Math 425
Introduction to Probability
Lecture 13

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

February 6, 2009

Example: Three dice

☞ Sometime before 1642 Galileo was asked whether it was more likely to throw a 9 or a 10 on three dice.

The problem is easy for us (although a bit tedious) by counting outcomes in a sample space.

☞ Our sample space is

$$S = \{(i,j,k) \mid 1 \le i,j,k \le 6\},$$

and each $s \in S$ is equiprobable, so

$$\mathbf{P}(\{s\}) = \frac{1}{|S|} = \frac{1}{6^3} = \frac{1}{216}.$$

Outcomes

k	Canonical outcomes						Total
3	(1,1,1)						1
4	(1,1,2)						3
5	(1,1,3)	(1,2,2)					6
6	(1,1,4)	(1,2,3)	(2,2,2)				10
7	(1,1,5)	(1,2,4)	(1,3,3)	(2,2,3)			15
8	(1,1,1)	(1,2,5)	(1,3,4)	(2,2,4)	(2,3,3)		21
9	(1,2,6)	(1,3,5)	(1,4,4)	(2,2,5)	(2,3,4)	(3,3,3)	25
10	(1,3,6)	(1,4,5)	(2,2,6)	(2,3,6)	(2,4,4)	(3,3,4)	27
11	(1,4,6)	(1,5,5)	(2,3,6)	(2,4,5)	(3,3,5)	(3,4,4)	27
12	(1,5,6)	(2,4,6)	(2,5,5)	(3,3,6)	(3,4,5)	(4,4,4)	25
13	(1,6,6)	(2,5,6)	(3,4,6)	(3,5,5)	(4,5,5)		21
14	(2,6,6)	(3,5,6)	(4,4,6)	(4,5,5)			15
15	(3,6,6)	(4,5,6)	(5,5,5)				10
16	(4,6,6)	(5,5,6)					6
17	(5,6,6)						3
18	(6,6,6)						1

Example: Three dice

☞ We only introduced the sample space to compute the probabilities of the events

- E_i: the sum of the three dice is i (so, $3 \le i \le 18$)

Example. The answer to Galileo's problem

$$\mathbf{P}(E_9) = \frac{25}{216} < \frac{27}{216} = \mathbf{P}(E_{10})$$

Example. Which is more likely, throwing a sum less than 9 or throwing a sum greater than 9 but less than 12?

$$\mathbf{P}(E_3 \cup E_4 \cup E_5 \cup E_6 \cup E_7 \cup E_8) = \frac{56}{216}$$

$$\mathbf{P}(E_{10} \cup E_{11}) = \frac{54}{216}$$

Example: Three dice

☞ Since we only care about the sum of the three dice, it is more convenient to think of the outcomes as random real number.

☞ It is customary and convenient to use a capital letter, such as X to denote a random real number, and use it to describe events.

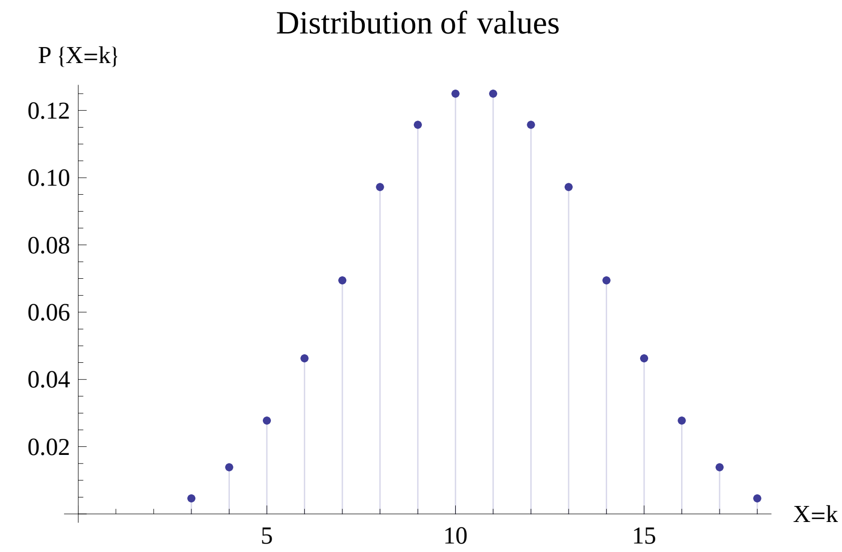
- $\{X = k\}$: the event that the sum of the dice is k.
- $\{X < 9\}$: the event that the sum of the dice is less than 9,
- $\{9 < X < 12\}$: the event that the sum of the dice is greater than 9, but less than 12.

