ENV 710: Lecture 1

discrete probability distributions

Bernoulli random variables

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

discrete probability distributions

Bernoulli random variables

learning goals

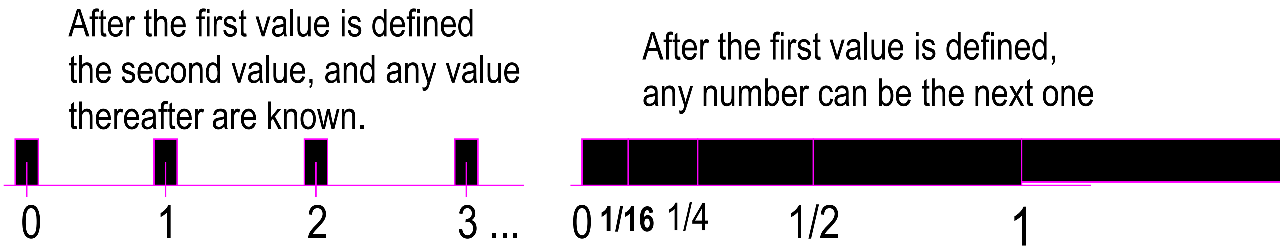
- what is a random variable?
- what is a probability distribution?
- what is a Bernoulli random variable?
- what are examples of Bernoulli random variables?

stuff you should know

random variable

variable whose value is a function of a random process

- discrete or continuous probability distributions



- if X is a random variable, then P(X=x) is the probability that the value x will occur

e.g., $P(X=2) = 0.25$

probability distribution

probability distribution is a table or equation describing the possible values of a random variable and their associated probabilities…

x	1	2	3	4	5	6
P(x)	0.16	0.16	0.16	0.16	0.16	0.16

e.g., $P(X=5) = 0.16$

discrete probability distributions can be represented in tabular form, and each value of the discrete variable has a non-zero probability

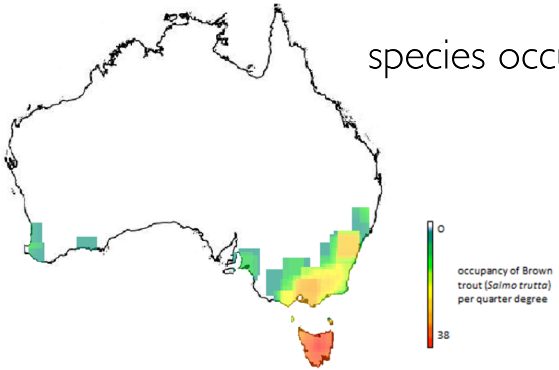
Bernoulli random variable

when an individual trial has only two possible outcomes, it is called a Bernoulli random variable

- type of discrete random variable

$$X \sim \text{Bernoulli}(p)$$

$$P(x) = \begin{cases} 1\text{-}p & \text{for } x = 0 \\ p & \text{for } x = 1 \end{cases}$$

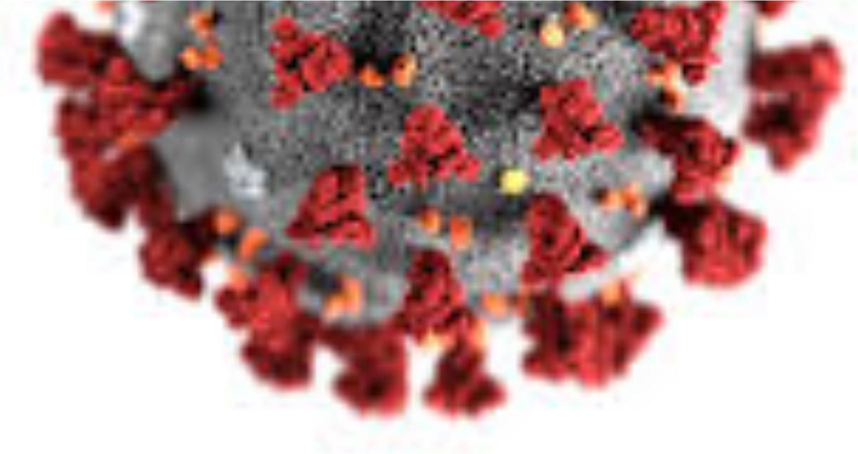


species occupancy

coin flips

Bernoulli trial

- always a single trial
- if a Bernoulli trial is repeated multiple times, the experiment is known as a binomial experiment
- Bernoulli distribution can model a single individual experiencing death or disease

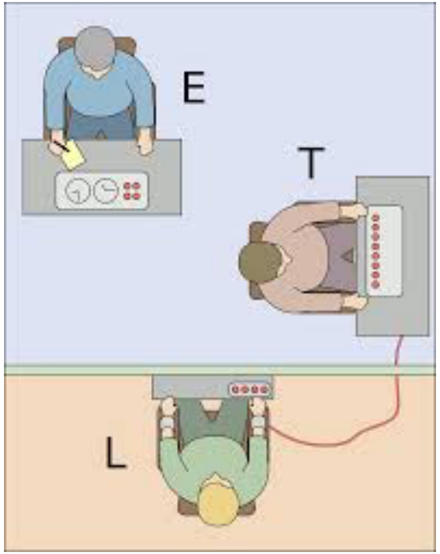


COVID-19

What if we tested a random set of 1000 subjects for COVID-19?

