WESTMINSTER
International University in Tashkent

Discrete Probability Distributions

By

Olmos Isakov

Office hours: Tuesday, 09:00 – 11:00 (ATB 216)

AGENDA

- understand the concept of a discrete random variable.
- be able to find expected value and variance of discrete distributions.
- learn about probability mass function of Binomial distribution.
- understand Geometric distribution.

DISCRETE RANDOM VARIABLE

A random variable is **discrete** if it has a finite or countable number of possible outcomes that can be listed.

Experiment	Random Variable	Possible Values
Make 100 sales Calls	# of sales	0, 1, 2, … , 100
Inspect 70 Cell phones	# of defects	0, 1, 2, … , 70
Answer 33 Questions	# of correct	0, 1, 2, … , 33
Count Cars at Toll between 11:00 & 13:00	# of cars arriving	0, 1, 2, … , n

DISCRETE PROBABILITY DISTRIBUTIONS

- A **discrete probability distribution** lists each possible value the random variable can assume, together with its probability.
- *Ex*: the number of heads you get when you flip a coin three times

Possible outcomes	Probability	Explanation
0 head	**1/8 = 12.5%**	**{(T,T,T)}**
1 head	**3/8 = 37.5%**	**{(H,T,T), (T,H,T), (T,T,H)}**
2 heads	**3/8 = 37.5%**	**{(H,HT), (T,H,H), (H,T,H)}**
3 heads	**1/8 = 12.5%**	**{(H,H,H)}**

DISCRETE PROBABILITY DISTRIBUTIONS

A probability distribution must satisfy the following conditions:

In Words	*In Symbols*
1. The probability of each value of the discrete random variable is between 0 and 1, inclusive.	$0 \leq P(x) \leq 1$
2. The sum of all the probabilities is 1.	$\Sigma P(x) = 1$

Possible outcomes	Probability	Explanation
0 head	**1/8 = 12.5%**	**{(T,T,T)}**
1 head	**3/8 = 37.5%**	**{(H,T,T), (T,H,T), (T,T,H)}**
2 heads	**3/8 = 37.5%**	**{(H,HT), (T,H,H), (H,T,H)}**
3 heads	**1/8 = 12.5%**	**{(H,H,H)}**

MEASURES FOR DISCRETE DISTRIBUTIONS

1. Expected Value (Mean of probability distribution)
   - Weighted average of all possible values

$$\mu = E(x) = \Sigma x\, p(x)$$

2. Variance
   - Weighted average of squared deviation about mean

$$\sigma^2 = E[(x - \mu)^2] = \Sigma\, (x - \mu)^2\, p(x) = \Sigma x^2 p(x) - \mu^2$$

3. Standard Deviation: Square root of variance

$$\sigma = \sqrt{\sigma^2}$$

EXAMPLE

At a raffle, 500 tickets are sold for $1 each for two prizes of $100 and $50.

a. What is the expected value of your gain?

b. What is the standard deviation of your gain?

Solution

1. Your gain for the $100 prize is $100 - $1 = $99
2. Your gain for the $50 prize is $50 - $1 = $49
3. Remaining times your gain is -$1 (loss).

Gain, x	$P(x)$
$99	$\frac{1}{500}$
$49	$\frac{1}{500}$
–$1	$\frac{498}{500}$

Winning no prize

a. E(X) = μ = $\sum x * p(x) = 99*\frac{1}{500} + 49*\frac{1}{500} + (-1)*\frac{498}{500}$ = -$0.70

b. Variance = $\sigma^2 = \Sigma\,(x-\mu)^2\, p(x) = (99 + 0.70)^2*\frac{1}{500} + (49+0.70)^2*\frac{1}{500} +$

$+ (-1 + 0.70)^2*\frac{498}{500}$ = $24.91

Standard deviation = $\sigma = \sqrt{24.91} = \$4.99$

BINOMIAL DISTRIBUTION

A **binomial experiment** is a probability experiment that satisfies the following conditions.

