Geometric Distribution

Definitions:

- Uniform Distribution: The probability of each valid outcome is the same.
- Geometric Distribution: The number of Bernoulli trials until you get a success
- Pascal Distribution: The number of Bernoulli trials until you get r successes
- Geometric Distribution The number of times you roll a die until you get a one.
- The number of trips you make a trip with a car until something fails.
- The number of days until you an accident happens at work...

Distribution	description	pdf	mgf	mean	variance
Bernoulli trial	flip a coin obtain m heads	$p^m q^{1-m}$	$q + p/z$	p	p(1-p)
Binomial	flip n coins obtain m heads	$\binom{n}{m} p^m q^{n-m}$	$(q+p/z)^n$	np	np(1-p)
Hyper Geometric	Bernoulli trial without replacement	$\frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}$			
Uniform range = (a,b)	toss an n-sided die	$1/n \quad a \le m \le b$ $0 \quad otherwise$	$\left(\frac{1}{n}\right)\left(\frac{1+z+z^2+...+z^{n-1}}{z^b}\right)$	$\left(\frac{a+b}{2}\right)$	$\left(\frac{(b-a+1)^2-1}{12}\right)$
Geometric	Bernoulli until 1st success	$p\, q^{k-1}$	$\left(\frac{p}{z-q}\right)$	$\left(\frac{1}{p}\right)$	$\left(\frac{q}{p^2}\right)$
Pascal	Bernoulli until rth success	$\binom{k-1}{r-1} p^r q^{k-r}$	$\left(\frac{p}{z-q}\right)^r$	$\left(\frac{r}{p}\right)$	$\left(\frac{rq}{p^2}\right)$

Geometric Distribution:

A geometric distribution is one where you conduct a Bernoulli trial (think: flip a coin) until you get a success. The pdf for a geometric distribution is:

$$f(k) = \begin{cases} p\, q^{k-1} & k = 1, 2, 3 \\ 0 & otherwise \end{cases}$$

where 'p' is the probability of a success and x is the number of flips it takes before you get a success.

To see this, consider the following with tossing a coin. Assume the probability of a heads is 'p'. The probability of getting a heads on nth flip is the probability of getting n-1 tails followed by a heads:

- x=1: f() = p
- x = 2: f() = p q
- x = 3: f() = p q^2
- etc.

Note that the '1' in the notation means the game is over after the first success. You might guess that there will be more general distribution where you look for m successes. You'd be right...

The pdf for a geometric distribution looks like the following:

p = 0.9

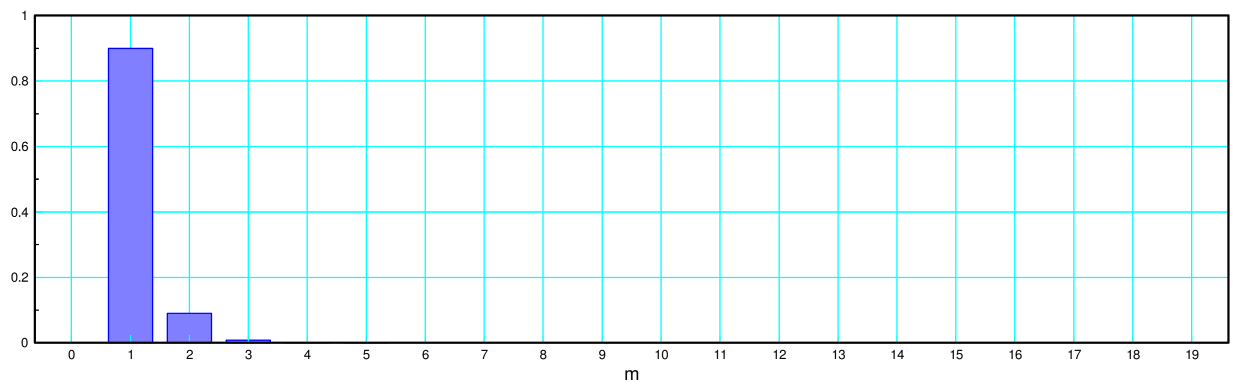
```
f = zeros(20,1);

p = 0.9;

for i=1:20
   f(i) = p * (1-p)^(i-1);
   end
```

