Geometric Distribution

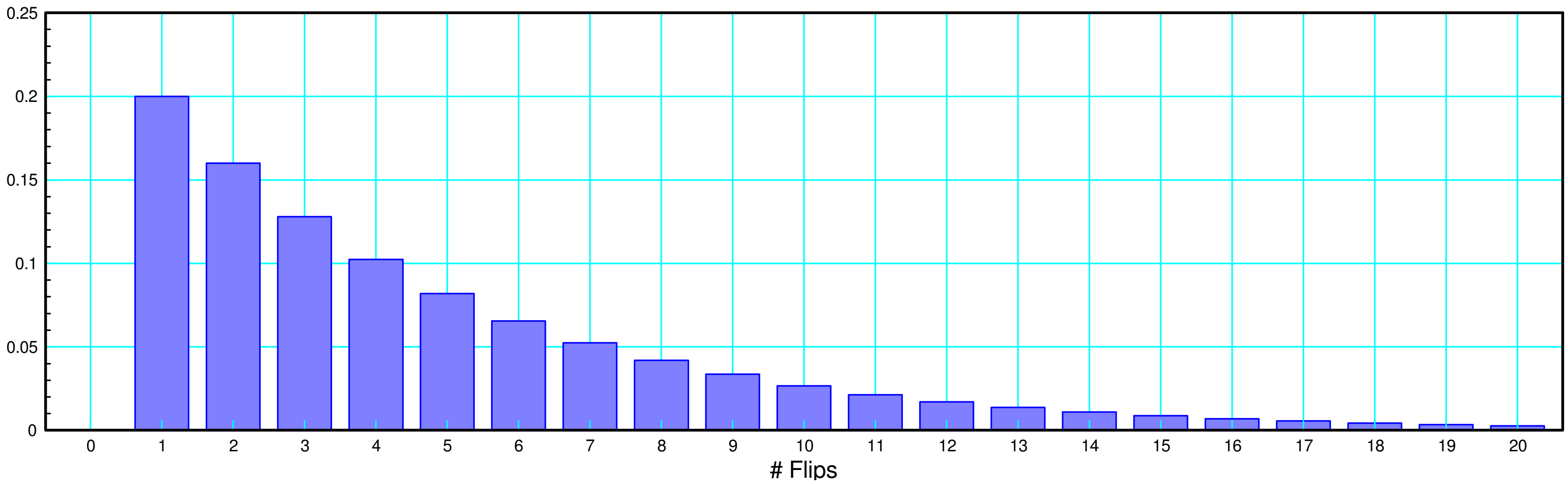
ECE 341: Random Processes

Lecture #10

note: All lecture notes, homework sets, and solutions are posted on www.BisonAcademy.com

Geometric Distribution

- The number of Bernoulli trials until you get a success
- # die rolls until you get a 1
- # times you do the dishes until someone notices
- # of car trips you tke until something fails
- # of days until you make a mistake at work your boss notices
- etc



pdf / mgf / mean / variance

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	$1/n \quad a \le m \le b$ $0 \quad otherwise$	$\left(\frac{1+z+z^2+\ldots+z^{n-1}}{n\, z^b}\right)$	$\left(\frac{a+b}{2}\right)$	$\left(\frac{(b+1-a)^2-1}{12}\right)$
Geometric	Bernoulli until 1st success	$p\, q^{k-1}$	$\left(\frac{p}{z-q}\right)$	$\left(\frac{1}{p}\right)$	$\left(\frac{q}{p^2}\right)$

Geometric Distribution:

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

pdf:

$$f(k) = p\, q^{k-1}\, u(k-1)$$

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

Example: Toss a coin.