1. Experiment consists of ***n*** identical and independent trials
2. Each trial results in one of two outcomes: success or failure

   P(success) = p

   P(failure) = q = 1 - p for all trials
3. The random variable of interest, X, is the number of successes in the n trials.
4. X has a binomial distribution with parameters ***n*** and ***p***

BINOMIAL PROBABILITY

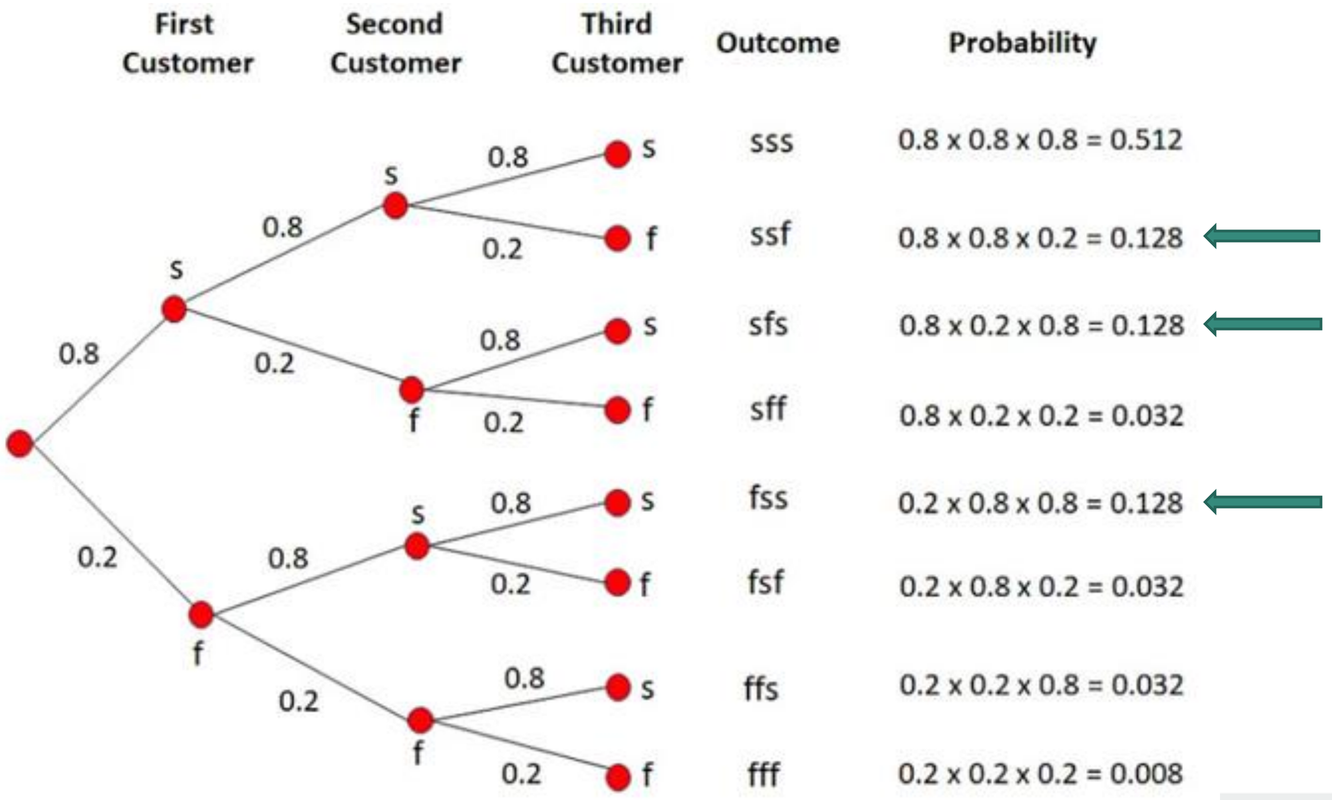
$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

Symbol	*Description*
$\boldsymbol{n}$	The number of times a trial is repeated.
$\boldsymbol{p = P(S)}$	The probability of success in a single trial.
$\boldsymbol{q = P(F)}$	The probability of failure in a single trial. $(q = 1 - p)$
$\boldsymbol{x}$	The random variable represents a count of the number of successes in n trials: $x = 0, 1, 2, 3, \ldots, n$.

EXAMPLE

Just imagine if the accuracy of taking order in a fast food restaurant A is approximately 80%. Suppose there are three customers who take order by drive-thru. What is the probability that two of the three customers' order will be taken correctly?

<u>This is a binomial experiment</u>: n = 3; p = 0.8; q = 0.2; x = 0, 1, 2, 3; X ~ Bin(3, 0.8)



EXAMPLE

Just imagine if the accuracy of taking order in a fast food restaurant A is approximately 80%. Suppose there are three customers who take order by drive-thru. What is the probability that two of the three customers' order will be taken correctly?