☞ These events have probabilities, which we write as follows

$$\mathbf{P}\{X = 9\} = \frac{25}{216} \quad \mathbf{P}\{X < 9\} = \frac{56}{216} \quad \mathbf{P}\{9 < X < 12\} = \frac{54}{216}$$

$$\mathbf{P}\{X = 9.5\} = 0 \quad \mathbf{P}\{4.5 \leq X \leq 5.5\} = \mathbf{P}\{X = 5\} = \frac{6}{216}$$

Distribution of values



Mean of throws

☞ The weighted average (or mean) on a throw of three dice is

$$\sum_{k=3}^{18} k \cdot \mathbf{P}\{X = k\} = 10.5$$

The weight for $X = k$ is determined by the probability, $\mathbf{P}\{X = k\}$.

☞ The arithmetic average of the possible values treats all values as the same weight:

$$\frac{\sum_{k=3}^{18} k}{16} = 10.5$$

Mean of throws

☞ I ran 1000 experiments, where each experiment threw three dice 1000 times. Here are the statistics:

- Mean: 10.5033 (average value on each experiment)
- Maximum: 10.789
- Minimum: 10.211
- Standard Deviation: 0.093 (I can be 95% certain that the true value is within 0.18 of 10.5033.)

Example Chuck-a-Luck

Example

The game of chuck-a-luck is played by throwing three dice. The gambler bets on one of the numbers 1 through 6. The gambler receives $\$2 \cdot i$ for each of the i times the number appears (so $i = 0, 1, 2, 3$). The cost to enter the game is \$1.

☞ Suppose the gambler bets on 1. Let Y be the amount the gambler wins. So,

- $\{Y = i\}$: the event that the gambler makes $\$i$.
- $\{Y > 0\}$: the event that the gambler makes money,
- $\{Y < 0\}$: the event that the gambler loses money.

Example Chuck-a-Luck

☞ What are the probable outcomes? Is this a good bet?

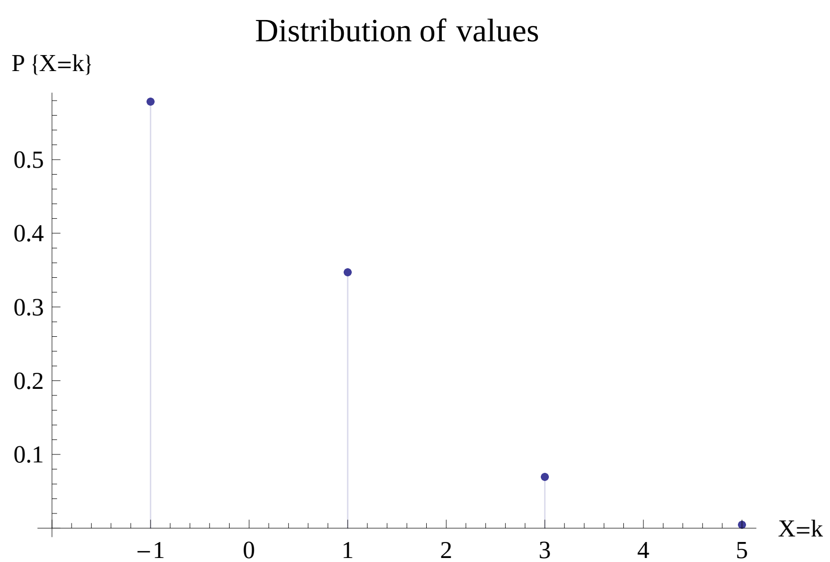
$$\begin{aligned}
\mathbf{P}\{Y = -1\} &= \binom{3}{0} \cdot \left(\frac{5}{6}\right)^3 \approx 0.579 \\
\mathbf{P}\{Y = 1\} &= \binom{3}{1} \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^2 \approx 0.347 \\
\mathbf{P}\{Y = 3\} &= \binom{3}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^1 \approx 0.069 \\
\mathbf{P}\{Y = 5\} &= \binom{3}{3} \cdot \left(\frac{1}{6}\right)^3 \approx 0.005 \\
\mathbf{P}\{Y > 0\} &\approx 0.421 \\
\mathbf{P}\{Y < 0\} &\approx 0.579
\end{aligned}$$

