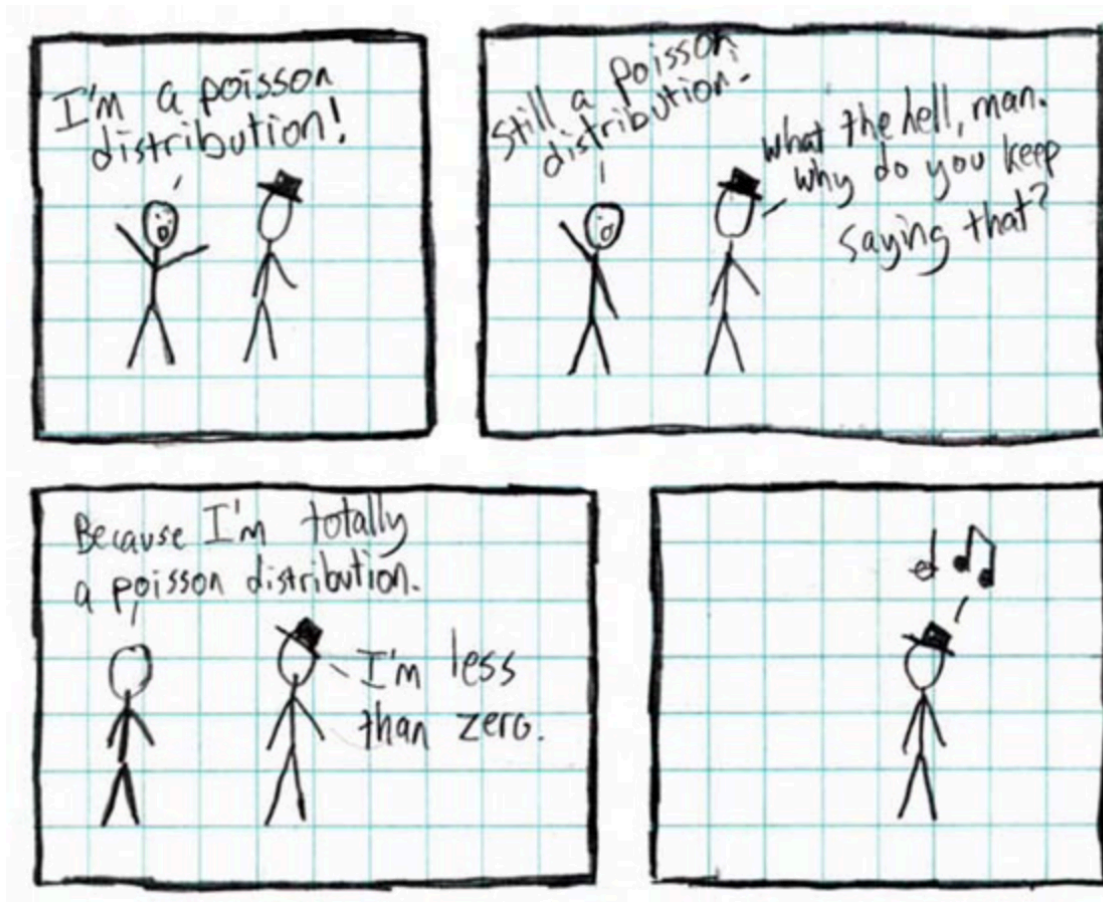


Stat 88: Probability & Mathematical Statistics in Data Science



<https://xkcd.com/12/>

Lecture 11: 2/10/2021

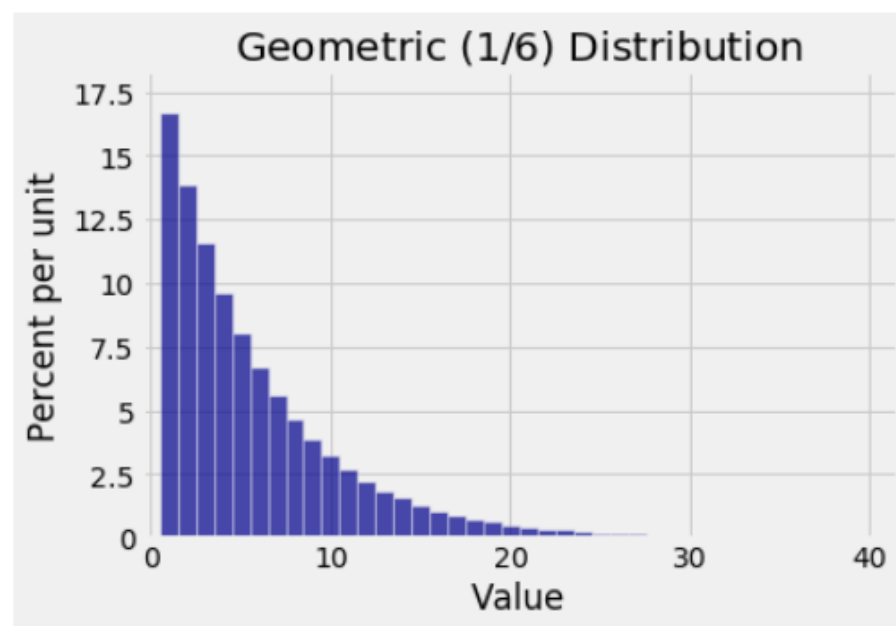
Waiting times, exponential approximations, the Poisson distribution

