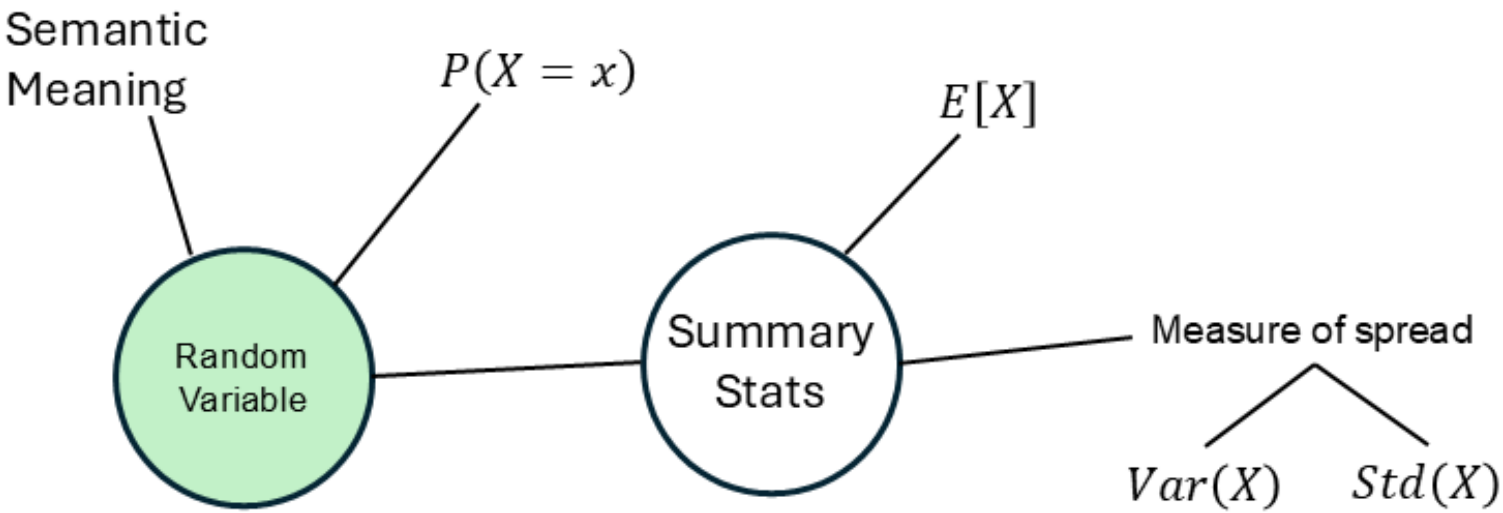
Probability and Statistic (for CS Students)

Lecture 10: Poisson

University of Mazandaran

Fall 1405

Recap: Fundamental Properties of Random Variables

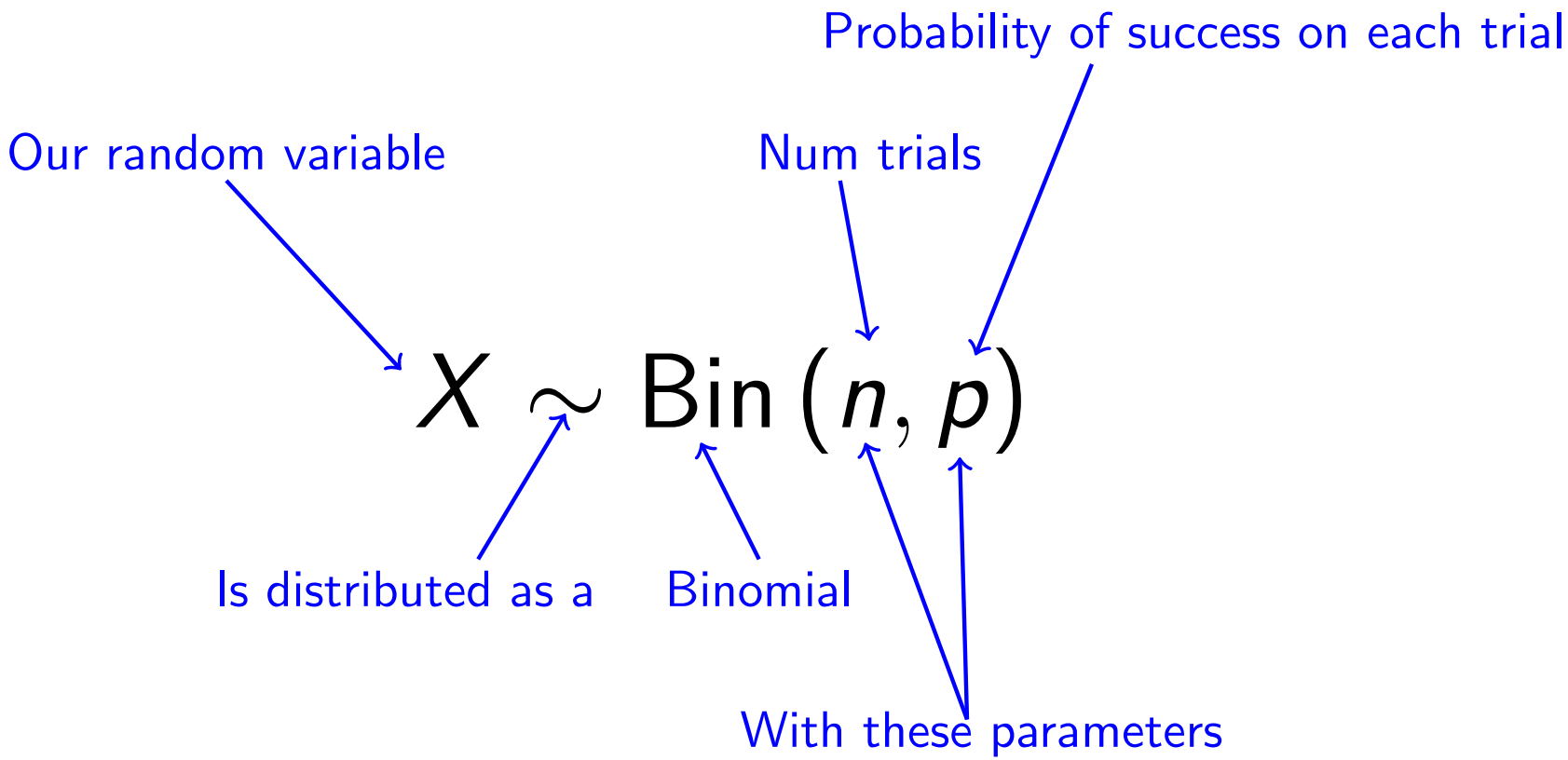


Recap: Binomial Random Variable

The number of successes, in n independent trials, where each trial is a success with probability p:

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

Recap: Binomial Random Variable



Recap: Binomial Random Variable

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

Probability that our variable takes on the value k

Recap: Binomial Random Variable

Binomial Random Variable

Notation: $X \sim \text{Bin}(n, p)$

Description: Number of "successes" in n identical, independent experiments each with probability of success p.