☞ Your expected winnings (i.e. the weighted average) in this game is

$$-1 \cdot \mathbf{P}\{Y = -1\} + 1 \cdot \mathbf{P}\{Y = 1\} + 3 \cdot \mathbf{P}\{Y = 3\} + 5 \cdot \mathbf{P}\{Y = 5\} \approx 0.$$

This is considered a fair game.

Distribution of values



Mean of Chuck-a-Luck

☞ I played chuck-a-luck over 1000 dice throws to compute a total monetary gain (or loss). I ran this experiment 1000 times. Here are the statistics:

- Mean: -1.402 (average loss!!)
- Maximum: 416
- Minimum: -460
- Standard Deviation: 133.766 (I can be 95% certain that the true value is within 267.5 of -1.402!!)

Random variables: What?

Definition

Let S be a sample space with probability function $\mathbf{P}$.
A random variable X is a function from S into the real numbers, $\mathbb{R}$. That is, $X : S \to \mathbb{R}$.

☞ Think of a random variable X as randomly chosen real number, which is determined by some experiment.
On a given experiment:

- X is more likely to take some values than others, and
- there may be many values for which it is impossible for X to take.

Example of a Random Variable

Example. We earlier looked at the experiment of throwing three dice. We care about the sum of the face values, not the face values themselves.

$$S = \{(e, f, g) \mid 1 \leq e, f, g \leq 6\}$$

Let X be the random variable

$$X(e, f, g) = e + f + g \qquad \text{for each } (e, f, g) \in S.$$

So,

$$X(1,1,1) = 3 \quad X(1,1,2) = 4 \quad X(1,2,2) = 5 \quad \text{etc.}$$

Example. In Chuck-a-Luck we introduced a random variable Y over the same sample space S:

$$Y(e, f, g) = 2 \cdot i - 1 \quad \text{where } i \text{ is the number of 1's in } e, f, g$$

Example of a Random Variable

Example. Consider the experiment: tossing a fair coin until 3 heads appears. We count the number of tosses required.
Sample space:

$$S = \{(t_1, t_2, \ldots, t_n) \mid \text{exactly 3 } t_i \text{ are heads}\}$$

Let X be the random variable

$$X(t_1, t_2, \ldots, t_n) = n \qquad \text{for each } (t_1, t_2, \ldots, t_n) \in S.$$

Example of a Random Variable

Example. Consider the experiment: choose a random person in a drug study involving 1000 people. We want to know information such as the age, height, weight, etc.

☞ What is the sample space S? Could be annoyingly complicated.

☞ We really care about random variables:

- $X(p)$: the age of person p,
- $Y(p)$: the height of person p,
- $Z(p)$: the weight of person p.

Example of a Random Variable

☞ Random variables can be useful, even when it is not (at first) apparent what values we should assign.