Each test is a Bernoulli trial.

example: Milgram experiment



(E) experimenter orders the (T) teacher to shock the (L) learner each time the learner incorrectly answers a question

- each teacher in the experiment is a trial
- "success" if a shock is administered
- "failure" if a shock is not administered

example

Suppose we select 4 random individuals to participate in the experiment. What is the probability that exactly one of them will administer a shock?

P(shock) = 0.65

example

Suppose we select 4 random individuals to participate in the experiment. What is the probability that exactly one of them will administer a shock?

P(shock) = 0.65

Experiment 1: $\underline{\text{shock}} \times \underline{\text{refuse}} \times \underline{\text{refuse}} \times \underline{\text{refuse}}$

$0.65 \times 0.35 \times 0.35 \times 0.35 = 0.0279$

joint probability 'and'

two events occurring together

example

Suppose we select 4 random individuals to participate in the experiment. What is the probability that exactly one of them will administer a shock?

P(shock) = 0.65

Experiment 1: $\underline{\text{shock}} \times \underline{\text{refuse}} \times \underline{\text{refuse}} \times \underline{\text{refuse}}$
$0.65 \times 0.35 \times 0.35 \times 0.35 = 0.0279$

Experiment 2: $\underline{\text{refuse}} \times \underline{\text{shock}} \times \underline{\text{refuse}} \times \underline{\text{refuse}}$
$0.35 \times 0.65 \times 0.35 \times 0.35 = 0.0279$

example

Suppose we select 4 random individuals to participate in the experiment. What is the probability that exactly one of them will administer a shock?

P(shock) = 0.65

Experiment 1: $\underline{\text{shock}} \times \underline{\text{refuse}} \times \underline{\text{refuse}} \times \underline{\text{refuse}}$ $= 0.0279$

$0.65 \times 0.35 \times 0.35 \times 0.35$

Experiment 2: $\underline{\text{refuse}} \times \underline{\text{shock}} \times \underline{\text{refuse}} \times \underline{\text{refuse}}$ $= 0.0279$

$0.35 \times 0.65 \times 0.35 \times 0.35$

Experiment3: $\underline{\text{refuse}} \times \text{refuse} \times \underline{\text{shock}} \times \underline{\text{refuse}}$ $= 0.0279$

$0.35 \times 0.35 \times 0.65 \times 0.35$

example

Suppose we select 4 random individuals to participate in the experiment. What is the probability that exactly one of them will administer a shock?

P(shock) = 0.65

Experiment 1: $\underline{\text{shock}} \times \underline{\text{refuse}} \times \underline{\text{refuse}} \times \underline{\text{refuse}}$ $= 0.0279$
$0.65 \times 0.35 \times 0.35 \times 0.35$

Experiment 2: $\underline{\text{refuse}} \times \underline{\text{shock}} \times \underline{\text{refuse}} \times \underline{\text{refuse}}$ $= 0.0279$
$0.35 \times 0.65 \times 0.35 \times 0.35$

Experiment 3: $\underline{\text{refuse}} \times \text{refuse} \times \underline{\text{shock}} \times \underline{\text{refuse}}$ $= 0.0279$
$0.35 \times 0.35 \times 0.65 \times 0.35$

Experiment 4: $\underline{\text{refuse}} \times \underline{\text{refuse}} \times \underline{\text{refuse}} \times \underline{\text{shock}}$ $= 0.0279$
$0.35 \times 0.35 \times 0.35 \times 0.65$

example

Suppose we select 4 random individuals to participate in the experiment. What is the probability that exactly one of them will administer a shock?