Parameters: $n \in \{0, 1, \ldots\}$, the number of experiments.
$p \in [0, 1]$, the probability that a single experiment gives a "success".

Support: $x \in \{0, 1, \ldots, n\}$

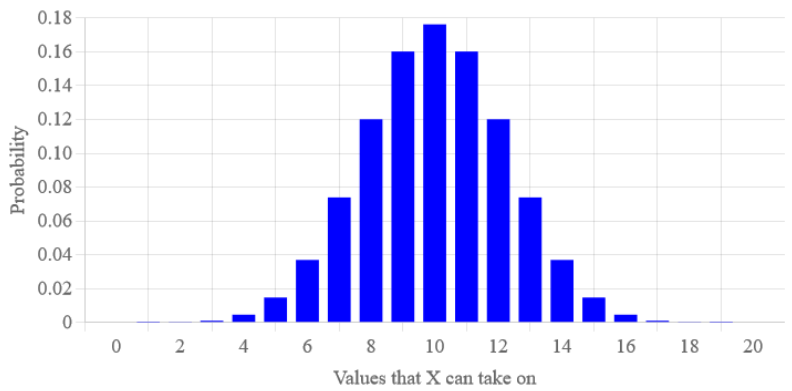
PMF equation: $\text{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

Expectation: $\text{E}[X] = n \cdot p$

Variance: $\text{Var}(X) = n \cdot p \cdot (1-p)$

PMF graph:

Parameter n: 20 Parameter p: 0.5



Bernoulli Random Variable

Notation: $X \sim \text{Bern}(p)$

Description: A boolean variable that is 1 with probability p

Parameters: p, the probability that $X = 1$.

Support: x is either 0 or 1

PMF equation: $\text{P}(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$

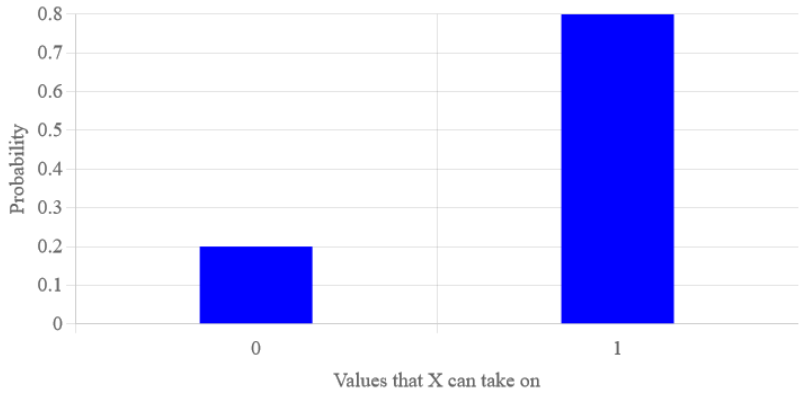
PMF (smooth): $\text{P}(X = x) = p^x (1-p)^{1-x}$

Expectation: $\text{E}[X] = p$

Variance: $\text{Var}(X) = p(1-p)$

PMF graph:

Parameter p: 0.80



Natural Exponent

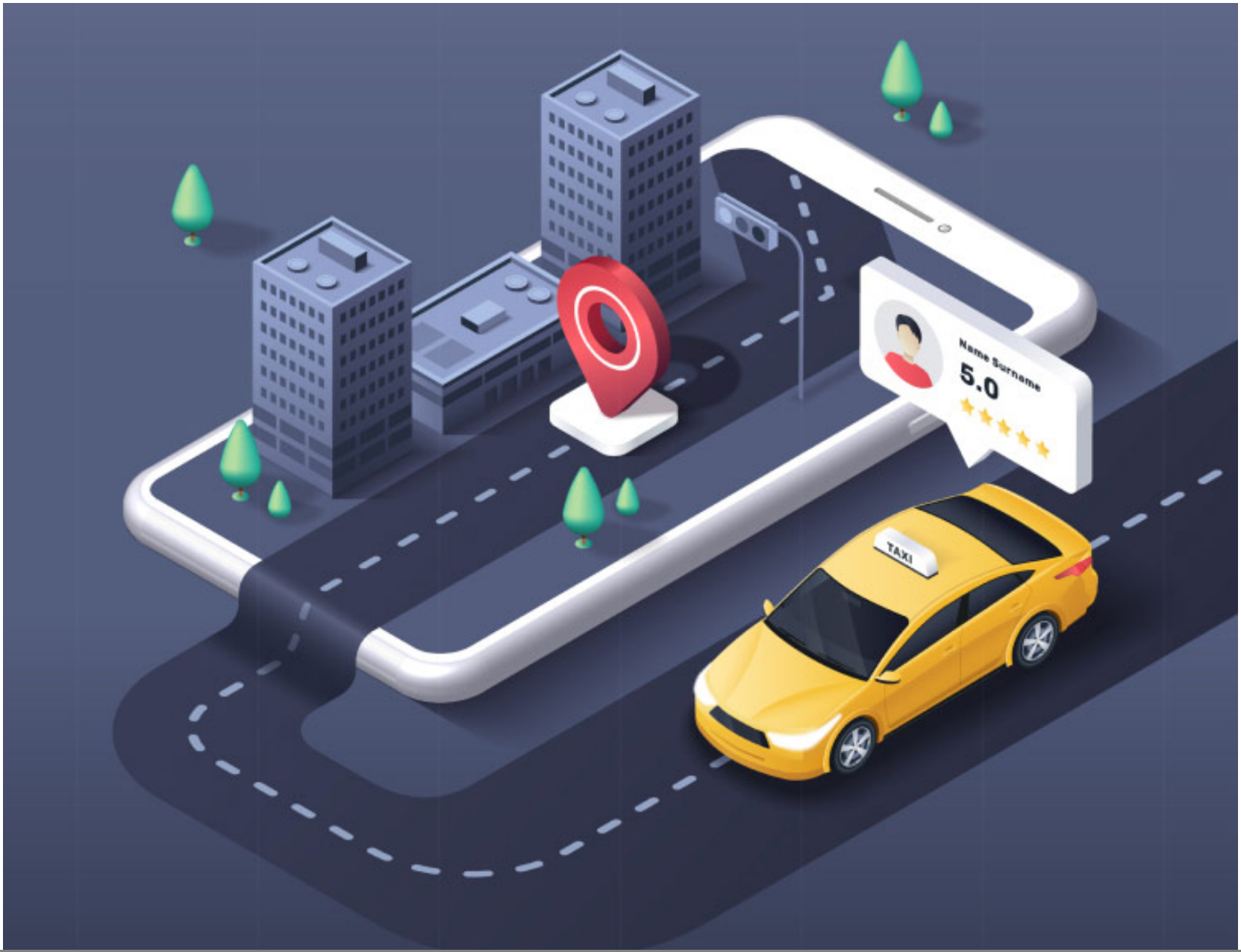
$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Jakob Bernoulli



