Variance, Bernoulli, Binomial

Noah Arthurs
CS109, Stanford University

Review

Conditional Independence

- Two events E and F are called **conditionally independent given G**, if

$$P(EF|G) = P(E|G)P(F|G)$$

- Or, equivalently if:

$$P(E|FG) = P(E|G)$$

Conditional Independence

What does conditional independence tell you about independence and vice versa?

Nothing

Random Variable

- A **Random Variable** is a variable will have a value. But there is uncertainty as to what value.
- Example:
  - 3 fair coins are flipped.
  - Y = number of “heads” on 3 coins
  - **Y is a random variable**
  - P(Y = 0) = 1/8 (T, T, T)
  - P(Y = 1) = 3/8 (H, T, T), (T, H, T), (T, T, H)
  - P(Y = 2) = 3/8 (H, H, T), (H, T, H), (T, H, H)
  - P(Y = 3) = 1/8 (H, H, H)
  - P(Y ≥ 4) = 0

Probability Mass Function

Let X be a random variable that represents the result of a **single dice roll**. X can take on the values {1, 2, 3, 4, 5, 6}

$$P(X = x)$$

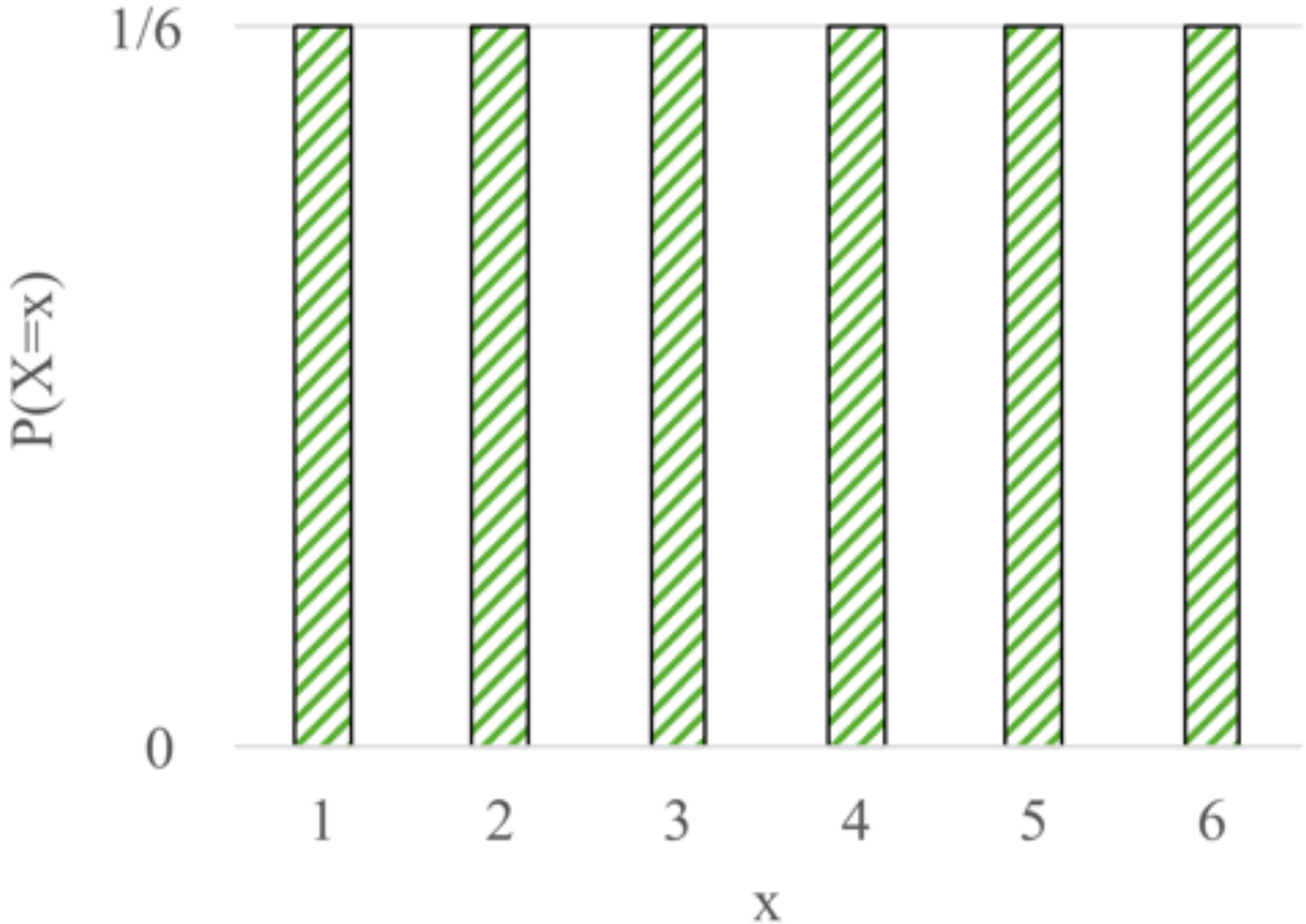
$$p(x)$$

This is shorthand notation for the PMF

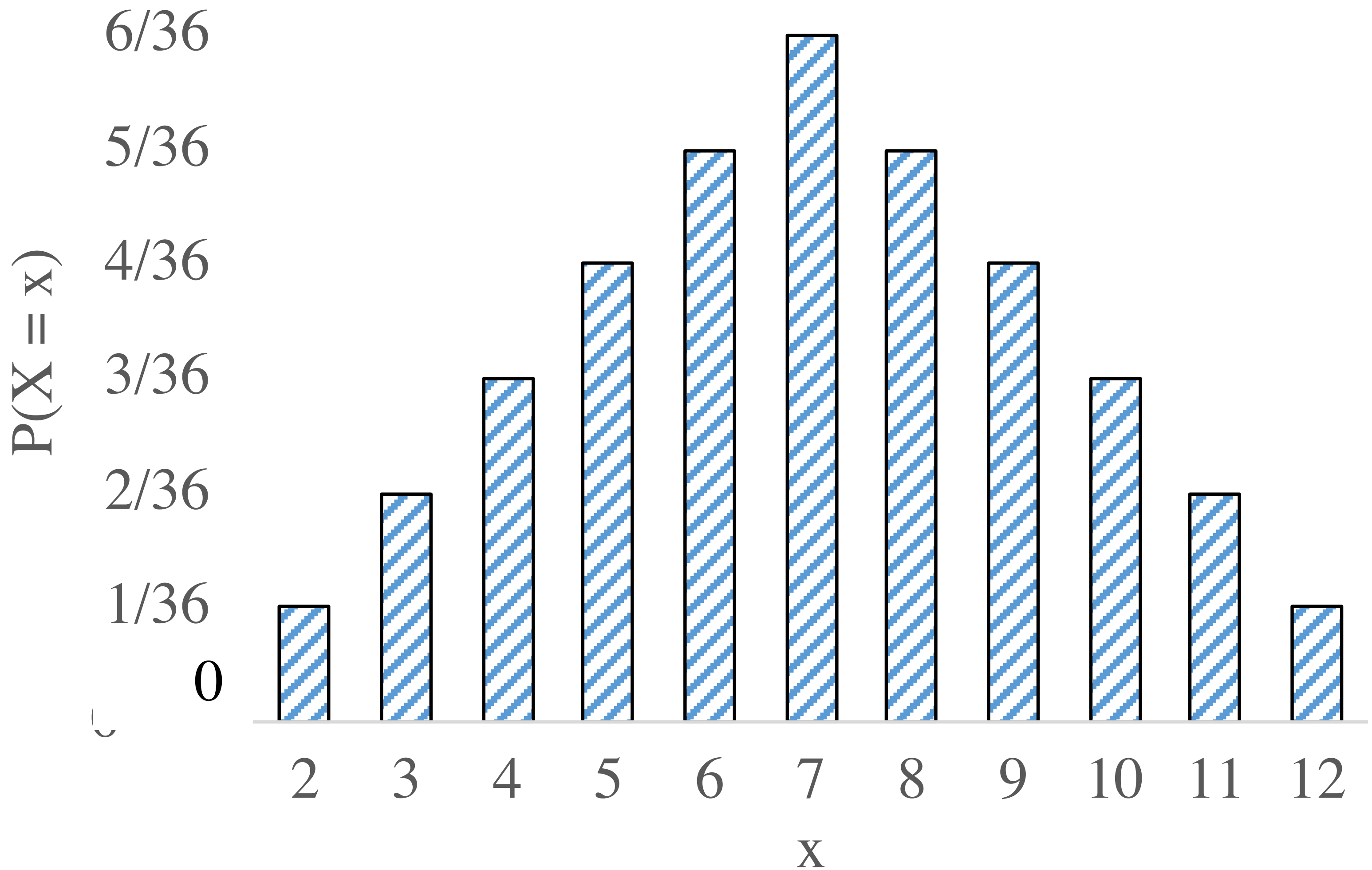
$$p_X(x)$$

This is also shorthand notation for the PMF

PMF For a Single 6 Sided Dice

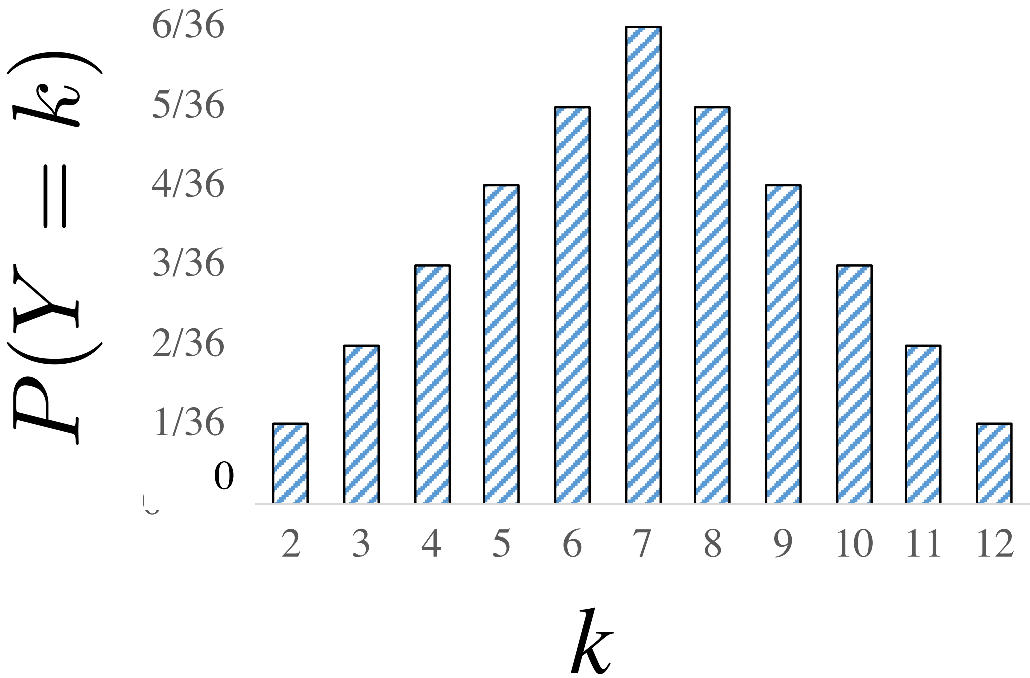


PMF for the sum of two dice



Sanity Check

$$\sum_k P(Y = k) = 1$$



Expected Value

- The Expected Values for a discrete random variable X is defined as:

$$E[X] = \sum_{x:p(x)>0} x \cdot p(x)$$

- Note: sum over all values of x that have $p(x) > 0$.

- Expected value also called: ***Mean***, *Expectation*, ***Weighted Average***, ***Center of Mass***, *1st Moment*

Properties of Expectation

- **Linearity**:

$$E[aX + b] = aE[X] + b$$

  - Consider X = 6-sided die roll, Y = 2X – 1.
  - E[X] = 3.5   E[Y] = 6

- **Expectation of a sum** is the sum of expectations

$$E[X + Y] = E[X] + E[Y]$$

- **Unconscious statistician**:

$$E[g(x)] = \sum_x g(x)p(x)$$

Deriving Linearity

Let's say we want to know $E[aX + b]$…

First let $g(x) = ax + b$ meaning that we want $E[g(X)]$

By Unconscious Statistian:

$$E[g(X)] = \sum_x g(x)P(X = x)$$

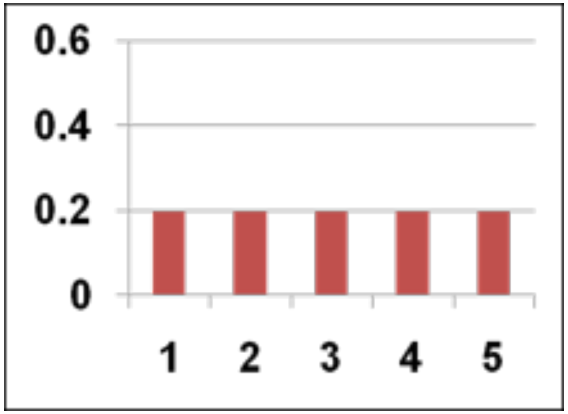
$$= \sum_x (ax + b)P(X = x)$$

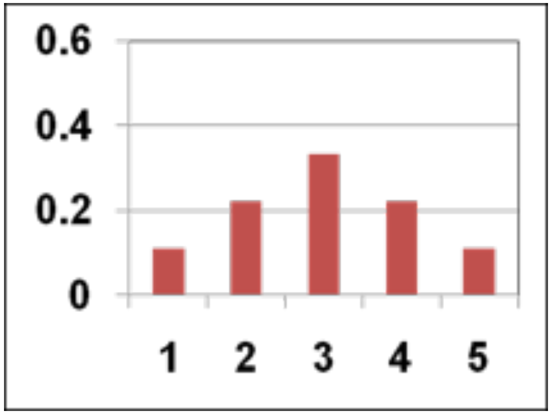
$$= a\sum_x xP(X = x) + b\sum_x P(X = x)$$

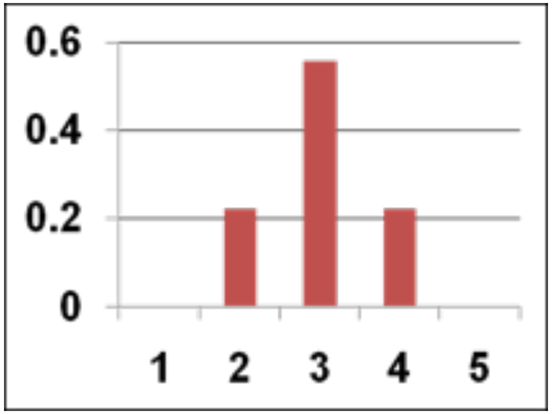
$$= aE[X] + b$$

Variance

- Consider the following 3 distributions (PMFs)







- All have the same expected value, E[X] = 3
- But "spread" in distributions is different
- Variance = a formal quantification of "spread"

Variance

- If X is a random variable with mean μ then the **variance** of X, denoted Var(X), is:

$$\mathrm{Var}(X) = E[(X - \mu)^2]$$

- Note: Var(X) ≥ 0

- Also known as the 2nd **Central** Moment, or square of the Standard Deviation

Computing Variance

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= \sum_x (x-\mu)^2 p(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\ &= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\ &= \boxed{E[X^2]} - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= \boxed{E[X^2] - (E[X])^2}\end{aligned}$$

Note: $\mu = E[X]$

Ladies and gentlemen, please welcome the 2nd moment!