P(shock) = 0.65

Experiment 1: shock × refuse × refuse × refuse = 0.0279

0.65 × 0.35 × 0.35 × 0.35

or

Experiment 2: refuse × shock × refuse × refuse = 0.0279

0.35 × 0.65 × 0.35 × 0.35

or

Experiment 3: refuse × refuse × shock × refuse = 0.0279

0.35 × 0.35 × 0.65 × 0.35

or

Experiment 4: refuse × refuse × refuse × shock = 0.0279

0.35 × 0.35 × 0.35 × 0.65

disjoint probability 'or'

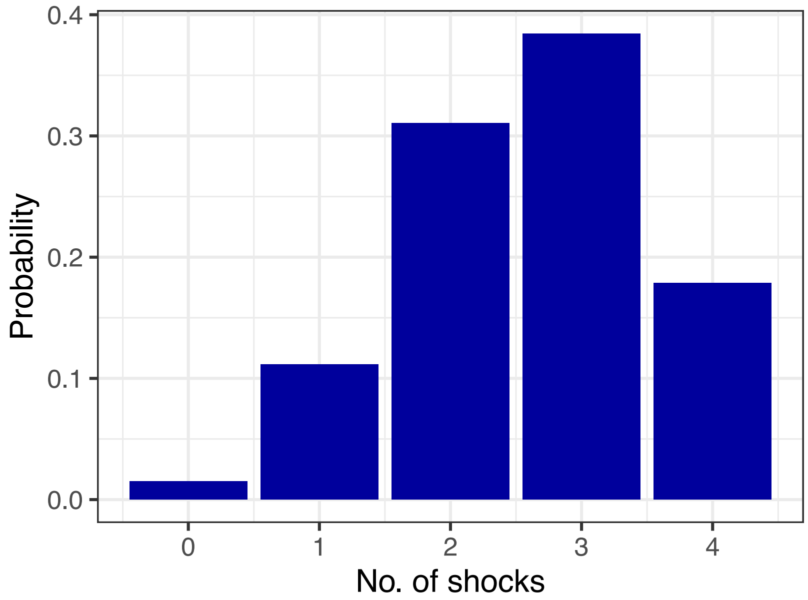
two events are mutually exclusive

4 x 0.0279 = 0.111

example

Binomial probability distribution

probability of 0 to 4 teachers delivering shocks to the learners



this graph represents a series of Bernoulli trials with every possible combination for obtaining 0, 1, ... 4 shocks in 4 experiments of 4 trials

ENV 710: Lecture 1

discrete probability distributions

binomial distribution

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

discrete probability distributions

binomial distribution

learning goals

- what is a probability mass distribution?
- what is the binomial distribution and when is it used?
- how do you calculate probabilities for specific binomial outcomes?

stuff
you
should
know

binomial distribution

the **binomial** distribution describes a series of Bernouilli trials.

a **binomial variable**, X, is the number of successful results in n trials.

$$X \sim Bin(n, p)$$

the **probability mass function** of X calculates the probability of k successes:

$$P(X = k) = \frac{n!}{k!(n-k)!} \cdot p^k \cdot (1-p)^{n-k}$$

binomial distribution

the **binomial** distribution describes the probability of having exactly ***k*** successes in ***n*** independent Bernoulli trials with probability of success ***p***

# of scenarios x P(1 scenario)

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

"n choose k"

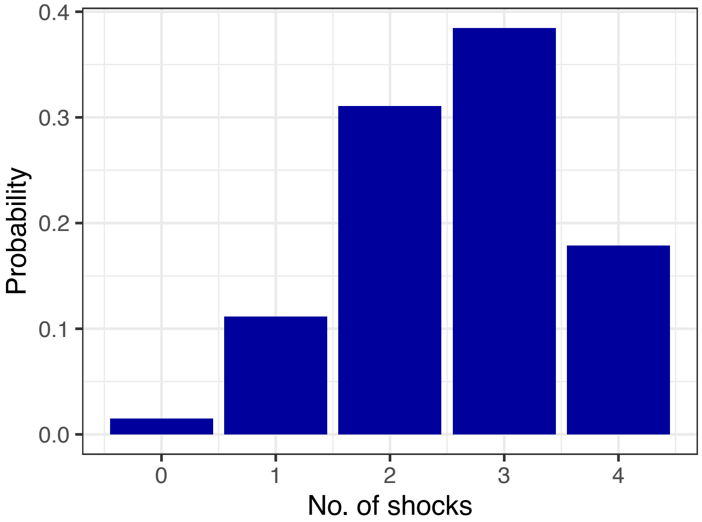
$$p^k (1-p)^{n-k}$$

P(success) to the power of successes

P(failure) to the power of failures

example

probability of 0 to 4 teachers delivering shocks to the learners



$$P(Shocks = 1) = \frac{4!}{1!(4-1)!} \cdot 0.65^1 \cdot (1-0.65)^{4-1} = 0.111$$

$$P(Shocks = 3) = \frac{4!}{3!(4-3)!} \cdot 0.65^3 \cdot (1-0.65)^{4-3} = 0.384$$

```
dbinom(x = 3, size = 4, prob = 0.65)
```

binomial conditions

- the trials must be independent
- the number of trials, n, must be fixed
- each trial outcome must be classified as a success or failure
- the probability of success, p, must be the same for each trial

mean & standard deviation

expected value (mean) of binomial distribution

$$\mu = np$$

n = number of trials, p = probability of success

standard deviation of binomial distribution

$$\sigma = \sqrt{np(1-p)}$$

what is the mean and standard deviation for the Milgram example where the probability of success is 0.65 and the number of trials is 4?

