Data collection, Analysis and Inference

Subject Code: CPE-RPE,

May 2021,
SRM Univeristy-AP, Andhrapradesh

Lecture- 4: Basic Statistical Distributions and their applications: Exponential, Geometric Distribution.

- Aim: To understand the random variables of exponential and geometric distributions.

Geometric random variable

- Experiment: flip a coin until heads occur

  Let X equal the number of flips required

  What are the values X can take?

  $$X = 1, 2, \ldots$$

  Let p be the probability of heads in a single flip

  Then, $P\{X = n\} = (1 - p)^{n-1}\, p \;\; ; \;\; n = 1, 2, 3, \ldots$

A random variable X which takes on values $1, 2, 3, \dots$ and whose probability mass function is given by

$$p(i) = \begin{cases} (1-p)^{i-1}p, & \text{if } i = 1, 2, 3, \dots, \\ 0, & \text{else}, \end{cases}$$

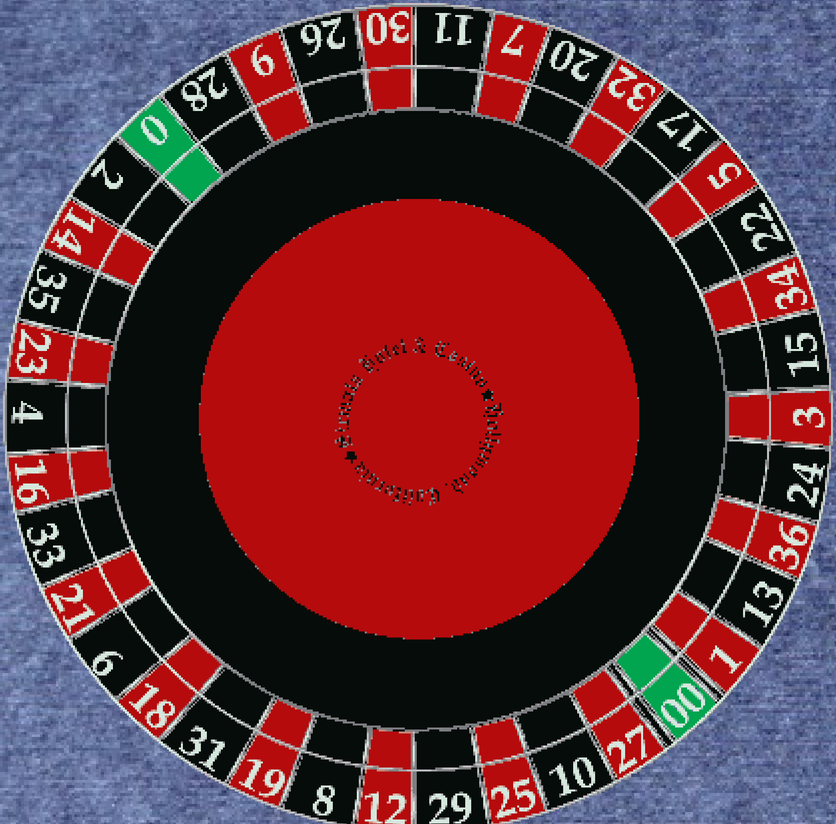
for some $p \in (0, 1)$, is called a **geometric random variable** with the parameter p.

$$\sum_{n=1}^{\infty} P\{X = n\} = p \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{p}{1-(1-p)} = 1$$

- Example: Consider a roulette wheel consisting of 38 numbers – 1 through 36, 0 and double 0. If Smith always bets that the outcome will be one of the numbers 1 through 12, what is the probability

  that (a) Smith will lose his first 5 bets;

  (b) his first win will occur on his fourth bet?



Smith always bets on the numbers 1 through 12, which occupy 12 spaces on the wheel

The probability of success is $p = 12/38 \approx 0.316$

- (a) The first five bets form a finite set of $n = 5$ trials and each spin of the roulette wheel is independent, and the probability of success p is constant

▸ X be the number of bets won by smith in 5 trials

$$\Rightarrow X \sim \text{Bin}(5, 0.316)$$

Losing all 5 bets $\Rightarrow X = 0$

- Hence $P\{X = 0\} = ({}^{5}c_{0})(0.316)^{0}(1 - 0.316)^{5}$

$$\approx 0.15$$

- (b) Now, let Y denote the number of bets for his first win

The question talks about the number of trials for the first win (success)

Hence Y follows geometric distribution with parameter $p = 0.316$

We have $P\{Y = i\} = (1 - p)^{i-1} p, i = 1, 2, 3, \ldots$

► We wish to find probability for the first win to occur on fourth bet

Hence required probability is

$P\{Y = 4\} = (1 - 0.316)^3 (0.316) \approx 0.1012$

Average number of trials required for the first success

$$E[X] = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Exponential random variable

- In practice, the exponential distribution often arises as the **distribution of the amount of time** until some specific event occurs

  For instance,

- the amount of time (starting from now) until an earthquake occurs, or
- a new war breaks out, or
- a telephone call you receive turns out to be a wrong number.

- Example: Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = 1/10$

  If someone arrives immediately ahead of you at a public telephone booth,

- Find the probability that you will have to wait

  (a) more than 10 minutes;

  (b) between 10 and 20 minutes.

A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

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

is said to be **an exponential random variable** (or, more simply, is said to be **exponentially distributed**) with parameter λ.

- Using the relation **$F(a) = \int_{-\infty}^{a} f(x)dx$,** we get the cumulative distribution function to be

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

- X denote the length of the call (in minutes) made by the person in booth

  Given that X is an exponential random variable with

  parameter $\lambda = 1/10$

We have,

$$F(a) = \begin{cases} 1 - e^{-a/10}, & \text{if } a \geq 0, \\ 0, & \text{else} \end{cases}$$

(a) $P\{X > 10\} = 1 - F(10) = e^{-1}$

(b) $P\{10 < X < 20\} = F(20) - F(10) = e^{-1} - e^{-2}$

Mean and variance

$$E[X] = \frac{1}{\lambda}$$

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

Memoryless property of exponential random variable

We say that a non-negative random variable X is **memoryless** if

$$P\{X > s + t | X > t\} = P\{X > s\} \text{ for all } s, t \geq 0$$

- Memoryless property means that the fact of having waited for t minutes gets “forgotten” and it does not affect the future waiting time.

- END