LECTURE 13

Discrete Random Variables

Numerical functions of random samples and their properties

CSCI 3022 @ CU Boulder

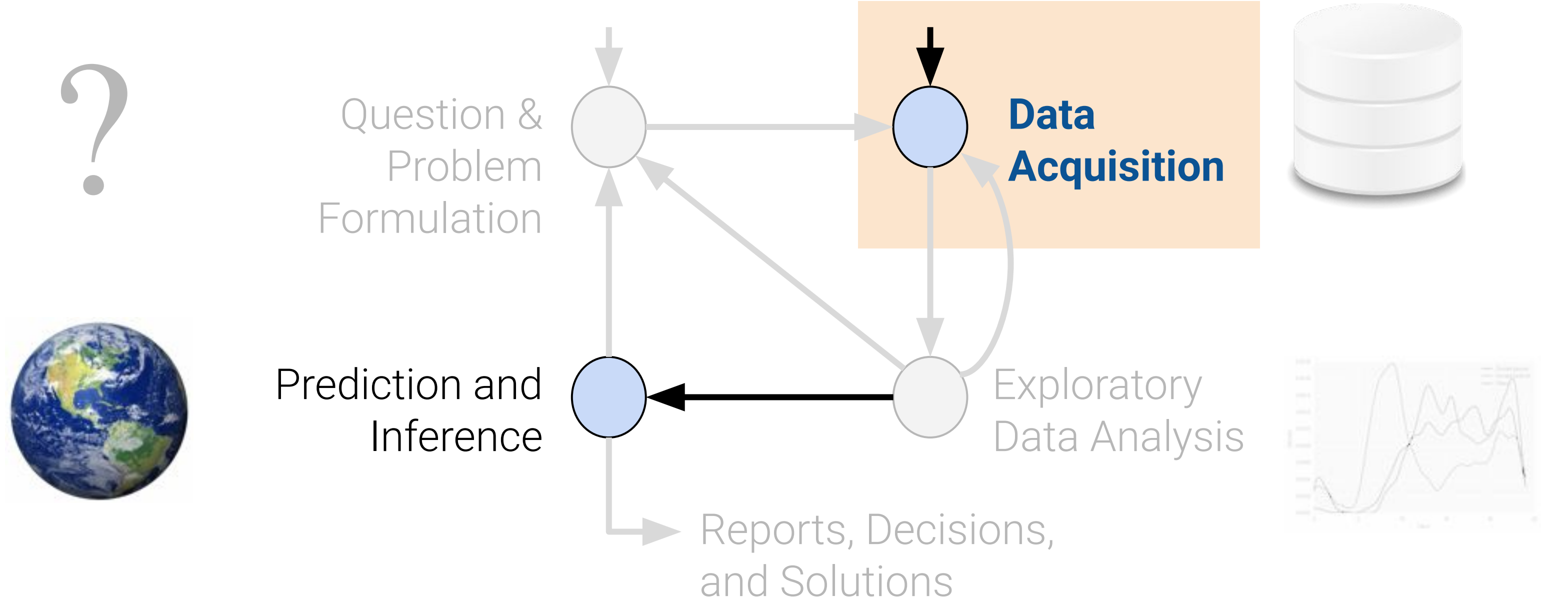
Maribeth Oscamou

BY NC SA

Announcements

- HW 5 due tomorrow (11:59pm MT)
- Quiz 5 Friday (scope: HW 4, Lessons 5-8)

Why Random Variables?



We will go over **just enough about Random Variables** to help you understand implications for modeling.

Today's Roadmap

- **Random Variables and Distributions**
- Common Discrete RV:
    - Bernoulli
    - Binomial
    - Poisson
- Expectation and Variance

[Extra Slides] Derivations

BY NC SA

[Terminology] Random Variable

Suppose we draw a random sample of size n from a population.

A **random variable** is a numerical function of a sample.

sample was drawn at random

value depends on how the sample came out

- Often denoted with uppercase “variable-like” letters (e.g. X, Y).
- Domain (input): all random samples of size n
- Range (output): number line

[Terminology] Random Variable

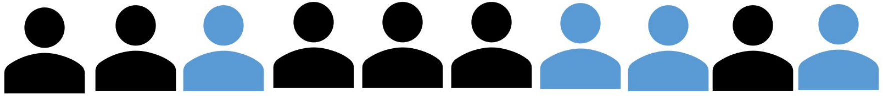
Suppose we draw a random sample of size n from a population.

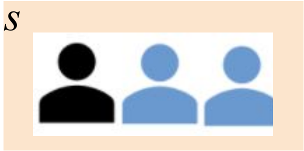
A **random variable** is a **numerical function** of a sample.

sample was drawn at random   value depends on how the sample came out

- Often denoted with uppercase “variable-like” letters (e.g. X, Y).
- Also known as a sample statistic, or **statistic**. (next lecture).
- Domain (input): all random samples of size n
- Range (output): number line

Suppose you draw a random sample s of size 3 from the following population:

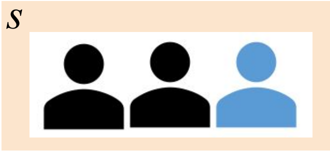




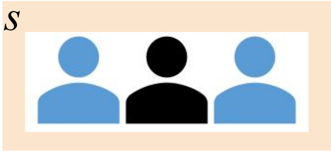
$X(s) = 2$

Define X = **# of blue people**.

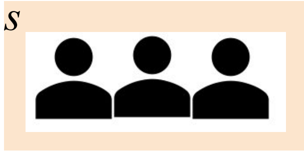
X is a random variable!