Variance of a 6 sided dice

- Let X = value on roll of 6 sided die
- Recall that E[X] = 7/2
- Compute $E[X^2]$

$$E[X^2] = (1^2)\frac{1}{6} + (2^2)\frac{1}{6} + (3^2)\frac{1}{6} + (4^2)\frac{1}{6} + (5^2)\frac{1}{6} + (6^2)\frac{1}{6} = \frac{91}{6}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

End Review

Properties of Variance

- $Var(aX + b) = a^2 Var(X)$
  - Proof:

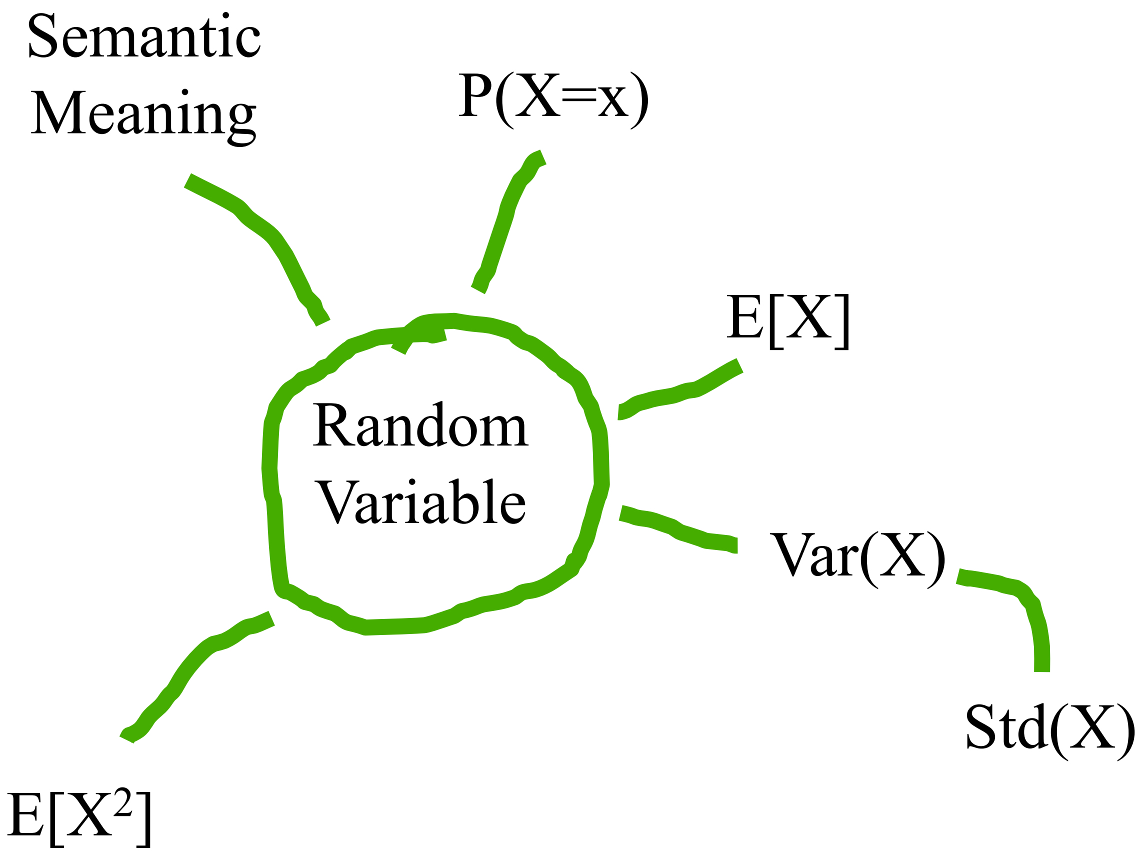
$$\begin{aligned}
Var(aX + b) &= E[(aX + b)^2] - (E[aX + b])^2 \\
&= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
&= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2) \\
&= a^2E[X^2] - a^2(E[X])^2 = a^2(E[X^2] - (E[X])^2) \\
&= a^2 Var(X)
\end{aligned}$$

- Standard Deviation of X, denoted SD(X), is:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

  - Var(X) is in units of X^2
  - SD(X) is in same units as X

Fundamental Properties



Classics

Jacob Bernoulli

- Jacob Bernoulli (1654-1705), also known as "James", was a Swiss mathematician

- One of many mathematicians in Bernoulli family
- The Bernoulli Random Variable is named for him
- He is my *academic* great^{13}-grandfather

Bernoulli Random Variable

- Experiment results in "Success" or "Failure"
  - X is random **indicator** variable (1 = success, 0 = failure)
  - $P(X = 1) = p(1) = p$ $\quad\quad$ $P(X = 0) = p(0) = 1 – p$
  - X is a **Bernoulli** Random Variable: $X \sim Ber(p)$
  - $E[X] = p$
  - $Var(X) = p(1 – p)$
- Examples
  - coin flip
  - random binary digit
  - whether a disk drive crashed
  - whether someone likes a netflix movie

Does a Program Crash?

Run a program, crashes with prob. p, works with prob. $(1 – p)$

X: 1 if program crashes

$P(X = 1) = p$
$P(X = 0) = 1 - p$

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

Does a User Click an Ad?

Serve an ad, clicked with prob. p, ignored with prob. $(1 - p)$

C: 1 if ad is clicked

$P(C = 1) = p$

$P(C = 0) = 1 - p$

<u>**C** ~ Ber(p)</u>

Binomial Random Variable

- Consider *n* independent trials of Ber(p) rand. var.
  - X is number of successes in *n* trials
  - X is a **Binomial** Random Variable: X ~ Bin(n, p)

$$P(X=i)=p(i)=\binom{n}{i}p^i(1-p)^{n-i} \qquad i=0,1,\ldots,n$$

  - By Binomial Theorem, we know that $\sum_{i=0}^{\infty} P(X=i)=1$
- Examples
  - # of heads in *n* coin flips
  - # of 1's in randomly generated length *n* bit string
  - # of disk drives crashed in 1000 computer cluster
    - Assuming disks crash independently

Our random variable

Num trials

Probability of success on each trial

$$X \sim \mathrm{Bin}(n, p)$$

Is distributed as a

Binomial

With these parameters

If X is a binomial with parameters n and p

Probability Mass Function
for a Binomial

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

Probability that our
variable takes on the
value k

Bernoulli vs Binomial

Bernoulli is an indicator RV

Binomial is the sum of *n* Bernoullis

Three Coin Flips

- Three fair ("heads" with p = 0.5) coins are flipped
  - X is number of heads
  - X ~ Bin(n = 3, p = 0.5)

$$P(X=0)=\binom{3}{0}p^0(1-p)^3=\frac{1}{8}$$

$$P(X=1)=\binom{3}{1}p^1(1-p)^2=\frac{3}{8}$$

$$P(X=2)=\binom{3}{2}p^2(1-p)^1=\frac{3}{8}$$

$$P(X=3)=\binom{3}{3}p^3(1-p)^0=\frac{1}{8}$$

Properties of Bin(n, p)

Consider: X ~ Bin(n, p)

- $P(X = i) = p(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0,1,\ldots,n$

- $\mathrm{E}[X] = np$

- $\mathrm{Var}(X) = np(1-p)$

- Note: $\mathrm{Ber}(p) = \mathrm{Bin}(1, p)$

I Really Want the Proof of Var :)

$E(X^2) = \sum_{k \geq 0}^{n} k^2 \binom{n}{k} p^k q^{n-k}$	Definition of Binomial Distribution: $p + q = 1$
$= \sum_{k=0}^{n} kn \binom{n-1}{k-1} p^k q^{n-k}$	Factors of Binomial Coefficient: $k\binom{n}{k} = n\binom{n-1}{k-1}$
$= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}$	Change of limit: term is zero when $k - 1 = 0$
$= np \sum_{j=0}^{m} (j+1) \binom{m}{j} p^j q^{m-j}$	putting $j = k - 1, m = n - 1$
$= np \left(\sum_{j=0}^{m} j \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)$	splitting sum up into two
$= np \left(\sum_{j=0}^{m} m \binom{m-1}{j-1} p^j q^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)$	Factors of Binomial Coefficient: $j\binom{m}{j} = m\binom{m-1}{j-1}$
$= np \left((n-1) p \sum_{j=1}^{m} \binom{m-1}{j-1} p^{j-1} q^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)$	Change of limit: term is zero when $j - 1 = 0$
$= np \left((n-1) p (p+q)^{m-1} + (p+q)^m \right)$	Binomial Theorem
$= np \left((n-1) p + 1 \right)$	as $p + q = 1$
$= n^2 p^2 + np(1-p)$	by algebra

How Many Program Crashes?

n runs of program, each crashes with prob. *p*, works with prob. (1 – *p*)

H: number of crashes

H ~ Bin(*n*, *p*)

P(H = k**)** = $\binom{n}{k}(p)^k(1-p)^{n-k}$

P(H $\geq$ z**)** = $\sum_{k=z}^{n}\binom{n}{k}(p)^k(1-p)^{n-k}$

How Many Ads Clicked?

1000 ads served, each clicked with p = 0.01, otherwise ignored.

H: number of clicks

H ~ Bin(n = 1000, p = 0.01)

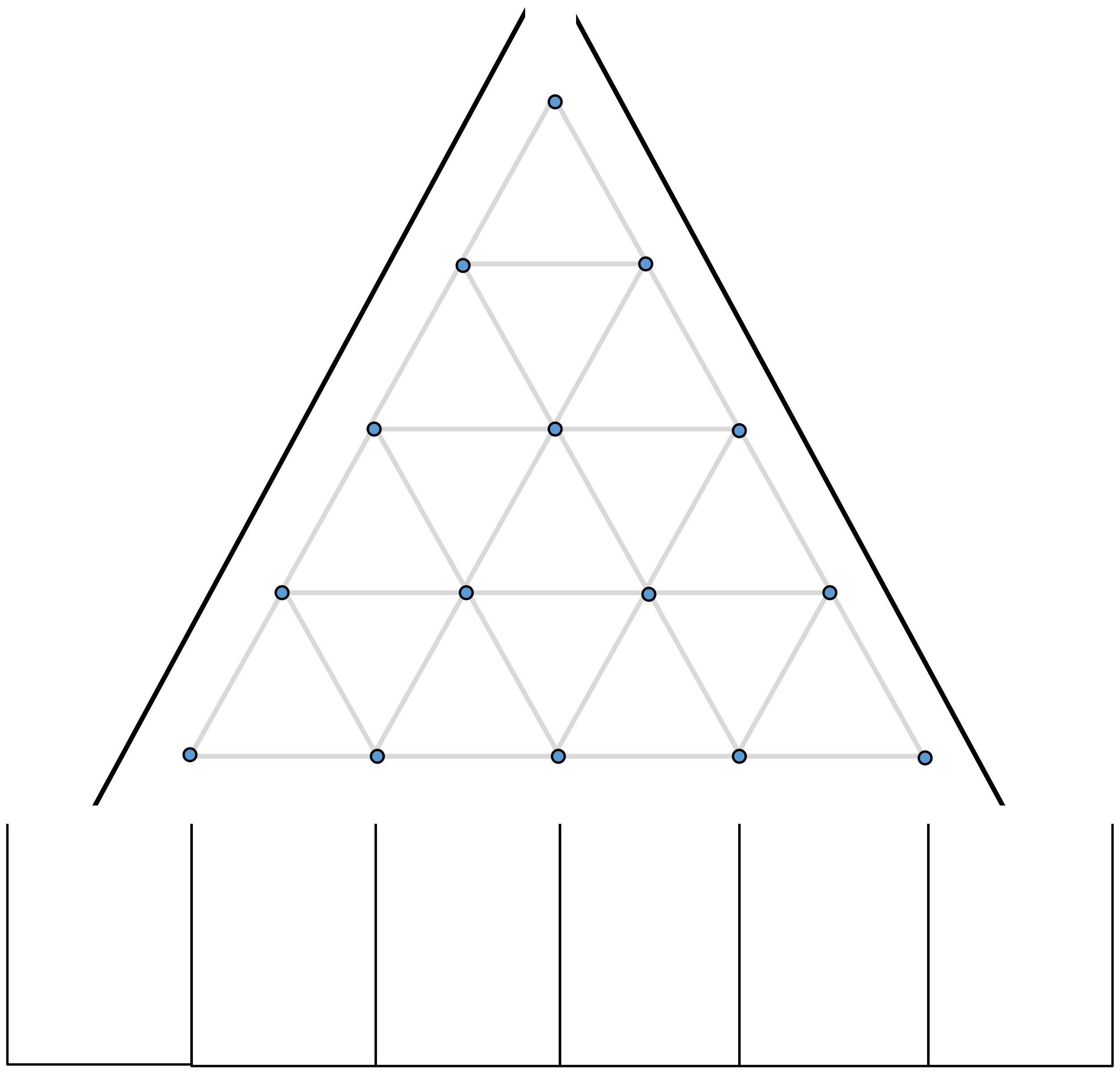
P(H = k**)** = $\binom{1000}{k}(0.01)^k(0.99)^{1000-k}$

Variance of number of ads clicked?

