CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 11:

- Random Variables (review)
- Probability Distributions of Random Variables
  - Bernoulli
  - Binomial
  - Geometric
  - Negative Binomial (Pascal)
  - ~~Hypergeometric~~
  - [We will delay the Poisson]

Discrete Random Variables: Probability Mass Function

The probability function of a **discrete** random variable X is a function

f_X = Probability Mass Function (PMF)

which assigns a probability to each real number in the range of X and follows the normal rules for a probability space:

$$f_X : R_x \to \mathcal{R}$$

$$\forall a \in R_x \quad f_X(a) \geq 0$$

$$\sum_{a \in R_x} f_X(a) = 1.0$$

Discrete RandomVariables: Probability Distributions

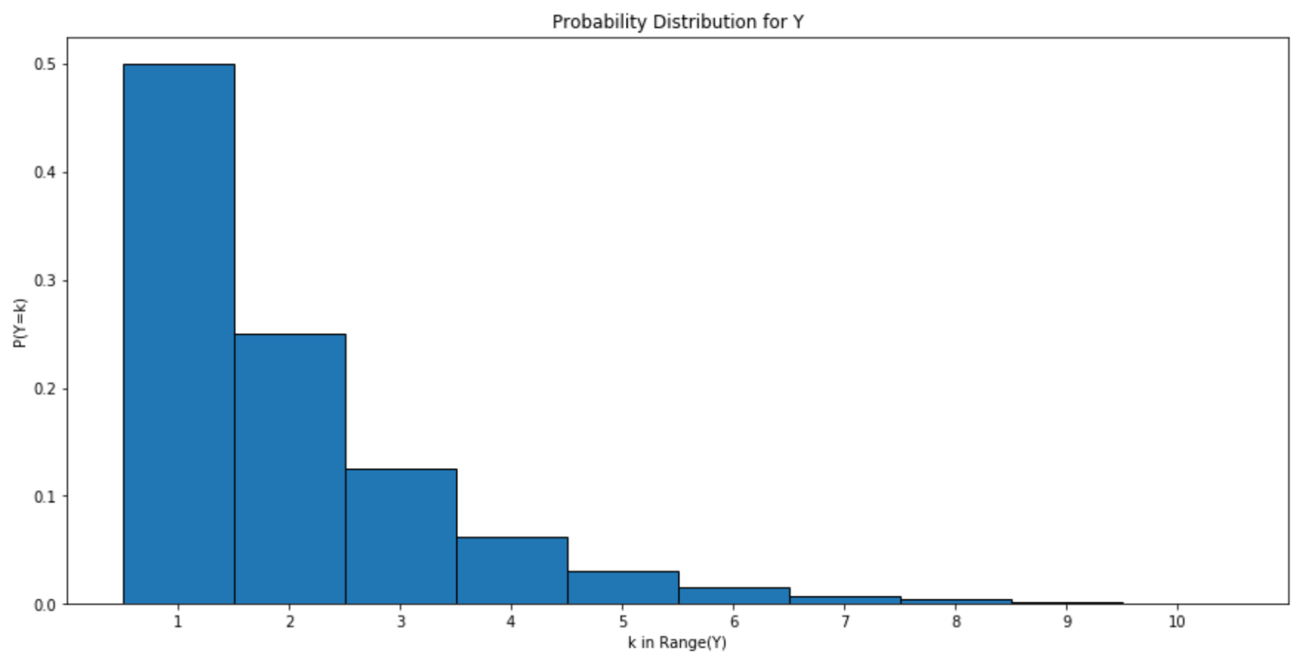
We will emphasize the distributions of random variables, using graphical representations to help our intuitions.

Example:

Y = "The number of tosses of a fair coin until a head appears"

$$R_Y = \{\ 1, 2, 3,\ \ldots\ \}$$

$$f_Y = \{\ \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8},\ \ldots \}$$



Discrete RandomVariables

Notation:

$$P(X = k) =_{def} \mathrm{f}_X(k)$$

$$P(X \neq k) =_{def} 1.0 - \mathrm{f}_X(k)$$

$$P(X \leq k) =_{def} \sum_{a \leq k} \mathrm{f}_X(a)$$

$$P(j \leq X \leq k) =_{def} \sum_{j \leq a \leq k} \mathrm{f}_X(a)$$

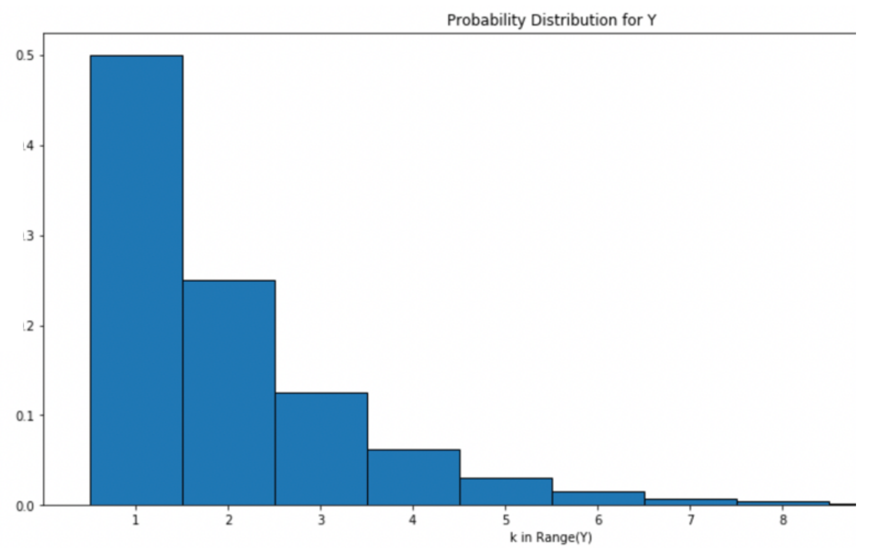
$$P(Y = 4) = \frac{1}{16}$$

$$P(Y < 4) = \frac{7}{8}$$

$$P(2 \leq Y \leq 4) = \frac{7}{16}$$

$$R_Y = \{\, 1, 2, 3, \ldots \,\}$$

$$\mathrm{f}_X = \{\, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \ldots \}$$



Functions of Discrete Random Variables

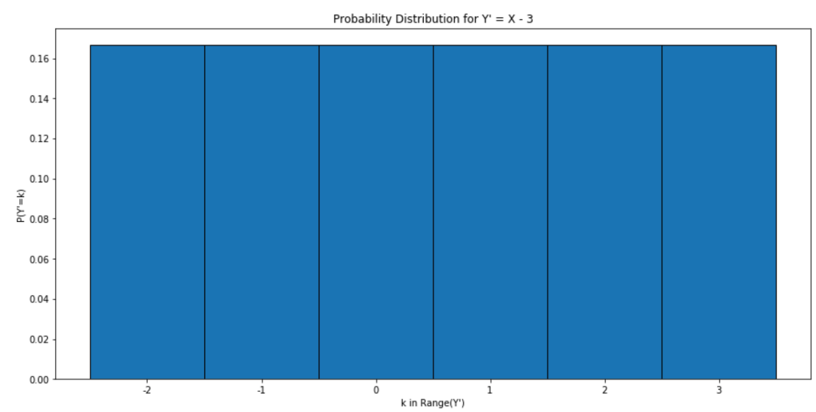
Why did I say you have to be careful?   Two main reasons…

One, the function of a random variable may combine outcomes…

Example: Let Y' = X – 3 and let Y = | X - 3 |

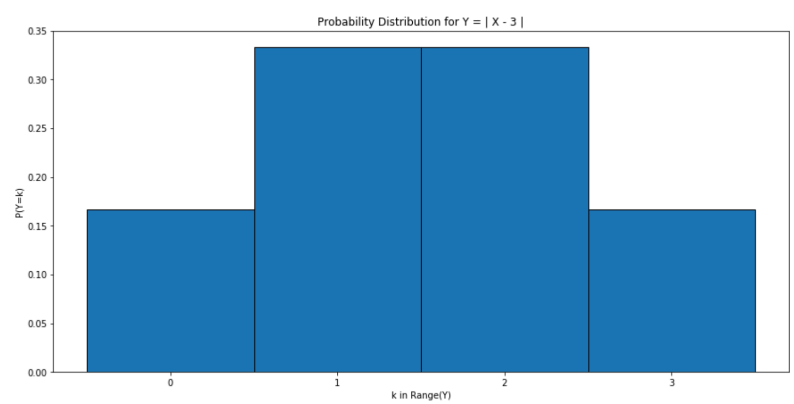
$$R_{Y'} = \{\, -2, -1, 0, 1, 2, 3 \,\}$$

$$f_{Y'} = \{\, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6} \,\}$$



$$R_Y = \{\, 0, 1, 2, 3 \,\}$$

$$f_Y = \{\, \tfrac{1}{6}, \tfrac{2}{6}, \tfrac{2}{6}, \tfrac{1}{6} \,\}$$



$$R_X = \{1, 2, 3, 4, 5, 6\}$$

Functions of Discrete RandomVariables

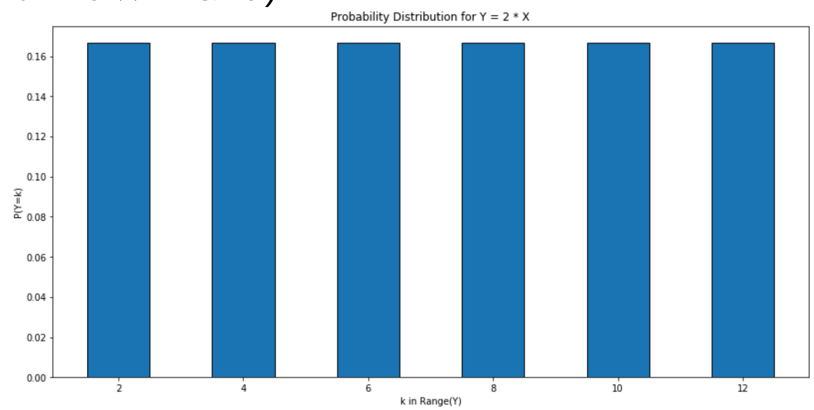
$$f_X = \{ \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6} \}$$