$X(s) = 1$



$X(s) = 2$



$X(s) = 0$

BY NC SA

Discrete Random Variable

A **discrete random variable** X is a function that maps the elements of the sample space Ω to a **finite number** of values $a_1, a_2, \dots, a_n$ **or** a ***countably*** infinite number of values.

Ex). Which of the following would typically be considered ***discrete*** random variables? Select all that apply.

A). The number of people who check out at a grocery line in a given hour.

B). The finish times of randomly chosen runners from the Bolder Boulder 10K.

C). The number of games played in the best of 7 NBA playoffs.

D). The weight of dogs taken from a random sample around Boulder.

E). The volume of water in randomly chosen Colorado lakes.

CC BY NC SA

Discrete Random Variables

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

<u>Example</u>: Suppose you roll two dice.

Sample Space:

$\Omega=\{(1, 1), (1, 2), (1, 3), \ldots,(2, 1),\ldots,(6, 5), (6, 6)\}$

What are some examples of discrete random variables whose domain is this sample space?

BY NC SA

Discrete Random Variables

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

<u>Example</u>: Suppose you roll two dice.

Sample Space:

$\Omega=\{(1, 1), (1, 2), (1, 3), \ldots,(2, 1),\ldots,(6, 5), (6, 6)\}$

Let Y be the discrete random variable that is the sum of the two dice

BY NC SA

Discrete Random Variables

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Example: Suppose you roll two dice.

Sample Space:

Ω={(1, 1), (1, 2), (1, 3), …,(2, 1),…,(6, 5), (6, 6)}

Let Y be the discrete random variable that is the sum of the two dice

$$\mathrm{Y}(\omega_1, \omega_2) = \omega_1 + \omega_2 = k \quad \text{for } (\omega_1, \omega_2) \in \Omega, \quad k \in \{2, 3, \dots, 12\}$$

cc BY NC SA

Discrete Random Variables

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Example: Suppose you roll two dice.

Sample Space:

$\Omega=\{(1, 1), (1, 2), (1, 3), \ldots,(2, 1),\ldots,(6, 5), (6, 6)\}$

Let Y be the discrete random variable that is the sum of the two dice. $Y(\omega_1, \omega_2) = \omega_1 + \omega_2 = k$ for $(\omega_1, \omega_2) \in \Omega,\ \ k \in \{2, 3, \ldots, 12\}$

cc BY NC SA

Discrete Random Variables

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

<u>Example</u>: Suppose you roll two dice.

Sample Space:

$\Omega=\{(1, 1), (1, 2), (1, 3), \ldots,(2, 1),\ldots,(6, 5), (6, 6)\}$

Let Y be the discrete random variable that is the sum of the two dice. $Y(\omega_1,\omega_2) = \omega_1 + \omega_2 = k \quad$ for $(\omega_1,\omega_2) \in \Omega, \;\; k \in \{2, 3, \ldots, 12\}$

What are all elements in the sample space that Y maps to an output of 8?

CC BY NC SA

Discrete Random Variables

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Example: Suppose you roll two dice.

Sample Space:

$\Omega=\{(1, 1), (1, 2), (1, 3), \ldots,(2, 1),\ldots,(6, 5), (6, 6)\}$

Let Y be the discrete random variable that is the sum of the two dice.

$$Y(\omega_1,\omega_2) = \omega_1 + \omega_2 = k \quad \text{for } (\omega_1,\omega_2) \in \Omega, \quad k \in \{2, 3, \ldots, 12\}$$

What are all elements in the sample space that Y maps to an output of 8?

$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

cc BY NC SA

Discrete Random Variables

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Example: Suppose you roll two dice.

Sample Space:

Ω={(1, 1), (1, 2), (1, 3), …,(2, 1),…,(6, 5), (6, 6)}

Let Y be the discrete random variable that is the sum of the two dice.

$$Y(\omega_1, \omega_2) = \omega_1 + \omega_2 = k \quad \text{for } (\omega_1, \omega_2) \in \Omega, \quad k \in \{2, 3, \dots, 12\}$$

What are all elements in the sample space that Y maps to an output of 8?

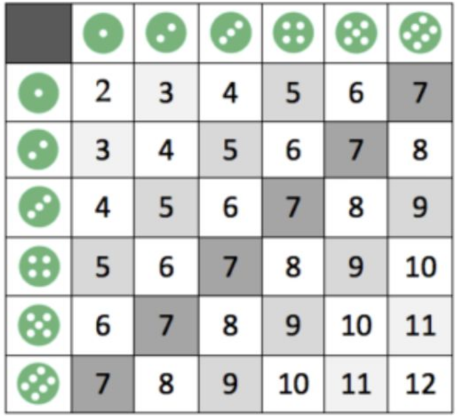
$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

What is P(Y=8)?