mean & standard deviation

expected value (mean) of binomial distribution

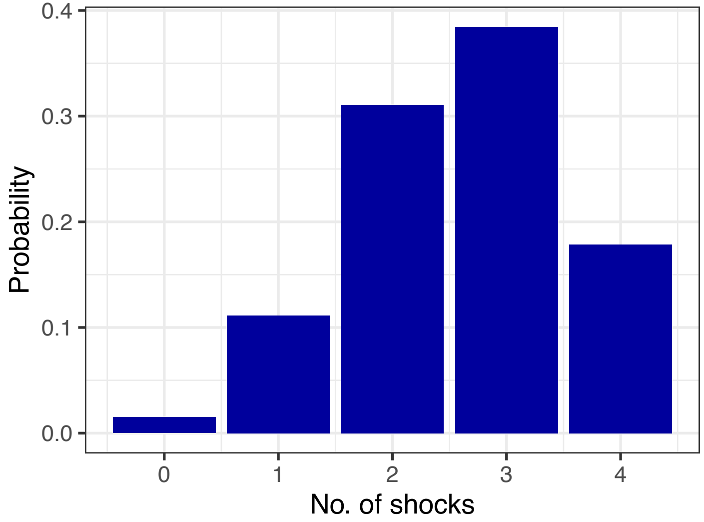
$$\mu = np$$

$$\mu = 4 \times 0.65 = 2.6$$

standard deviation of binomial distribution

$$\sigma = \sqrt{np(1-p)}$$

$$\sigma = \sqrt{(4 \times 0.65 \times 0.35)} = 0.95$$



example

The probability of finding a monkey during a survey of a site is 0.41.

In a survey of 3 independent sites, what is the probability of finding a monkey at all three sites?

In a survey of 20 independent sites, what is the probability that monkeys occur in the first 15 sites, but not the last 5 sites?

https://www.pinterest.com/pin/2674081008139126/?nic_v1=1a2WAxFRU

example

What is the probability of finding a monkey in each of three site censuses?

$0.41 \times 0.41 \times 0.41 = 0.069$

```
dbinom(x = 3, size = 3, prob = 0.41)

[1] 0.068921
```

example

In a census of 20 independent sites, what is the probability that monkeys occur in the first 15 sites, but not the last 5 sites?

$$(0.41)^{15}\ (0.59)^{5} = 1.11 \times 10^{-7}$$

example

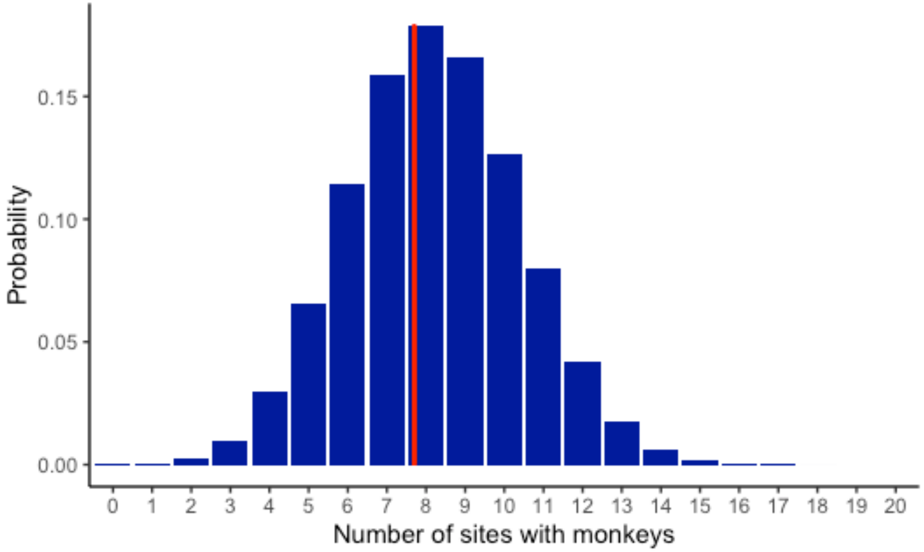
What is the probability that monkeys occur in **any** 15 sites, and not 5 sites?

example

What is the probability that monkeys occur in **any** 15 sites, but not 5 sites?

$$P(X = 15) = \frac{20!}{15!(20-15)!} \cdot 0.41^{15} \cdot 0.59^{5} = 0.002$$

```
dbinom(x = 15, size = 20,
       prob = 0.41)
```



working with the binomial…

According to a 2014 Gallup poll, 31% of Americans indicate that they worry "a great deal" about the quality of the environment, marking the lowest level of worry about the environment since Gallup began measuring it in 2001. What is the probability that in a random sample of 10 people, **exactly 6** are worried a great deal about the environment?