- p(success) = p

$f(0) = 0$

$f(1) = p$

$f(2) = p\ q$

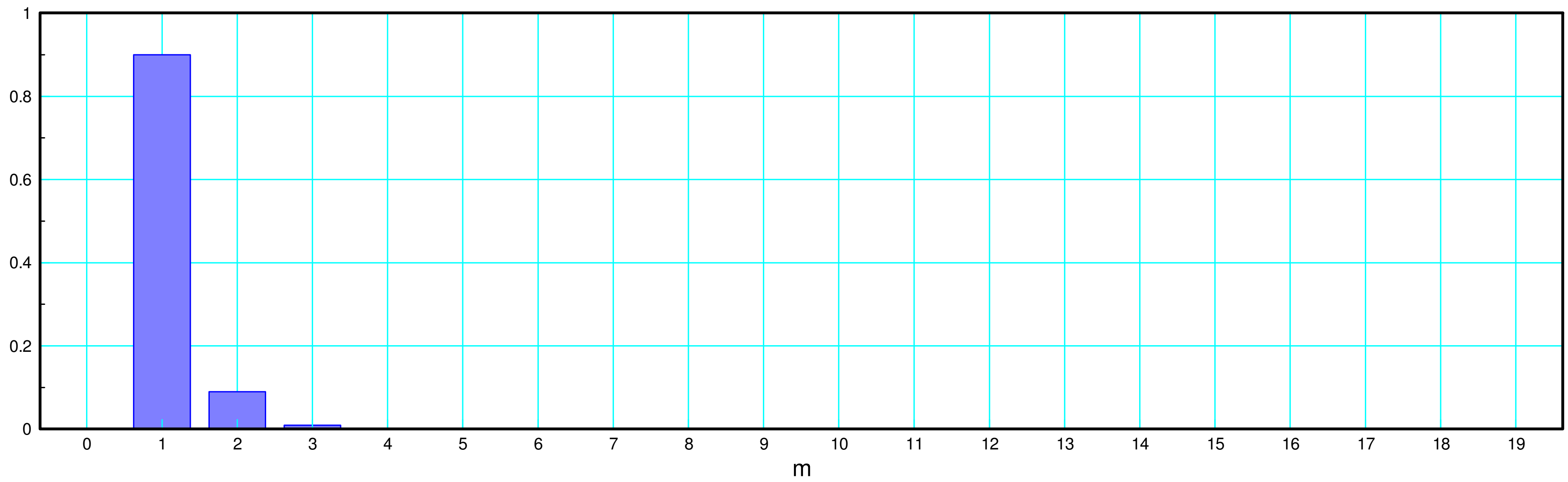
$f(3) = p\ q^2$

$f(4) = p\ q^3$

etc.

Geometric with p = 0.9

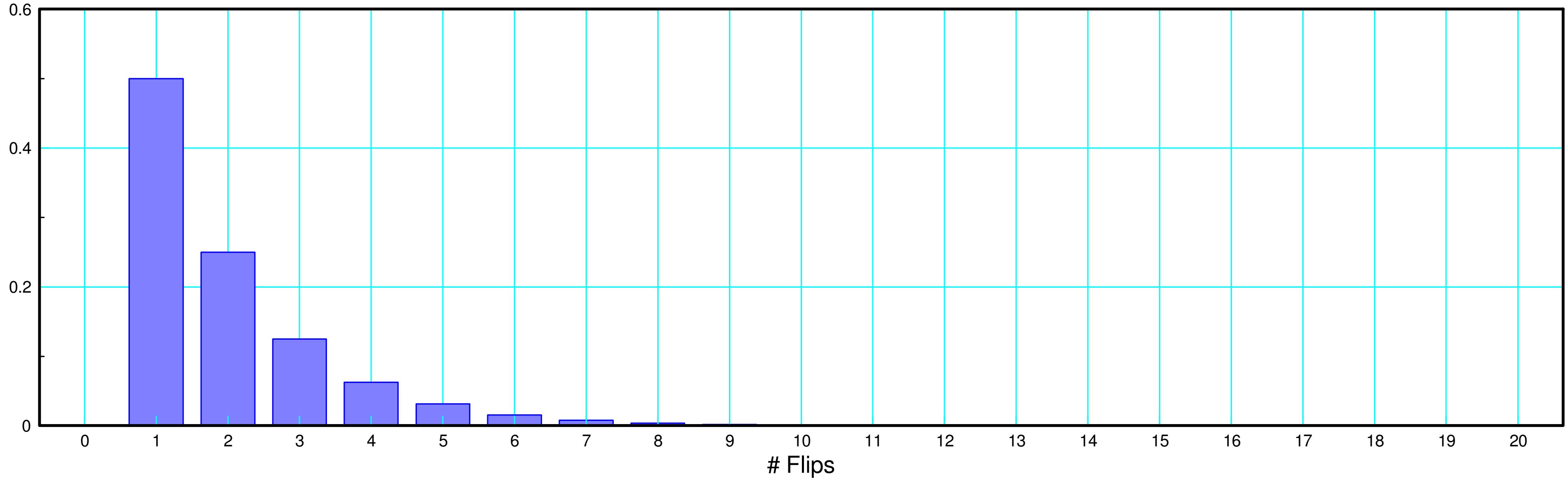
- $f(k) = (0.9)\ (0.1)^{k-1}\ u(k-1)$
- mean = 1.111
- variance = 0.123