BY NC SA

We can define the probability of the event $\{Y = k\}$ denoted $P(Y = k)$, for any k:



$P(Y = 2) = P(\{(1,1)\}) = 1/36$

$P(Y = 3) = P(\{(1,2),(2,1)\}) = 1/18$

$P(Y = 4) = P(\{(1,3),(2,2),(3,1)\}) = 1/12$

$P(Y = 5) = P(\{(1,4),(2,3),(3,2),(4,1)\}) = 1/9$

$P(Y = 6) = P(\{(1,5),(2,4),(3,3),(4,2),(5,1)\}) = 5/36$

$P(Y = 7) = P(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}) = 1/6$

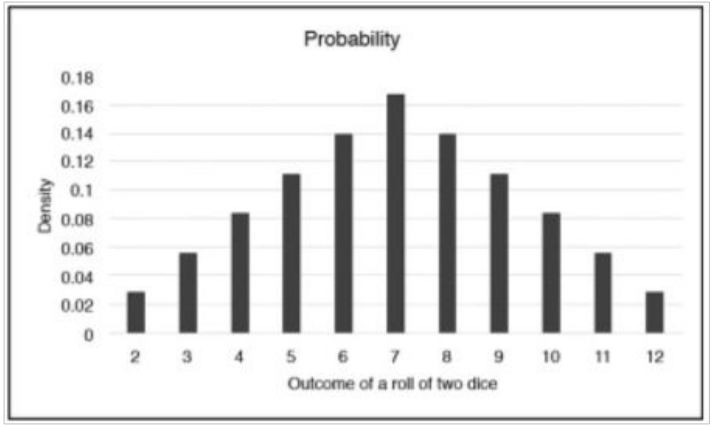
$P(Y = 8) = P(\{(2,6),(3,5),(4,4),(5,3),(6,2)\}) = 5/36$

$P(Y = 9) = P(\{(3,6),(4,5),(5,4),(6,3)\}) = 1/9$

$P(Y = 10) = P(\{(4,6),(5,5),(6,4)\}) = 1/12$

$P(Y = 11) = P(\{(5,6),(5,6),\}) = 1/18$

$P(Y = 12) = P(\{(6,6)\}) = 1/36$



BY NC SA

Probability Distribution: Discrete Random Variable

A **probability distribution** of a **discrete** random variable is a function, f, that maps a random variable's values $a_1, a_2, a_3, \ldots$ to the probabilities of those values:

$$f(a_k) = P(X = a_k) \quad \text{for } k = 1, 2, 3, \ldots$$

A **probability distribution of a discrete** random variable satisfies the following two conditions:

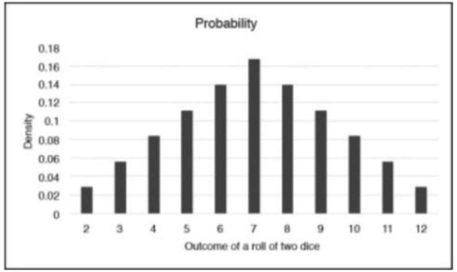
1. for $k = 1, 2, 3, \ldots$ 2.

We can represent the probability distribution in the following ways:

a). Distribution Table

a	2	3	4	5	6	7	8	9	10	11	12
f(a)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

B). Histogram



C). * Sometimes* as a closed-form discrete function

[Terminology] Distribution

The **distribution** of a random variable X is a description of how the total probability of 100% is split over all the possible values of X.

A distribution fully defines a random variable.

Assuming (for now) that X is **discrete**, i.e., has a finite number or countably infinite possible values:

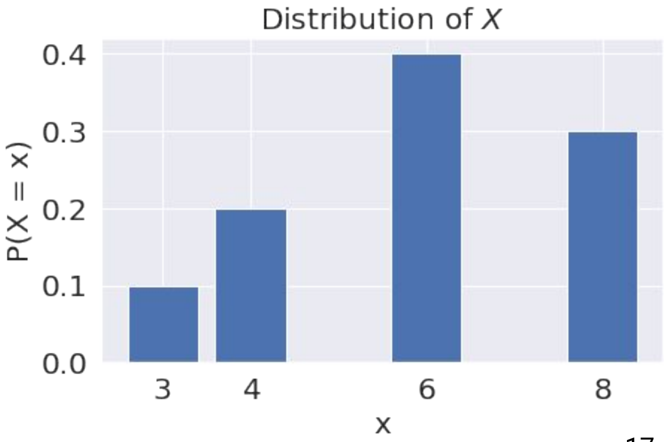
$$\boxed{P(X = x)}$$

The probability that random variable X takes on the value x.

$$\sum_{\text{all } x} P(X = x) = 1$$

Probabilities must sum to 1.

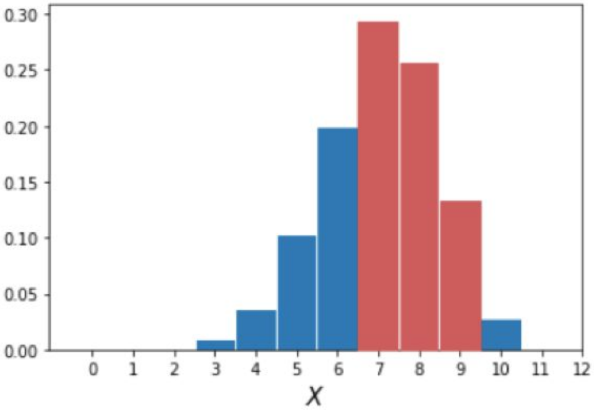
x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3



CC BY NC SA

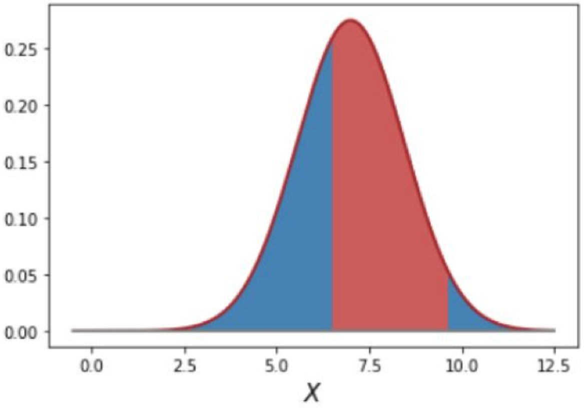
Probabilities are Areas of Histograms

Distribution of **discrete** random variable X



The area of the red bars is **P(7 <= X <= 9)**.