Example. You ask people whether they approve of the present administration. The results could be

disapprove strongly (-2), disapprove (-1), indifferent (0), approve (1), approve strongly (2)

Example. The outcomes of a Bernoulli trial are success (1) and failure (0). This is useful when we want to count the number of successes.
Alternatively, success (1) and failure (-1). This is useful when we want to bet of success.

Types of random variables

☞ A random variable X for an experiment is a random real number determined by an outcome of the experiment.

- If we can list all possible values of X (where the list may be infinite), we say X is a discrete random variable.
  Chapter 4
- If we cannot list all possible values X, we say that X is a continuous random variable.
  Chapter 5

Example of a continuous Random Variable

Example. Any random variable which takes only finitely many values is discrete.

Example. Here is an example of an infinite discrete random variable. Consider a sequence of Bernoulli trials (such as tosses of a coin), and let X record the number of trials until the first success. Then,

$$X \in \{0, 1, 2, 3, \ldots\}$$

$X(f, f, f, \ldots) = 0,\ X(s) = 1,\ X(f, s) = 2,\ X(f, f, s) = 3$ etc.

Example of a continuous Random Variable

☞ A continuous random variable can typically take any real value in some interval.

Example. A weathercock can point in any direction as measured clockwise from North in radians.

☞ Let X be the continuous random variable which records the direction of the wind. Then, $0 \leq X < 2\pi$ radians, and can take any value in this interval.

☞ Let Y be the discrete random variable which records the direction in to the nearest $\frac{n\pi}{4}$ where n is an integer with $0 \leq n \leq 7$.
We are recording the direction as North, Northeast, East, Southeast, etc.

Random Variables: Why?

☞ Why do we torment you with random variables?

(i) Most of the time we care about certain numeric values, not how the underlying sample space is represented. Without random variables, treating numeric values can be quite messy

(ii) Random variables provides a uniform treatment of the many ways we use random values which arise from experiments.

(iii) We want statistics which give us a better sense of the contours of the population of interest: means, variances, and related quantities. These are best handled using random variables.

(iv) Many numeric values are best thought of as functions of more basic random values which arise from an experiment. Random variables is by far the best machinery for handling these values.

Random Variables and events

☞ Random variables provide a convenient framework for talking about events.

Example. Let X be the random variable giving the sum of the faces of three dice.

- $\{X = i\}$: the event that the sum of the faces is i (where $3 \le i \le 18$).
- $\{i \le X\}$: is the event that the sum of the faces is at least i.
- $\{X \le i\}$: the event that the sum of the faces is no more than i.
- $\{i \le X \le j\}$: the event that the sum of the faces is between i and j.

Probability mass function for a random variable

☞ The most important property of random variables, is that they determine a probability mass function on the real numbers.

This is simply a weighted distribution of real value outcomes.

Definition

Let X be a discrete random variable.

The probability mass function (pmf) $p_X(x)$ for X is given by

$$p_X(x) = \mathbf{P}\{X = x\} \qquad \text{for any real number } x$$

If x is not a possible value of X, then $p_X(x) = 0$.

We will drop the subscript X when the random variable is understood.

Example of pmf

Example. Consider the experiment of throwing a single die. Let X be the discrete random variable which gives the face value.

The pmf for X is given by

$$p_X(x) = \begin{cases} \frac{1}{6} & \text{if } x = 1,2,3,4,5,6 \\ 0 & \text{otherwise} \end{cases}$$

☞ Here is the probability of some events determined by X:

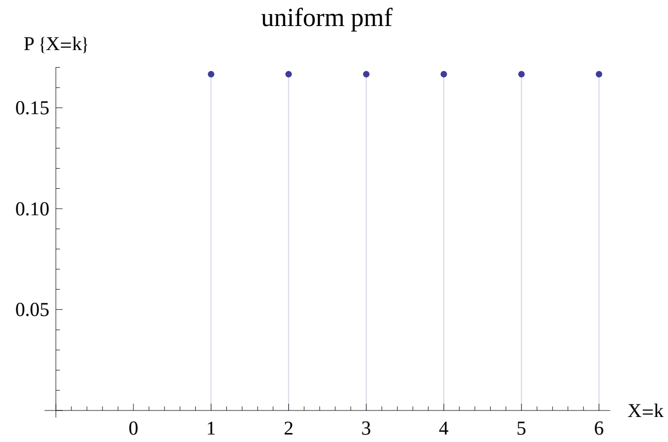
$$\mathbf{P}\{X \ge 0\} = 1 \quad \mathbf{P}\{X < 0\} = 0 \quad \mathbf{P}\{X > 6\} = 0$$

$$\mathbf{P}\{X = 7\} = p_X(7) = 0 \quad \mathbf{P}\{3.5 < X < 4.5\} = \frac{1}{6}$$

Uniform pmf

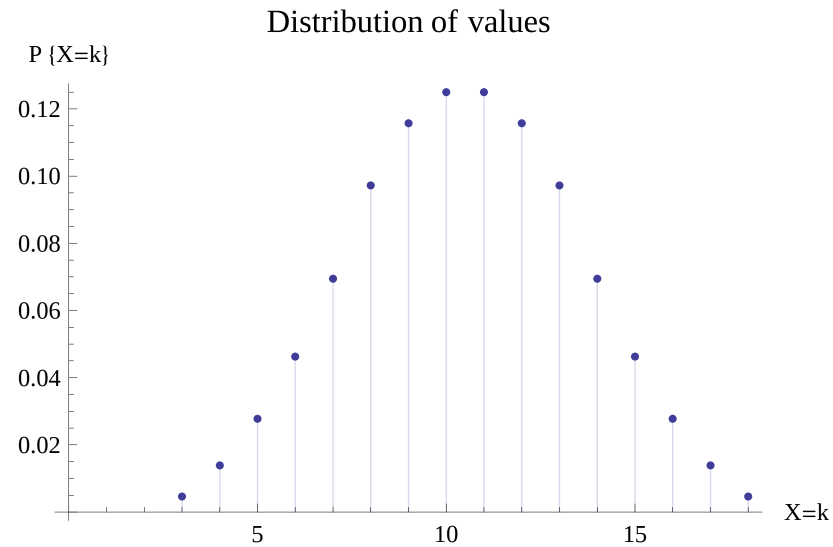
☞ Graph of the pmf for the random variable X giving the face value of a single die.

Note that the nonzero probability masses are uniformly spread.



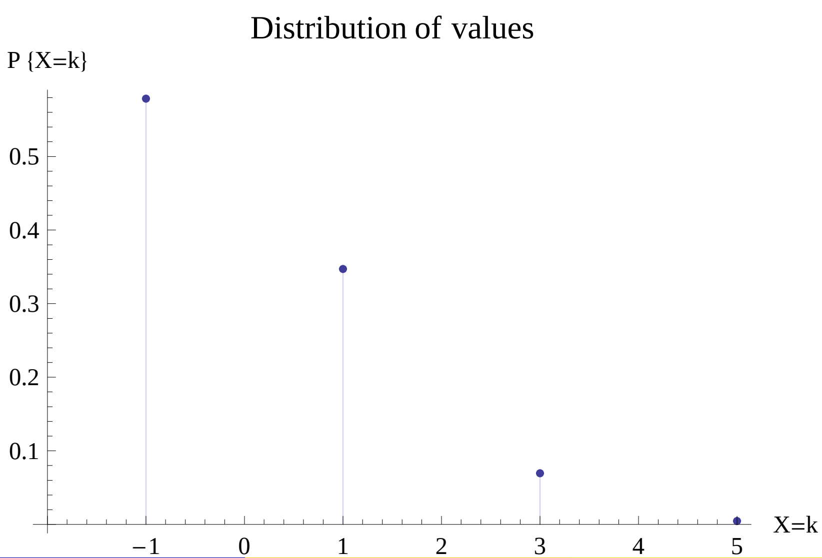
Distribution of values

☞ Graph of the pmf for the sum of values on three dice.