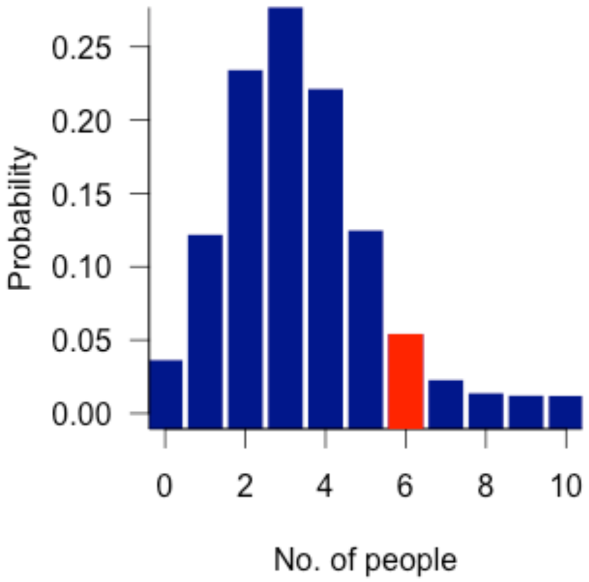
$P(X=6) = ???$

working with the binomial…

According to a 2014 Gallup poll, 31% of Americans indicate that they worry "a great deal" about the quality of the environment, marking the lowest level of worry about the environment since Gallup began measuring it in 2001. What is the probability that in a random sample of 10 people, **exactly 6** are worried a great deal about the environment?

P(X=6) = 0.042

```
> dbinom(x = 6, size = 10,
         prob = 0.31)
```



working with the binomial…

According to a 2014 Gallup poll, 31% of Americans indicate that they worry "a great deal" about the quality of the environment, marking the lowest level of worry about the environment since Gallup began measuring it in 2001. What is the probability that in a random sample of 10 people, **6 or more** are worried a great deal about the environment?

$P(X \geq 6) = ???$

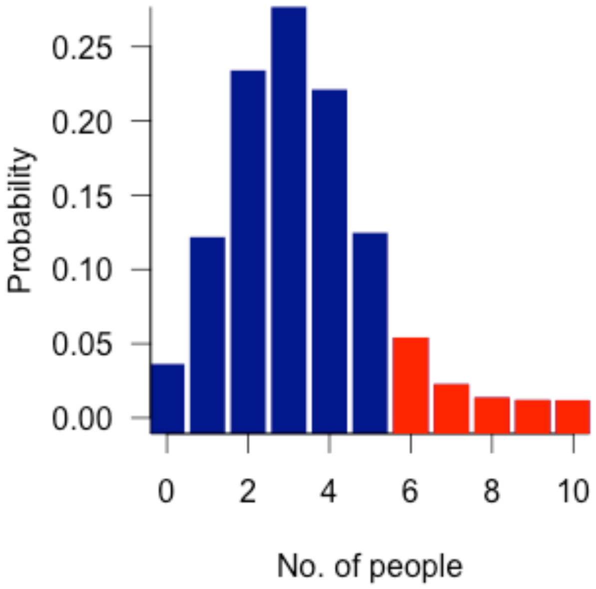
working with the binomial…

According to a 2014 Gallup poll, 31% of Americans indicate that they worry "a great deal" about the quality of the environment, marking the lowest level of worry about the environment since Gallup began measuring it in 2001. What is the probability that in a random sample of 10 people, **6 or more** are worried a great deal about the environment?

$P(X \geq 6) = 0.055$

```
> dbinom(x = 6, size = 10, prob = 0.31)
+ dbinom(x = 7, size = 10, prob = 0.31)
+ dbinom(x = 8, size = 10, prob = 0.31)
+ dbinom(x = 9, size = 10, prob = 0.31)
+ dbinom(x = 10, size = 10, prob = 0.31)
```

```
> sum(dbinom(x = 6:10,
      size = 10, prob = 0.31))
```



ENV 710: Lecture 1

discrete probability distributions

Poisson distribution

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

discrete probability distributions

Poisson distribution

learning goals

- what is a Poisson random variable?
- what is the Poisson distribution? when is it used?
- calculate the probability of getting specific numbers of counts…

stuff
you
should
know

Poisson random variable

- useful for count data (e.g. counting events)
- X is the number of occurrences of an event recorded in a sample area or during a fixed interval of time

$$X \sim \text{Poisson}(\lambda)$$

rate parameter
average & variance

Poisson random variable

- discrete probability distribution for counts of events that occur randomly in an interval of time (or space)
- ideal for events that have a low probability
  - annual deaths from lightening strikes in US
  - number of fish caught per hour
  - number of mutations in regions of a chromosome
- only the occurrences of the event are counted, not the non-occurrences