Note that the probability that something happens is one:

```
-->sum(f)
    1.
```

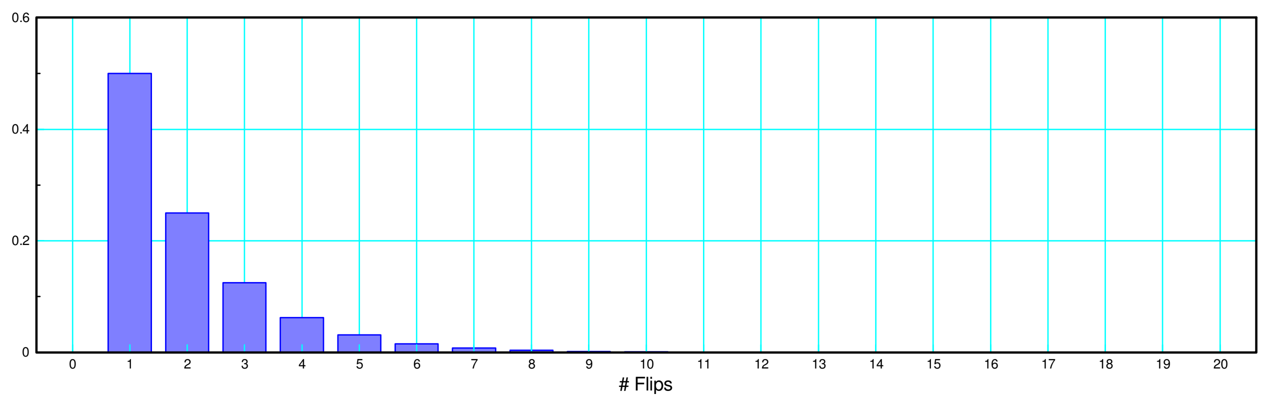
Plotting the pdf shows what the distribution looks like:

```
-->bar(f)
-->xlabel('Number of Flips')
```

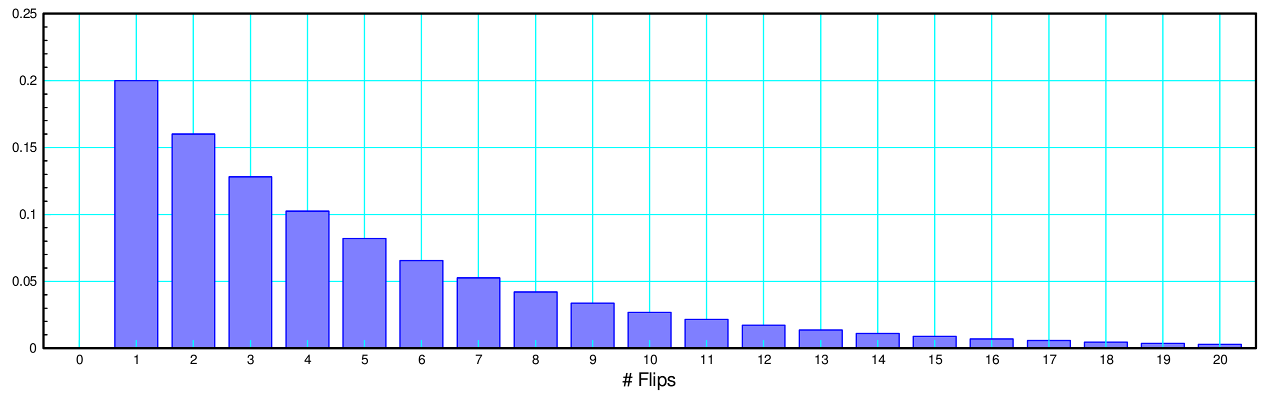


pdf for a geometric distribution with p = 0.9

Repeating for p = 0.5 and 0,2:



pdf for a geometric distribution with p = 0.5



pdf for a geometric distribution with p = 0.2

Note that for a geometric distribution, the probability of a success for each toss is the same. Examples of this would be:

- Tossing a coin until you get a heads
- Betting on 10-black in Roulette until you finally win
- Buying a lottery ticket each week until you finally win
- Trying to open a door with n keys where you replace the key after each trial and try again (and again and again..) This is called sampling with replacement.