Two, you have to be careful when a random variable is used more than once, since each occurrence refers to a potentially different random outcome!

Let Y = 2 * X (twice the dots showing on a thrown die)

$$R_Y = \{ 2, 4, 6, 8, 10, 12 \}$$

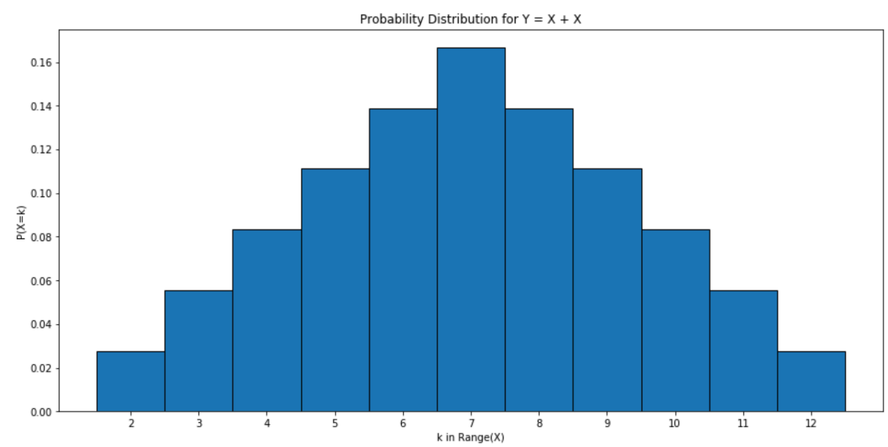
$$f_Y = \{ \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6} \}$$



Let Y = X + X (sum of the dots showing on two thrown dice)

$$R_Y = \{ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \}$$

$$f_Y = \{ \tfrac{1}{36}, \tfrac{2}{36}, \tfrac{3}{36}, \tfrac{4}{36}, \tfrac{5}{36}, \tfrac{6}{36}, \tfrac{5}{36}, \tfrac{4}{36}, \tfrac{3}{36}, \tfrac{2}{36}, \tfrac{1}{36} \}$$



Standard Distributions

We will look at the following discrete distributions:

Discrete Distributions

- Bernoulli
- Binomial (Iterated version of Bernoulli)
- Geometric
- Negative Binomial (Pascal) (Iterated version of Geometric)
- ~~Hypergeometric~~ (in your textbook, but we won't cover it)
- Poisson (We'll do this one separately in about a week).

These are summarized, with useful code for displaying the PMF and CDF in the notebook Distributions.ipynb on the class web site. Wikipedia has very good pages on all these distributions.

We will study these by considering the canonical problems which they describe….

Special Distributions

Any random variable X has a Probability Distribution which can be characterized by

- R_X -- The range of the random variable
- f_X -- The Probability Mass Function (PMF)
- F_X -- The CDF
- $E(X)$ -- Expected value
- $Var(X)$, σ_X -- Variance, Standard Deviation

Next week...

In addition, we are interested in

- The canonical experiment which defines it
- Formulae for calculating f_X and F_X, if such exist (hopefully efficient!)
- Algorithms for generating random variates from the distribution
- Any special properties of the distribution (e.g., the "memoryless property")
- Applications (random experiments which follow that distribution)

Bernoulli Distribution

Suppose you have a coin where the probability of a heads is p and we define the random variable

X = “the number of heads showing on one flipped coin”

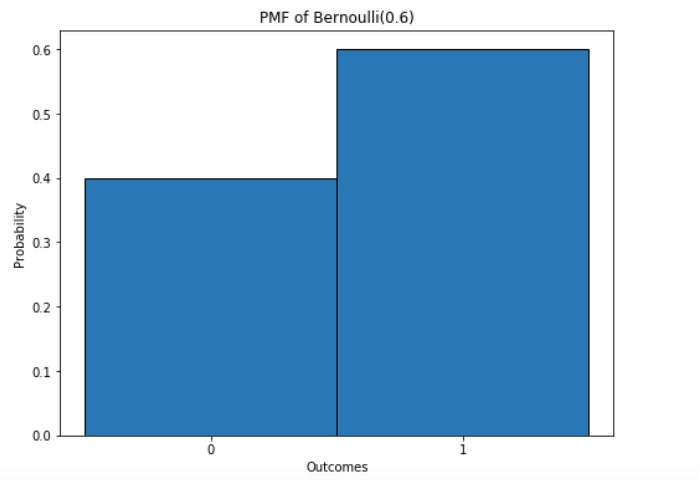
Then we say that X is distributed according to the Bernoulli Distribution with parameter p, and write this as:

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

where

$$R_X = \{0, 1\}$$

$$f_X = \{1 - p, p\}$$



Jacob Bernoulli

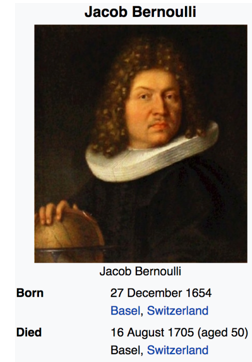
Jacob Bernoulli

Born	27 December 1654 Basel, Switzerland
Died	16 August 1705 (aged 50) Basel, Switzerland

Among other accomplishments, Bernoulli discovered the number ***e*** (but Euler got the credit for “Euler’s Number”).

$$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n$$

Bernoulli Distribution



Each "poke" of such a random variable is called a Bernoulli Trial, and the outcomes are often labelled as

1 = Success   0 = Failure

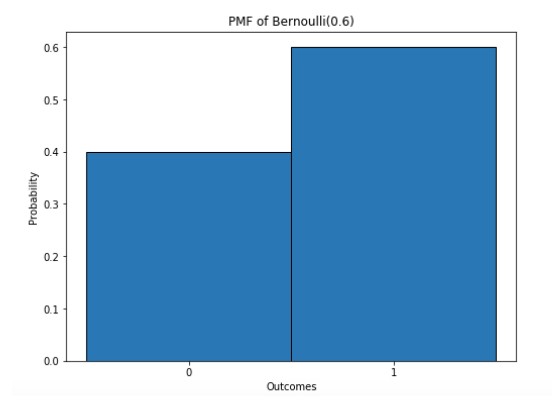
$$X \sim \text{Bernoulli}(\,p\,)$$

Bernoulli Trials, and Bernoulli random variables, describe simple random experiments where there are two possible outcomes, and the probability of the outcomes is fixed by p and (1-p); the notion of "success" and "failure" is just a convenience and does not always correspond to the desirability of the outcome:

$$R_X = \{\,0, 1\,\}$$

$$\mathrm{f}_X = \{\,1 - p, p\}$$

- Will I pass this course or not?
- Am I pregnant or not?
- Do I have cancer or not?



Binomial Distribution

The Binomial Distribution occurs when you count the number of successes in N independent and identically distributed Bernoulli Trials (i.e., p is the same each time).

Formally, if Y ~ Bernoulli(p), and

$$X = \text{“The number of successes in N trials of Y”} = \overbrace{Y + Y + \ldots + Y}^{\text{N times}}$$

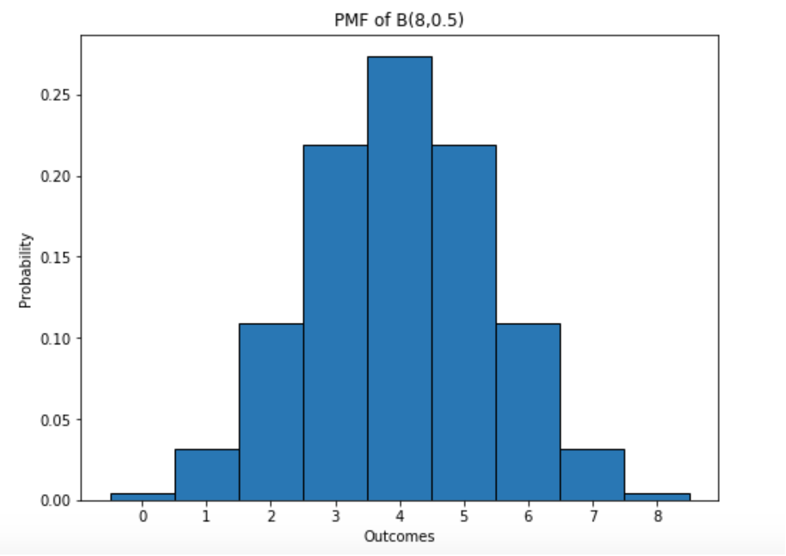
then we say that X is distributed according to the Binomial Distribution with parameters N and p, and write this as:

$$X \sim B(N, p)$$

Where

$$R_X = \{\, 0, \ldots, N \,\}$$

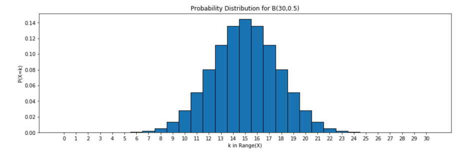
$$f_X(k) = \binom{N}{k} p^k (1-p)^{N-k}$$



Note: k successes and N-k failures SSS…FFFF…

has probability $p^k (1-p)^{N-k}$ and there are $\binom{N}{k}$ such sequences.