pdf for a geometric distribution with p = 0.9

Geometric with p = 0.5

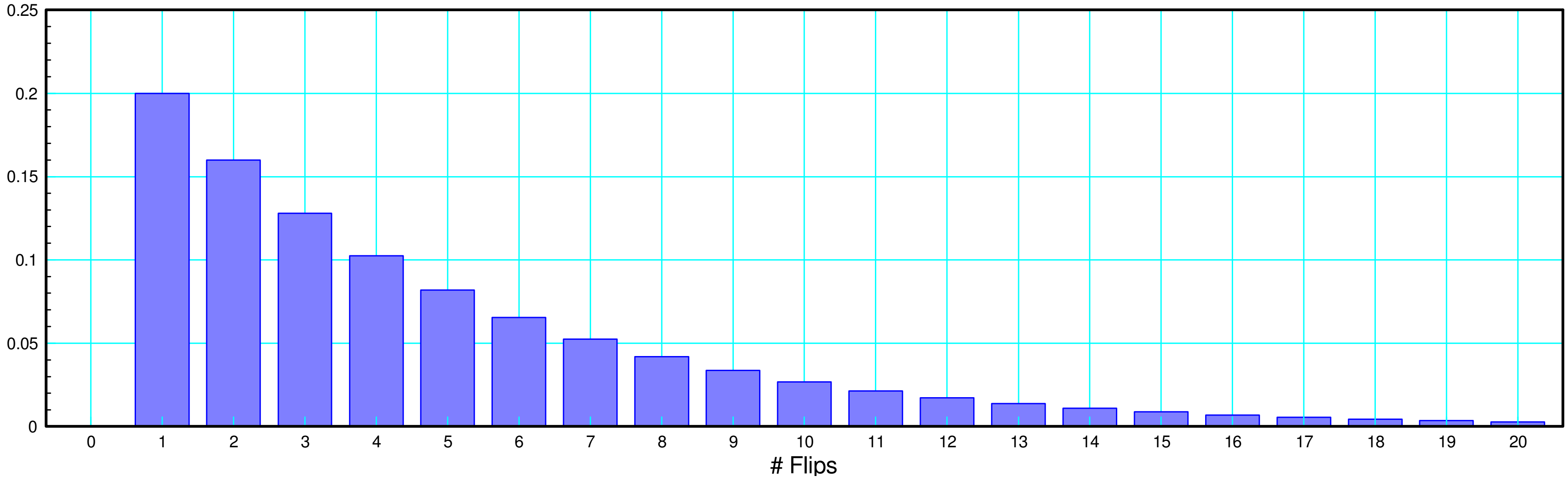
- $f(k) = (0.5)\,(0.5)^{k-1}\;u(k-1)$
- mean = 2.00
- variance = 2.000



pdf for a geometric distribution with p = 0.5

Geometric with p = 0.2

- $f(k) = (0.2)\ (0.8)^{k-1}\ u(k-1)$
- mean = 5.00
- variance = 20.000



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 and Variance (take 1)

Mean for a Geometric Distribution:

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

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

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.

Moment Generating Function

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

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

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

Moments

- Moment generating functions are useful for generating moments
- These allow you to compute the mean and standard deviation.

Zeroth Moment: (valid pdf)

$$m_0 = \psi(z)_{z=1}$$

$$m_0 = 1$$ *for this to be a valid distribution*

1st Moment (mean)

$$m_1 = -\psi'(z)_{z=1}$$ *m1 is the mean of the pdf*

2nd Moment

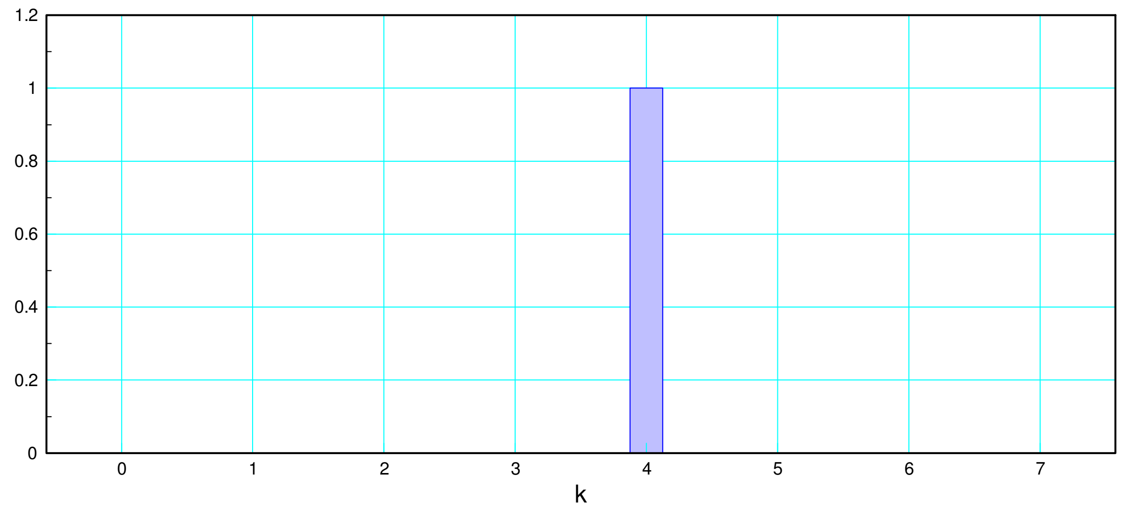
$$m_2 = \psi''(z)_{z=1}$$

Variance

$$\sigma^2 = m_2 - m_1 - m_1^2$$

Example #1: $y(k) = \delta(k-4)$

$$\psi(z) = \frac{1}{z^4}$$



Zeroth moment

$$m_0 = \psi(z=1) = 1$$

1st Moment

$$\psi'(z) = \frac{-4}{z^5}$$

$$m_1 = -\psi'(z)_{z=1} = 4$$

2nd Moment

$$\psi''(z) = \frac{20}{z^6}$$

$$m_2 = \psi''(z=1) = 20$$

$$\sigma^2 = m_2 - m_1 - m_1^2 = 0$$

Example: 6-sided die

$$\psi(z) = \frac{1}{6}\left(\frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} + \frac{1}{z^4} + \frac{1}{z^5} + \frac{1}{z^6}\right)$$