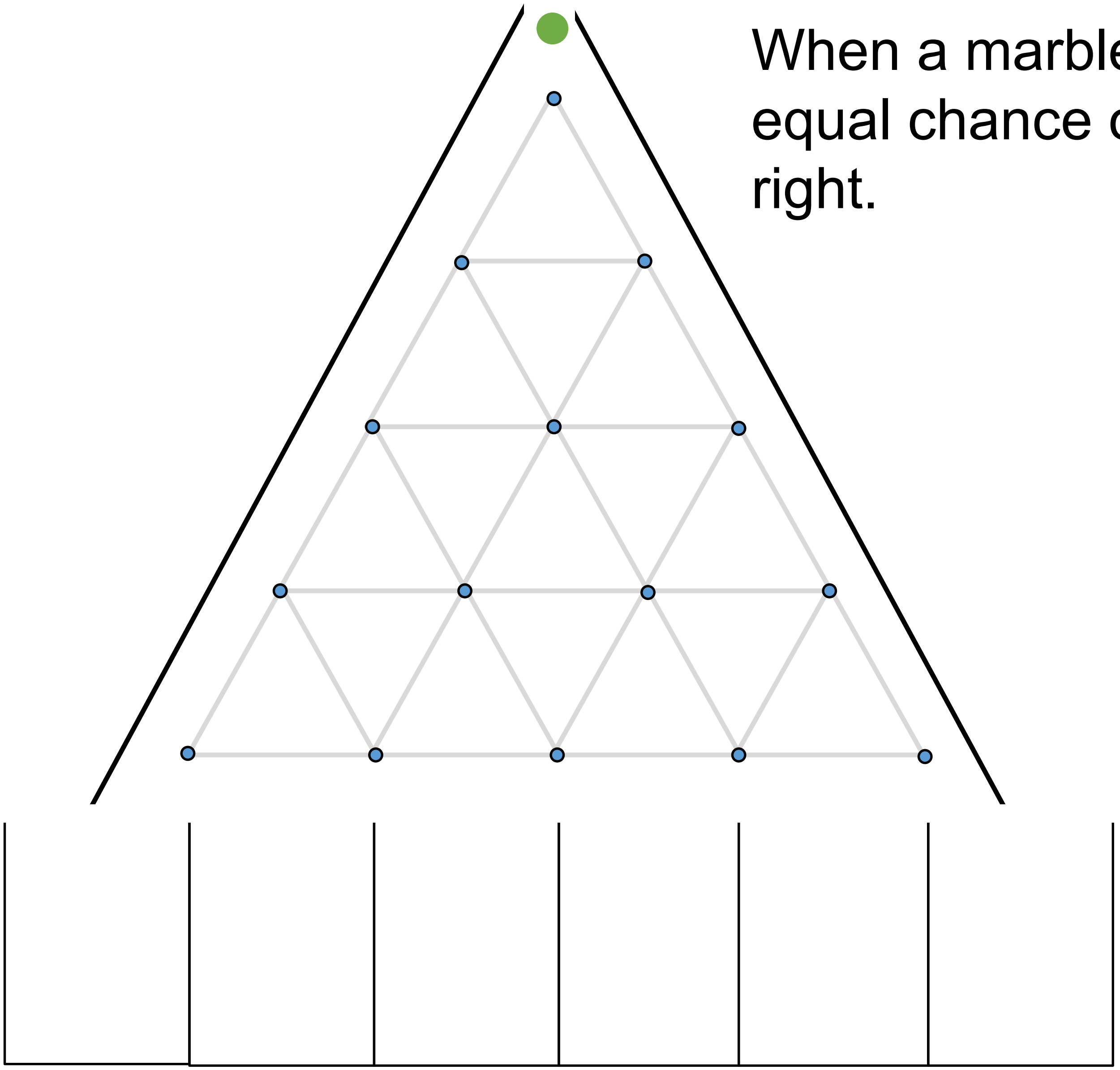
E[**H**] = np = 10

Var(**H**) = $np(1-p)$ = 9.9 Std(**H**) = 3.15

Galton Board

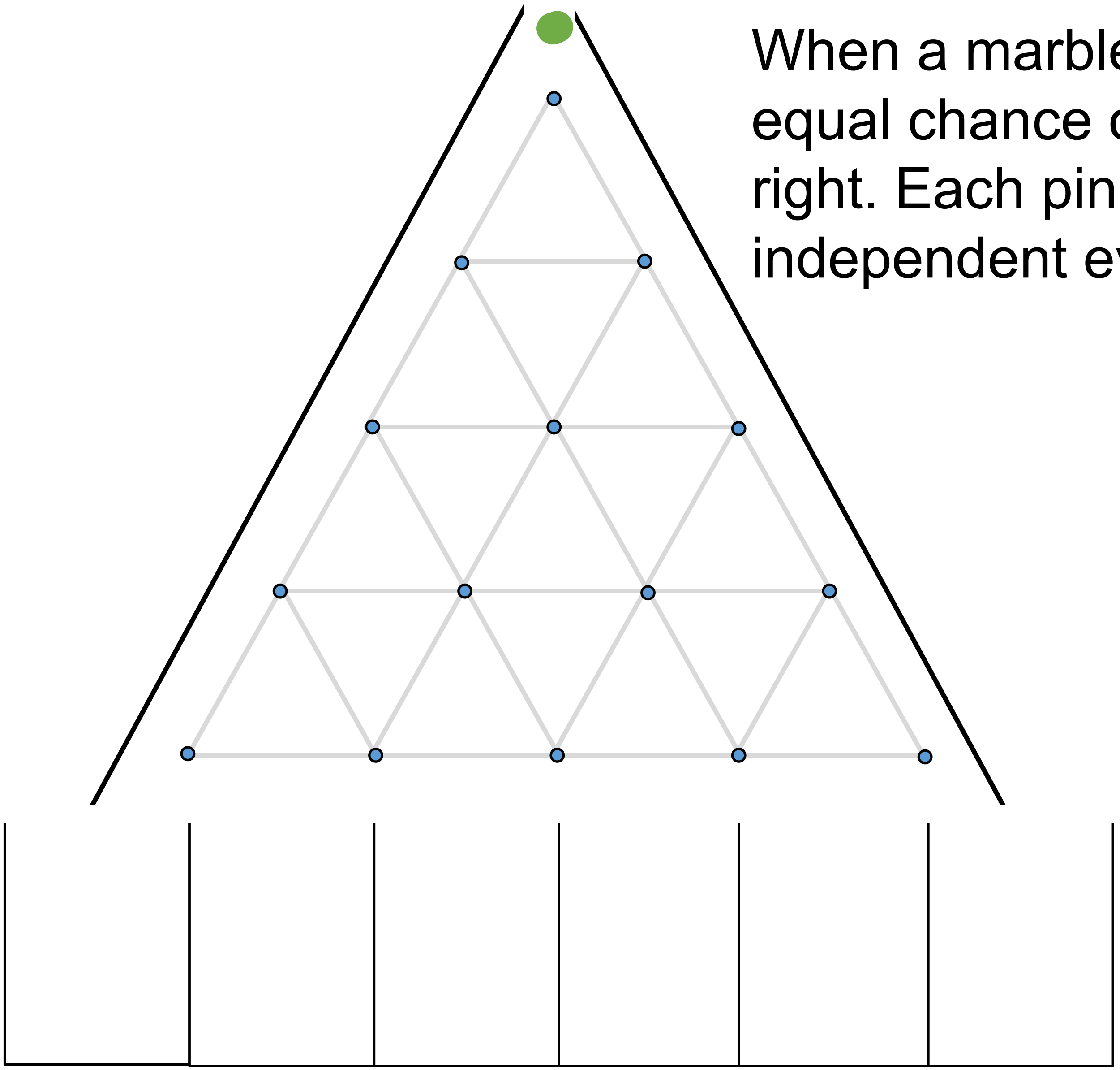


Galton Board



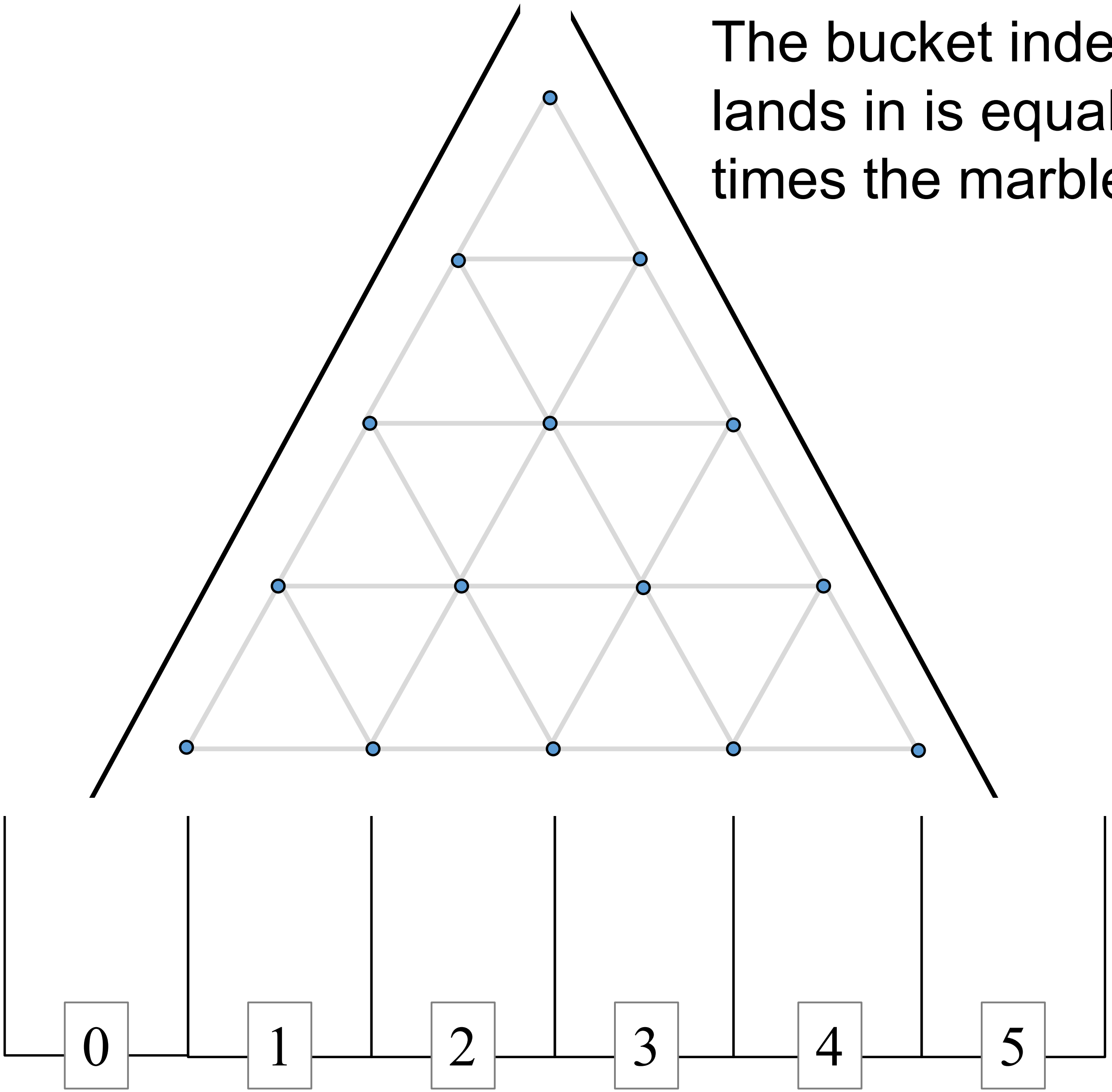
When a marble hits a pin, it has equal chance of going left or right.