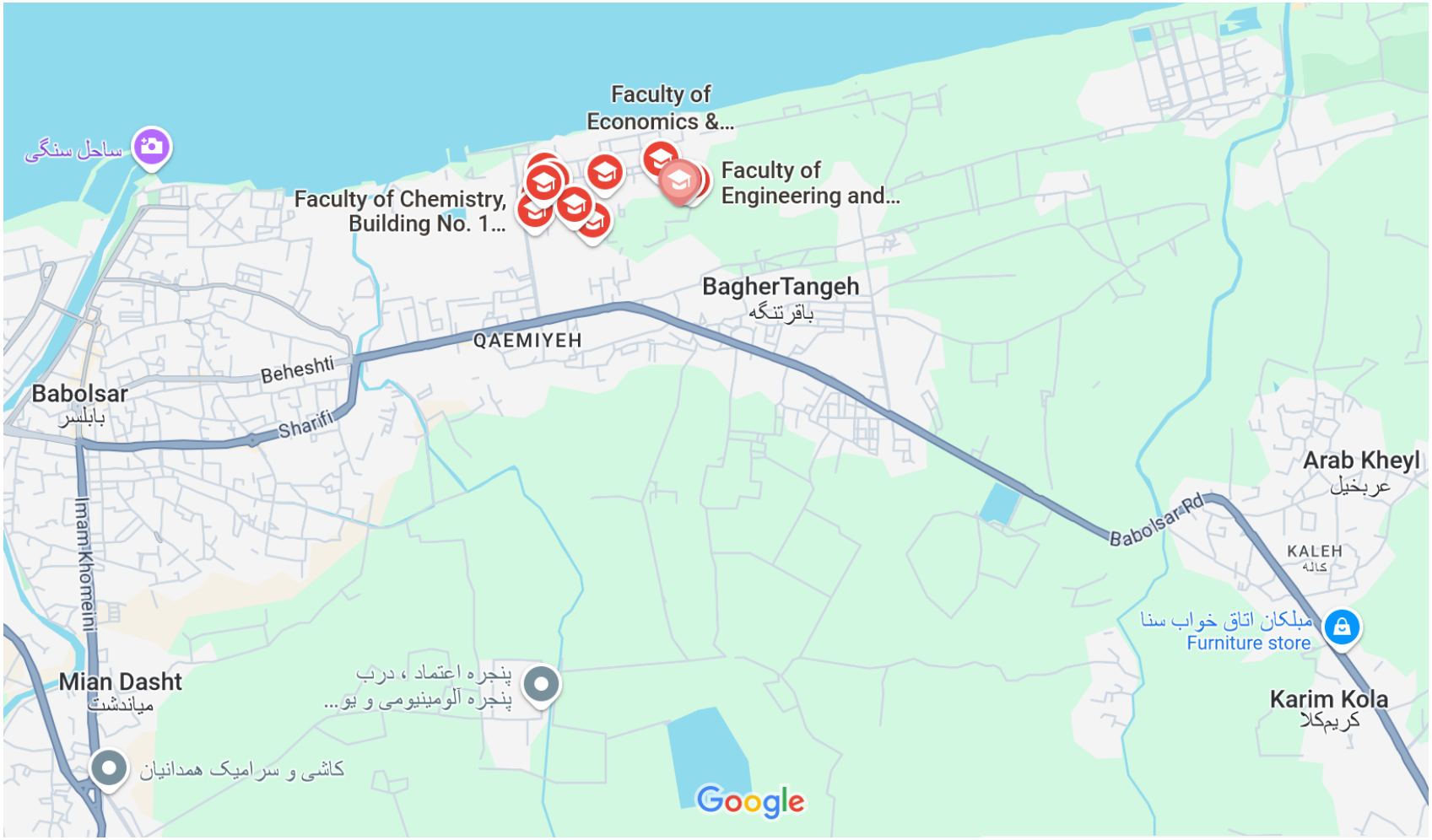
https://de.wikipedia.org/wiki/Jakob_I_Bernoulli

How Snapp Works?