Binomial Distribution Formula

$$P\,(X = x) = \binom{n}{x}\ p^x\ (1 - p)^{n-x}$$

Outcome	Probability	
sss	$0.8 \times 0.8 \times 0.8 = 0.512$	
ssf	$\boxed{0.8 \times 0.8 \times 0.2 = 0.128}$	$\rightarrow (0.8)^2 \text{ X } (0.2)^1$
sfs	$\boxed{0.8 \times 0.2 \times 0.8 = 0.128}$	$\rightarrow (0.8)^2 \text{ X } (0.2)^1$
sff	$0.8 \times 0.2 \times 0.2 = 0.032$	
fss	$\boxed{0.2 \times 0.8 \times 0.8 = 0.128}$	$\rightarrow (0.8)^2 \text{ X } (0.2)^1$
fsf	$0.2 \times 0.8 \times 0.2 = 0.032$	
ffs	$0.2 \times 0.2 \times 0.8 = 0.032$	
fff	$0.2 \times 0.2 \times 0.2 = 0.008$	

$$3 \times (0.8)^2 \text{ X } (0.2)^1 \rightarrow \binom{3}{2} \times (0.8)^2 \text{ X } (0.2)^1$$

Binomial coefficient

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Number of x items that can be selected from a set of n items

Example

Suppose that a real estate agent, Jeanette Nelson, has 5 contacts, and she believes that for each contact the probability of making a sale is 0.40. Using binomial formula, do the following:

a) Find the probability that she makes at most 1 sale.

b) Find the probability that she makes between 2 and 4 sales (inclusive).

c) Graph the probability distribution function.

Example

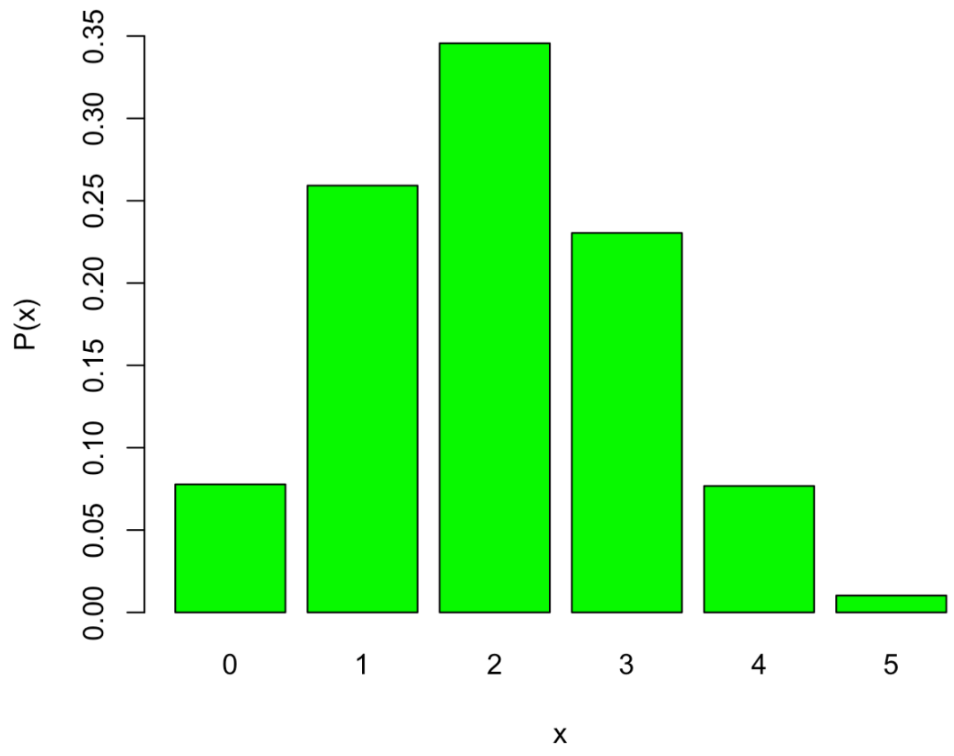
a) Find the probability that she makes at most 1 sale.

$P(x \leq 1) = P(x = 0) + P(x = 1) =$
${}_5C_0 * 0.4^0 * 0.6^5 + {}_5C_1 * 0.4^1 * 0.6^4 = 0.337$

b) Find the probability that she makes between 2 and 4 sales (inclusive).

$P(2 \leq x \leq 4) = P(x \leq 4) – P(x \leq 1) = 0.99 – 0.337 = 0.653$

c) Graph the probability distribution function.