Binomial Distribution



For the visual display of this distribution, we will look briefly at the Distributions notebook to get a sense for this.....

The motivation for this distribution comes from the fact that many complex phenomena are composed of the additive effect of many small binary choices or events (Bernoulli Trials!); a vivid illustration of this can be seen in the Galton Board or Quicunx:

https://www.mathsisfun.com/data/quincunx.html

https://www.youtube.com/watch?v=J7AGOptcR1E

Binomial Distribution

Example:

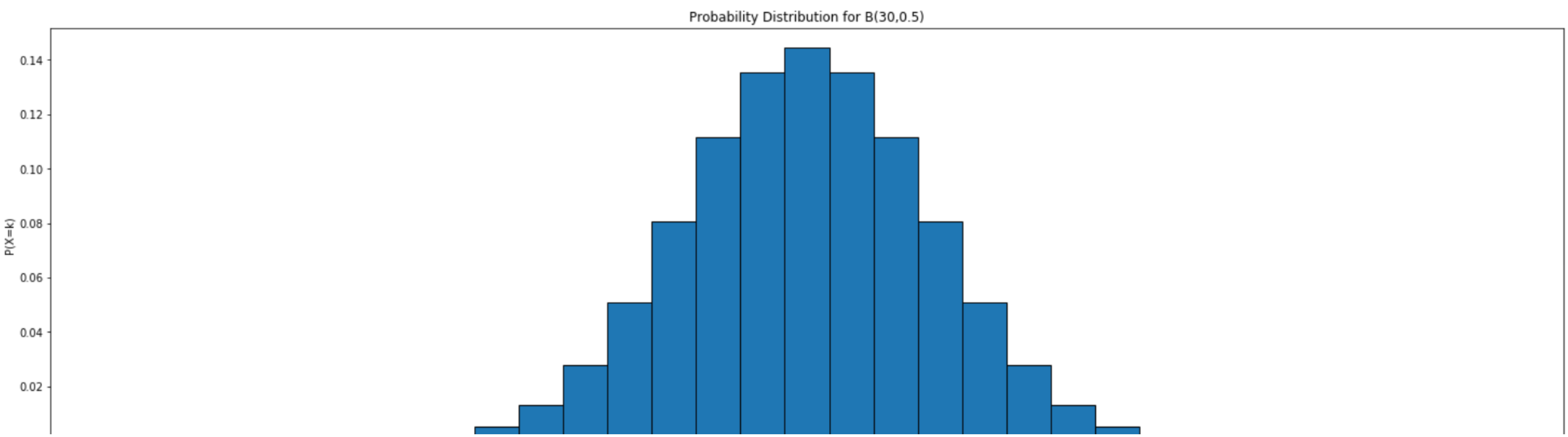
Suppose you have not studied for a True/False test and you randomly guess at each problem, and your probability of getting any particular problem correct is 50%; if there are 30 questions on the test and passing is a 60, what is your probability of passing?

$$P(X \geq 18) = \sum_{18 \leq k \leq 30} \binom{30}{k} 0.5^k \cdot (1 - 0.5)^{30-k}$$

In [7]:
```
def f(N,p,k):
    return comb(N,k) * (p ** k) * ((1-p) ** (N-k))

sum([f(30,0.5,x) for x in range(18,31)])
```

Out[7]: 0.1807973040267825



Binomial Distribution

Phenomena explained by the binomial are widespread throughout ordinary life, biology, engineering, and business:

- You go through 10 traffic lights; what is the probability that you stop at 4 of them?
- The probability of any individual in this class having a tattoo is 0.2; what is the probability that at least 40 people have a tattoo?
- Suppose 5% of tax returns are submitted with fraudulent data and the IRS examines 1% of returns; what is the probability that they will detect 3% of all fraudulent returns?
- There are about 700 gene variants which have been observed to have some influence on height; what is the probability that at least ¾'s of these genes will be dominant and have an influence on a person's height?