$$m_0 = \frac{1}{6}(1 + 1 + 1 + 1 + 1 + 1)$$

$$m_0 = 1$$

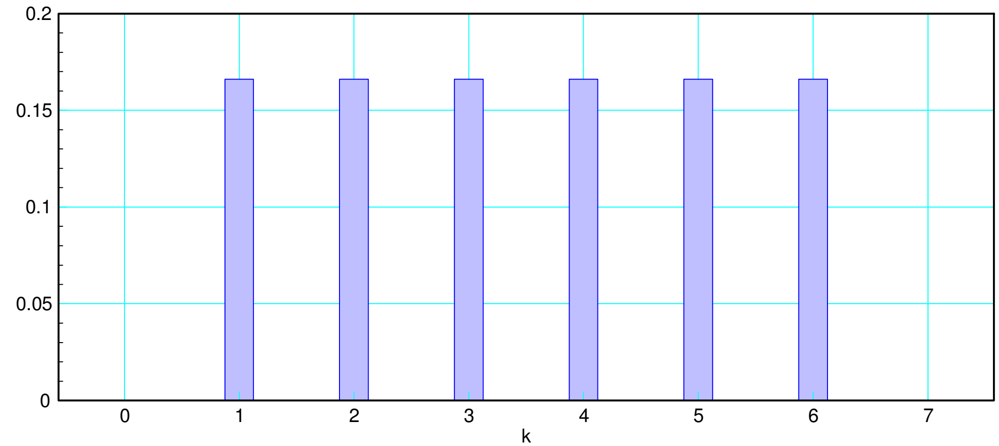
$$\psi'(z) = \frac{1}{6}\left(\frac{-1}{z^2} + \frac{-2}{z^3} + \frac{-3}{z^4} + \frac{-4}{z^5} + \frac{-5}{z^6} + \frac{-6}{z^7}\right)$$

$$m_1 = -\psi'(z = 1) = 3.500$$

$$\psi''(z) = \frac{1}{6}\left(\frac{1\cdot 2}{z^3} + \frac{2\cdot 3}{z^4} + \frac{3\cdot 4}{z^5} + \frac{4\cdot 5}{z^6} + \frac{5\cdot 6}{z^7} + \frac{6\cdot 7}{z^8}\right)$$

$$m_2 = \frac{112}{6}$$

$$\sigma^2 = m_2 - m_1 - m_1^2 = 2.91667$$



Example: Geometric Distribution

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

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

$$\psi'(z) = \left(\frac{-p}{(z-q)^2}\right)$$

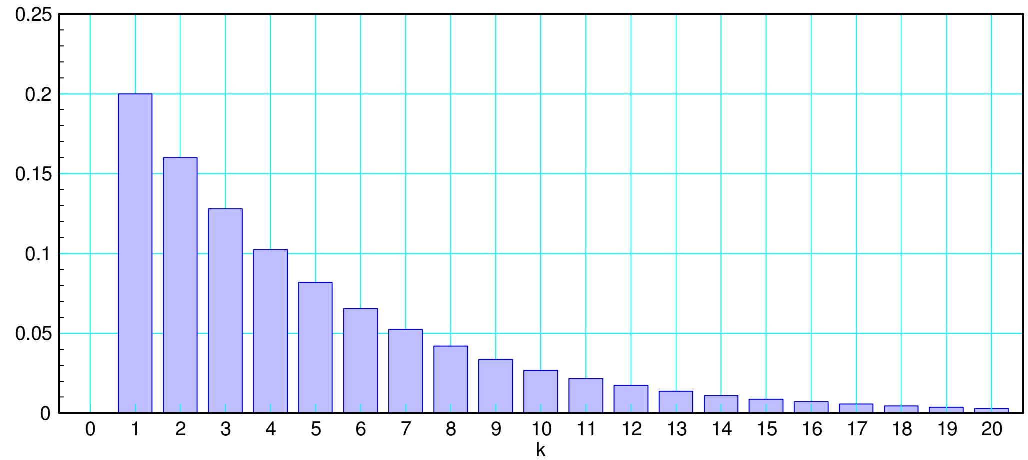
$$m_1 = -\psi'(z=1) = \left(\frac{p}{(1-q)^2}\right) = \left(\frac{1}{p}\right)$$

$$\psi''(z) = \left(\frac{2p}{(z-q)^3}\right)$$

$$m_2 = \psi''(z=1) = \left(\frac{2}{p^2}\right)$$

$$\sigma^2 = m_2 - m_1 - m_1^2$$

$$\sigma^2 = \left(\frac{2}{p^2}\right) - \left(\frac{1}{p}\right) - \left(\frac{1}{p}\right)^2 = \left(\frac{1-p}{p^2}\right) = \left(\frac{q}{p^2}\right)$$



Matlab Example:

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

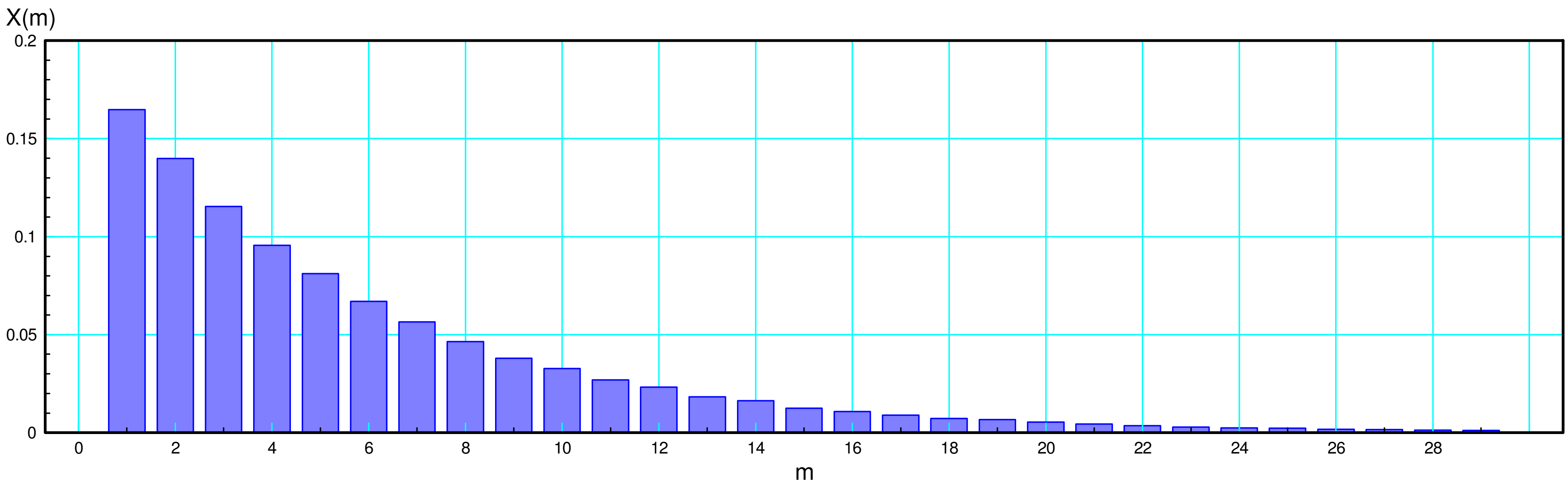
```
N = 1e5;
X = zeros(100,1);
p = 1/6;
for i=1:N
   n = 1;
   while(rand > p)
       n = n + 1;
   end
   X(n) = X(n) + 1;
end
X = X / N;

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

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

Result:

	Sim	Calc
x	6.0179	6.0000
var	30.0712	30.0000

pdf and cdf:

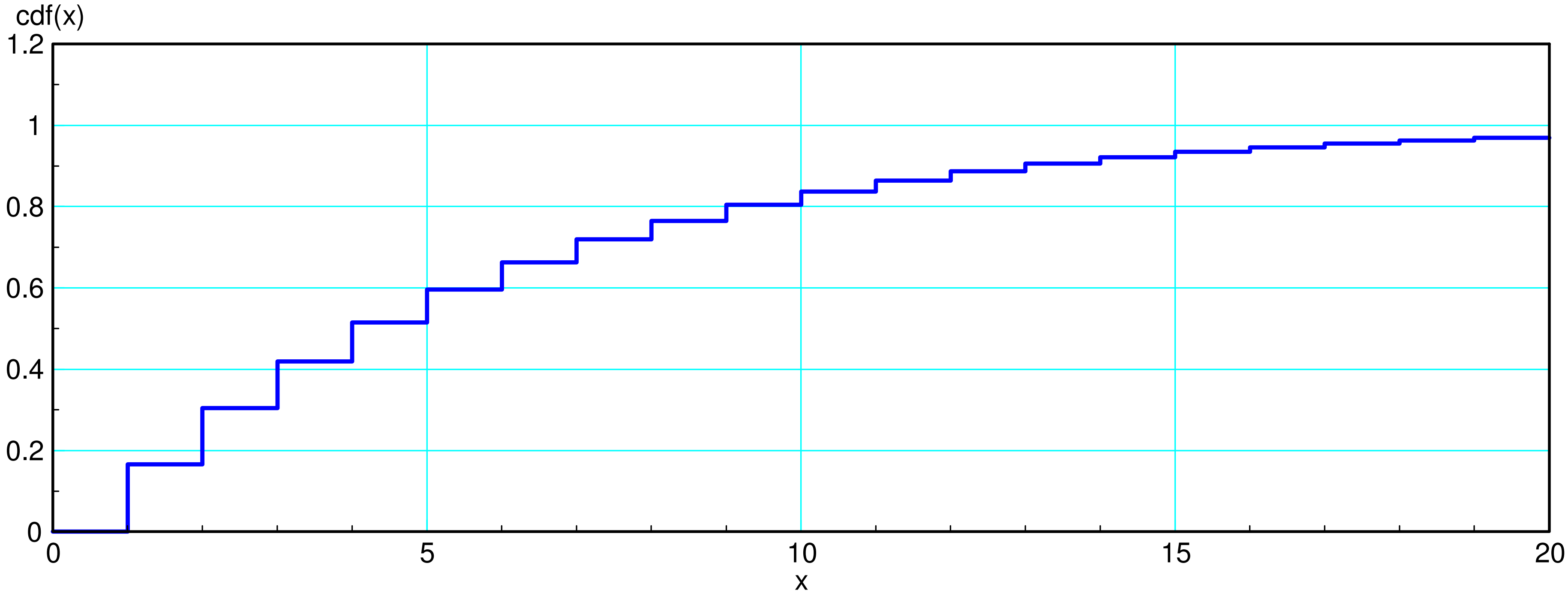
The pdf is the probability of k tosses



Experimental pdf for tossing a die until you roll a 6

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

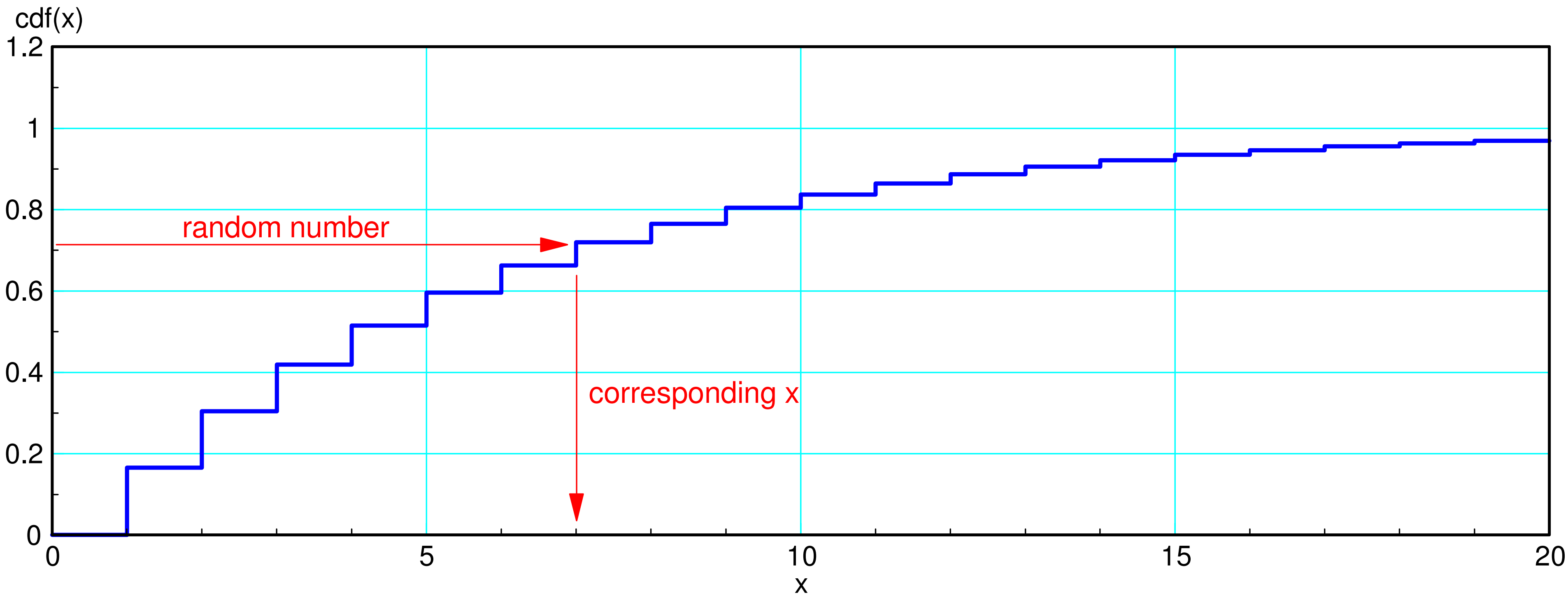
```
cdf = 0*X;
for i=1:length(cdf)
   cdf(i) = sum(pdf(1:i));
   end
```



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



Finding the cdf using z-transforms

- cdf is the integral of the pdf:

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

$$cdf = 1 - q^x$$

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

Expected Return

- The mean of a distribution is important.

Example: Dice game:

- It costs $N to play the game.
- Roll until you get a 1
- Payout is $1 x number of rolls

In this game

$$p = \frac{1}{6}$$

$$\mu = \frac{1}{p} = 6$$

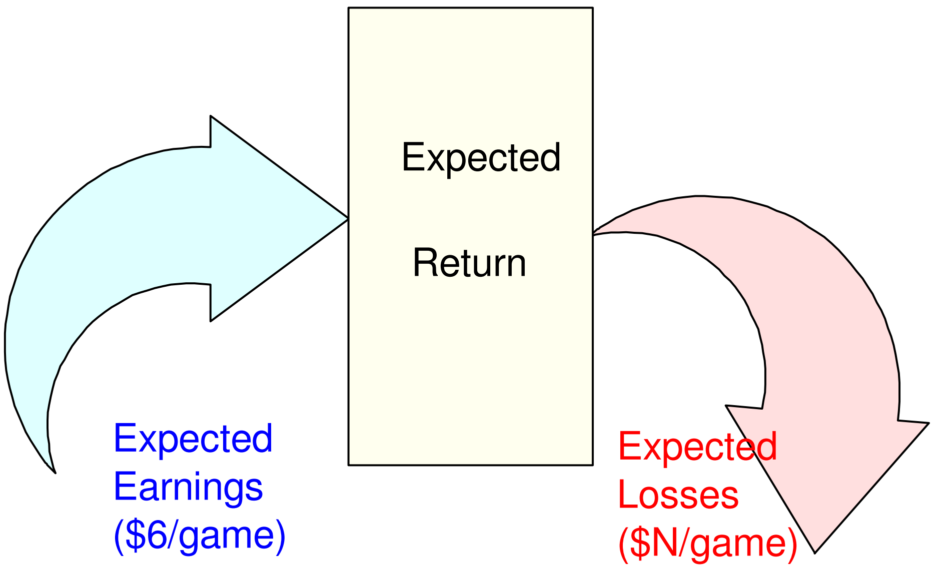
This means

- You expect to get paid $6 on average
- Every time you play the game.