MEAN, VARIANCE AND STANDARD DEVIATION

Population Parameters of a Binomial Distribution

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard deviation: $\sigma = \sqrt{npq}$

Example: One out of 5 students at a local college say that they skip breakfast in the morning. Find the mean, variance and standard deviation if 10 students are randomly selected.

$$n = 10$$
$$p = \frac{1}{5} = 0.2$$
$$q = 0.8$$

mean

$$\mu = np$$
$$= 10(0.2)$$
$$= 2$$

variance

$$\sigma^2 = npq$$
$$= (10)(0.2)(0.8)$$
$$= 1.6$$

st. deviation

$$\sigma = \sqrt{npq}$$
$$= \sqrt{1.6}$$
$$\approx 1.3$$

Geometric probability distribution

A random variable X that has a geometric distribution is denoted by

X ~ Geo(p), and the probability that first success occurs on the k^{th} trial is

$$\mathbf{P(X = k) = (1\text{-}p)^{k-1}p}, \quad k = 1, 2, \ldots$$

$P(\ X \le k) = 1 - q^k$ and $P(X > k) = q^k$ where $q = 1\text{-}p$

Example. In a particular country, 22% of cars are equipped with an anti-radar device to detect speed cameras. Cars are randomly selected and inspected for anti-radar device. Find the probability that the first car with anti-radar device is:

a. found on the fourth car inspected.
b. not one of the first 5 inspected.
c. one of the first 5 inspected.

Solution.

In a particular country, 22% of cars are equipped with an anti-radar system to detect speed cameras. Cars are randomly selected and inspected for anti-radar system. Find the probability that the first car with anti-radar system is:

Let X be the number of cars inspected up to and including the first car equipped with anti-radar system, then $X \sim Geo(0.22)$, where $p = 0.22$ and $q = 0.78$

a. found on the fourth car inspected.

$P(X = 4) = q^3 p = 0.78^3 * 0.22 = 0.104$

b. not one of the first 5 inspected.

$P(X > 5) = 0.78^5 = 0.289$

c. one of the first 5 inspected.

$P(X \leq 5) = 1 - 0.78^5 = 1 – 0.289 = 0.711$

Mean, Variance and Standard Deviation of Geometric distribution

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{1-p}{p^2} \qquad \sigma = \sqrt{\frac{1-p}{p^2}}$$

Geometric probability distribution is the only **memoryless** discrete probablity distribution.

$$\mathbf{P(X > m+n \mid P(X > m) = P(X > n), \quad m, n \geq 0}$$

Example. Suppose a computer server crashes randomly with probability p = 0.01 on each request. Then the number of requests until the first crash, X ~ Geo(0.01).

- Probability of no crash in the first 100 requests:

$P(X > 100) = (1 - 0.01)^{100} \approx 0.366$ (≈ 37% chance)

- Probability of no crash in the next 100 more requests given survival of 100 requests.

$$P(X > 200 \mid X > 100) = \frac{P(X > 200 \cap X > 100)}{P(X>100)} = \frac{P(X>200)}{P(X>100)} = \frac{0.99^{200}}{0.99^{100}} = 0.99^{100} \approx 0.366$$

A fair six-sided die is rolled repeatedly until the face “3” appears for the first time. Let X be the number of rolls required to obtain the first “3.” Which value of X has the highest probability of occurring?

x	1	2	3	4	5	6
p(x)	$\frac{1}{6} =$ 0.17	$\frac{5}{6} * \frac{1}{6} =$ 0.14	$(\frac{5}{6})^2 * \frac{1}{6} =$ 0.12	$(\frac{5}{6})^3 * \frac{1}{6} =$ 0.10	$(\frac{5}{6})^4 * \frac{1}{6} =$ 0.08	$(\frac{5}{6})^5 * \frac{1}{6} =$ 0.07

Excel codebook

= binom.dist(X, n, p, cumulative)

Example. $n = 3, p = 0.3$

$P(X = 1) = binom.dist(1, 3, 0.3, 0) = 0.441$

$P(X \leq 1) = binom.dist(1, 3, 0.3, 1) = 0.784$

	A	B	C
1			
2	0.441	=BINOM.DIST(1,3,0.3, 0)	
3	0.784	=BINOM.DIST(1,3,0.3, 1)	
4			

REFERENCES

1. Textbook (McClave). Chapter 4.
2. Textbook (Lind et al). Chapter 6.
3. https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Mostly_Harmless_Statistics_(Webb)/05%3A_Discrete_Probability_Distributions/5.03%3A_Geometric_Distributions

Thank You!