08: Poisson and More

Lisa Yan
April 22, 2020
# Quick slide reference

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Poisson</td>
<td>08a_poisson</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Poisson Paradigm</td>
<td>08b_poisson_paradigm</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Other Discrete RVs</td>
<td>08c_other_discrete</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Exercises</td>
<td>LIVE</td>
<td></td>
</tr>
</tbody>
</table>
Poisson RV
Before we start

The natural exponent $e$:

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

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

Jacob Bernoulli while studying compound interest in 1683.
Algorithmic ride sharing

Probability of $k$ requests from this area in the next 1 minute?

Suppose we know: On average, $\lambda = 5$ requests per minute
Algorithmic ride sharing, approximately

Probability of $k$ requests from this area in the next 1 minute?

On average, $\lambda = 5$ requests per minute

Break a minute down into 60 seconds:

At each second:
- Independent trial
- You get a request (1) or you don’t (0).

Let $X = \#$ of requests in minute.

$E[X] = \lambda = 5 = n \cdot p$

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

$P(X = k) = \binom{60}{k} \left( \frac{5}{60} \right)^k \left( 1 - \frac{5}{60} \right)^{n-k}$

😊 But what if there are two requests in the same second?
Algorithmic ride sharing, approximately

Probability of $k$ requests from this area in the next 1 minute?

On average, $\lambda = 5$ requests per minute

Break a minute down into 60,000 milliseconds:

At each millisecond:
- Independent trial
- You get a request (1) or you don’t (0).

Let $X = \#$ of requests in minute.

$E[X] = \lambda = 5 = np$

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

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

But what if there are two requests in the same millisecond?
Algorithmic ride sharing, approximately

Probability of $k$ requests from this area in the next 1 minute?

On average, $\lambda = 5$ requests per minute

Break a minute down into infinitely small buckets:

For each time bucket:
• Independent trial
• You get a request (1) or you don’t (0).

Let $X = \#$ of requests in minute.

$E[X] = \lambda = 5 = n\rho$

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

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

Who wants to see some cool math?
Binomial in the limit

\[ P(X = k) = \lim_{n \to \infty} \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} \]

Expand:

\[ = \lim_{n \to \infty} \frac{n!}{k!(n-k)!} \frac{\lambda^k}{(1 - \frac{\lambda}{n})^k} \frac{(1 - \frac{\lambda}{n})^n}{n^k} \]

Rearrange:

\[ = \lim_{n \to \infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left( 1 - \frac{\lambda}{n} \right)^{n-k} \]

Def natural exponent:

\[ = \lim_{n \to \infty} \frac{n!}{k!(n-k)!} \frac{\lambda^k}{k!} \left( 1 - \frac{\lambda}{n} \right)^k \]

Expand:

\[ = \lim_{n \to \infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{(n-k)!} \frac{(n-k)!}{k!} \frac{e^{-\lambda}}{(1 - \frac{\lambda}{n})^k} \]

+ Cancel:

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

Simplify:

\[ = \frac{\lambda^k}{k!} e^{-\lambda} \]

Lim analysis:

\[ \lim_{n \to \infty} \left( 1 - \frac{\lambda}{n} \right)^n = e^{-\lambda} \]
Algorithmic ride sharing

Probability of $k$ requests from this area in the next 1 minute?

On average, $\lambda = 5$ requests per minute

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

French mathematician (1781 – 1840)
- Published his first paper at age 18
- Professor at age 21
- Published over 300 papers

“Life is only good for two things: doing mathematics and teaching it.”
Consider an experiment that lasts a fixed interval of time.

**def** A **Poisson** random variable $X$ is the number of successes over the experiment duration.

**Examples:**
- # earthquakes per year
- # server hits per second
- # of emails per day

Yes, expectation == variance for Poisson RV! More later.
Earthquakes

There are an average of 2.79 major earthquakes in the world each year. What is the probability of 3 major earthquakes happening next year?