Expected Return

The expected return is

- The expected earnings
  - The mean
  - $6/game
- Minus the expected losses
  - the cost to play
  - $N/game



Should you play a game?

- If the expected return is positive, yes
- If the expected return is negative, no

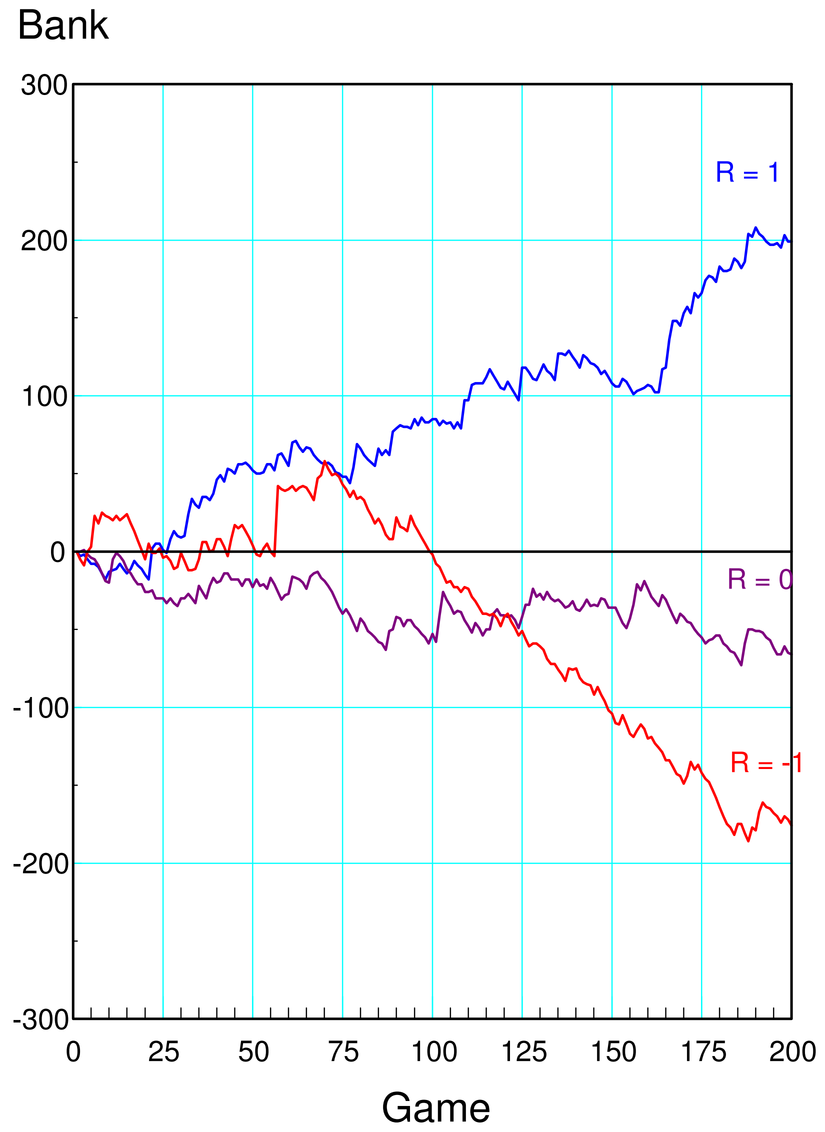
Expected Return Example

Play the previous game 200 times when

- N = 5 (you expect to make money)
- N = 6 (you expect to break even), and
- N = 7 (you expect to lost money)

The Monte Carlo results are

- You tend to make money when the expected return is positive
- You tend to lose money when the expected return is negative



State Lotteries

Most state lotteries return 50 cents for every dollar you bet.

For example, North Dakota 2-by-2

- Each ticket costs $1
- Pick two red numbers from 1..26
- Pick two white numbes from 1..26

The daily payout is:

Match	Prize	Prize Tuesdays
4 Numbers	$22,000	$44,000
3 Numbers	$100	$200
2 Numbers	$3	$6
1 Number	1 ticket	2 tickets

State Lottery: Odds of Winning

The total number winning numbers

$$N = \binom{26}{2}\binom{26}{2} = 105625$$

Number of combinations with 4 winning numbers

$$M = \binom{2}{2}\binom{24}{0} \cdot \binom{2}{2}\binom{24}{0} = 1$$

Number of combinations with 3 winning numbers

$$M = 2 \cdot \binom{2}{2}\binom{24}{0} \cdot \binom{2}{1}\binom{24}{1} = 96$$