Galton Board



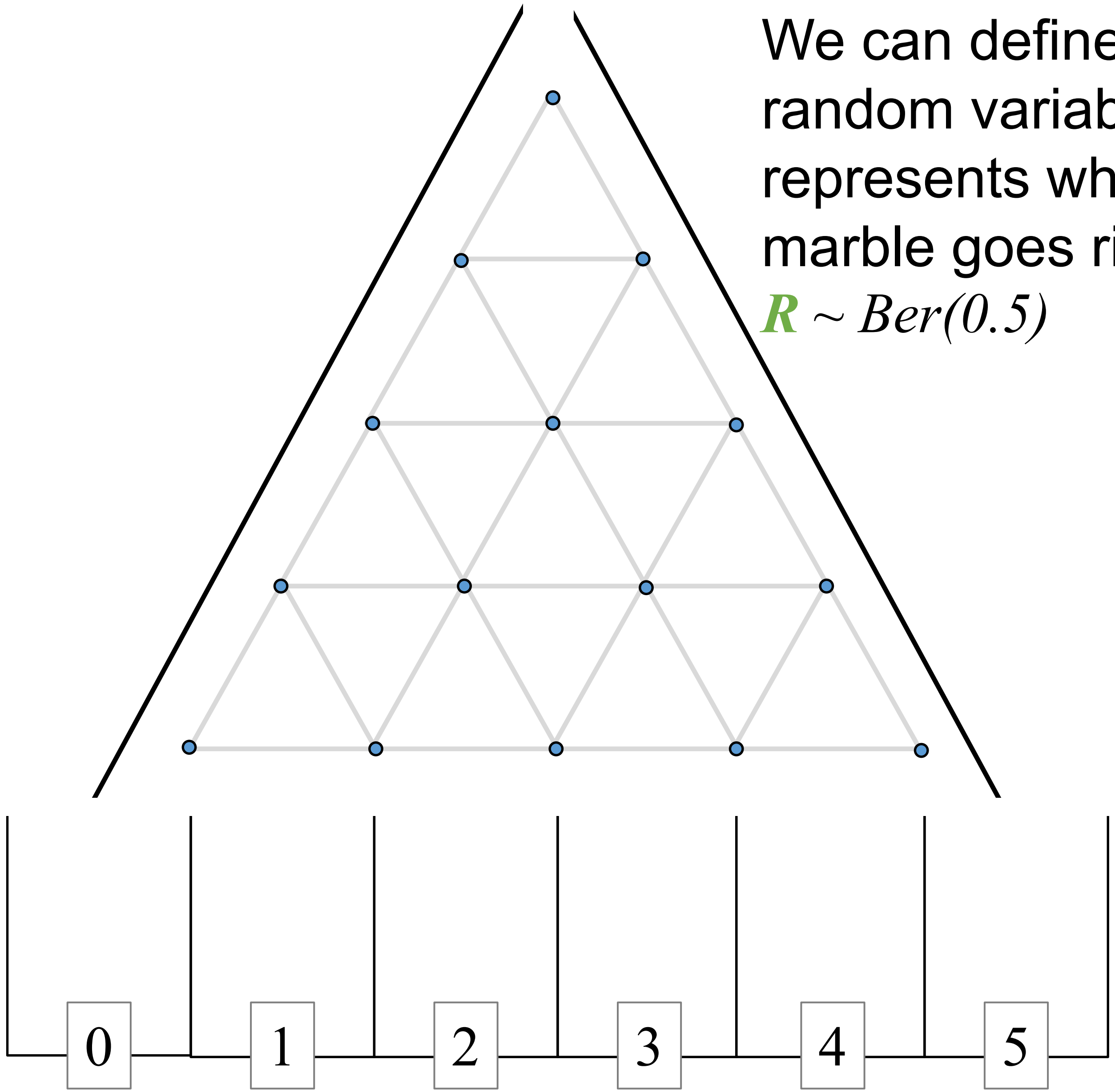
When a marble hits a pin, it has equal chance of going left or right. Each pin represents an independent event.

Galton Board



The bucket index that a marble lands in is equal to the number of times the marble went right

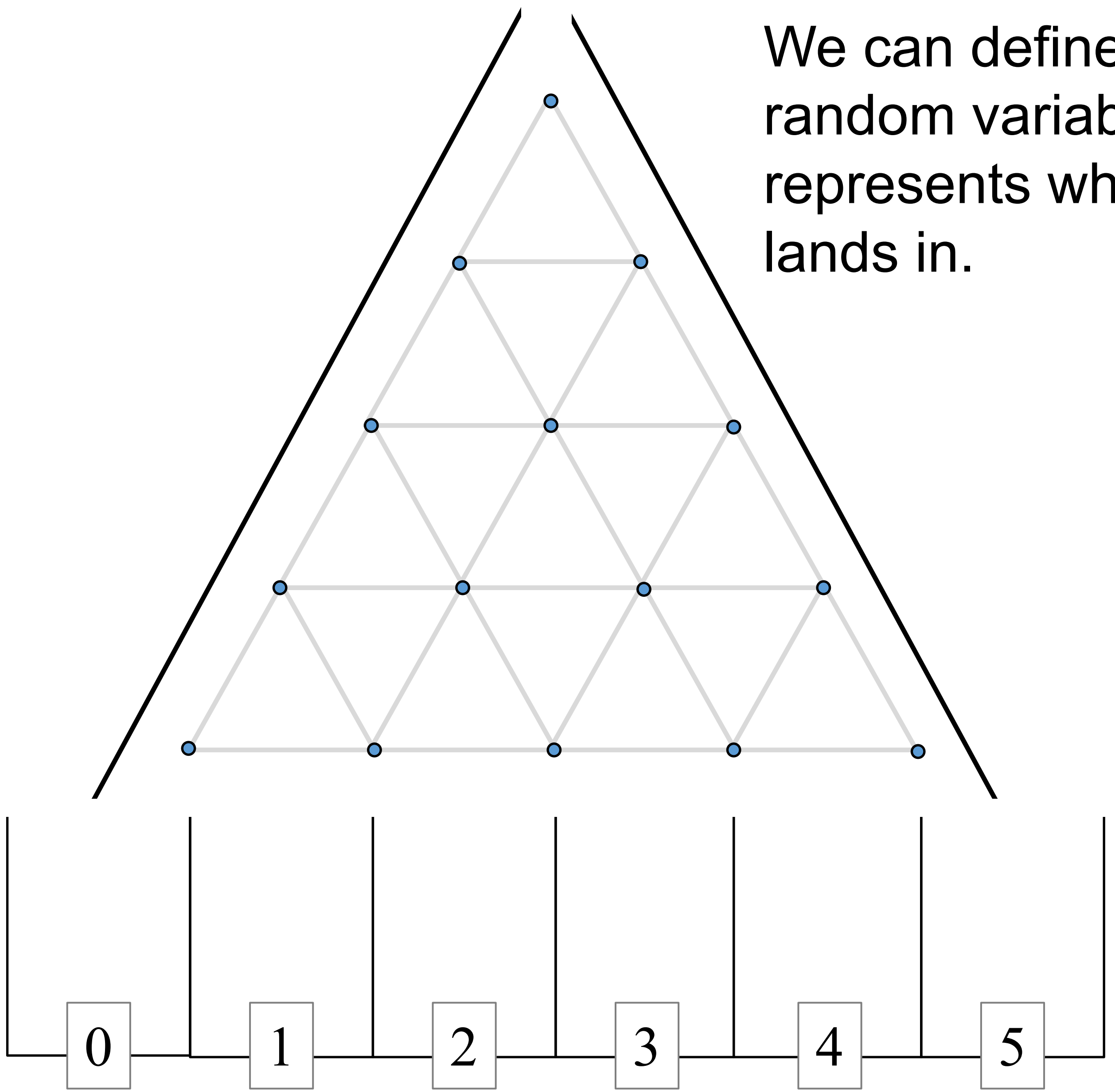
Galton Board



We can define an indicator random variable (***R***) which represents whether a particular marble goes right as a Bernoulli

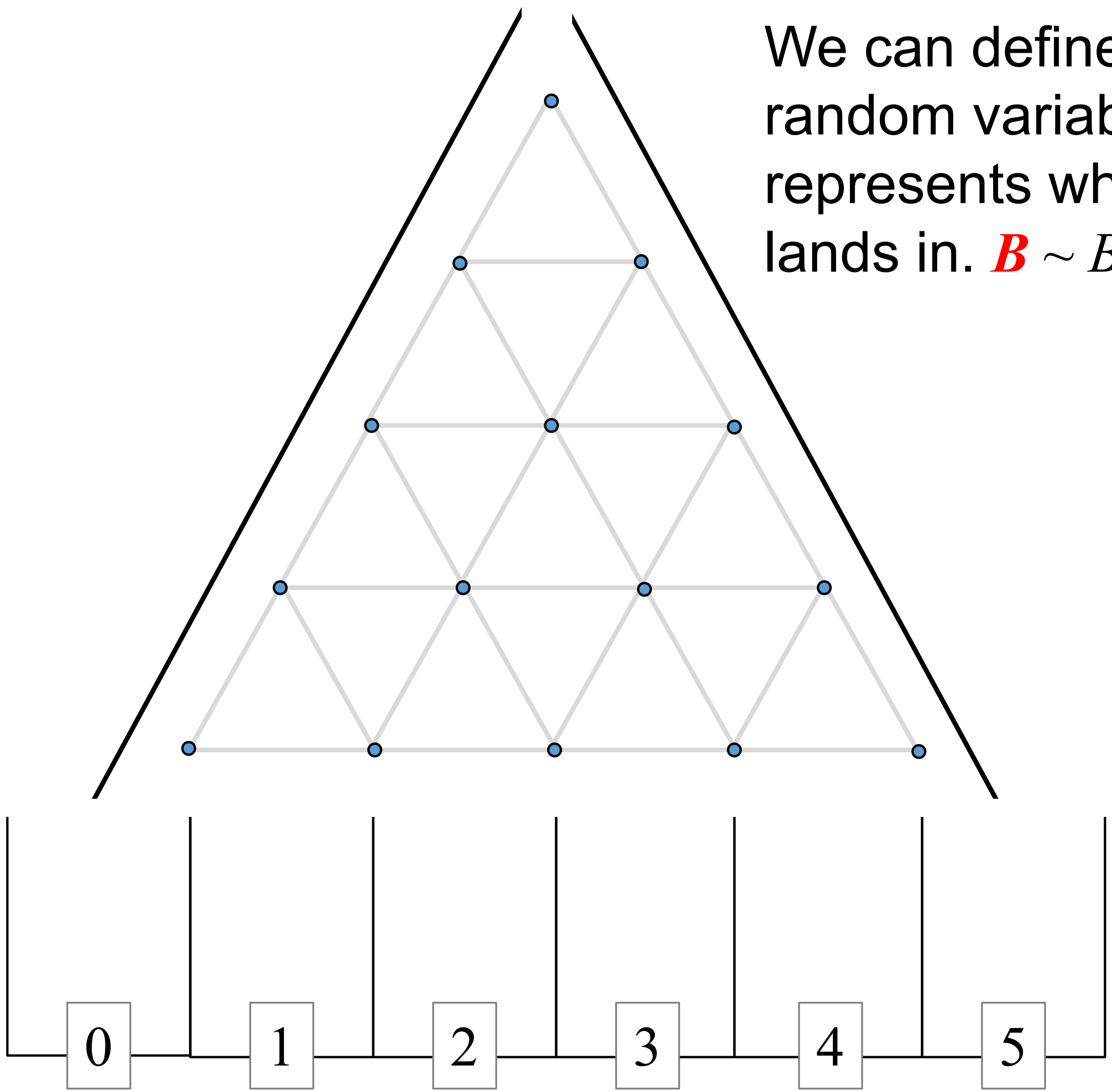
$R \sim Ber(0.5)$

Galton Board



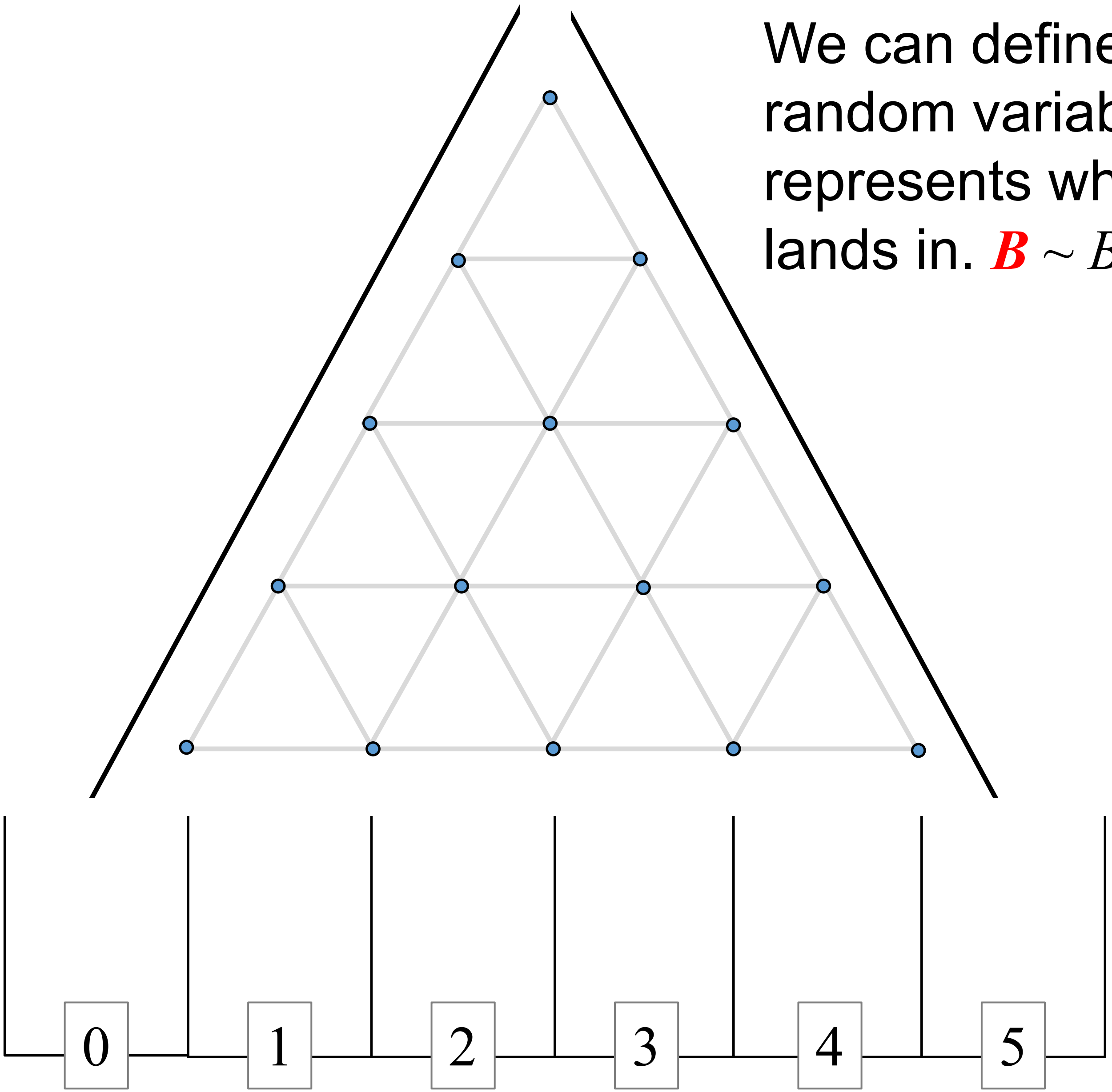
We can define an indicator random variable (***B***) which represents what bucket a marble lands in.