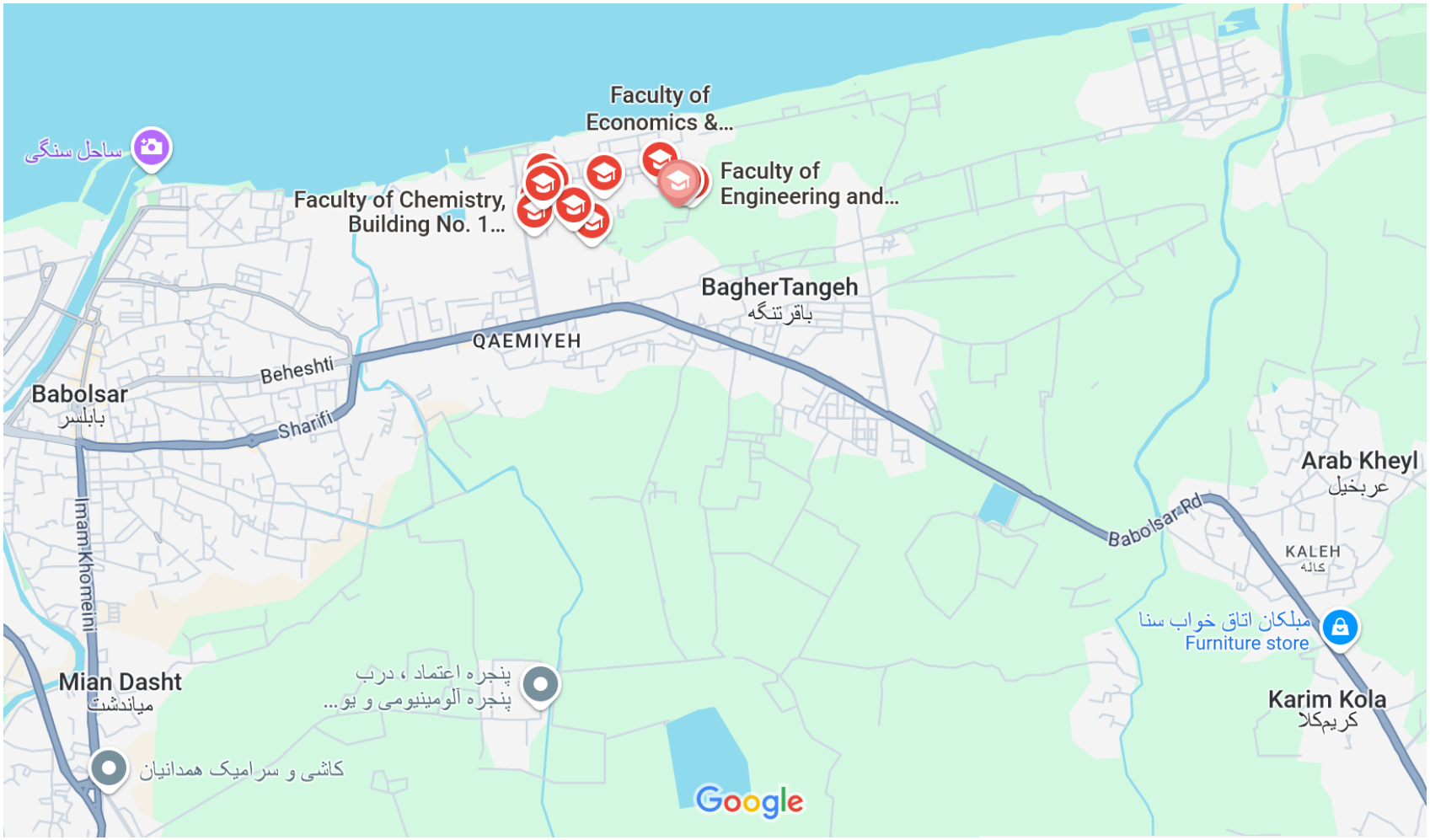
How Snapp Works?

Probability of k ride request from this area in the next 1 min



How Snapp Works?

Probability of k ride request from this area in the next 1 min



Attention! Attention! On average, $\lambda = 5$ requests per minute comes from this area ;)

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds



How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds



At each second, either get a request or not!

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds



At each second, either get a request or not!

Let X = Number of requests in the next minute

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds



At each second, either get a request or not!

Let X = Number of requests in the next minute

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

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds



At each second, either get a request or not!
Let X = Number of requests in the next minute

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

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

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into seconds



At each second, either get a request or not!
Let $X =$ Number of requests in the next 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 request in the same second!?

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into milliseconds



At each millisecond, either get a request or not!

Let X = Number of requests in the next minute

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into milliseconds



At each millisecond, either get a request or not!

Let X = Number of requests in the next minute

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into milliseconds



At each millisecond, either get a request or not!
Let X = Number of requests in the next minute

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

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into milliseconds



At each millisecond, either get a request or not!
Let X = Number of requests in the next minute

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

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

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into milliseconds



At each millisecond, either get a request or not!
Let X = Number of requests in the next 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 millisecond?

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into infinitely small buckets



How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into infinitely small buckets



Let X = Number of requests in the minute

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into infinitely small buckets



Let X = Number of requests in the minute

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

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into infinitely small buckets



Let X = Number of requests in the minute

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

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

How Snapp Works?

On average $\lambda = 5$ requests per minute

We can break the next minute down into infinitely small buckets



Let X = Number of requests in the minute

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

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

Let's see some beautiful math!

How Snapp Works?

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

How Snapp Works?

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

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^k} \quad \text{By exapnding each term}$$

How Snapp Works?

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

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k} \quad \text{By exapnding each term}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{e^{-\lambda}}{1} \quad \text{By definition of natural exp}$$

How Snapp Works?

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

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k} \quad \text{By exapnding each term}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{e^{-\lambda}}{1} \quad \text{By definition of natural exp}$$

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

How Snapp Works?

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

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k} \quad \text{By exapnding each term}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{e^{-\lambda}}{1} \quad \text{By definition of natural exp}$$

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

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

How Snapp Works?

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

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} . \frac{\lambda^k}{n^k} . \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k} \quad \text{By exapnding each term}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} . \frac{\lambda^k}{n^k} . \frac{e^{-\lambda}}{1} \quad \text{By definition of natural exp}$$