Number of combinations with 2 winning numbers

$$M = 2 \cdot \binom{2}{2}\binom{24}{0} \cdot \binom{2}{0}\binom{24}{2} + 1 \cdot \binom{2}{1}\binom{24}{1} \cdot \binom{2}{1}\binom{24}{1} = 285$$

State Lottery: Return on Investment

You expect $0.51 for every $1 bet

- You lose 49 cents for every dollar you bet

This is pretty bad

- You would expect that noone would every play this game
- Yet people do.

Why?

Match	Prize x	Combos M	Return x * M/N
4 Numbers	22,000	1	0.21
3 Numbers	100	96	0.09
2 Numbers	3	2,856	0.08
1 Number	1	13,248	0.13
		Total	0.51

State Lottery: Is Money Linear?

One thought is that money isn't linear:

- Losing $1 on a lottery ticket won't change your life, but
- Winning $22,000 can change your life.

Linear:

$$f(a+b) = f(a) + f(b)$$

If you're poor, $22,000, could change your life

If you're rich, it makes little difference

This suggests playing the lottery only makes sense if you're poor

- Which is what happens
- Lotteries are a very regressive tax.

Gauss' Dilemma:

Along these lines, here's a game that

- No-one will play
  - you (almost) always lose, and
- No cassino will ever offer
  - 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	7	8	...
Probability (p)	1/2	1/4	1/8	1/16	1/32	1/64	1/128	1/256	...
Pot (x)	1	2	4	8	16	32	64	128	...
Winnings (p*x)	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	...