Galton Board



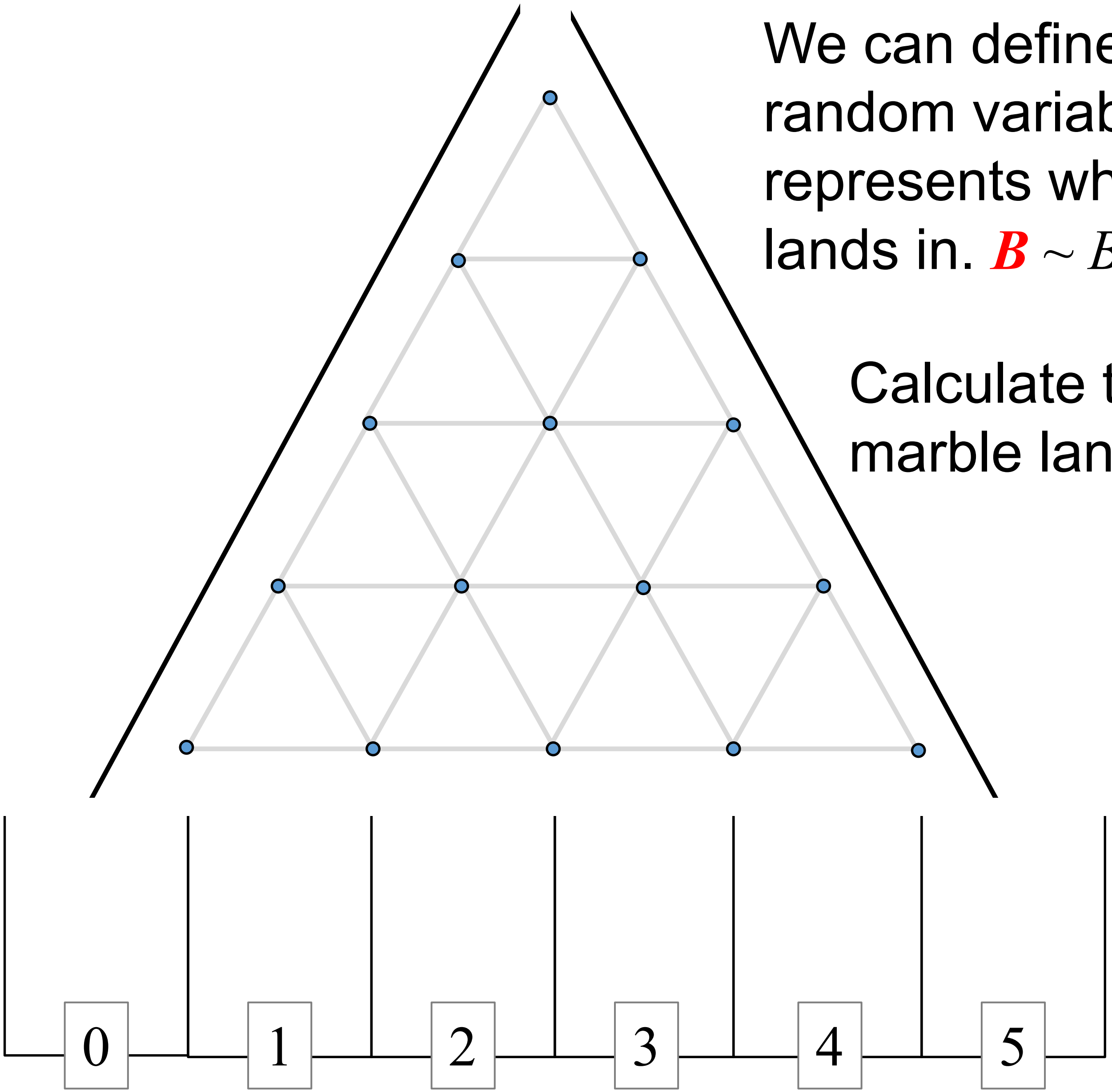
We can define an indicator random variable (***B***) which represents what bucket a marble lands in. $\boldsymbol{B} \sim Bin(\text{levels}, 0.5)$

Galton Board



We can define an indicator random variable (***B***) which represents what bucket a marble lands in. $B \sim Bin(5, 0.5)$

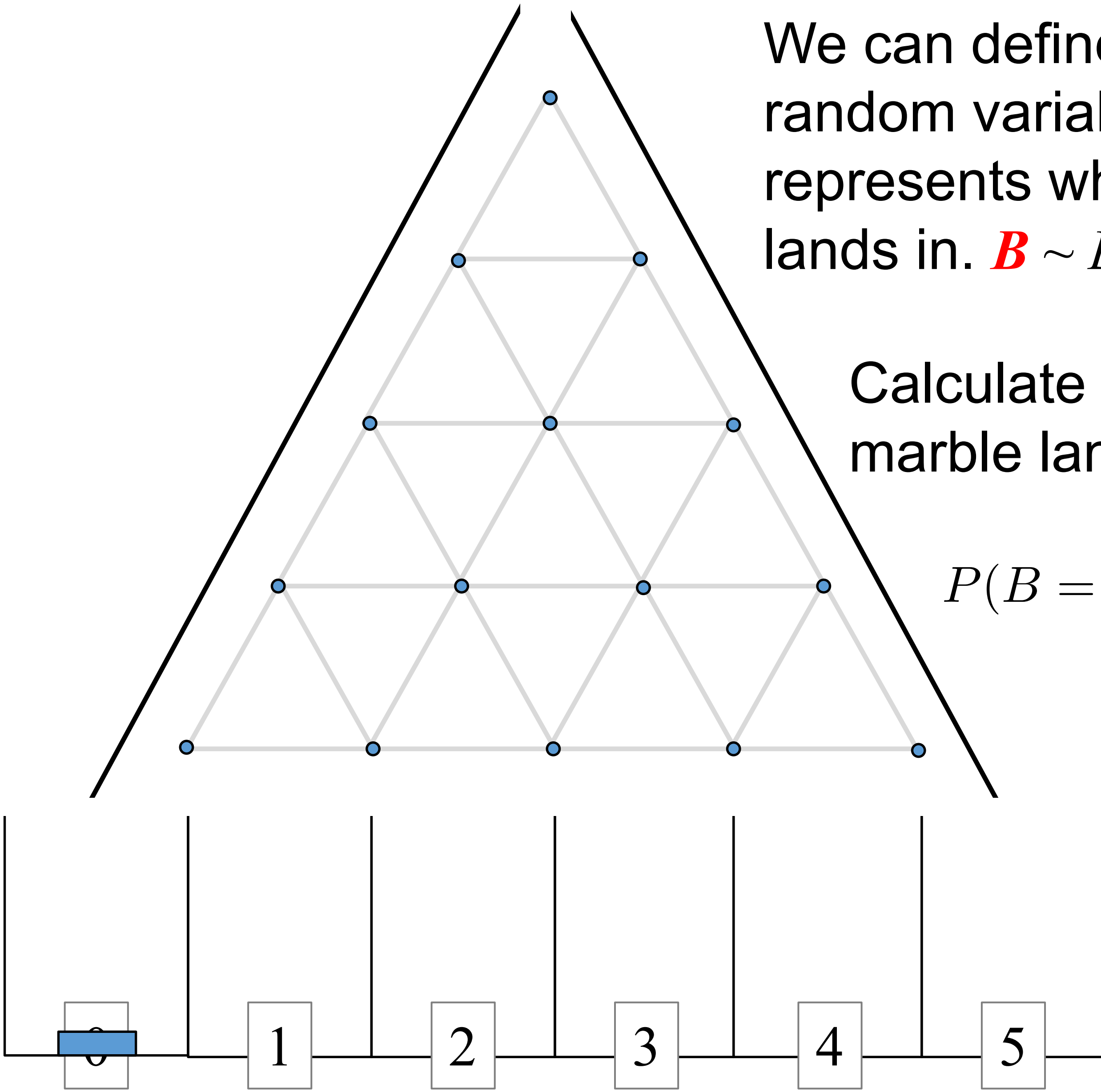
Galton Board



We can define an indicator random variable (***B***) which represents what bucket a marble lands in. $B \sim Bin(5, 0.5)$

Calculate the probability of a marble landing in a bucket.

Galton Board

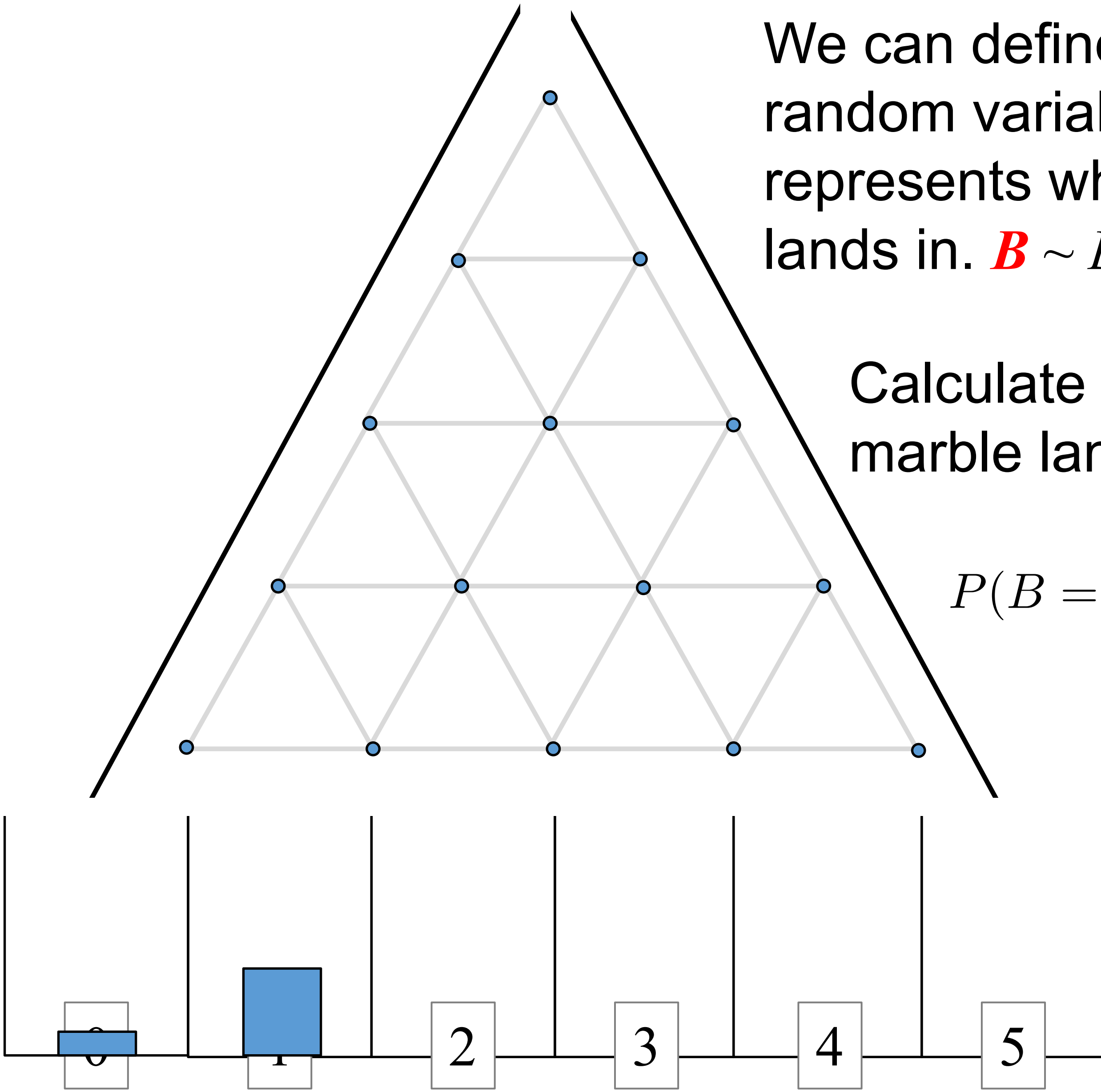


We can define an indicator random variable (***B***) which represents what bucket a marble lands in. ***B*** $\sim Bin(5, 0.5)$

Calculate the probability of a marble landing in a bucket.