1. Define RVs
   \[ X \sim \text{Poi}(\lambda = 2.79) \]

2. Solve
   \[
P(X = 3) = e^{-\lambda} \frac{\lambda^k}{k!} = e^{-2.79} \frac{(2.79)^3}{3!} \approx 0.23
   \]

\[ X \sim \text{Poi}(\lambda) \]
\[ E[X] = \lambda \]
\[ p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \]
Are earthquakes really Poissonian?

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 Paradigm
DNA

All the movies, images, emails and other digital data from more than 600 smartphones (10,000 GB) can be stored in the faint pink smear of DNA at the end of this test tube.

What is the probability that DNA storage stays uncorrupted?
DNA

What is the probability that DNA storage stays uncorrupted?

- In DNA (and real networks), we store large strings.
- Let string length be long, e.g., $n \approx 10^4$
- Probability of corruption of each base pair is very small, e.g., $p = 10^{-6}$
- Let $X = \#$ of corruptions.

What is $P(\text{DNA storage is uncorrupted}) = P(X = 0)$?

1. Approach 1:

   $X \sim \text{Bin}(n = 10^4, p = 10^{-6})$

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

   ![unwieldy!]

   $= \binom{10^4}{0} (10^{-6})^0 (1-10^{-6})^{10^4-0}$

   $\approx 0.99049829$

2. Approach 2:

   $X \sim \text{Poi}(\lambda = 10^4 \cdot 10^{-6} = 0.01)$

   $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} = e^{-0.01} \frac{0.01^0}{0!}$

   $= e^{-0.01}$

   $\approx 0.99049834$ a good approximation!
The Poisson Paradigm, part 1

Poisson approximates Binomial when $n$ is large, $p$ is small, and $\lambda = np$ is “moderate.”

Different interpretations of “moderate”:
- $n > 20$ and $p < 0.05$
- $n > 100$ and $p < 0.1$

Poisson is Binomial in the limit:
- $\lambda = np$, where $n \to \infty, p \to 0$

Poisson can approximate Binomial.
Poisson Random Variable

Consider an experiment that lasts a fixed interval of time.

**def** A **Poisson** random variable $X$ is the number of occurrences over the experiment duration.

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

**PMF**

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

**Support:** \{0, 1, 2, ... \}

**Expectation**

$$E[X] = \lambda$$

**Variance**

$$\text{Var}(X) = \lambda$$

Examples:

- # earthquakes per year
- # server hits per second
- # of emails per day

Time to show intuition for why expectation == variance!
Properties of $\text{Poi}(\lambda)$ with the Poisson paradigm

Recall the Binomial:

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

<table>
<thead>
<tr>
<th>Expectation</th>
<th>$E[Y] = np$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>$\text{Var}(Y) = np(1 - p)$</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Expectation</th>
<th>$E[X] = \lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>$\text{Var}(X) = \lambda$</td>
</tr>
</tbody>
</table>

Proof:

$$E[X] = np = \lambda$$
$$\text{Var}(X) = np(1 - p) \to \lambda(1 - 0) = \lambda$$
A Real License Plate Seen at Stanford

No, it’s not mine... but I kind of wish it was.
Poisson Paradigm, part 2

Poisson can still provide a good approximation of the Binomial, even when assumptions are “mildly” violated.

You can apply the Poisson approximation when:

- "Successes” in trials are not entirely independent e.g.: # entries in each bucket in large hash table.
- Probability of “Success” in each trial varies (slightly), like a small relative change in a very small p e.g.: Average # requests to web server/sec may fluctuate slightly due to load on network

We won’t explore this too much, but I want you to know it exists.
Other Discrete RVs
## Grid of random variables

<table>
<thead>
<tr>
<th>Number of successes</th>
<th>Time until success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ber($p$)</td>
<td>One success</td>
</tr>
<tr>
<td>Bin($n, p$)</td>
<td>Several successes</td>
</tr>
<tr>
<td>Poi($\lambda$)</td>
<td>Interval of time to first success</td>
</tr>
</tbody>
</table>

- **One trial**: $n = 1$

- **Interval of time**: (tomorrow)
Consider an experiment: independent trials of $\text{Ber}(p)$ random variables. A geometric random variable $X$ is the number of trials until the first success.