Poisson probability mass function

$$P(X) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \qquad x = 0, 1, 2, 3, \ldots$$

- λ is the average number of occurrences in each sample
- e is the base of the natural logarithm
- probability mass distribution determines probability of obtaining any value of X

e is written as exp in R

mean & standard deviation

$$P(X) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \qquad x = 0, 1, 2, 3, \ldots$$

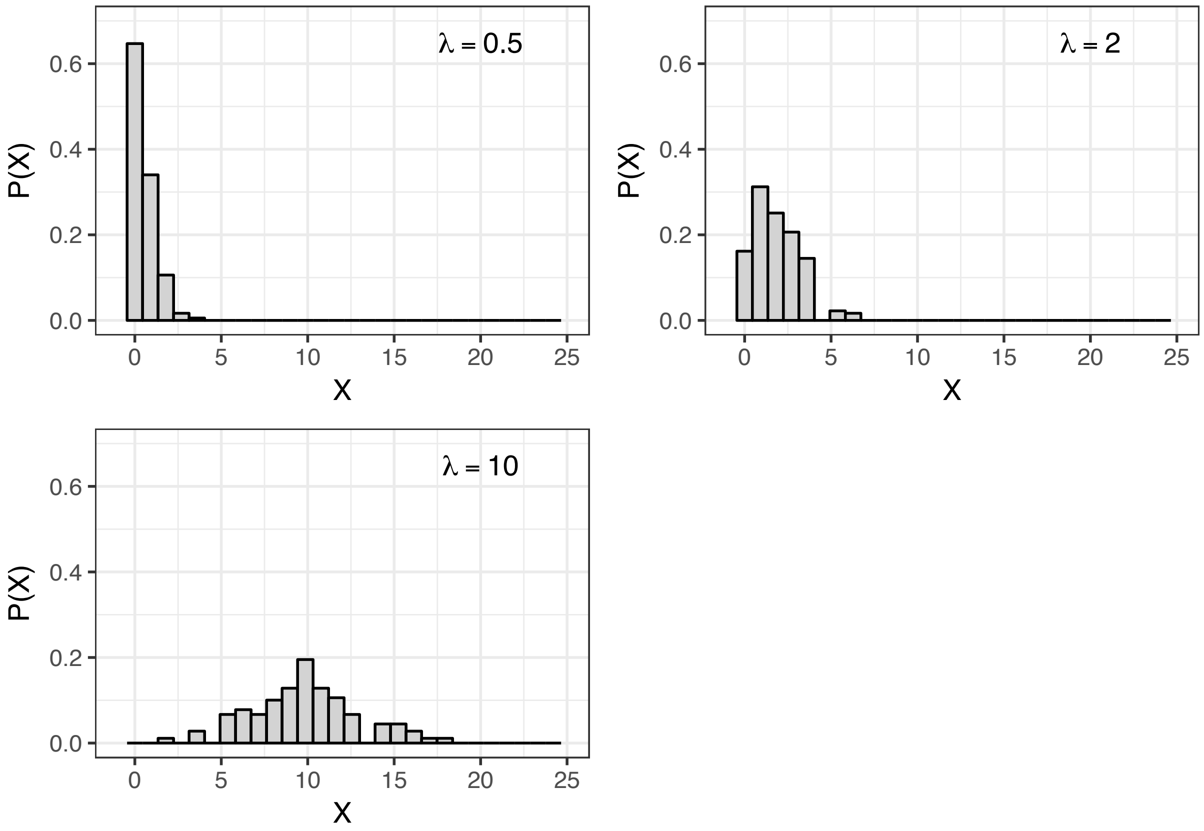
expected value (mean) of Poisson distribution

$$\mu = \lambda$$

variance of Poisson distribution

$$\sigma^2 = \lambda$$

Poisson probability distributions



example

The average rate of births in a hospital is 1.8 births per hour. What is the probability that we observe 5 births in a 1-hour interval?

example

The average rate of births in a hospital is 1.8 births per hour. What is the probability that we observe 5 births in a 1-hour interval?

$$P(X = 5) \sim Pois(1.8)$$

$$P(X = 5) = \frac{1.8^5}{5!} e^{-1.8} = 0.026$$

example

The average rate of births in a hospital is 1.8 births per hour. What is the probability that we observe 5 births in a 1-hour interval?