$$P(B = 0) = \binom{5}{0} \frac{1}{2}^{5} \approx 0.03$$

Galton Board

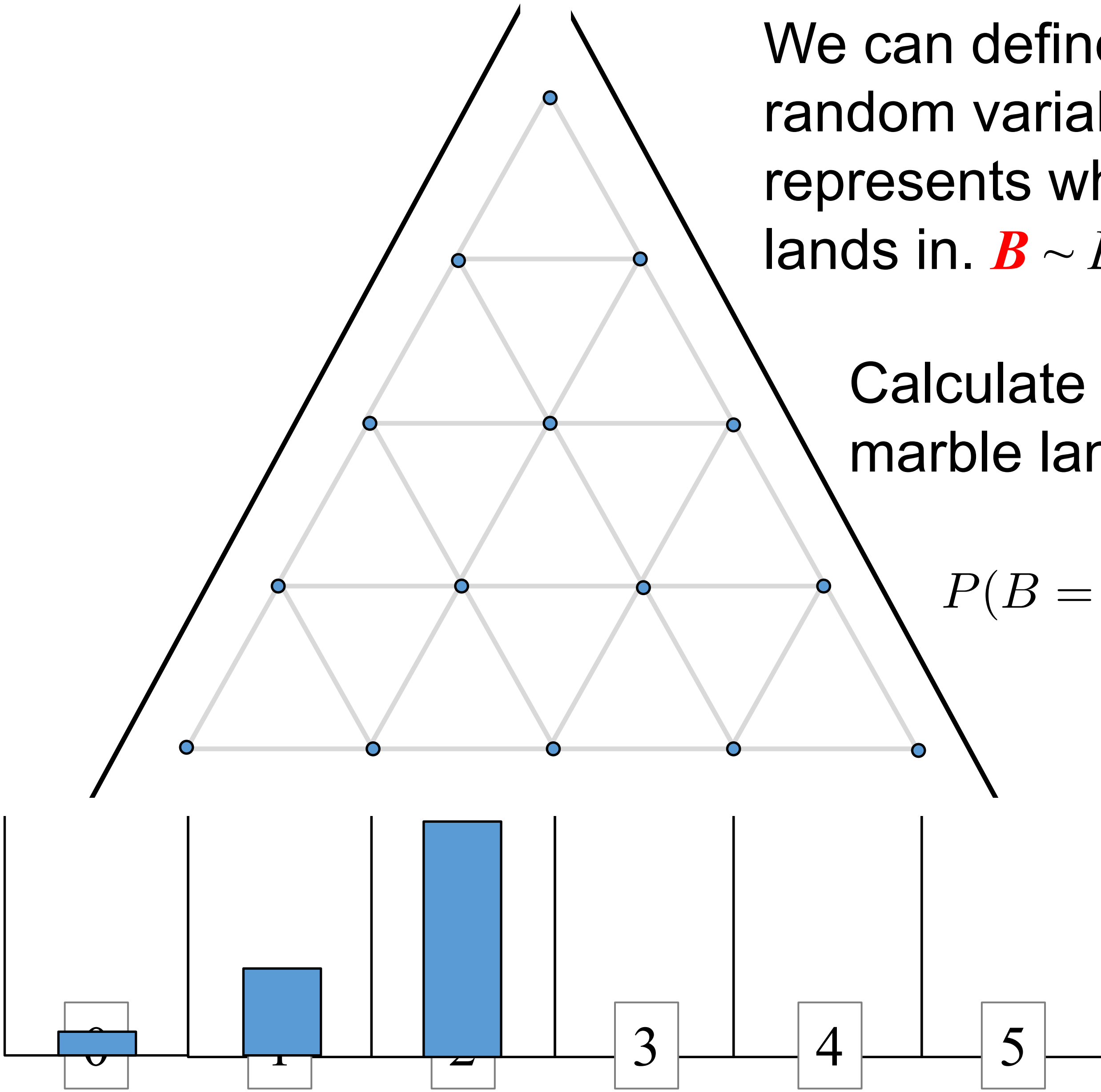


We can define an indicator random variable (***B***) which represents what bucket a marble lands in. ***B*** ~ *Bin(5, 0.5)*

Calculate the probability of a marble landing in a bucket.

$$P(B = 1) = \binom{5}{1} \frac{1}{2}^{5} \approx 0.16$$

Galton Board

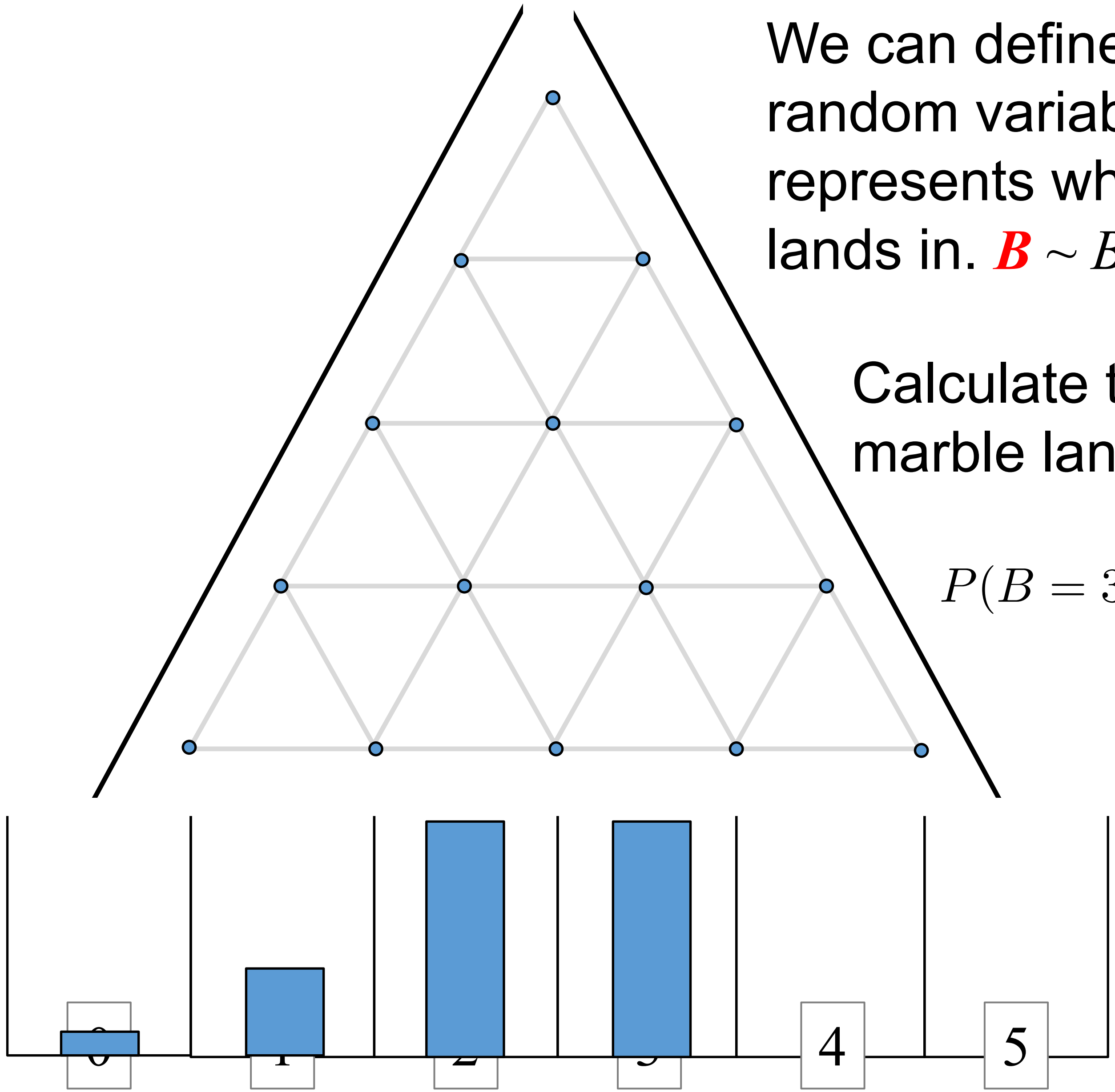


We can define an indicator random variable (***B***) which represents what bucket a marble lands in. ***B*** $\sim Bin(5, 0.5)$

Calculate the probability of a marble landing in a bucket.

$$P(B = 2) = \binom{5}{2} \frac{1}{2}^5 \approx 0.31$$

Galton Board



We can define an indicator random variable (***B***) which represents what bucket a marble lands in. ***B*** $\sim Bin(5,\ 0.5)$

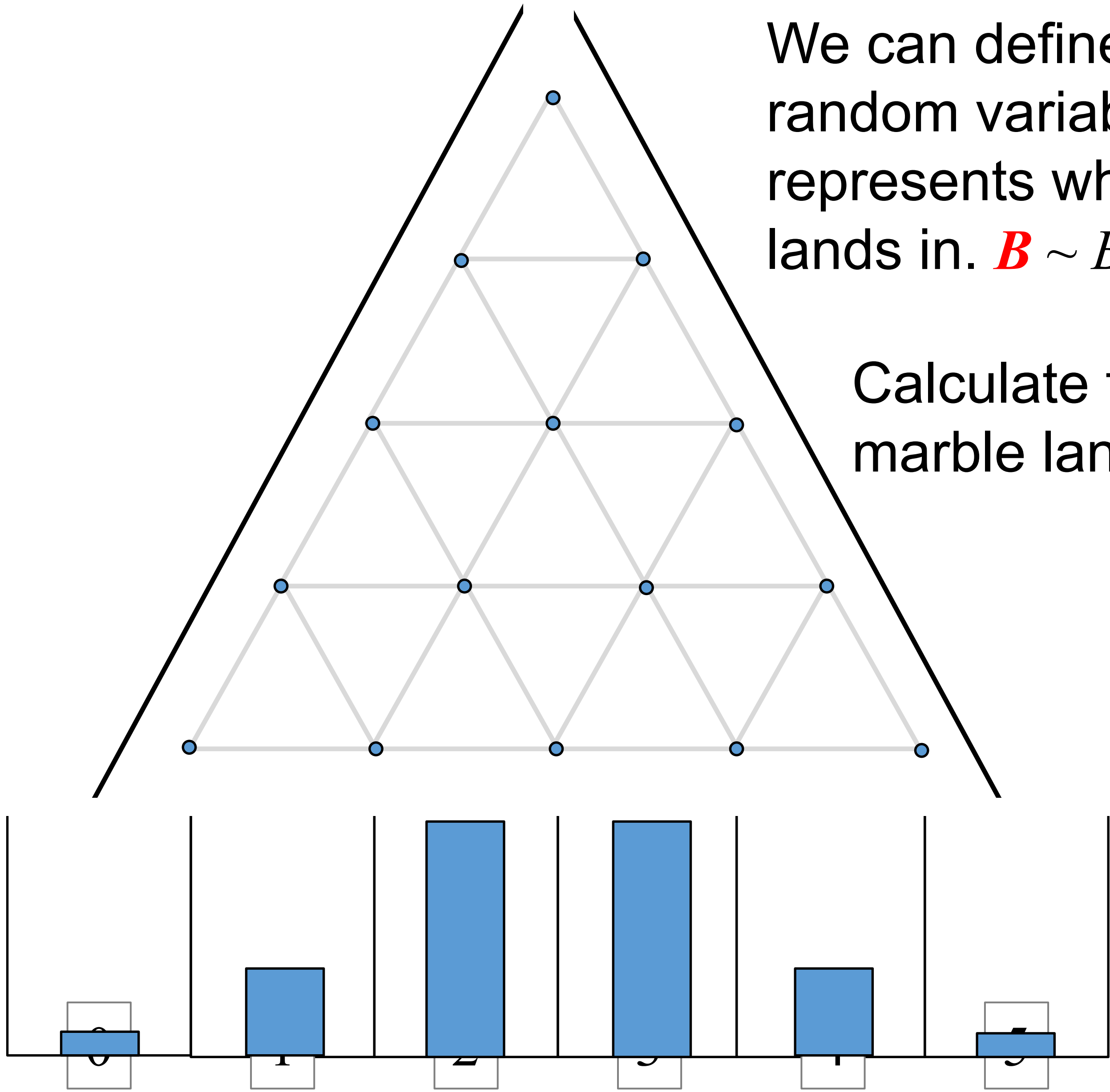
Calculate the probability of a marble landing in a bucket.

$$P(B = 3) = \binom{5}{2} \frac{1}{2}^5 \approx 0.31$$

Galton Board

We can define an indicator random variable (***B***) which represents what bucket a marble lands in. $B \sim Bin(5, 0.5)$

