Computer Science

CSC196: Analyzing Data

Some Discrete Probability Distributions

Jason Pacheco and Cesim Erten

Outline

- Bernoulli and Binomial Distributions
- Multinomial Distribution
- Negative Binomial and Geometric Distributions
- Poisson Distribution

Outline

- Bernoulli and Binomial Distributions
- Multinomial Distribution
- Negative Binomial and Geometric Distributions
- Poisson Distribution

Discrete Probability Distributions

Can represent discrete probability distributions in a number of ways,

- Graphical representation
- Tabular format
- Formula

- A small collection of formulas can represent many instances

- We assign names to these formulas to have *named distributions*

- A handful of distributions describe many real-life phenomena

Discrete Probability Distributions

Example A study involving testing drug efficacy--we measure the number of cured patients. This follows a **binomial distribution.**

Example In an industrial setting we sample items from a production batch and test them. The number of defective items in the batch often follows a **hypergeometric distribution.**

Example In a sample of a patient's blood, the number of white blood cells follows a **Poisson distribution.**

Independent Binary Outcomes

- An experiment often consists of repeated trials where the outcomes are either *success* or *failure* (e.g. binary outcomes).
- Obvious application: Testing items as they come off an assembly line—they may be *defective* or *nondefective*.
- Process is referred to as a **Bernoulli process**
- Each trial is a **Bernoulli trial**

Bernoulli Distribution

*A.k.a. the **coinflip** distribution on <u>binary</u> RVs* $X \in \{0,1\}$ *is given by:*

$$P(X = x) = \text{Bernoulli}(X; p) = p^X (1-p)^{(1-X)}$$

Where $0 \leq p \leq 1$ *is the probability of **success** (e.g. heads), and also the mean:*

$$\mathbf{E}[X] = p \cdot 1 + (1-p) \cdot 0 = p$$

The variance is given by:

$$\text{Var}[X] = p(1-p)$$

Bernoulli Process

Must posses the following properties:

1. The experiment consists of repeated trials.
2. Each trial results in an outcome that may be classified as a *success* or *failure*.
3. The probability of successes, denoted p, remains constant from trial-to-trial.
4. The repeated trials are independent.

Bernoulli Process?

An urn contains 10 red balls and 10 green balls. A sequence of 5 balls are drawn from the urn **without replacement**. Clearly these are binary outcomes. Is this a Bernoulli process? Why or why not?

What if we select the balls **with replacement**?

Bernoulli Process

Example Consider the set of Bernoulli trials where three items are selected at random from a manufacturing process, inspected, and classified as defective or nondefective. The number of successes is a random variable X assuming integer values from 0 through 3. The possible outcomes and the corresponding values of X are:

Outcome	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
x	0	1	1	1	2	2	2	3

Bernoulli Process

Outcome	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
$\boldsymbol{x}$	0	1	1	1	2	2	2	3

Since the items are selected independently and we assume that the process produces 25% defectives, we have

$$P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{64}.$$

Similar calculations yield the probabilities for the other possible outcomes. The probability distribution of X is therefore

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Binomial Distribution

The number of successes X in n independent Bernoulli trials follows a **Binomial distribution**, denoted:

$$P(X = x) = \text{Binomial}(X; n, p)$$

- If there are X successes then there are *(n-X)* failures
- Since trials are independent we can multiply all probabilities
- Each success occurs with probability p
- Each failure occurs with probability *(1-p)*
- Therefore, the probability of any sequence in order is:

$$p^X (1 - p)^{(n-X)}$$

Binomial Distribution

We don't care about order. How many combinations of X successes in n trials?

$$\binom{n}{X}$$

Therefore, we have that the Binomial PMF is given by:

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

With mean and variance given by:

$$\mathbf{E}[X] = np, \qquad \text{Var}[X] = np(1-p)$$

Example

Example 5.1: The probability that a certain kind of component will survive a shock test is 3/4. Find the probability that exactly 2 of the next 4 components tested survive.

***Solution*:** Assuming that the tests are independent and $p = 3/4$ for each of the 4 tests, we obtain

$$b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \left(\frac{4!}{2!\,2!}\right)\left(\frac{3^2}{4^4}\right) = \frac{27}{128}.$$

Example

Example 5.2: The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive

***Solution*:** Let X be the number of people who survive.

(a) $$P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662$$

$$= 0.0338$$

Outline

- Bernoulli and Binomial Distributions
- Multinomial Distribution
- Negative Binomial and Geometric Distributions
- Poisson Distribution

