} \end{cases}$	$f_X(a) = \begin{cases} \lambda e^{-\lambda a} & \text{if } a \geq 0 \\ 0 & \text{o/w} \end{cases}$	$f_X(a) = \begin{cases} \frac{\lambda(\lambda a)^{n-1} e^{-\lambda a}}{(n-1)!} & \text{if } a \geq 0 \\ 0 & \text{o/w} \end{cases}$
cdf	$F_X(a) = \sum_{k=0}^{\lfloor a \rfloor} \frac{e^{-\mu}\mu^k}{k!}$	$F_X(a) = \begin{cases} 1 - e^{-\lambda a} & \text{if } a \geq 0 \\ 0 & \text{o/w} \end{cases}$	$F_X(a) = \begin{cases} 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda a}(\lambda a)^k}{k!} & \text{if } a \geq 0 \\ 0 & \text{o/w} \end{cases}$
expected value	$E[X] = \mu$	$E[X] = \frac{1}{\lambda}$	$E[X] = \frac{n}{\lambda}$
variance	$\text{Var}(x) = \mu$	$\text{Var}(X) = \frac{1}{\lambda^2}$	$\text{Var}(X) = \frac{n}{\lambda^2}$

7 Exercises

Problem 1. Let $Y \sim \text{Poisson}(3)$. Compute the following:

a. $\Pr\{3 \le Y \le 5\}$, using the pmf
b. $\Pr\{3 \le Y \le 5\}$, using the cdf
c. $\Pr\{Y > 5\}$
d. $E[Y]$
e. $\text{Var}(Y)$

Problem 2. Let $G \sim \text{Exponential}(1/3)$. Compute the following:

a. $\Pr\{2 \le G \le 4\}$, using the pdf
b. $\Pr\{2 \le G \le 4\}$, using the cdf
c. $\Pr\{G > 4\}$
d. $E[G]$
e. $\text{Var}(G)$

Problem 3. Let $T \sim \text{Erlang}(4, 2/5)$. Compute the following:

a. $\Pr\{1 \le T \le 2\}$
b. $\Pr\{T = 2\}$
c. $\Pr\{T > 2\}$
d. $E[T]$
e. $\text{Var}(T)$