Distribution of **continuous** random variable Y



The red area under the curve is **P(6.8 <= Y <= 9.5)**.

BY NC SA

Understanding Discrete Random Variables

Compute the following probabilities for the random variable X.

1. $P(X = 4) =$

2. $P(X < 6) =$

3. $P(X \leq 6) =$

4. $P(X = 7) =$

5. $P(X \leq 8) =$

x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3

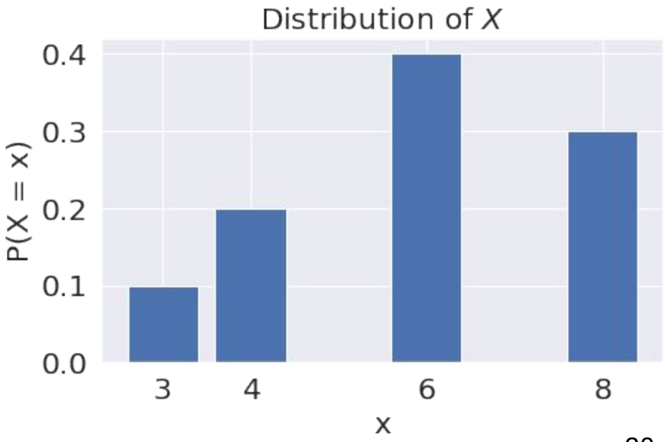
cc BY NC SA

Understanding Discrete Random Variables

Compute the following probabilities for the random variable X.

1. $P(X = 4) =$ 0.2
2. $P(X < 6) =$ $0.1 + 0.2 = 0.3$
3. $P(X \leq 6) =$ $0.1 + 0.2 + 0.4 = 0.7$
4. $P(X = 7) =$ 0
5. $P(X \leq 8) =$ 1

x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3



cc BY NC SA

Some Common Discrete Random Variables

Common Discrete Random Variables

Bernoulli(p)

- Takes on value 1 with probability p, and 0 with probability 1 - p
- AKA the "indicator" random variable.

Binomial(n, p)

- Number of 1s in n independent Bernoulli(p) trials
- Probabilities given by the binomial formula

Poisson(u)

- Number of random events per unit of measurement

We'll go over these in detail.
The rest are provided for your reference.

Uniform on a finite set of values

- Probability of each value is 1 / (size of set)
- For example, a standard die

Geometric(p)

The numbers in parentheses are the **parameters** of a random variable, which are constants. Parameters define a random variable's shape (i.e., distribution) and its values.

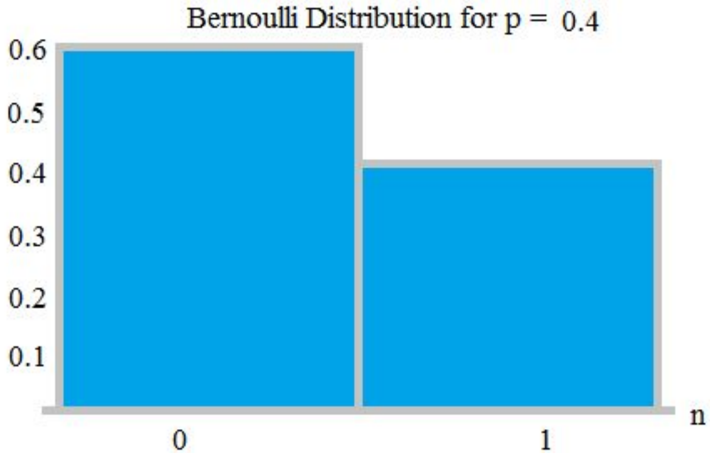
CC BY NC SA

Bernoulli Distribution

The Bernoulli distribution is used to model experiments with only two possible outcomes; often referred to as “success” and “failure”, and encoded as 1 and 0, respectively.

A discrete random variable X has a **Bernoulli distribution,** denoted $X \sim Ber(p)$, where $0 \leq p \leq 1$, if its probability distribution is given by

$$\begin{aligned} P(X = 1) &= p \\ P(X = 0) &= 1 - p \end{aligned}$$



Examples of Bernoulli Random Variables:

- Outcome from flipping a coin
- Getting an answer correct when guessing on a multiple choice question
- Buying a winning lottery ticket
- Asking someone if they’re left-handed

Binomial Distribution

A discrete random variable X has a **Binomial Distribution,** denoted $X \sim Bin(n, p)$, with $n = 1, 2, \ldots$ and $0 \leq p \leq 1$, if its probability distribution is given by

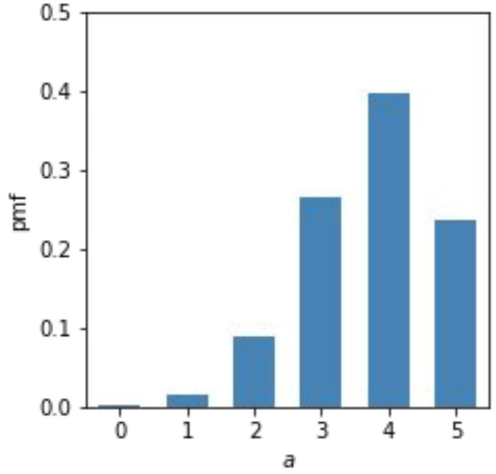
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } 0 \leq k \leq n$$

p **= probability of success for each of the n trials**

k = number of successful trials out of n total

s

Example: n=5, p=0.75



Examples of Binomial Random Variables:

- Number of correct guesses on a multiple choice test when you randomly guess all answers
- Number of winning lottery tickets when you buy 10 tickets of the same kind
- Number of left-handers in a randomly selected sample of 100 unrelated people

BY NC SA

Binomial Distribution

Ex: Suppose a coin is weighted such that its 3 times as likely to land on heads than tails. You flip this coin 4 times in a row. Let X be the random variable that describes the number of times the coin lands on tails.

a). State the values that the random variable can equal with non-zero probability.

b). Give the distribution for X

cc BY NC SA

Binomial Distribution

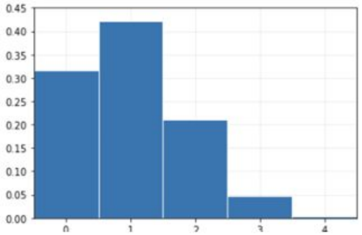
Ex: Suppose a coin is weighted such that its 3 times as likely to land on heads than tails. You flip this coin 4 times in a row. Let X be the random variable that describes the number of times the coin lands on tails.

a). State the values that the random variable can equal with non-zero probability.

b). Give the distribution for X

- As a closed-form function: $f(k) = \binom{4}{k}(.25)^k(.75)^{4-k}$ for $k = 0, 1, 2, 3, 4$
- As a table:

a	0	1	2	3	4
$f(a)$	$\frac{81}{256}$	$\frac{108}{256}$	$\frac{54}{256}$	$\frac{12}{256}$	$\frac{1}{256}$



The probability of 0 Tails out of 4 flips = $\binom{4}{0}(.25)^0(.75)^4 = \frac{81}{256} \approx .316$

The probability of 1 Tail out of 4 flips = $\binom{4}{1}(.25)^1(.75)^3 = \frac{108}{256} \approx .422$

The probability of 2 Tails out of 4 flips= $\binom{4}{2}(.25)^2(.75)^2 = \frac{54}{256} \approx .211$

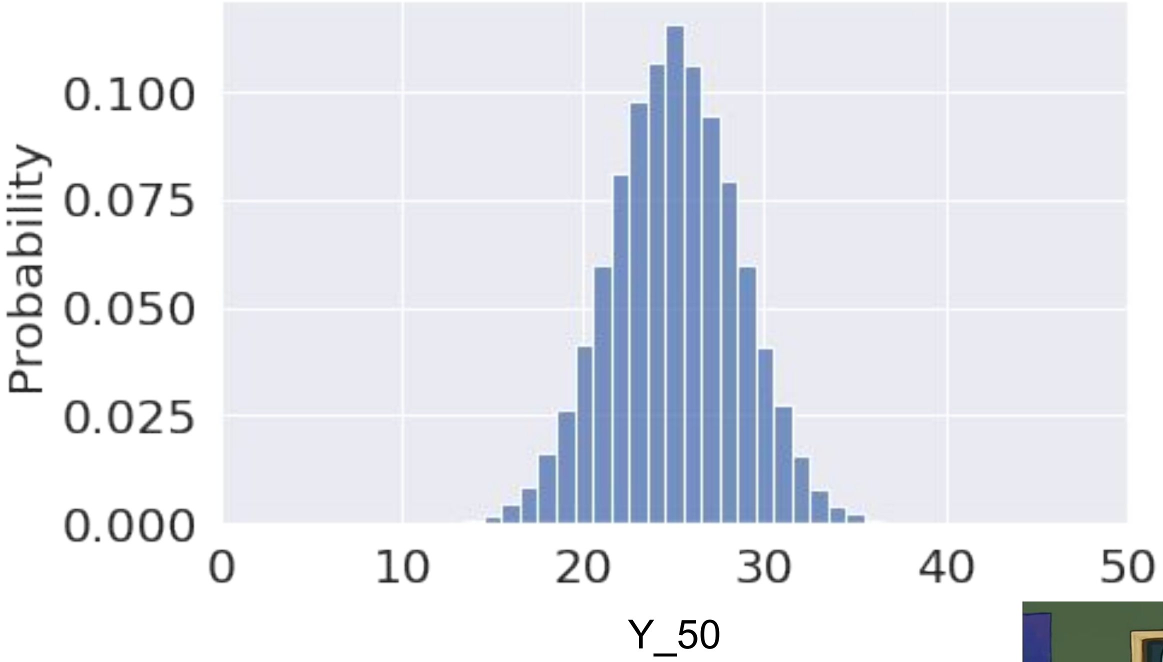
The probability of 3 Tails out of 4 flips = $\binom{4}{3}(.25)^3(.75)^1 = \frac{12}{256} \approx .047$

The probability of 4 Tails out of 4 flips = $\binom{4}{4}(.25)^4(.75)^0 = \frac{1}{256} \approx .004$

CC BY NC SA

Binomial(n, p) for large n

For p = 0.5, n = 50 (i.e. number of heads in 50 fair coin flips):



CC BY NC SA

Poisson Distribution

A discrete random variable X has a **Poisson distribution** denoted $X \sim Pois(\mu)$ with parameter $\mu > 0$, if its distribution is given by

$$P(X = k) = \frac{\mu^k e^{-\mu}}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

(i.e. probability of k events per unit of measurement)