The expected winnings are

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

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

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

$$E = \infty$$

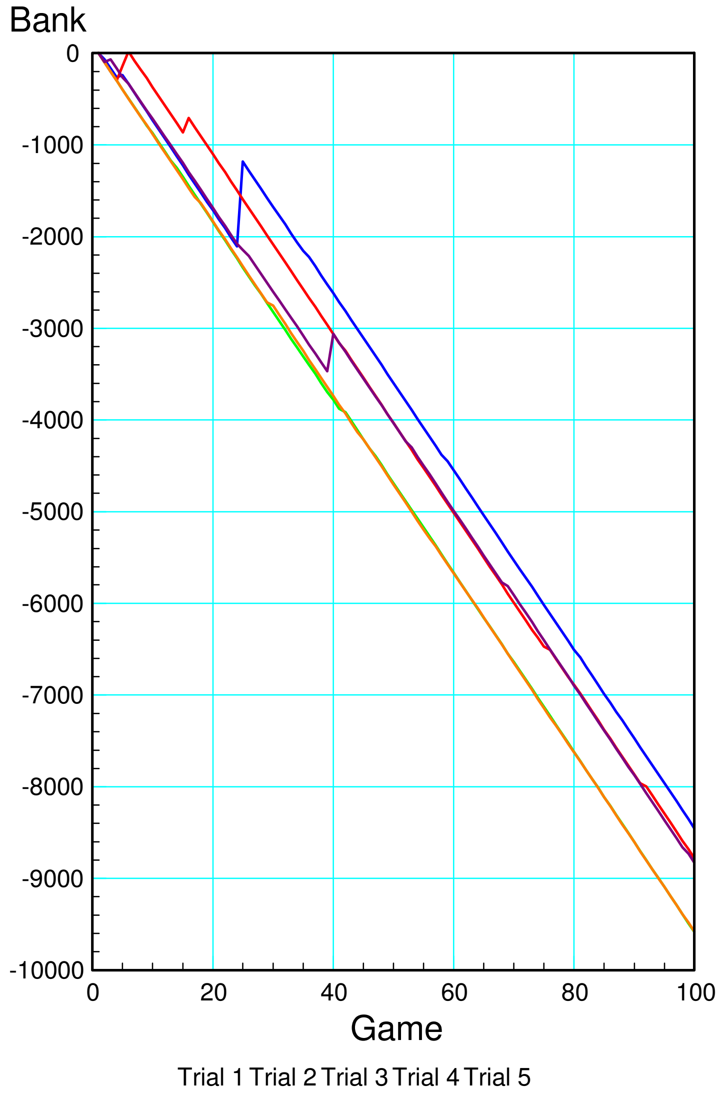
With infinite expected winnings, this sounds like a good game to play.

Monte-Carlo Simulation

```
N = 100;
Winnings = 0;
p = 0.5;

for i=1:N
   Pot = 1;
    while(rand > p)
       Pot = Pot * 2;
   end
    Winnings = Winnings + Pot - 100;
end
```

Every time you play this game, you end up losing

- 5 trials
- Down \$8300 to \$9500 after 100 games



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
```

This is a game where everyone loses

- The players lose
  - Pretty much every single time
- The casino loses
  - The expected winnings are infinite

Summary

Geometric distributions describe events where you continue playing until an event happens

- Toss a die until you roll a one
- Keep plugging away until your boss notices you
- Keep going to parties until you get Covid

Moment Generating Functions are useful for finding

- The mean (1st moment)
- The variance

Cumulative Density Functions (cdf's) are useful for coverting a probability to a number

Expected Returns are useful when deciding to play a game or not

- Pretty much all casino and lottery games have a negative return