**Geometric RV**

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

- **PMF**
  \[ P(X = k) = (1 - p)^{k-1}p \]

- **Expectation**
  \[ E[X] = \frac{1}{p} \]

- **Variance**
  \[ \text{Var}(X) = \frac{1-p}{p^2} \]

**Examples:**
- Flipping a coin ($P(\text{heads}) = p$) until first heads appears
- Generate bits with $P(\text{bit} = 1) = p$ until first 1 generated

\[ \text{Ber}(\frac{1}{2}) \Rightarrow \text{geo}(p = \frac{1}{2}) \]

\[ E[X] = 2 \]
Consider an experiment: independent trials of $\text{Ber}(p)$ random variables.  

**Def** A **Negative Binomial** random variable $X$ is the # of trials until $r$ successes.

\[ X \sim \text{NegBin}(r, p) \]

Support: \{r, r + 1, ... \}  

**PMF** \[ P(X = k) = \binom{k - 1}{r - 1} (1 - p)^{k-r} p^r \]

**Expectation** \[ E[X] = \frac{r}{p} \]

**Variance** \[ \text{Var}(X) = \frac{r(1-p)}{p^2} \]

**Examples:**  
- Flipping a coin until $r^{th}$ heads appears  
- # of strings to hash into table until bucket 1 has $r$ entries

\[ \text{Geo}(p) = \text{NegBin}(1, p) \]
Grid of random variables

Number of successes

<table>
<thead>
<tr>
<th>One trial</th>
<th>Ber($p$)</th>
<th>Bin($n, p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Several trials</td>
<td>Geo($p$)</td>
<td>NegBin($r, p$)</td>
</tr>
</tbody>
</table>

Time until success

<table>
<thead>
<tr>
<th>Interval of time</th>
<th>Poi($\lambda$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(tomorrow)</td>
<td></td>
</tr>
</tbody>
</table>

One success

Several successes

Interval of time to first success

Poisson distribution

Binomial distribution

Negative binomial distribution
Catching Pokemon

Wild Pokemon are captured by throwing Pokeballs at them.
- Each ball has probability $p = 0.1$ of capturing the Pokemon.
- Each ball is an independent trial.

What is the probability that you catch the Pokemon on the 5th try?

1. Define events/ RVs & state goal
   \[ X \sim \text{some distribution} \]
   Want: $P(X = 5)$

2. Solve
   A. $X \sim \text{Bin}(5, 0.1)$
   B. $X \sim \text{Poi}(0.5)$
   C. $X \sim \text{NegBin}(5, 0.1)$
   D. $X \sim \text{NegBin}(1, 0.1)$
   E. $X \sim \text{Geo}(0.1)$
   F. None/other
Catching Pokemon

Wild Pokemon are captured by throwing Pokeballs at them.
• Each ball has probability $p = 0.1$ of capturing the Pokemon.
• Each ball is an independent trial.

What is the probability that you catch the Pokemon on the 5th try?

1. Define events/ RVs & state goal

   $X \sim $ some distribution

   Want: $P(X = 5)$

2. Solve

   A. $X \sim $ Bin(5, 0.1)
   B. $X \sim $ Poi(0.5)  # must be r: #successes
   C. $X \sim $ NegBin(5, 0.1)
   D. $X \sim $ NegBin(1, 0.1)
   E. $X \sim $ Geo(0.1)
   F. None/other
Catching Pokemon

Wild Pokemon are captured by throwing Pokeballs at them.

- Each ball has probability $p = 0.1$ of capturing the Pokemon.
- Each ball is an independent trial.

What is the probability that you catch the Pokemon on the 5th try?

1. Define events/RVs & state goal

$x \sim \text{Geo}(0.1)$

Want: $P(X = 5)$

$P(X = 5) = (1-p)^{k-1}p$

$= (0.9)^4 \cdot 0.1$

$\approx 0.066$