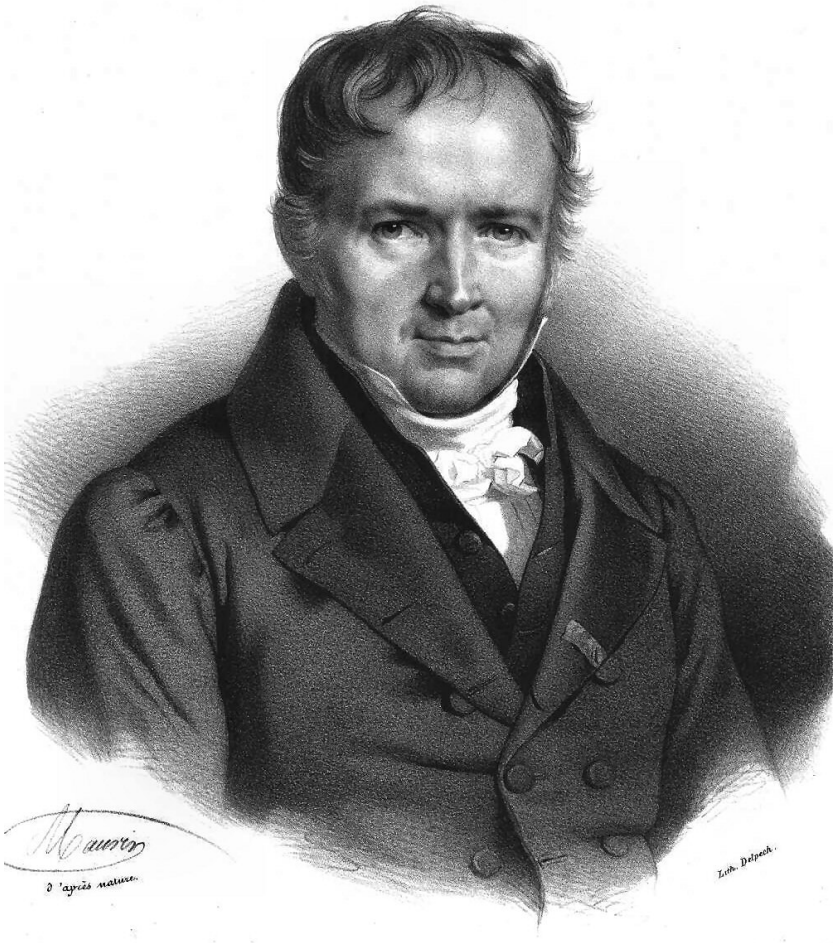
$$= \lim_{n\to\infty} \frac{n!}{(n-k)!n^k} . \frac{\lambda^k}{k!} . \frac{e^{-\lambda}}{1} \quad \text{Rearrange terms}$$

$$= \lim_{n\to\infty} \frac{n^k}{n^k} . \frac{\lambda^k}{k!} . \frac{e^{-\lambda}}{1} \quad \text{Limit analysis}$$

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

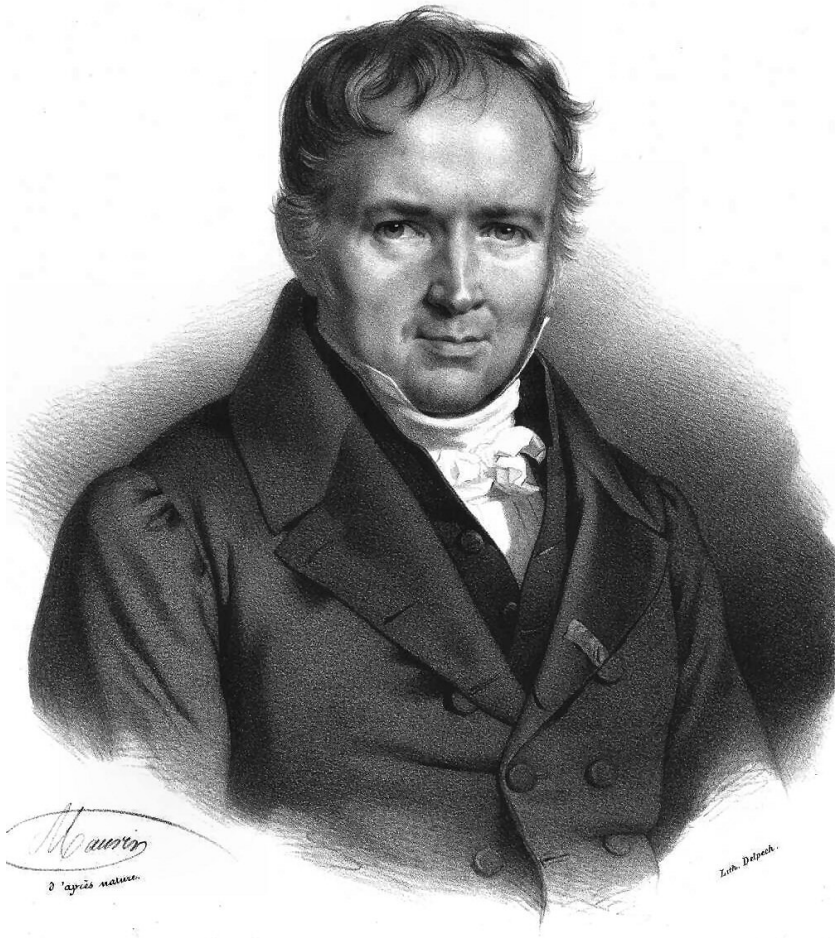
Simeon-Denis Poisson

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



Simeon-Denis Poisson

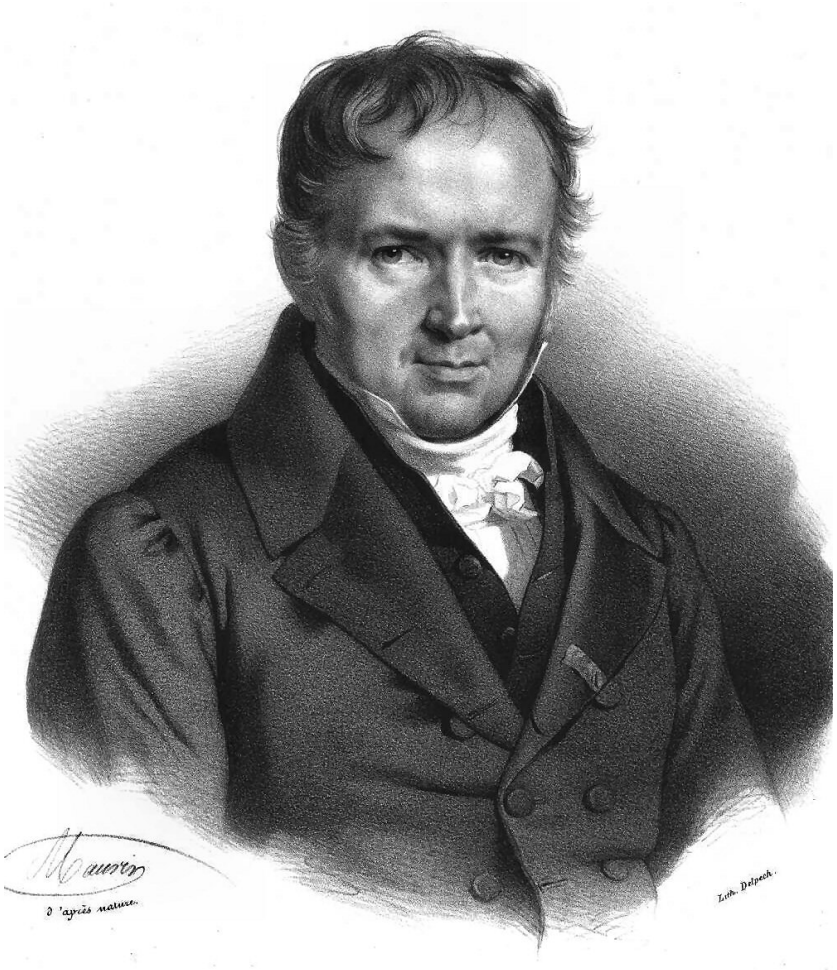
Simeon-Denis Poisson (1781-1840) was a French mathematician