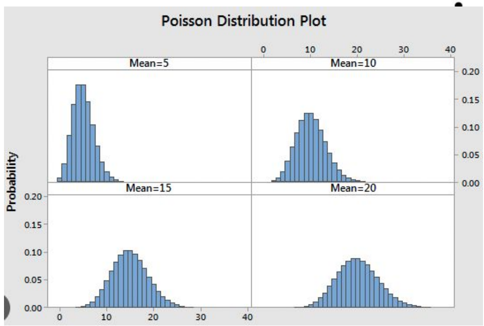
$\boldsymbol{\mu}$ is the expected number of events per unit of measurement (time period, length, space, volume,etc)

Examples that are modeled with a Poisson Distribution:

- The number of car accidents at a given intersection in a month
- The number of pieces of mail received in a day
- The number of people arriving at a restaurant during a shift
- The number of phone calls per hour received by a call center
- The number of meteors striking the earth each year

Assumptions used when deriving Poisson distribution:

- Probability of observing a single event over a small interval of time is proportional to the size of the interval.
- Each event/arrival is independent.



cc BY NC SA

Poisson Distribution: Derivation

Example: Suppose between 5pm-6pm on a weekday, a given check-out line a King Soopers checks out on average 10 people. What is the probability that on a given day exactly 8 people are checked out at that check-out line between 5pm-6pm?

CC BY NC SA

Poisson Distribution: Derivation

<u>Example</u>: Suppose between 5pm-6pm on a weekday, a given check-out line a King Soopers checks out on average 10 people. What is the probability that on a given day exactly 8 people are checked out at that check-out line between 5pm-6pm?

Solution:

Let X=the number who are checked at that check-out line between 5pm-6pm (note X is a discrete random variable)

Break the hour into **n subintervals** with n small enough that it's unlikely that more than one person will go through the checkout line during that time interval.

P(one person is checked out during a subinterval) = $\frac{10}{n}$

P(no one is checked out during a subinterval) = $1-\frac{10}{n}$

CC BY NC SA

Poisson Distribution: Derivation

<u>Example</u>: Suppose between 5pm-6pm on a weekday, a given check-out line a King Soopers checks out on average 10 people. What is the probability that on a given day exactly 8 people are checked out at that check-out line between 5pm-6pm?

Solution:

Let X=the number who are checked at that check-out line between 5pm-6pm (note X is a discrete random variable)

Break the hour into **n subintervals** with n small enough that it's unlikely that more than one person will go through the checkout line during that time interval.

P(one person is checked out during a subinterval) = $\frac{10}{n}$

P(no one is checked out during a subinterval) = $1 - \frac{10}{n}$

Derivation:

$$P(X=8) = \lim_{n\to\infty} \binom{n}{8} \left(\frac{10}{n}\right)^8 \left(1-\frac{10}{n}\right)^{n-8} = \lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-7)}{8!} \frac{10^8}{n^8} \frac{\left(1-\frac{10}{n}\right)^n}{\left(1-\frac{10}{n}\right)^8}$$

$$= \frac{10^8}{8!} \lim_{n\to\infty} 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots.\left(1-\frac{7}{n}\right) \frac{\left(1-\frac{10}{n}\right)^n}{\left(1-\frac{10}{n}\right)^8} =$$

CC BY NC SA

Poisson Distribution: Derivation

Example: Suppose between 5pm-6pm on a weekday, a given check-out line a King Soopers checks out on average 10 people. What is the probability that on a given day exactly 8 people are checked out at that check-out line between 5pm-6pm?

Solution:

Let X=the number who are checked at that check-out line between 5pm-6pm (note X is a discrete random variable)

Break the hour into **n subintervals** with n small enough that it's unlikely that more than one person will go through the checkout line during that time interval.

P(one person is checked out during a subinterval) = $\frac{10}{n}$

P(no one is checked out during a subinterval) = $1-\frac{10}{n}$

Derivation:

$$P(X=8)=\lim_{n\to\infty}\binom{n}{8}\left(\frac{10}{n}\right)^8\left(1-\frac{10}{n}\right)^{n-8}=\lim_{n\to\infty}\frac{n(n-1)(n-2)\ldots(n-7)}{8!}\,\frac{10^8}{n^8}\,\frac{\left(1-\frac{10}{n}\right)^n}{\left(1-\frac{10}{n}\right)^8}$$

$$=\frac{10^8}{8!}\lim_{n\to\infty}1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots.\left(1-\frac{7}{n}\right)\frac{\left(1-\frac{10}{n}\right)^n}{\left(1-\frac{10}{n}\right)^8}=\frac{10^8e^{-10}}{8!}$$

Thus, in general if the **average number** of occurrences during an interval is μ
and X = *the actual number of observed occurrences in an interval*
then X has a **Poisson Distribution** given by: $P(X=k)=\frac{\mu^k e^{-\mu}}{k!}$ for $k=0,1,2,\ldots$

CC BY NC SA

Expectation and Variance

Random Variables and Distributions

Expectation and Variance

BY NC SA

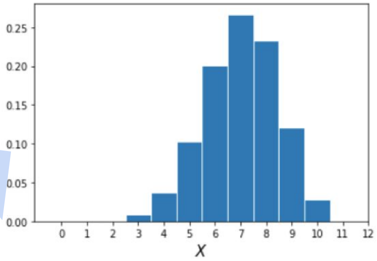
Descriptive Properties of Random Variables

There are several ways to describe a random variable:

Closed form formula (sometimes)

Distribution Table
$P(X = x)$

Histogram



Definition (fully describes)

Properties (summarizes)

Random Variable X

Expectation
the "average value" of X
$$\mathbb{E}[X]$$

Variance
the "spread" of X
$$\mathrm{Var}(X)$$

The expectation and variance of a random variable are **numerical summaries** of X.