Distribution of values

☞ Graph of the pmf for chuck-a-luck bets.



Principal properties of pmfs

☞ We turn to the principal properties of a pmf $p_X(x)$ for some discrete random variable X.

❶ From the definition of a pmf,

$$0 \leq p_X(x) = \mathbf{P}\{X = x\} \leq 1 \qquad \text{for all real numbers } x.$$

❷ Let $D = \{x_1, x_2, \ldots\}$ include all the possible values of X. Whenever $x \neq y$, the events $\{X = x\}$ and $\{X = y\}$ are disjoint.

Since D includes all possible values of X, by the Addition Rule

$$\sum_{x \in D} p_X(x) = 1$$

Benford's distribution

Example. Take any large collection of numbers, such as the Census report or an almanac. Offer to bet at even odds that a number picked at random from the book will have first significant digit less than 5.

☺The more people you can find to accept this bet, the more you will win!!

Benford's distribution

☞ You might think that all nine digits are equally likely, but this is not so. Actual experiments show the distribution is closer to

$$p(k) = \log_{10}\left(1 + \frac{1}{k}\right) \qquad 1 \le k \le 9$$

This is Benford's distribution, and the actual values are approximately

k	$p(k)$	k	$p(k)$
1	0.301	6	0.067
2	0.176	7	0.058
3	0.125	8	0.051
4	0.097	9	0.046
5	0.079		

☞ Note that

$$p(1) + p(2) + p(3) + p(4) \approx 0.7$$

Benford's distribution

☞ The same rule applies if you look at larger number of a significant digits.

If you look at the first two significant digits, $\{10, 11, \ldots, 99\}$, it is determined by the same probability distribution:

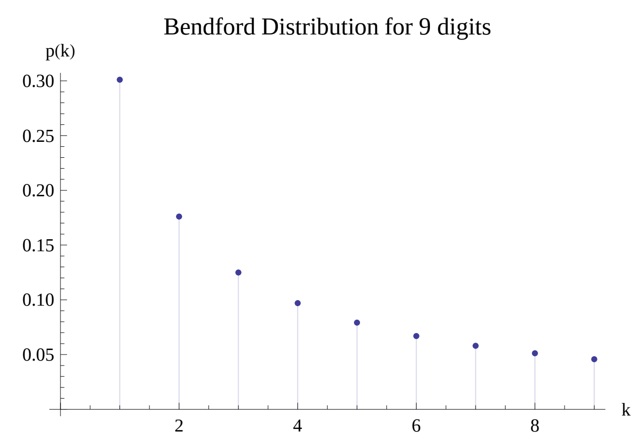
$$p(k) = \log_{10}\left(1 + \frac{1}{k}\right) \qquad 10 \le k \le 99$$

☞ There is an explanation. The tables are printed in base 10 digits, but if you are silicon based (base 2 digits) or from Mars (where I hear then use base 16), you might also expect that your first significant digit is also uniformly distributed.

Benford's distribution is the only pmf which is base-invariant – that is, it is the logarithm of the digits (to the appropriate base) which is uniformly distributed:

$$p(k) = \log_{10}(k+1) - \log_{10}(k).$$

Benford's distribution



Examples of pmfs

☞ One of the great values of random variables is that we begin to see some pmfs occurring with regularity, even though they are defined over random variables from very different sample spaces.

☞ The most important discrete pmf's are those associated with a sequence of Bernoulli trials with success probability p and failure probability $1 - p$.

- Bernoulli: probability of success/failure on one Bernoulli trial.
- Binomial: probability of k successes in n trials.
- Geometric: probability that first success is on kth trial.
- Negative Binomial: probability kth success first occurs on nth trial.