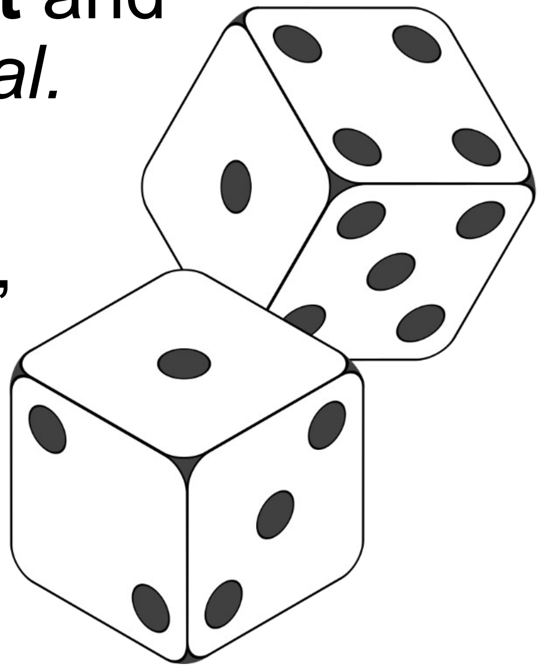
Multinomial Experiment

- If there are more than 2 outcomes then we have a **multinomial experiment.**

- Can think of this as consecutive rolls of a die.

- Suppose we draw a card from a deck **with replacement** and count the number of each suit we see. *This is multinomial.*

- Suppose we have an urn with 10 balls of each color red, black, green, blue. The number of each color balls drawn **with replacement** is *multinomial*.



Multinomial Distribution

Suppose event E_1 occurs X_1 times, E_2 occurs X_2 times, and so on until E_k occurs X_k times. Suppose also that there are n total trials, e.g.:

$$X_1 + X_2 + \ldots + X_k = n$$

Assume events E_1, E_2, ..., E_k occur independently with probabilities p_1, p_2, ..., p_k such that:

$$p_1 + p_2 + \ldots + p_k = 1$$

Then we say that X_1, X_2, ..., X_k follow a Multinomial distribution:

$$P(X_1, X_2, \ldots, X_k; p_1, p_2, \ldots, p_k, n) = \text{Multinomial}(X_1, X_2, \ldots, X_k; p_1, p_2, \ldots, p_k, n)$$

Counting Sample Points

Recall that the number of permutations of n things of which x_1 are of one kind, x_2 are another kind, and so on until x_k are of the last kind is given by:

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! x_2! \ldots x_k!}$$

Example 2.20: In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?

***Solution*:** Directly using Theorem 2.4, we find that the total number of arrangements is

$$\frac{10!}{1!\ 2!\ 4!\ 3!} = 12,600.$$

Multinomial Distribution

Often, we use the shorthand:

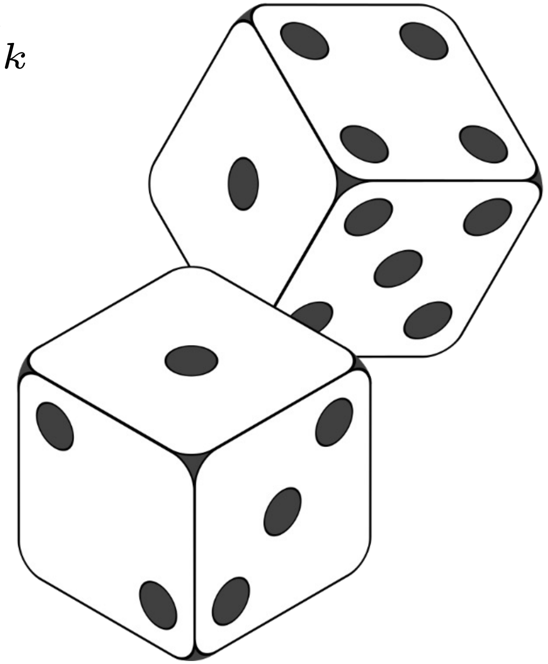
$$P(X; p, n) = \text{Multinomial}(X; p, n)$$

The PMF is given by:

$$\text{Multinomial}(X; p, n) = \binom{n}{X_1, X_2, \ldots, X_k} p_1^{X_1} p_2^{X_2} \ldots p_k^{X_k}$$

Where:

$$\sum_{i=1}^{k} X_k = n, \qquad \text{and} \qquad \sum_{i=1}^{k} p_i = 1$$



Example

Example 5.7: The complexity of arrivals and departures of planes at an airport is such that computer simulation is often used to model the “ideal” conditions. For a certain airport with three runways, it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet:

Runway 1: $p_1 = 2/9$,
Runway 2: $p_2 = 1/6$,
Runway 3: $p_3 = 11/18$.