They are numbers and are not random!

CC BY NC SA

Definition of Expectation: Discrete Random Variables

The **expectation** of a random variable X is the **weighted average** of the values of X, where the weights are the probabilities of the values.

$$\mathbb{E}[X] = \sum_{\substack{\text{all possible} \\ x}} x P(X = x)$$

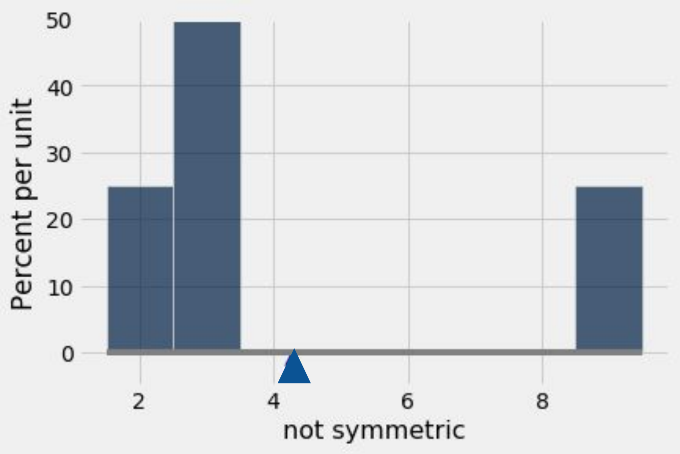
Expectation is a number, not a random variable!

- It is analogous to the **average** (same units as the random variable).
- It is the center of gravity of the probability histogram.
- It is the long run average of the random variable, if you simulate the variable many times.

CC BY NC SA

Don’t Panic! You’ve Seen This Before

> The **expectation** (mean) is the center of gravity or **balance point** of the **probability distribution**



When doing Exploratory Data Analysis, you computed this from the datapoints themselves (i.e., the sample of data).

Now we redefine these terms with respect to probability distributions.

(there is a small subtlety here regarding what the histogram represents—we’ll revisit this later in lecture)

CC BY NC SA

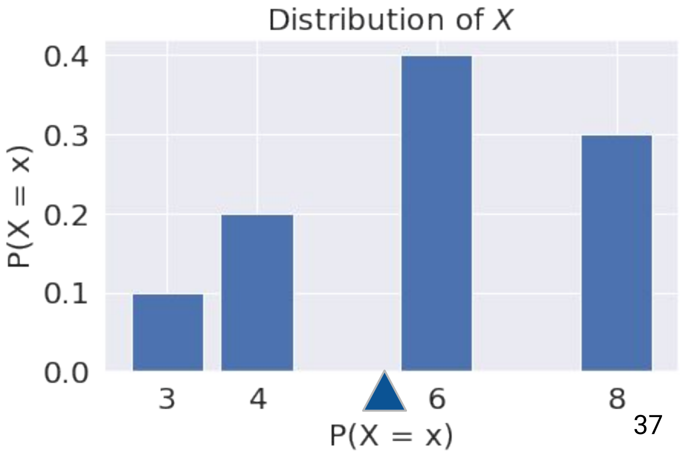
Example

$$\mathbb{E}[X] = \sum_x xP(X = x)$$

Consider the random variable X we defined earlier.

x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3

$$\begin{aligned}\mathbb{E}[X] &= \sum_x x \cdot P(X = x) \\ &= 3 \cdot 0.1 + 4 \cdot 0.2 + 6 \cdot 0.4 + 8 \cdot 0.3 \\ &= 0.3 + 0.8 + 2.4 + 2.4 \\ &= 5.9\end{aligned}$$



Note, $E[X] = 5.9$ is not a possible value of X!
It is an average.

The expectation of X does not need to be a value of X.

CC BY NC SA

Definition of Variance

Variance is the **expected squared deviation from the expectation** of X.

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

- The units of the variance are the square of the units of X.
- To get back to the right scale, use the **standard deviation** of X: $\text{SD}(X) = \sqrt{\text{Var}(X)}$

Variance is a number, not a random variable!

- The main use of variance is to **quantify chance error**. How far away from the expectation could X be, just by chance?

By Chebyshev's inequality

No matter what the shape of the distribution of X is, the vast majority of the probability lies in the interval "expectation plus or minus a few SDs."

CC BY NC SA

Chebyshev's Inequality

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

By Chebyshev's inequality

No matter what the shape of the distribution of X is, the vast majority of the probability lies in the interval "expectation plus or minus a few SDs."

cc BY NC SA

Variance: Alternate Calculation

There's a more convenient form of variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- Proof (involves expanding the square and properties of expectation/summations): link
- Useful in Mean Squared Error calculations
  - If X is centered (i.e. E[X] = 0), then $E[X^2] = \text{Var}(X)$
- When computing variance by hand, often used instead of definition.