The first two are finite pmf's, the last two are infinite pmf's.

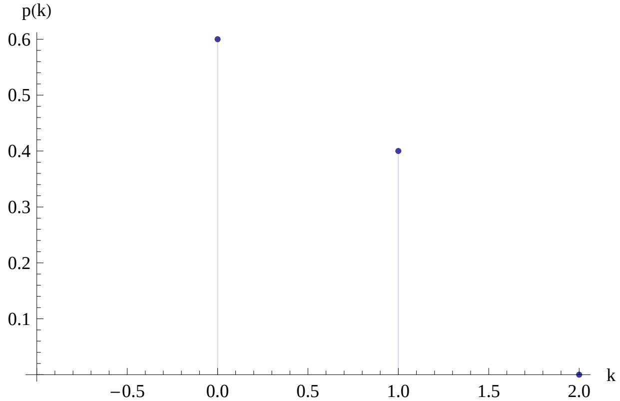
Bernoulli pmf

☞ A Bernoulli pmf is one given by

$$p(1) = p \qquad p(0) = 1 - p$$

and is the probability of success on a single Bernoulli trial.

Bernoulli Distribution for p = 0.4



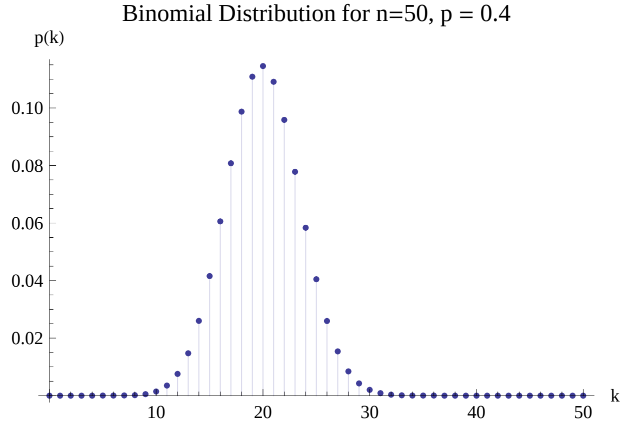
Binomial pmf

☞ A Binomial pmf on n Bernoulli trials is one given by

$$p(k) = \binom{n}{k} p^k (1 - p)^{n-k} \qquad 0 \leq k \leq n$$

and is the probability of k successes in n Bernoulli trials.

Binomial Distribution for n=50, p = 0.4



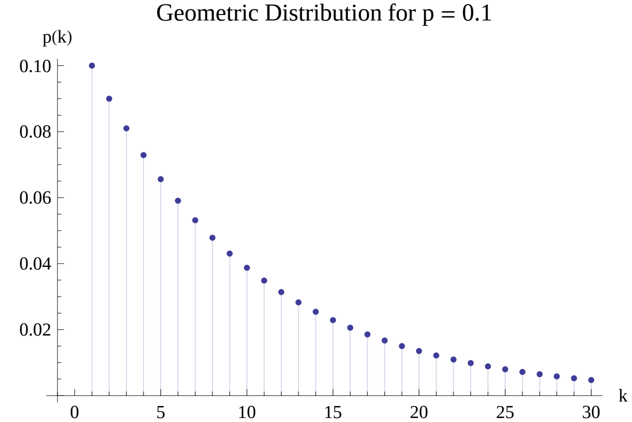
Geometric pmf

☞ A Geometric pmf on Bernoulli trials is one given by

$$p(k) = p(1 - p)^{k-1} \qquad k = 1, 2, 3, \ldots$$

and is the probability of that the first success is on the kth trial.

Geometric Distribution for p = 0.1



Negative Binomial pmf

☞ A Negative binomial pmf is the probability of requiring n trials before we get k successes:

$$p(n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad n = k, k+1, k+2, \ldots$$