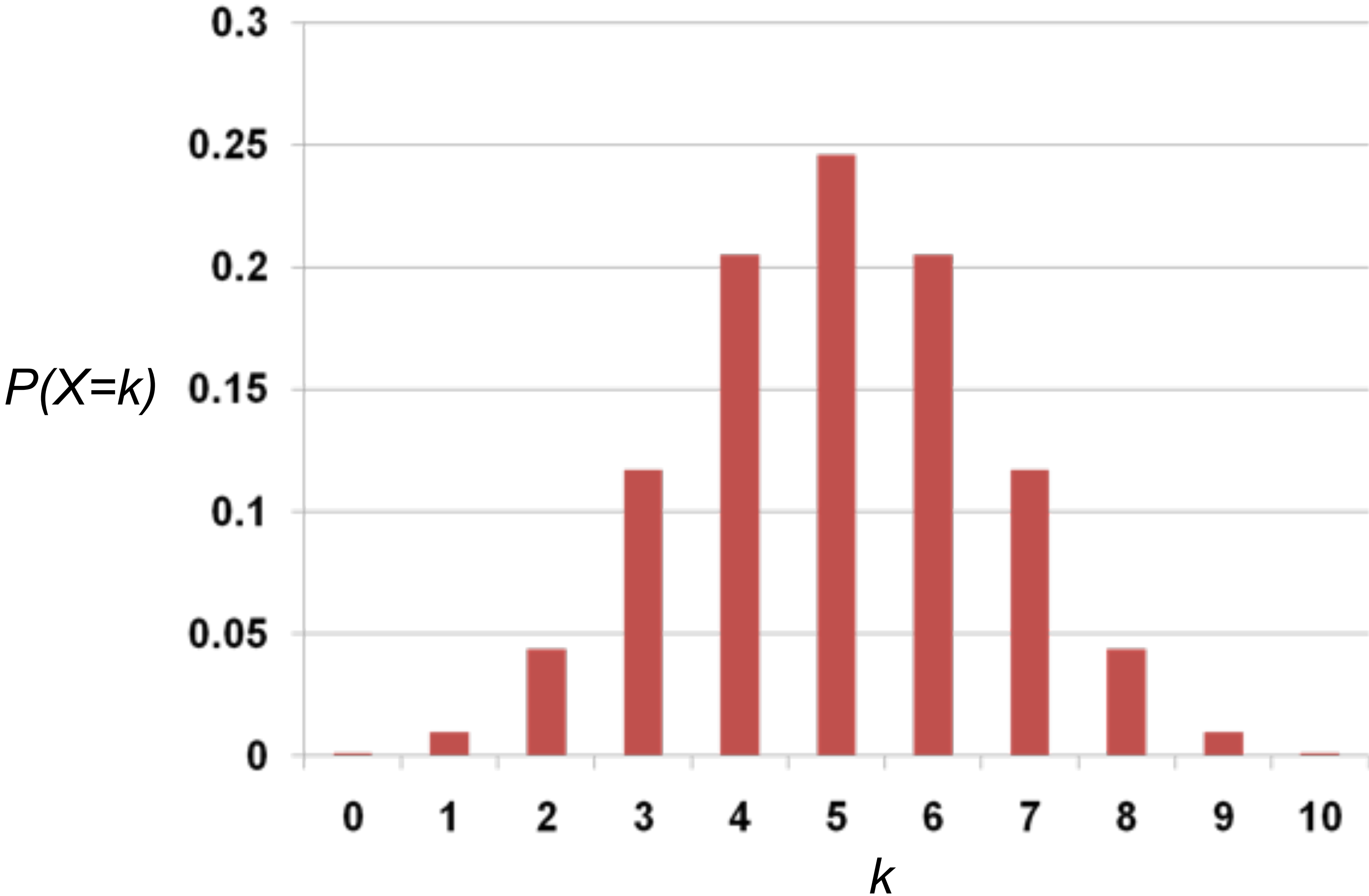
Calculate the probability of a marble landing in a bucket.



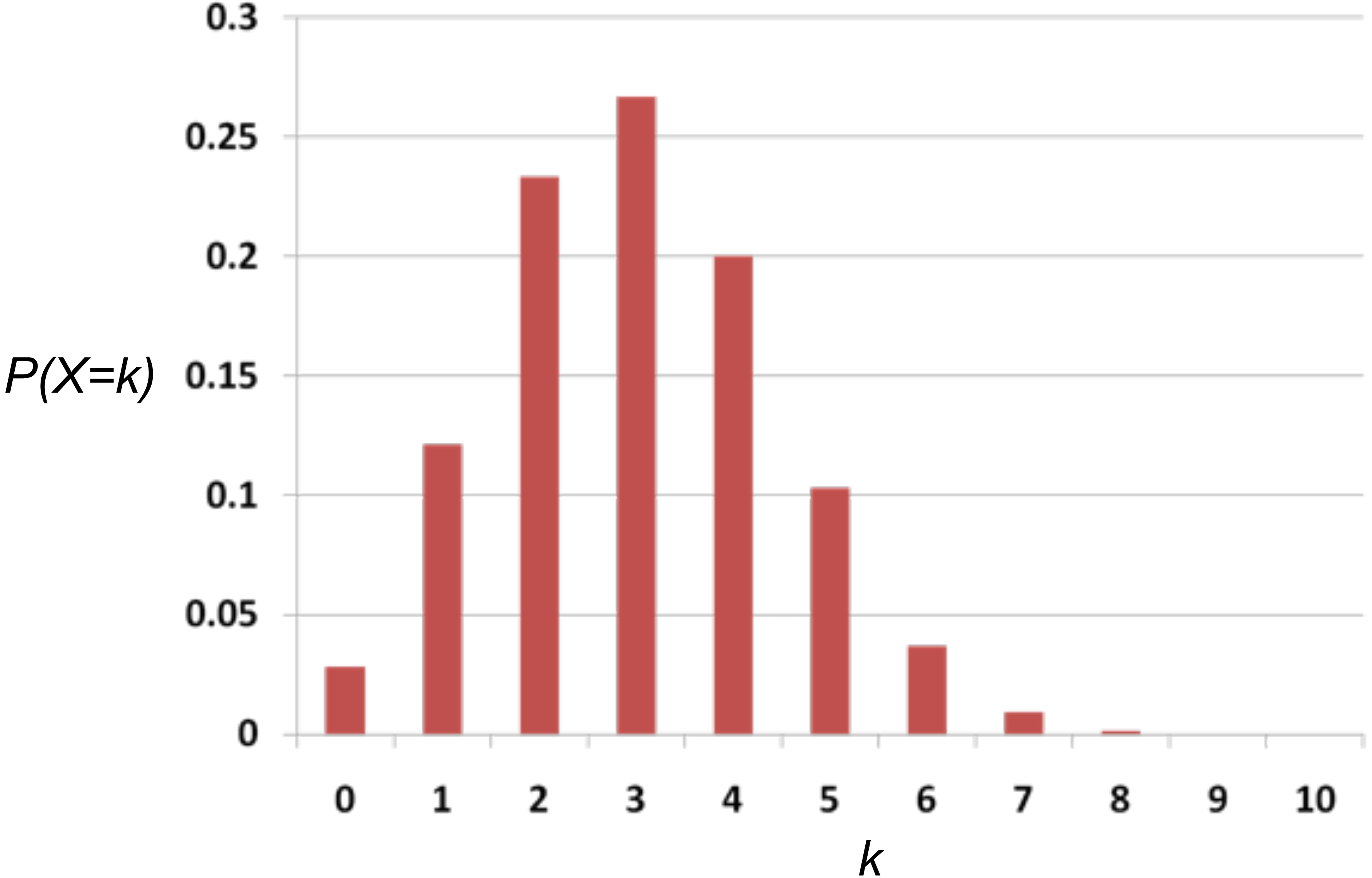
PMF

From Chaos to Order

PMF for X ~ Bin(10, 0.5)



PMF for X ~ Bin(10, 0.3)



Genetic Inheritance

- Person has 2 genes for trait (eye color)
  - Child receives 1 gene (equally likely) from each parent
  - Child has brown eyes if either (or both) genes brown
  - Child only has blue eyes if both genes blue
  - Brown is "dominant" (d) , Blue is "recessive" (r)
  - Parents each have 1 brown and 1 blue gene
- 4 children, what is P(3 children with brown eyes)?
  - Child has blue eyes: p = (½) (½) = ¼ (2 blue genes)
  - P(child has brown eyes) = 1 – (¼) = 0.75
  - X = # of children with brown eyes. X ~ Bin(4, 0.75)

$$P(X=3)=\binom{4}{3}(0.75)^3(0.25)^1 \approx 0.4219$$


GOLDEN STATE
35
WARRIORS
GOLDEN STATE

Probability you win a series?

Warriors played the Raptors in a best of 7 series during the 2019 NBA finals. What was the probability (going in) that the warriors would win the series? Each game is **independent**. Each game, the warriors have a 0.55 probability of winning? Win series if you win at least 4 games.

Let X be the number of games won. X ~ Bin(n= 7, p = 0.55).
P(X > 3)?

$$P(X \geq 4) = \sum_{i=4}^{7} P(X = i)$$

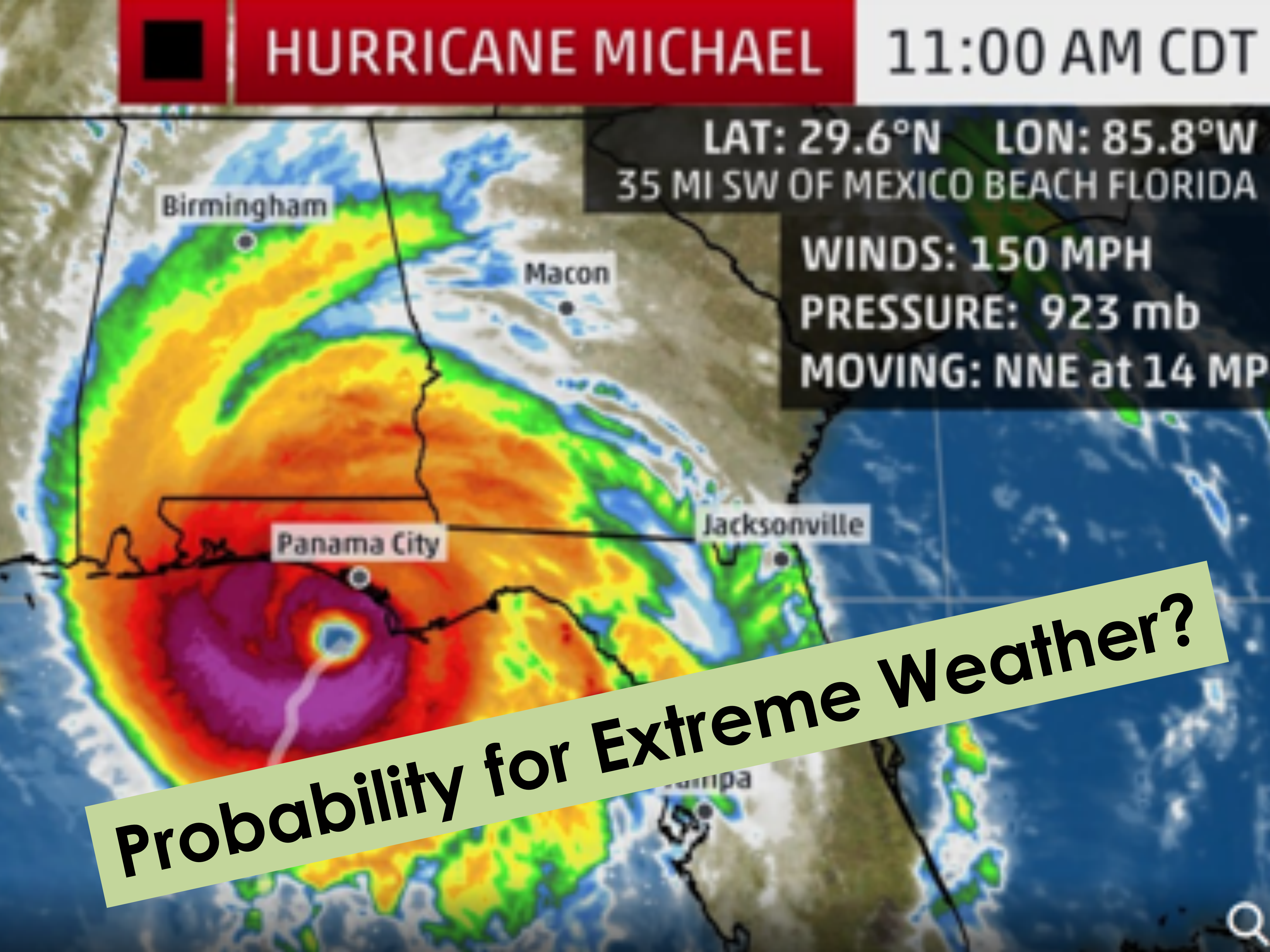
$$= \sum_{i=4}^{7} \binom{7}{i} p^i (1-p)^{7-i}$$

$$= \sum_{i=4}^{7} \binom{7}{i} 0.55^i (0.45)^{7-i} \quad \approx 0.61$$

Stretch!

Poisson

Noah Arthurs
CS109, Stanford University


HURRICANE MICHAEL
11:00 AM CDT
LAT: 29.6°N LON: 85.8°W
35 MI SW OF MEXICO BEACH FLORIDA
WINDS: 150 MPH
PRESSURE: 923 mb
MOVING: NNE at 14 MP
Birmingham
Macon
Panama City
Jacksonville
Probability for Extreme Weather?