CC BY NC SA

Dice Is the Plural; Die Is the Singular

Let X be the outcome of a single die roll.
X is a random variable.

$$P(X=x) = \begin{cases} 1/6 & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases}$$

1. What is the expectation, $E[X]$?

$$\mathbb{E}[X] = \sum_x x P(X=x)$$
$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

(definitions/properties)

2. What is the variance, $\mathrm{Var}(X)$?

Dice Is the Plural; Die Is the Singular

Let X be the outcome of a single die roll.
X is a random variable.

$$P(X = x) = \begin{cases} 1/6 & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \sum_x x P(X = x)$$

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned}$$

1. What is the expectation, $E[X]$?

$$\begin{aligned} \mathbb{E}[X] &= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) \\ &= (1/6)(1+2+3+4+5+6) = \frac{7}{2} \end{aligned}$$

2. What is the variance, $\mathrm{Var}(X)$?

Dice Is the Plural; Die Is the Singular

Let X be the outcome of a single die roll.
X is a random variable.

$$P(X = x) = \begin{cases} 1/6 & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \sum_x x P(X = x)$$

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned}$$

1. What is the expectation, $E[X]$?

$$\begin{aligned} \mathbb{E}[X] &= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) \\ &= (1/6)(1+2+3+4+5+6) = \frac{7}{2} \end{aligned}$$

2. What is the variance, $\mathrm{Var}(X)$?

Approach 1: Definition

$$\begin{aligned} \mathrm{Var}(X) = (1/6)\big((1 - 7/2)^2 &+ (2 - 7/2)^2 \\ + (3 - 7/2)^2 &+ (4 - 7/2)^2 \\ + (5 - 7/2)^2 &+ (6 - 7/2)^2\big) \end{aligned}$$

$$= 35/12$$

Approach 2: Property

$$\mathbb{E}[X^2] = \sum_x x^2 P(X = x)$$

1^2*(⅙) + 2^2 * (⅙) +... 6^2*(⅙) = 91/6

$$\mathrm{Var}(X) = 91/6 - (7/2)^2 = 35/12$$

cc BY NC SA

Properties of Expectation #1

Recall definition of expectation: $\mathbb{E}[X] = \sum_x xP(X = x)$ $\mathbb{E}[X] = \sum_{\substack{\text{all samples} \\ s}} X(s)P(s)$

1. **Expectation is linear**:

(intuition: summations are linear)

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

Proof:

$$\begin{aligned}\mathbb{E}[aX + b] &= \sum_x (ax + b)P(X = x) = \sum_x \big(axP(X = x) + bP(X = x)\big) \\ &= a\sum_x xP(X = x) + b\sum_x P(X = x) \\ &= a\mathbb{E}[X] + b \cdot 1\end{aligned}$$

CC BY NC SA

Recall definitions of expectation:

$$\mathbb{E}[X] = \sum_x x P(X = x)$$

$$\mathbb{E}[X] = \sum_{\substack{\text{all samples} \\ s}} X(s) P(s)$$

3. **Expectation is linear in sums of RVs**:

For any relationship between X and Y.

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

Proof:

$$\begin{aligned}
\mathbb{E}[X + Y] &= \sum_s (X + Y)(s) P(s) = \sum_s \big(X(s) + Y(s)\big) P(s) \\
&= \sum_s \big(X(s)P(s) + Y(s)P(s)\big) \\
&= \sum_s X(s)\, P(s) + \sum_s Y(s) P(s) \\
&= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

CC BY NC SA

Jump back: link

We know that $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$

In order to compute Var(aX + b), consider:

- A shift by **b** units **does not** affect spread. Thus, Var(aX + b) = Var(aX).
- The multiplication by **a does** affect spread!

Then,

$$\begin{aligned} Var(aX+b) = Var(aX) &= E((aX)^2) - (E(aX))^2 \\ &= E(a^2X^2) - (aE(X))^2 \\ &= a^2\left(E(X^2) - (E(X))^2\right) \\ &= a^2Var(X) \end{aligned}$$

In summary:

$$\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$$

$$\mathbb{V}ar(aX+b) = a^2\mathbb{V}ar(X)$$
$$\mathbb{SD}(aX+b) = |a|\mathbb{SD}(X)$$

Don't forget the absolute values and squares!

cc BY NC SA

Properties of Variance #2

Jump back: link

The variance of a sum is affected by the dependence between the two random variables that are being added. Let's expand out the definition of Var(X + Y) to see what's going on.

Let $\mu_x = E[X], \mu_y = E[Y]$

By the linearity of expectation, and the substitution.

We see

$$
\begin{aligned}
Var(X+Y) &= E\left[(X+Y-E(X+Y))^2\right] \\
&= E\left[((X-\mu_x)+(Y-\mu_y))^2\right] \\
&= E\left[(X-\mu_x)^2 + 2(X-\mu_x)(Y-\mu_y) + (Y-\mu_y)^2\right] \\
&= E\left[(X-\mu_x)^2\right] + E\left[(Y-\mu_y)^2\right] + 2E\left[(X-\mu_x)(Y-\mu_y)\right] \\
&= Var(X) + Var(Y) + 2E\left[(X-E(X))(Y-E(Y))\right]
\end{aligned}
$$

CC BY NC SA

Addition rule for variance

If X and Y are **uncorrelated** (in particular, if they are **independent**), then

$$\mathbb{V}ar(X+Y) \;=\; \mathbb{V}ar(X) + \mathbb{V}ar(Y)$$

Therefore, under the same conditions,

$$\mathbb{SD}(X+Y) \;=\; \sqrt{\mathbb{V}ar(X) + \mathbb{V}ar(Y)} \;=\; \sqrt{(\mathbb{SD}(X))^2 + (\mathbb{SD}(Y))^2}$$

- Think of this as "Pythagorean theorem" for random variables.
- Uncorrelated random variables are like orthogonal vectors.

CC BY NC SA

LECTURE 16

Random Variables

Content credit: Lisa Yan, Anthony D. Joseph, Suraj Rampure, Ani Adhikari