published his first paper at 18, became professor at 21!

Simeon-Denis Poisson

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



published his first paper at 18, became professor at 21!

"Life is good for 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:

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

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

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

Poisson Process

Consider events that occur over time:

- Earthquake, cars passing a street, 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

Poisson Process

Consider events that occur over time:

- Earthquake, cars passing a street, 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 $n \to \infty$ sub-intervals

- Assume at most 1 event per sub-interval
- Events occurrences in sub-intervals are independent
- With many sub-intervals, probability of event occurring in any sub-interval is small

Poisson Process

Consider events that occur over time:

- Earthquake, cars passing a street, 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 $n \to \infty$ sub-intervals

- Assume at most 1 event per sub-interval
- Events occurrences in sub-intervals are independent
- With many sub-intervals, probability of event occurring in any sub-interval is small

$N(t) = \#$ events in original time interval $\sim \text{Poi}(\lambda)$

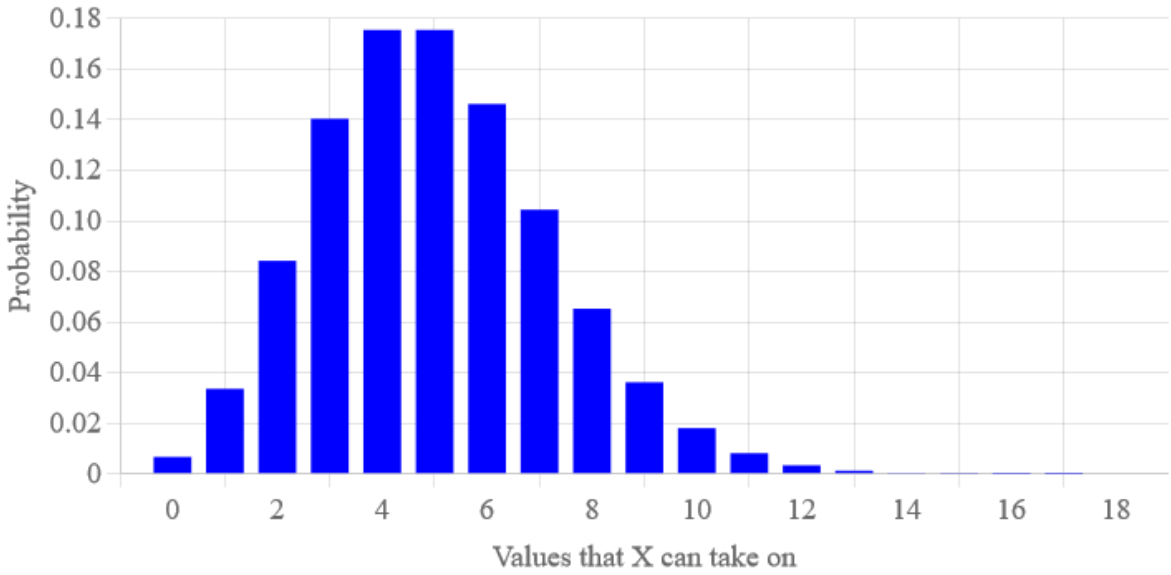
Poisson Process

Poisson Random Variable

Notation:	$X \sim \text{Poi}(\lambda)$
Description:	Number of events in a fixed time frame if (a) the events occur with a constant mean rate and (b) they occur independently of time since last event.
Parameters:	$\lambda \in \mathbb{R}^+$, the constant average rate.
Support:	$x \in \{0, 1, \ldots\}$
PMF equation:	$\text{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$
Expectation:	$\text{E}[X] = \lambda$
Variance:	$\text{Var}(X) = \lambda$

PMF graph:

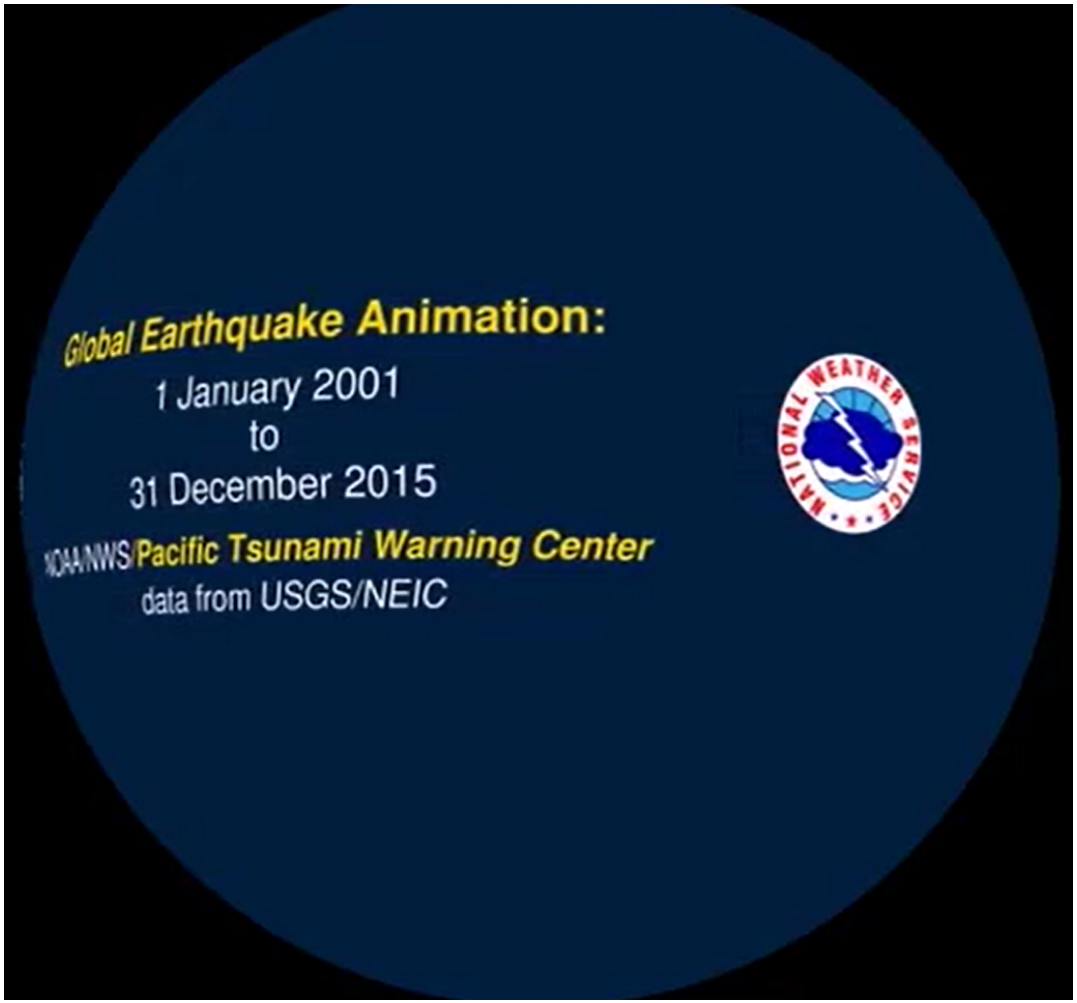
Parameter λ: 5



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

Make sure that the time unit for "rate" matches the unit for the probability question

Earthquakes



https://sos.noaa.gov/catalog/datasets/earthquakes-2001-2015/

Average of 2.79 major earthquakes per year.
What is the probability of 3 major earthquakes next year?

Earthquakes

Let $X =$ number of earthquakes next year

$$X \sim \text{Poi}\,(2.79)$$

$$P\,(X = 3) = \frac{\lambda^k e^{-\lambda}}{k!} \approx \frac{2.79^3 e^{-2.79}}{3!}$$

Bulletin of the
Seismological Society of America

Vol. 64 October 1974 No. 5

IS THE SEQUENCE OF EARTHQUAKES IN SOUTHERN CALIFORNIA, WITH AFTERSHOCKS REMOVED, POISSONIAN?

By J. K. Gardner and L. Knopoff

Bulletin of the
Seismological Society of America

Vol. 64 October 1974 No. 5

IS THE SEQUENCE OF EARTHQUAKES IN SOUTHERN CALIFORNIA, WITH AFTERSHOCKS REMOVED, POISSONIAN?

By J. K. Gardner and L. Knopoff

Abstract

Yes.

Poisson can approximate a Binomial!

Wait, why would you want to do that?

- Binomial can be expensive to compute
- Connections help build math intuition

Poisson Distribution and Social Studies

Japan is among countries with highest suicides. In 2022, Japan population was $\sim$ 124 million and in total 21881 suicides were reported (that maps to 17.5 person per 100000 people!). what is the probability that 800 suicides happened in Tokyo (14 million population)?

Poisson Distribution and Social Studies

Japan is among countries with highest suicides. In 2022, Japan population was $\sim$ 124 million and in total 21881 suicides were reported (that maps to 17.5 person per 100000 people!). what is the probability that 800 suicides happened in Tokyo (14 million population)?

Using Binomial:

Poisson Distribution and Social Studies

Japan is among countries with highest suicides. In 2022, Japan population was ~ 124 million and in total 21881 suicides were reported (that maps to 17.5 person per 100000 people!). what is the probability that 800 suicides happened in Tokyo (14 million population)?

Using Binomial:

X: number of people in Tokyo who committed suicide

$$X \sim Bin(n, p)$$

$$n = 14 \times 10^6 \qquad p = \frac{21881}{124 \times 10^6}$$

$$P(X = 800) = \binom{n}{800} p^{800} (1-p)^{n-800}$$

Poisson Distribution and Social Studies

Japan is among countries with highest suicides. In 2022, Japan population was ~ 124 million and in total 21881 suicides were reported (that maps to 17.5 person per 100000 people!). what is the probability that 800 suicides happened in Tokyo (14 million population)?

Using Poisson:

Poisson Distribution and Social Studies

Japan is among countries with highest suicides. In 2022, Japan population was ~ 124 million and in total 21881 suicides were reported (that maps to 17.5 person per 100000 people!). what is the probability that 800 suicides happened in Tokyo (14 million population)?

Using Poisson:

X: number of people in Tokyo who committed suicide in 2022
Compute the suicide rate in Tokyo:

$$\lambda = p_{\text{suicide}} \times n_{\text{Tokyo}} = \frac{21881}{124 \times 10^6} \times 14 \times 10^6 \approx 247.04$$

$$X \sim \text{Poi}\,(\lambda = 247.04)$$

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

Poisson is a Binomial in the Limit

Poisson approximates Binomial where n is large, p is small, and $\lambda = np$ is "moderate"

Poisson is a Binomial in the Limit

Poisson approximates Binomial where n is large, p is small, and $\lambda = np$ is "moderate"

Different interpretation of "moderate"

- $n > 20$ and $p < 0.05$
- $n > 100$ and $p < 0.1$

Poisson is a Binomial in the Limit

Poisson approximates Binomial where n is large, p is small, and $\lambda = np$ is "moderate"