Good to know

Natural Exponent def:

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

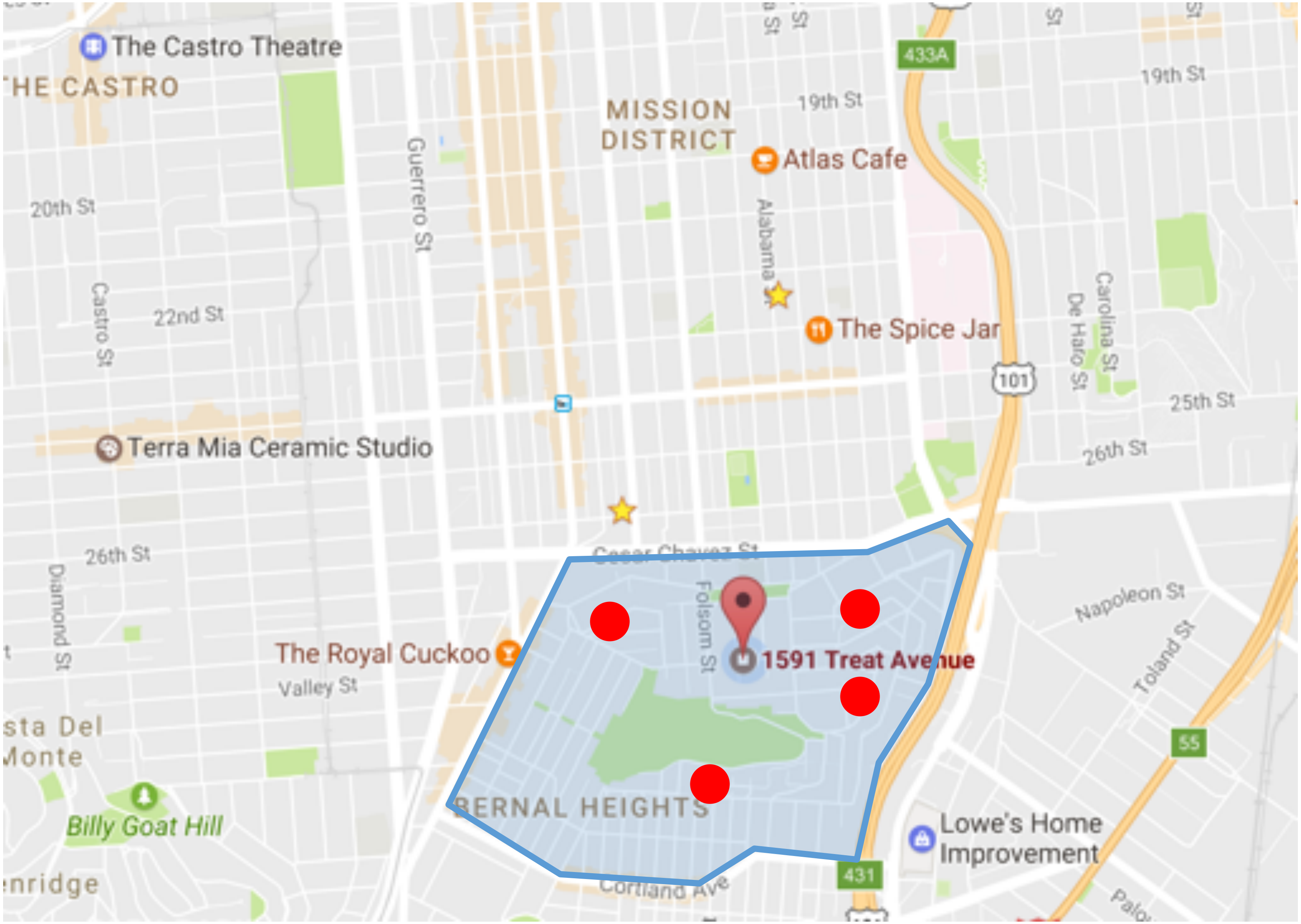
Jacob Bernoulli

https://en.wikipedia.org/wiki/E_(mathematical_constant)

Algorithmic Ride Sharing

Probability of *k* requests from this area in the next 1 min

Probability of *k* requests from this area in the next 1 min



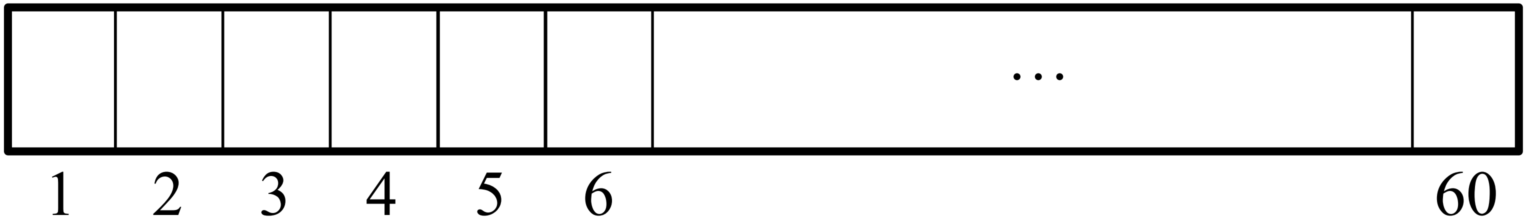
Probability of *k* requests from this area in the next 1 min

On average $\lambda = 5$ requests per minute

Probability of *k* requests from this area in the next 1 min

On average $\lambda = 5$ requests per minute

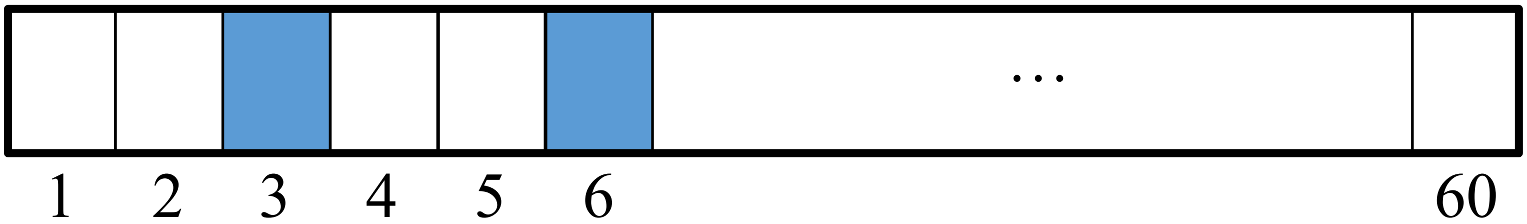
We can break the next minute down into seconds



Probability of *k* requests from this area in the next 1 min

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds



At each second either get a request or you don't.

Probability of *k* requests from this area in the next 1 min

On average λ = 5 requests per minute

We can break the next minute down into seconds

At each second either get a request or you don't.
Let X = Number of requests in the minute

$$X \sim \text{Bin}(n = 60, p = 5/60)$$

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

$$P(X = 3) = \binom{60}{3} (5/60)^3 (1 - 5/60)^{57}$$

Probability of *k* requests from this area in the next 1 min

On average λ = 5 requests per minute

We can break the next minute down into seconds

At each second either get a request or you don't.

Let X = Number of requests in the minute

$$X \sim \text{Bin}(n = 60, p = 5/60)$$

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

But what if there are two requests in the same second?

Probability of *k* requests from this area in the next 1 min

On average $\lambda = 5$ requests per minute

We can break that next minute down into ***milli***-seconds

At each ***milli***-second either get a request or you don't.

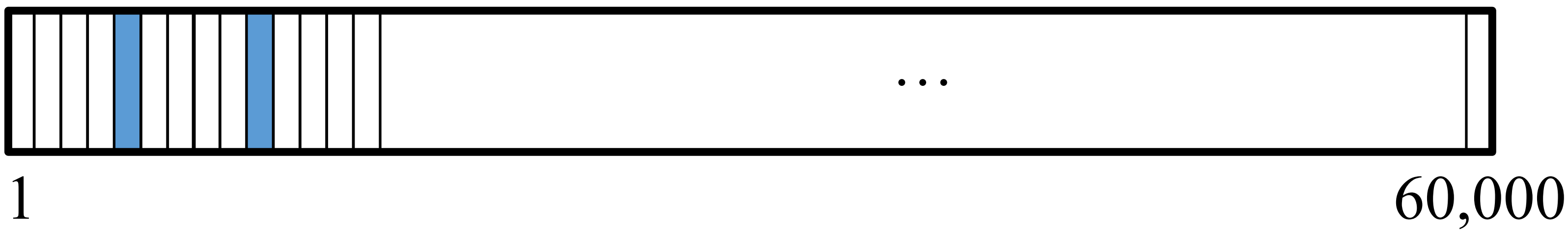
Let X = Number of requests in the minute

But what if there are two requests in the same milli-second?

Probability of *k* requests from this area in the next 1 min

On average λ = 5 requests per minute

We can break that next minute down into ***milli***-seconds



At each ***milli***-second either get a request or you don't.
Let X = Number of requests in the minute

$$X \sim \text{Bin}(n = 60000, p = \lambda/n)$$

$$P(X = k) = \binom{n}{k} (\lambda/n)^k (1 - \lambda/n)^{n-k}$$

Can we do any better than milli-seconds?

Binomial in the Limit

On average λ = 5 requests per minute

We can break that minute down into ***infinitely small*** buckets

OMG so small

1 ∞

Let X = Number of requests in the minute

$$X \sim \text{Bin}(n, p = \lambda/n)$$

$$P(X = k) = \lim_{n \to \infty} \binom{n}{k} (\lambda/n)^k (1 - \lambda/n)^{n-k}$$

Who wants to see some cool math?

Binomial in the Limit

$$P(X = k) = \lim_{n\to\infty} \binom{n}{k} (\lambda/n)^k (1 - \lambda/n)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}$$

By expanding each term

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{e^{-\lambda}}{1}$$

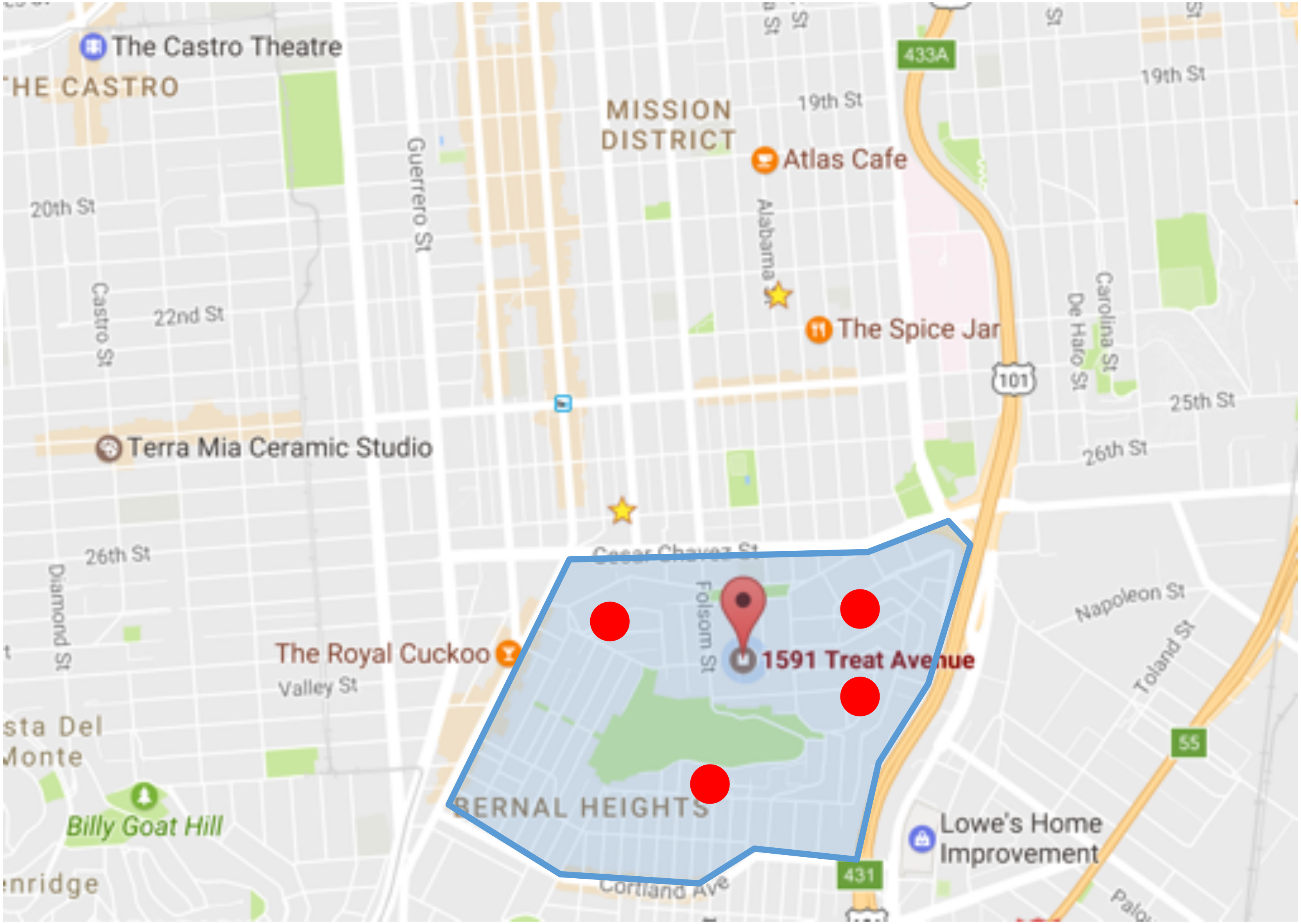
ral exp

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1}$$

$$= \lim_{n\to\infty} \frac{n^k}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1}$$

$$= \frac{\lambda^k e^{-\lambda}}{k!}$$

Probability of *k* requests from this area in the next 1 min



Simeon-Denis Poisson

- Simeon-Denis Poisson (1781-1840) was a prolific French mathematician

- Published his first paper at 18, became professor at 21, and published over 300 papers in his life
  - He reportedly said *“Life is good for only two things, discovering mathematics and teaching mathematics.”*

Poisson Random Variable

- X is a **Poisson** Random Variable: the number of occurrences in a fixed interval of time.

$$X \sim \mathrm{Poi}(\lambda)$$

  - λ is the "rate"
  - X takes on values 0, 1, 2…
  - has distribution (PMF):

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Poisson Process

- Consider events that occur over time
  - Earthquakes, radioactive decay, hits to web server, etc.
  - Have time interval for events (1 year, 1 sec, whatever...)
  - Events arrive at rate: λ events per interval of time
- Split time interval into $n \rightarrow \infty$ sub-intervals
  - Assume at most one event per sub-interval
  - Event occurrences in sub-intervals are independent
  - With many sub-intervals, probability of event occurring in any given sub-interval is small
- $N(t)$ = # events in original time interval ~ $\text{Poi}(\lambda)$

Poisson is great when you have a rate and you care about # of occurrences!

Make sure that the rate is the expected number of occurrences within the timeframe you care about.