$$P(X = 5) \sim Pois(1.8)$$

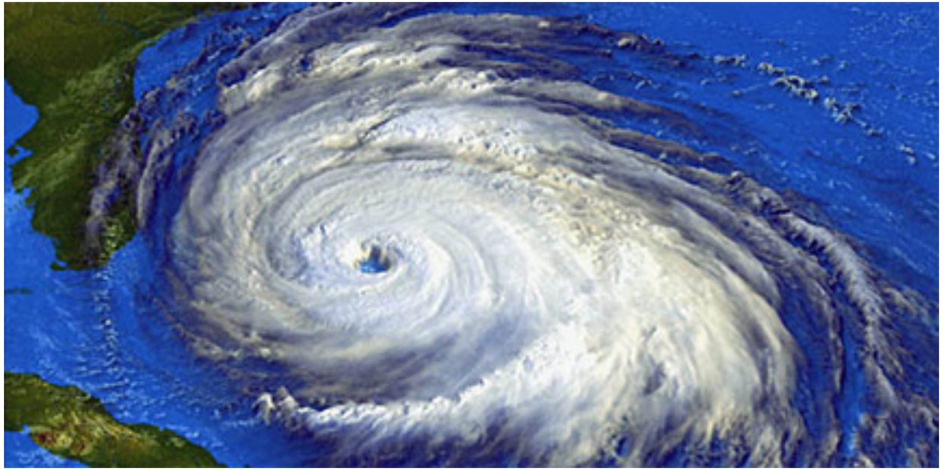
$$P(X = 5) = \frac{1.8^5}{5!} e^{-1.8} = 0.026$$

```
1.8^5/factorial(5)* exp(-1.8)

dpois(x = 5, lambda = 1.8)
```

What if λ in unknown? What is the expected number of hurricanes to hit the US coastline in a year?

- a probability distribution specifies the values, *x*, of a discrete random variable, *X*, and their probabilities, *P(x)*



Category 3 hurricanes

x	P(x)
0	0.30
1	0.40
2	0.25
3	0.05

expected value

expected value: the predicted value of a variable, calculated as the sum of all possible values, each multiplied by the probability of its occurrence

$$0 \cdot (0.3) + 1 \cdot (0.4) + 2 \cdot (0.25) + 3 \cdot (0.05) = 1.05$$

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

Category 3 hurricanes

x	P(x)
0	0.30
1	0.40
2	0.25
3	0.05

Is *Senecio jacobaea* randomly distributed?

If the # of plants in each of 100 quadrats follows a Poisson distribution, then it is randomly distributed.

0	1	2	3	4	5	6
46	36	15	3	0	0	0
0.46	0.36	0.15	0.03	0	0	0

Is *Senecio jacobaea* randomly distributed?

If the # of plants in each of 100 quadrats follows a Poisson distribution, then it is randomly distributed.

0	**1**	**2**	**3**	**4**	**5**	**6**
46	36	15	3	0	0	0
0.46	0.36	0.15	0.03	0	0	0

Is *Senecio jacobaea* randomly distributed?

If the # of plants in each of 100 quadrats follows a Poisson distribution, then it is randomly distributed.

0	1	2	3	4	5	6
46	36	15	3	0	0	0
0.46	0.36	0.15	0.03	0	0	0

$E(X) = 0(.46) + 1(.36)+2(0.15) + 3(0.03) + 0 + 0 = 0.75$

Is *Senecio jacobaea* randomly distributed?