Binomial Distribution

The binomial distribution is of widespread applicability, but it has a disadvantage: the only way to compute probabilities is to use the formula

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

and this can involve some very large numbers.... for example:

Input:

$\binom{10000}{5000}$

$\binom{n}{m}$ is the binomial coefficient

Result:

Fewer digits | More digits

15 917 902 635 324 389 483 375 972 736 415 211 886 530 058 374 576 145 504 283 ⋱
191 035 177 726 371 200 957 986 632 628 539 442 222 177 433 585 982 993 226 ⋱
205 580 463 290 870 802 073 985 087 987 219 595 848 962 041 757 866 458 580 ⋱
184 099 587 512 068 914 331 597 813 531 740 514 534 731 996 705 213 945 025 ⋱
384 772 773 360 083 120 537 844 882 395 127 432 175 550 288 318 092 736 464 ⋱
428 179 545 934 936 890 023 546 288 054 736 628 292 721 322 091 972 680 306 ⋱
215 783 976 985 524 868 345 084 786 889 499 461 126 202 336 023 529 898 945 ⋱
892 848 842 759 111 037 432 164 623 520 292 909 554 584 530 402 349 292 778 ⋱
714 312 397 841 036 259 290 830 007 542 173 305 536 549 242 536 830 628 153 ⋱
072 965 334 088 925 565 069 087 515 064 761 594 462 237 620 432 685 223 006 ⋱
267 821 125 937 595 165 711 534 284 824 533 318 106 868 409 528 400 428 469 ⋱
950 435 925 781 799 643 074 138 942 264 944 758 662 628 186 218 375 754 128 ⋱
036 254 688 138 854 475 912 595 618 587 146 845 438 186 146 366 235 072 846 ⋱
821 144 165 546 574 399 328 400 579 417 002 212 869 168 618 937 974 722 788...

Decimal approximation:

More digits

$1.5917902635324389483375972736415211886530058374576145504 28\ldots \times 10^{3008}$

Number length:

3009 decimal digits

For comparison, there are about 10^{87} atoms in the universe....

Geometric Distribution

The Geometric Distribution occurs when you count the number of independent and identically distributed Bernoulli trials until the first success.

Formally, if Y ~ Bernoulli(p), and

X = "The number of trials of Y until the first success"

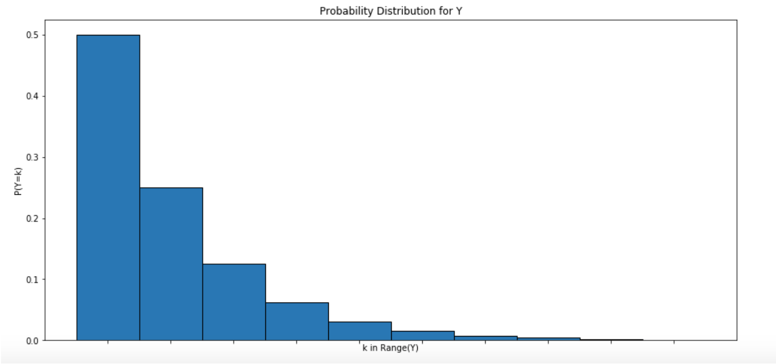
then we say that X is distributed according to the Geometric Distribution with parameter p, and write this as:

$$X \sim G(p)$$

where

$$R_X = \{1, 2, 3, \ldots\}$$

$$f_X(k) = (1-p)^{k-1}p$$



For k, we have k-1 failures and 1 success (FFF... FS), which has probability $(1-p)^{k-1}$ p .

Geometric Distribution

Example

An absent-minded professor has 6 keys on his key ring and does not always remember which of his keys opens his office door. He chooses keys randomly and with replacement to try to open his door.

What is the probability that he opens it on the 3rd try?

Geometric Distribution

Example

An absent-minded professor has 6 keys on his key ring and does not always remember which of his keys opens his office door. He chooses keys randomly and with replacement to try to open his door.

What is the probability that he opens it on the 3rd try?

Solution. This is G(1/6).

$P(X=3) = (5/6)^2 (1/6) = 0.1157$

Geometric Distribution

$$
\begin{array}{llllllll}
R_X = \{ & 1, & 2, & 3, & \ldots & k, & \ldots & \} \\
S = \{ & S, & FS, & FFS, & \ldots & FFF..S, & \ldots & \} \\
f_X = \{ & p, & (1-p)p, & (1-p)^2 p, & \ldots & (1-p)^{k-1} p, & \ldots & \}
\end{array}
$$

But how do we know this is even a distribution? The only question is: Does f_X sum to 1.0?

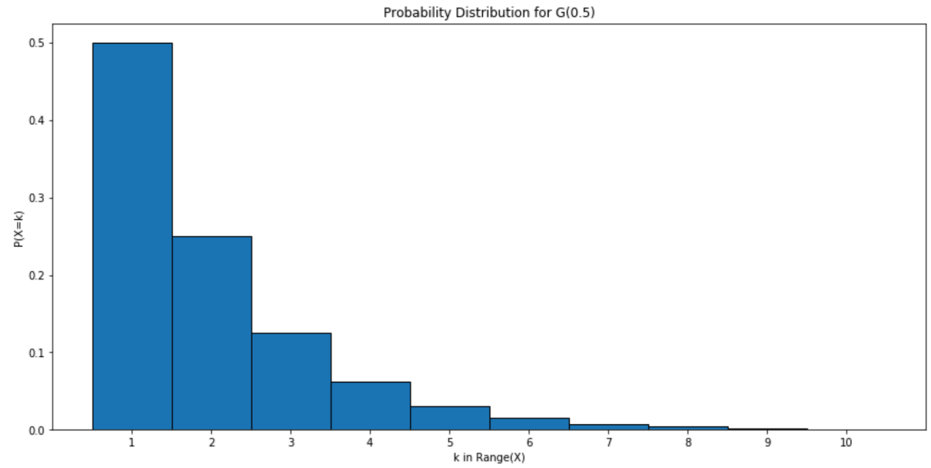
Yes, no worries.... Suppose α is the sum of f_X. Then:

$$
\begin{aligned}
\alpha &= p + (1-p)\,p + (1-p)^2\,p + (1-p)^3\,p + \ldots \\
&= p + (1-p)\,(p + (1-p)\,p + (1-p)^2\,p + \ldots) \\
&= p + (1-p)\,\alpha \\
&= p + \alpha - p\alpha
\end{aligned}
$$

Subtracting α from both sides we have:

$$
\begin{aligned}
& 0 = p - p\alpha \\
\Leftrightarrow \quad & p\alpha = p \\
\Leftrightarrow \quad & \alpha = p/p = 1.0
\end{aligned}
$$

Geometric Distribution



Fortunately, the PMF is easy to compute, and there is are convenient formulae for inequalities:

$$R_X = \{ 1, 2, 3, \ldots \}$$

$$\mathrm{f}_X(k) = (1-p)^{k-1} p$$

$$\begin{aligned} P(X > k) &= (1-p)^k p + (1-p)^{k+1} p + (1-p)^{k+2} p + \ldots \\ &= (1-p)^k \left(p + (1-p)p + (1-p)^2 p + (1-p)^3 p + \ldots \right) \\ &= (1-p)^k \end{aligned}$$

$$P_X(X \leq k) = 1.0 - P_X(X > k) = 1.0 - (1-p)^k$$

Geometric Distribution

Example

From an ordinary deck of 52 cards we draw cards at random and *with replacement*, and successively until an ace is drawn. What is the probability that at least 10 draws are needed?

Geometric Distribution

Example

From an ordinary deck of 52 cards we draw cards at random and *with replacement*, and successively until an ace is drawn. What is the probability that at least 10 draws are needed?

Solution: The probability of an ace is 4/52 = 1/13. Thus:

$$P(X > 9) = (1 - 1/13)^9 = 0.4866$$

Negative Binomial (Pascal) Distribution

The Negative Binomial is simply an "iterated" version of the Geometric.

Formally, if Y ~ Bernoulli(p) and

X = "The number of trials of Y until m successes occur"

$$= \underbrace{Y + \dots + Y}_{\text{m times}}$$

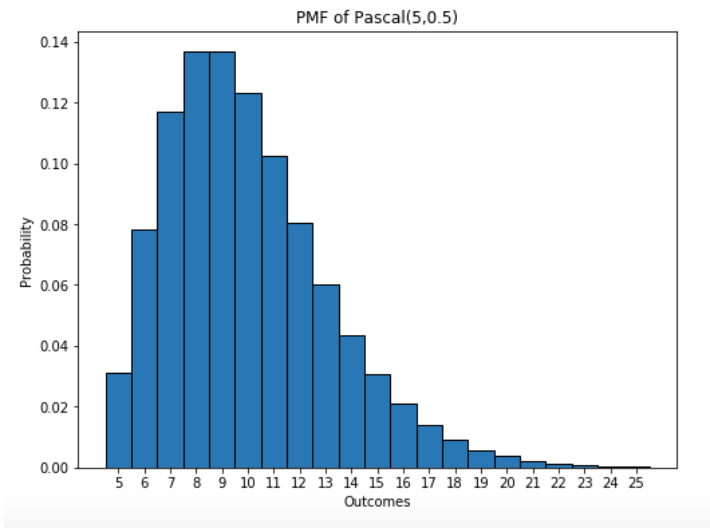
then we say that X is distributed according to the Pascal Distribution with parameters m and p, and write:

$$X \sim Pascal(m, p)$$

where

$$R_X = \{m, m+1, m+2, \dots\}$$

$$f_X(k) = \binom{k-1}{m-1} p^m (1-p)^{k-m}$$



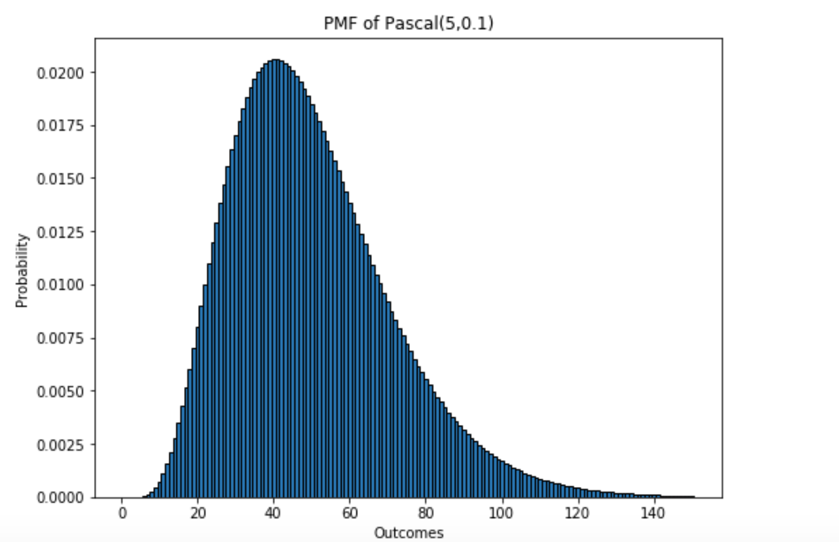
Pascal Distribution

Example

Suppose you are throwing darts at a target for practice, and you decide that you will keep throwing until you hit the bull's eye 5 times. Suppose your probability of hitting the bullseye is 10%. What is the probability it takes exactly 10 throws?

Solution:

$$f_X(10) = \binom{9}{4} 0.1^5 \cdot 0.9^5 = 7.44 \times 10^{-4}$$



C(9,4) * 0.1^(5) * 0.9 ^(5)

Extended Keyboard   Upload

Assuming "C" is a math function | Use as a

Input:

$\binom{9}{4} \times 0.1^5 \times 0.9^5$

Result:

0.0007440174

```
# Just a more convenient syntax:

from scipy.special import comb

def C(N,K):
    return comb(N,K,exact=True)
```

```
In [6]: C(9,4)*0.1**5 * 0.9 ** 5
Out[6]: 0.0007440174000000003
```

Pascal Distribution: Digression

CORRESPONDENCE Open Access

Low dispersion in the infectiousness of COVID-19 cases implies difficulty in control

Daihai He[1*], Shi Zhao[2,3], Xiaoke Xu[4], Qiangying Lin[5], Zian Zhuang[1], Peihua Cao[6], Maggie H. Wang[2,3], Yijun Lou[1], Li Xiao[7], Ye Wu[8,9*] and Lin Yang[10*]

Abstract

The individual infectiousness of coronavirus disease 2019 (COVID-19), quantified by the number of secondary cases of a typical index case, is conventionally modelled by a negative-binomial (NB) distribution. Based on patient data of 9120 confirmed cases in China, we calculated the variation of the individual infectiousness, i.e., the dispersion parameter k of the NB distribution, at 0.70 (95% confidence interval: 0.59, 0.98). This suggests that the dispersion in the individual infectiousness is probably low, thus COVID-19 infection is relatively easy to sustain in the population and more challenging to control. Instead of focusing on the much fewer super spreading events, we also need to focus on almost every case to effectively reduce transmission.

Keywords: COVID-19, Basic reproductive number, Dispersion, Negative binomial, Mitigation

Biometrical Journal

Article

On the Use of the Negative Binomial in Epidemiology

Prof. B. M. Bennett

First published: 1981 | https://doi.org/10.1002/bimj.4710230109 | Citations: 26

Cumulative Distribution Functions

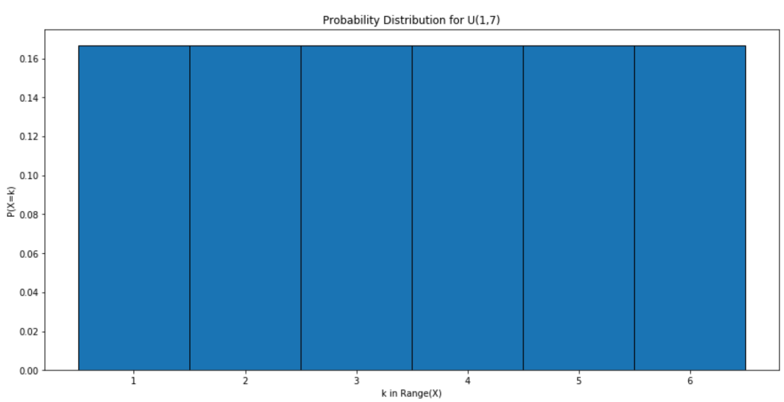
One more topic before the lab tomorrow!

The Cumulative Distribution Function (CDF) for a random variable X shows what happens when we keep track of the sum of the probability distribution from left to right over its range:

$$F_X(k) = P(X \leq k) = \sum_{a \leq k} f_X(a)$$

Example: X = "The number of dots shown on a thrown die"

Probability Distribution Function P_X



Cumulative Distribution Function F_X