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

Runway 1: 2 airplanes,
Runway 2: 1 airplane,
Runway 3: 3 airplanes

Example

***Solution*:** Using the multinomial distribution, we have

$$f\left(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6\right)=\binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3$$

$$=\frac{6!}{2!\,1!\,3!}\cdot\frac{2^2}{9^2}\cdot\frac{1}{6}\cdot\frac{11^3}{18^3}=0.1127.$$

Outline

- Bernoulli and Binomial Distributions
- Multinomial Distribution
- Negative Binomial and Geometric Distributions
- Poisson Distribution

Thought Experiment…

What is the probability of seeing 3 heads in 4 flips?

What are the possible outcomes?

Last flip is always heads…

$$HHTH, HTHH, THHH$$

How many outcome (combinations) are there?

$$\binom{3}{2}$$

What is the probability of each outcome?

$$p^3(1-p) = (0.5)^3(1-0.5)$$

Example

Example A drug is effective 60% of the time that it is used. What is the probability that we have to test 7 patients to cure 5?

Let *S* designate a success, and *F* a failure. What is the probability of the sequence, "*SFSSSFS*":

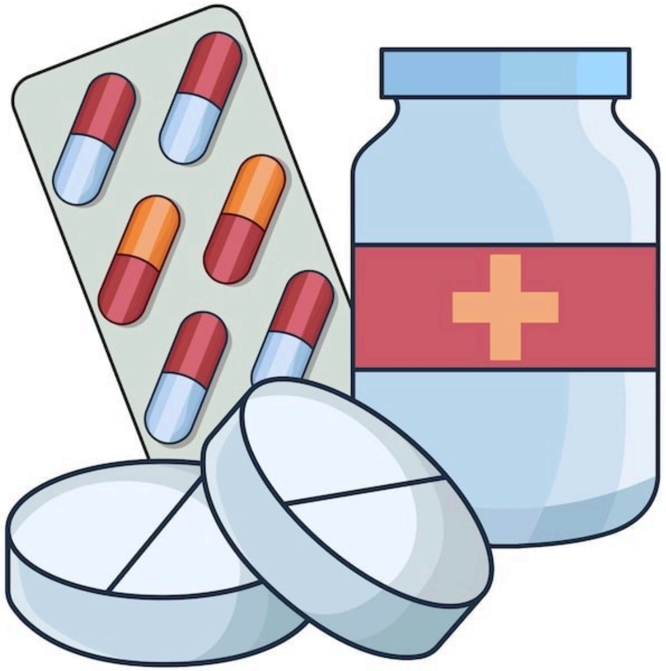
$$(0.6)(0.4)(0.6)(0.6)(0.6)(0.4)(0.6)$$

How many combinations of 5 successes in 7 trials?

$$\binom{6}{4}$$

Putting this all together we have the probability:

$$\binom{6}{4}(0.6)^5(0.4)^2$$



Thought Experiment…

What is the probability that the k^{th} head (success) occurs on the x^{th} flip?

How many combinations of k successes in x flips?

$$\binom{x-1}{k-1}$$

What is the probability of each outcome?

$$p^k(1-p)^{(x-k)}$$

Put this all together we have the probability…

$$\binom{x-1}{k-1}p^k(1-p)^{(x-k)}$$

Negative Binomial Distribution

The number *X* of trials to produce *k* successes is called a **negative binomial random variable**. Its distribution is denoted,

$$\mathrm{NB}(X;k,p) = \binom{x-1}{k-1} p^k (1-p)^{(x-k)}$$

Note that for the sequence “SFFS” the probability is,

$$p(1-p)^2 p = p^2 (1-p)^2$$

Likewise, the probability of *any* sequence with k successes is:

$$p^k (1-p)^{(x-k)}$$

Example

Example 5.14: In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B.

(a) What is the probability that team A will win the series in 6 games?

$$\begin{aligned} \text{NB}(X = 6; k = 4, p = 0.55) &= \binom{x-1}{k-1} p^k (1-p)^{(x-k)} \\ &= \binom{5}{3} (0.55)^4 (1-0.55)^2 = 0.1853 \end{aligned}$$

(b) What is the probability that team A will win the series?

Example

(b) What is the probability that team A will win the series?

They can either win in 4, 5, 6, or 7 games so…

$$P(\text{Team A wins the championship}) = \ldots$$
$$= \text{NB}(4; 4, 0.55) + \text{NB}(5; 4, 0.55) + \text{NB}(6; 4, 0.55) + \text{NB}(7; 4, 0.55)$$

Note: The book uses the notation,

$$b^*(X; k, p)$$

A slightly different question…

What is the probability that we have to flip X times until the first heads (success)?