$$P_{0.75}(0) = \frac{(0.75)^0}{0!} e^{-0.75} = 0.472$$

$$P_{0.75}(1) = \frac{(0.75)^1}{1!} e^{-0.75} = 0.354$$

$$P_{0.75}(2) = \frac{(0.75)^2}{2!} e^{-0.75} = 0.133$$

$$P_{0.75}(3) = \frac{(0.75)^3}{3!} e^{-0.75} = 0.033$$

$$P_{0.75}(4) = \frac{(0.75)^4}{4!} e^{-0.75} = 0.006$$

$$P_{0.75}(0 \leq x \leq 4) = 0.99$$

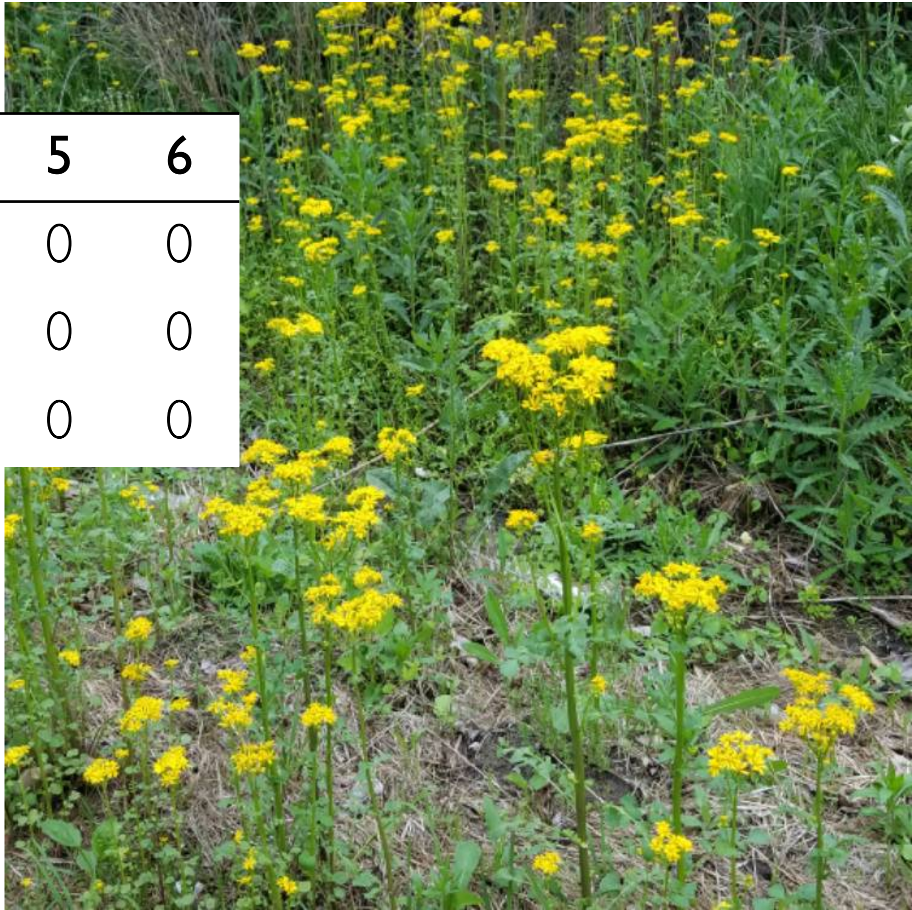
```
> dpois(x = 0, lambda = 0.75)
[1] 0.4723666
```

```
> dpois(x = 0:4, lambda = 0.75)
[1] ???
```

```
> sum(dpois(x = 0:4, lambda = 0.75))
[1] ???
```

Is *Senecio jacobaea* randomly distributed?

	0	1	2	3	4	5	6
	46	36	15	3	0	0	0
Observed	0.46	0.36	0.15	0.03	0	0	0
Predicted	0.47	0.35	0.13	0.03	0.01	0	0



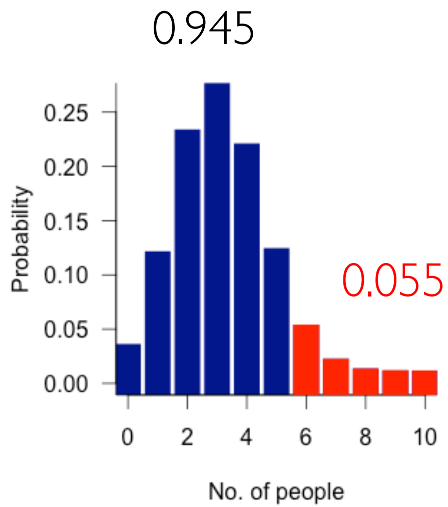
R functions for probability

- `dbinom()` – calculates the probability for a particular value of a random variable

```
sum(dbinom(x = 0:5, size = 10, p = 0.31)
```

- `pbinom()` – calculates the cumulative probability of a binomial distribution for a particular quantile
  - area of the histogram up to *q*

```
pbinom(q = 5, size = 10, p = 0.31)
```




```
sum(dpois(x = 0:4, lambda = 0.75))
```

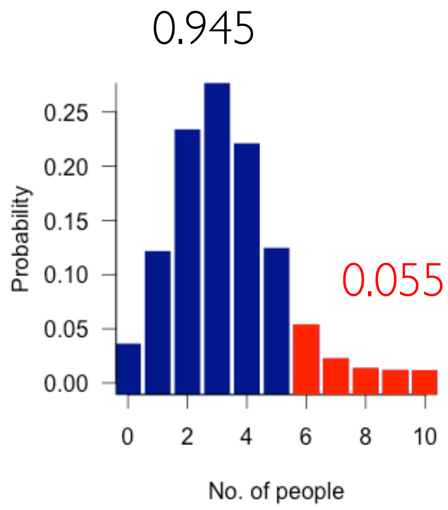
cumulative probability

- `dbinom()` – calculates the probability for a particular value of a random variable

  `sum(dbinom(x = 0:5, size = 10, p = 0.31))`

- `pbinom()` – calculates the cumulative probability of a binomial distribution for a particular quantile
  - area of the histogram up to *q*

  `pbinom(q = 5, size = 10, p = 0.31)`



Post your questions to be answered during lecture