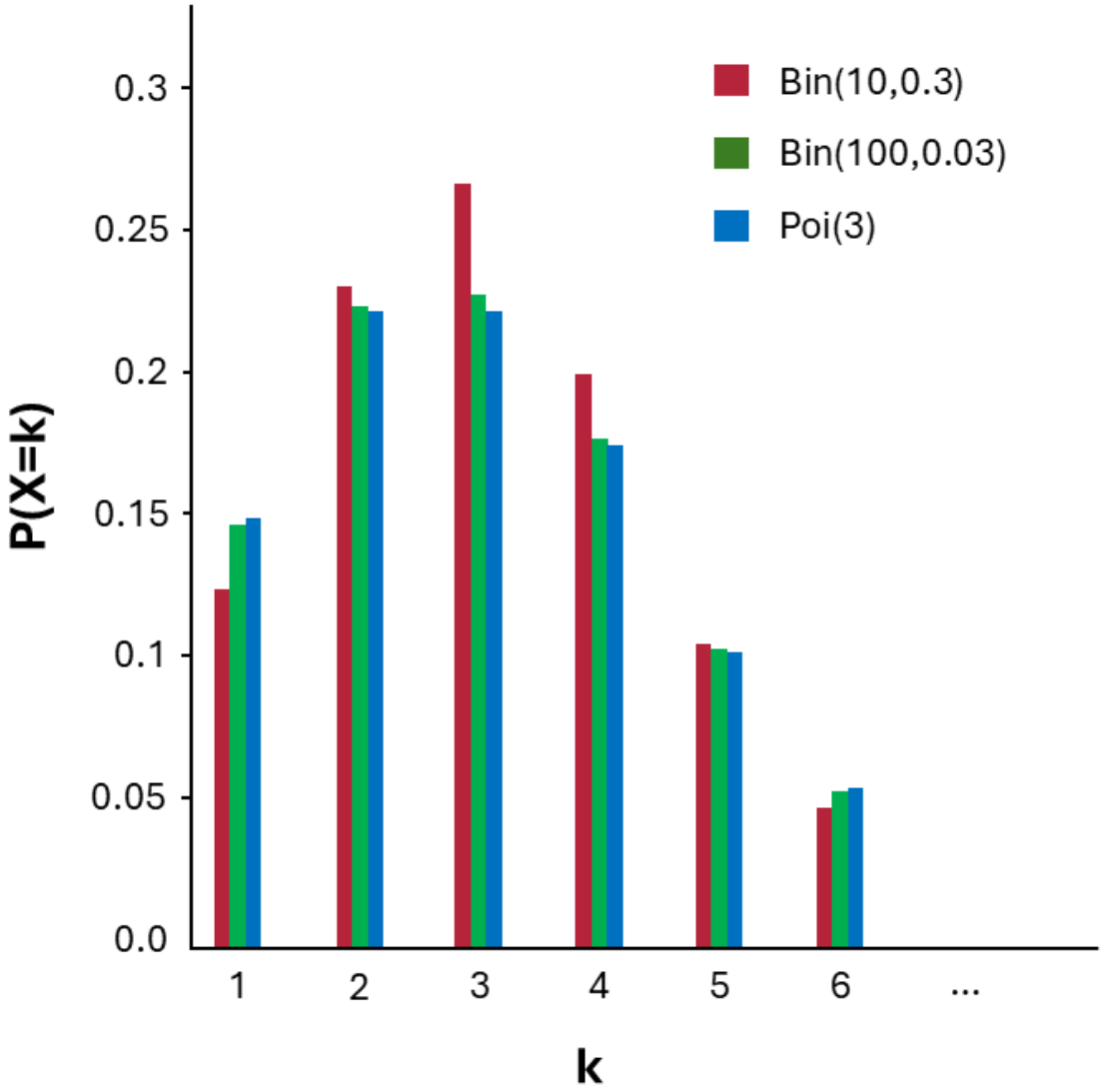
Different interpretation of "moderate"

- $n > 20$ and $p < 0.05$
- $n > 100$ and $p < 0.1$

In fact, Poisson is Binomial as

$$n \to \infty \text{ and } p \to 0, \text{ where } np = \lambda$$

Bin(10,0.3) vs Bin(100,0.03) vs Poi(3)



Moments With Poisson

Recall: $Y \sim \text{Bin}(n, p)$

- $E[Y] = np$
- $\text{Var}(Y) = np(1-p)$

Moments With Poisson

Recall: $Y \sim \text{Bin}(n, p)$

- $E[Y] = np$
- $\text{Var}(Y) = np(1-p)$

$X \sim \text{Poi}(\lambda)$ where $\lambda = np$ $(n \to \infty$ and $p \to 0)$

Moments With Poisson

Recall: $Y \sim \text{Bin}(n, p)$

- $E[Y] = np$
- $\text{Var}(Y) = np(1-p)$

$X \sim \text{Poi}(\lambda)$ where $\lambda = np$ $(n \to \infty$ and $p \to 0)$

- $E[X] = np = \lambda$

Moments With Poisson

Recall: $Y \sim \text{Bin}(n, p)$

- $E[Y] = np$
- $\text{Var}(Y) = np(1-p)$

$X \sim \text{Poi}(\lambda)$ where $\lambda = np$ $(n \to \infty$ and $p \to 0)$

- $E[X] = np = \lambda$
- $\text{Var}(X) = np(1-p) = \lambda(1-0) = \lambda$

Moments With Poisson

Recall: $Y \sim \text{Bin}(n, p)$

- $E[Y] = np$
- $\text{Var}(Y) = np(1-p)$

$X \sim \text{Poi}(\lambda)$ where $\lambda = np$ ($n \to \infty$ and $p \to 0$)

- $E[X] = np = \lambda$
- $\text{Var}(X) = np(1-p) = \lambda(1-0) = \lambda$
- **Expectation and variance of Poisson are same**

Web Server Load

Consider requests to a web server in 1 second

- In past, server load averages 2 hits/second
- $X = \#$ hits server receives in a second
- What is $P(X < 5)$

Solution:

Web Server Load

Consider requests to a web server in 1 second

- In past, server load averages 2 hits/second
- $X = \#$ hits server receives in a second
- What is $P(X < 5)$

Solution:

$$X \sim \text{Poi}(\lambda = 2)$$

$$P(X < 5) = \sum_{k=0}^{4} P(X < k)$$

$$= \sum_{k=0}^{4} e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{Since } X \text{ is Poisson}$$

$$= \sum_{k=0}^{4} e^{-2} \frac{2^k}{k!} \approx 0.95$$