There is only one outcome that satisfies this event:

$$TTT\dots TTH$$

X-1 tails, 1 heads…

The probability of this event is:

$$\mathrm{NB}(X;1,p) = (1-p)^{(X-1)}p$$

We call this the ***Geometric distribution****…*

Example

Example 5.15: For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Outline

- Bernoulli and Binomial Distributions
- Multinomial Distribution
- Negative Binomial and Geometric Distributions
- Poisson Distribution

Poisson Distribution

A **Poisson random variable** *X* represents the number of outcomes in a specified time interval or region.

- The given interval may be any length: minute, hour, day, month, …
- The region may be: line segment, an area, a volume, …

Example The number of telephone calls received per hour at an office.

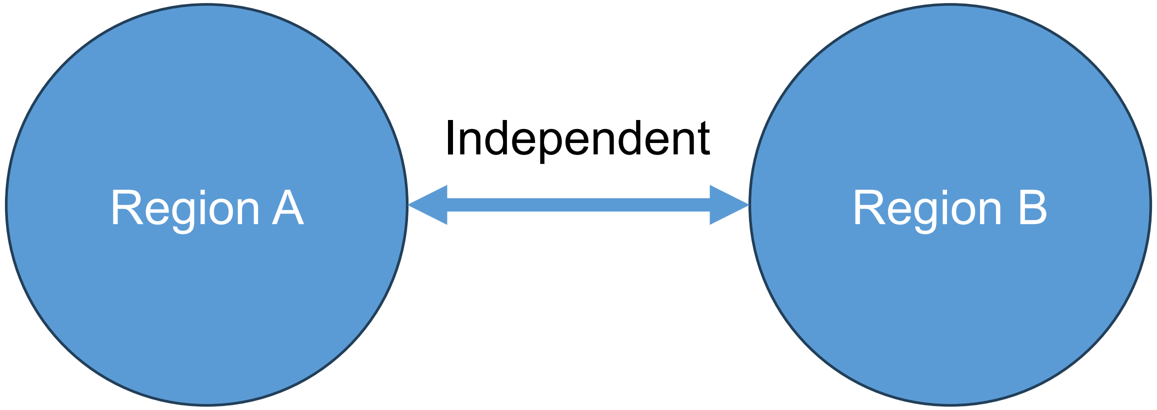
Example The number of games postponed due to rain per baseball season.

Example The number of field mice per acre of land.

Properties of the Poisson Process

Such an experiment follows a **Poisson process**, defined by the following three properties:

1. The number of outcomes occurring in one time interval or specified region of space is <u>independent</u> of the number that occur in any other disjoint time interval or region. In this sense we say that **the Poisson process has no memory**.



Properties of the Poisson Process

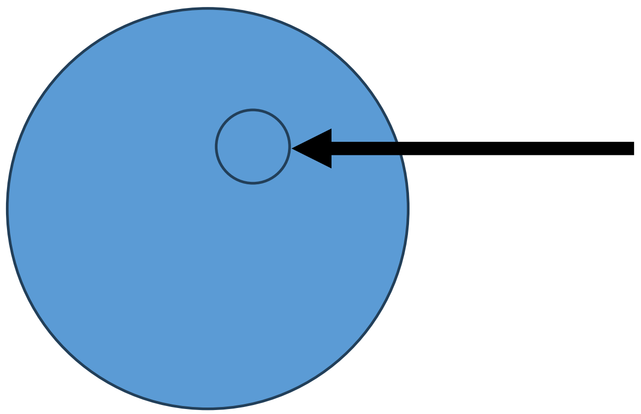
Such an experiment follows a **Poisson process**, defined by the following three properties:

2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.

Properties of the Poisson Process

Such an experiment follows a **Poisson process**, defined by the following three properties:

3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.



Poisson Distribution

Definition A discrete random variable *X* has a Poisson distribution with *rate parameter* $\lambda > 0$ if it has probability mass function (PMF):

$$\mathrm{Pois}(X; \lambda) = \frac{\lambda^{X} e^{-\lambda}}{x!}$$

Where,

- *X* is the number of ocurrences(*X* = 0, 1, 2, …)
- *e* is **Euler's number** *e* = *2.71828…*
- *X! = X(X-1)(X-2)…* is the **factorial** operator

Poisson Distribution

The rate parameter controls, both, the expected value and the variance:

$$\mathbf{E}[X] = \lambda \qquad \mathrm{Var}[X] = \lambda$$

Note: the book uses slightly different notation for the PMF,

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

I will largely avoid this notation in lecture…

And its moments,

$$\mathbf{E}[X] = \lambda t \qquad \mathrm{Var}[X] = \lambda t$$

Example

Example 5.17: During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

$$\text{Pois}(X = 6; \lambda = 4) = \frac{4^6 e^{-6}}{4!} = 0.1042$$

Example

Example 5.18: Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} \text{Pois}(x; 10) = 0.0487$$