Mean for a Geometric Distribution:

$$\mu = \sum_{k=1}^{\infty} k \cdot p \cdot q^{x-1}$$

$$\mu = p(1 + q + 2q^2 + 3q^3 + 4q^4 + ...)$$

Variance for a Geometric Distribution:

$$\sigma^2 = \sum_{k=1}^{\infty} (k - \mu)^2 \cdot p \cdot q^{k-1}$$

You can kind of see that we need a better tool. That will be moment generating functions (coming in a lecture shortly.). The net result is going to be....

$$\mu = \frac{1}{p}$$

$$\sigma^2 = \frac{q}{p^2}$$

Moment Generating Function for an Exponential Distribution:

The time-series (where m means time) is

$$x(k) = q \cdot x(k-1)$$

$$x(1) = p$$

Taking the z-transform

$$x(k) = q \cdot x(k-1) + p\ \delta(k-1)$$

$$X = q\ z^{-1} X + p\ z^{-1}$$

Solve for X

$$(z-q)X = p$$

$$\psi = \left(\frac{p}{z-q}\right)$$

Using this, you can find moments as well as the mean and variance[1]

Zeroth Moment:

- m_o must be 1.000 to be a valid pdf (all probabilities add to 1)

$$m_0 = \psi(z=1) = \left(\frac{p}{z-q}\right)_{z=1} = \left(\frac{p}{1-q}\right) = \left(\frac{p}{p}\right) = 1$$

1st-Moment (mean)

- The first moment is the mean

$$m_1 = -\frac{d}{dz}(\psi(z))_{t=1} = -\psi'(z=1)$$

$$m_1 = -\frac{d}{dz}\left(\frac{p}{z-q}\right)_{z=1} = -\left(\frac{-p}{(z-q)^2}\right)_{t=1} = \left(\frac{p}{(1-q)^2}\right) = \left(\frac{p}{p^2}\right) = \left(\frac{1}{p}\right)$$

$$\mu = m_1 = \left(\frac{1}{p}\right)$$

Second Moment:

- $m_2 = \frac{d^2}{dz^2}(\psi(z))_{z=1} = \psi''(z=1)$
- $m_2 = \frac{d}{dz}\left(\frac{p}{(z-q)^2}\right) = \left(\frac{-2p(z-q)}{(z-q)^4}\right)_{z=1} = \left(\frac{2p(1-q)}{(1-q)^4}\right) = \left(\frac{2p^2}{p^4}\right) = \left(\frac{2}{p^2}\right)$

[1] Probability and Statistics, Morris DeGroot

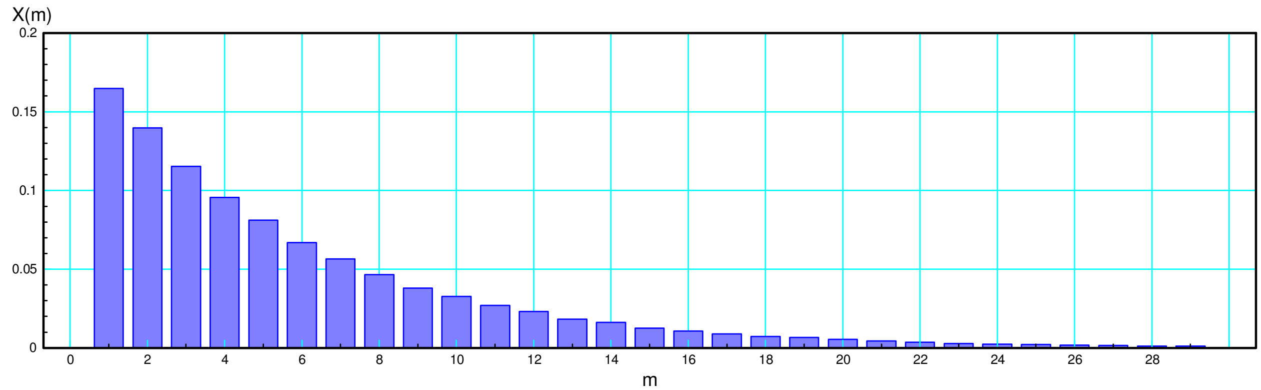
Variance:

- $\sigma^2 = m_2 - m_1^2$
- $\sigma^2 = \left(\frac{2}{p^2}\right) - \left(\frac{1}{p}\right)^2 = \left(\frac{1}{p^2}\right)$
- $\sigma^2 = \left(\frac{q}{p^1}\right)$ *actual variance: not sure why I'm off by q*

That was a *lot* easier than applying the definition. z-transforms are really useful

Matlab Example: Toss a die until you roll a 6 (p = 1/6). Determine the mean and standard deviation after 10,000 games

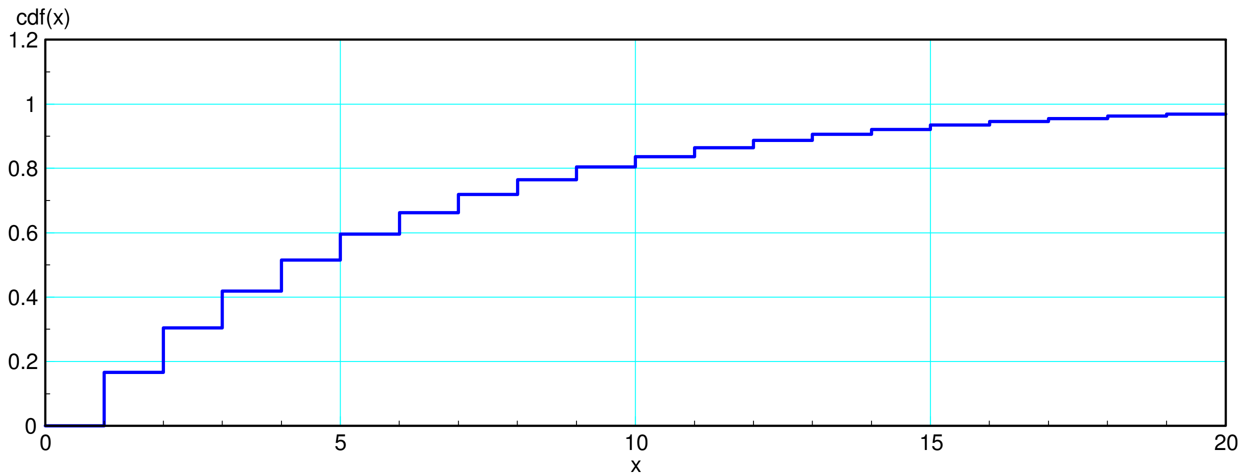
```
N = 1e5;
X = zeros(100,1);
p = 1/6;
q = 1-p;

for i=1:N

   n = 1;

   while(rand > p)
       n = n + 1;
   end

   X(n) = X(n) + 1;
end

X = X / N;

M = [1:100]';
x = sum(M .* X);
s2 = sum(X .* (M-x).*(M-x));

disp([x,1/p])
disp([s2,q/(p*p)])


          Sim        Calc
x       6.0179    6.0000
var    30.0712   30.0000
```



Experimental pdf for tossing a die until you roll a 6

The cdf is the integral (sum) of the pdf from 0 to x:

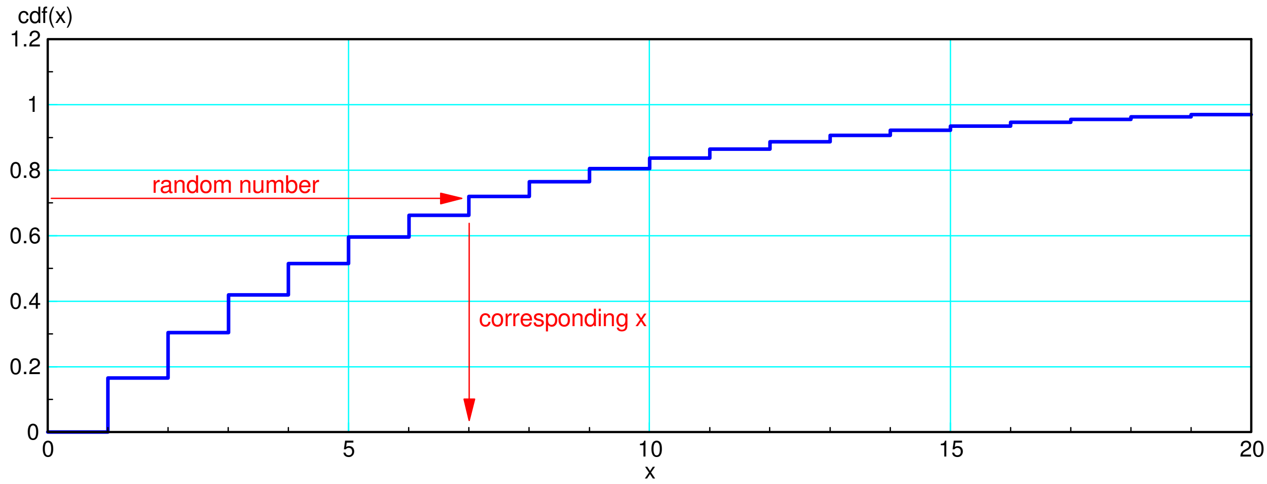
```
cdf = 0*X;
for i=1:length(cdf)
   cdf(i) = sum(pdf(1:i));
   end
```



Experimental cdf for a geometric distribution

The cdf is a more useful way of generating x

- Pick a random number in the interval of (0, 1)
  - This is the y-coordinate
- Find the corresponding x



You can also compute the cdf using z-transforms (with a lot less work). The cdf is the integral of the pdf:

$$cdf = pdf \cdot \left(\frac{z}{z-1}\right)$$

or

$$cdf = \left(\frac{p}{z-q}\right)\left(\frac{z}{z-1}\right)$$

$$= \left(\frac{p}{(z-q)(z-1)}\right)z = \left(\frac{1}{z-1} + \frac{-1}{z-q}\right)z$$

$$cdf = 1 - q^x$$

Solving backwards

$$x = ceil\left(\frac{\ln(1-cdf)}{\ln(q)}\right)$$

To find x:

- Pick a random number in the range of (0, 1)
- Convert to x using the above formula

Gauss' Dilemma:

This is a game which

- No-one will play because you (almost) always lose, and
- No-one will offer because the expected winnings are infinite.

Pay some amount, like $100 to play. Start with $1 in the pot.

Toss a coin. If it comes up tails, double the pot.

Keep playing until the coin comes up heads. Once that happens, the game ends and you collect your winnings.

This is a geometric distribution with the probability density function being

# Tosses (m)	1	2	3	4	5	6
Probability (p)	1/2	1/4	1/8	1/16	1/32	1/64
Pot (x)	1	2	4	8	16	32

The expected winnings are the cost to play (-$100) plus the sum of the pots times their probabilities:

$$E = \sum p(m) \cdot x(m) - 100$$

$$E = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots - 100$$

$$E = \infty$$

With infinite expected winnings, this sounds like a good game to play. If you play it, you almost always lose.

For example, play 10 games in Matlab:

```
% Gauss' Dilemma

N = 10;
Winnings = 0;
p = 0.5;

for i=1:N

   Pot = 1;

   while(rand > p)
       Pot = Pot * 2;
   end

   Winnings = Winnings + Pot - 100;
end

Winnings / N

-98.2
```

Each time you play, you lose on average $98.2

Play the game 1000 times and you lose $95 each time you play (meaning you're now down $95,000):

```
Winnings / N  =  -95.0180
```

Play 1 million times, and you're down $89 each time you play (meaning you're down $89 million)

```
Winnings / N  =  -89.7185
```

Likewise, it's a really bad game to play. With an expected winnings of infinity, it's also a really bad game to offer.

Hence the name Gauss' Dilemma