Sections 4.3, 4.4 (and finish 4.2)

Geometric distribution

- Recall that T_1 has the **geometric distribution**, denoted $T_1 \sim \text{Geom}(p)$ on $\{1, 2, 3, \dots\}$, when we have $k-1$ failures, and then first success is on k^{th} trial.
- $f(k) = P(T_1 = k) = P(FFF \dots FS) =$
- $F(k) = P(T_1 \leq k) = 1 - P(T_1 > k) =$

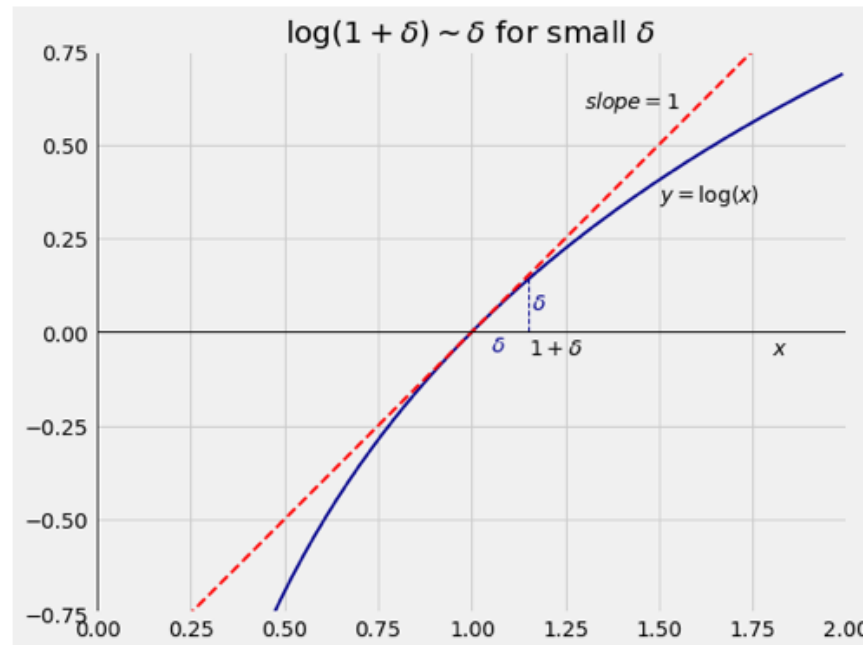
Roll a die until *first* ace (1 spot):



Waiting time until r^{th} success

- Say we roll a 8 sided die.
- What is the chance that the first time we roll an eight is on the 11th try?
 $= P(\text{FFFFFFFFFS}) =$
- What is the chance that it takes us 15 times until the 4th time we roll eight?
(That is, the waiting time until the 4th time we roll an eight is 15)
 $= P(\text{_____S})$
- What is the chance that we need **more** than 15 rolls to roll an eight 4 times?
- Notice that the **right-tail** probability of T_4 is a left hand (cdf) of the Binomial distribution for $(15, 1/8)$, and where $k=3$.
- In general, $P(T_r = k) =$
- And $P(T_r > k) =$

4.3 Exponential Approximations



Very useful approximation: $\log(1 + \delta) \approx \delta$, for δ close to 0

How to use this approximation

- Approximate the value of $x = \left(1 - \frac{3}{100}\right)^{100}$
- $x = \left(1 - \frac{2}{1000}\right)^{5000}$
- $x = (1 - p)^n$, for large n and small p

Example

- A book chapter $n = 100,000$ words and the chance that a word in the chapter has a typo (independently of all other words) is very small : $p = 1/1,000,000 = 10^{-6}$. Give an approximation of the chance the chapter *doesn't* have a typo. (Note: A typo is a *rare event*)

Bootstraps and probabilities

- Bootstrap sample: sample of size n drawn with replacement from original sample of n individuals
- Suppose one particular individual in the original sample is called Ali. What is the probability that Ali is chosen at least once in the bootstrap sample?
- Use the complement.

The Poisson Distribution

- Used to model rare events. X is the number of times a rare event occurs, $X = 0, 1, 2, \dots$

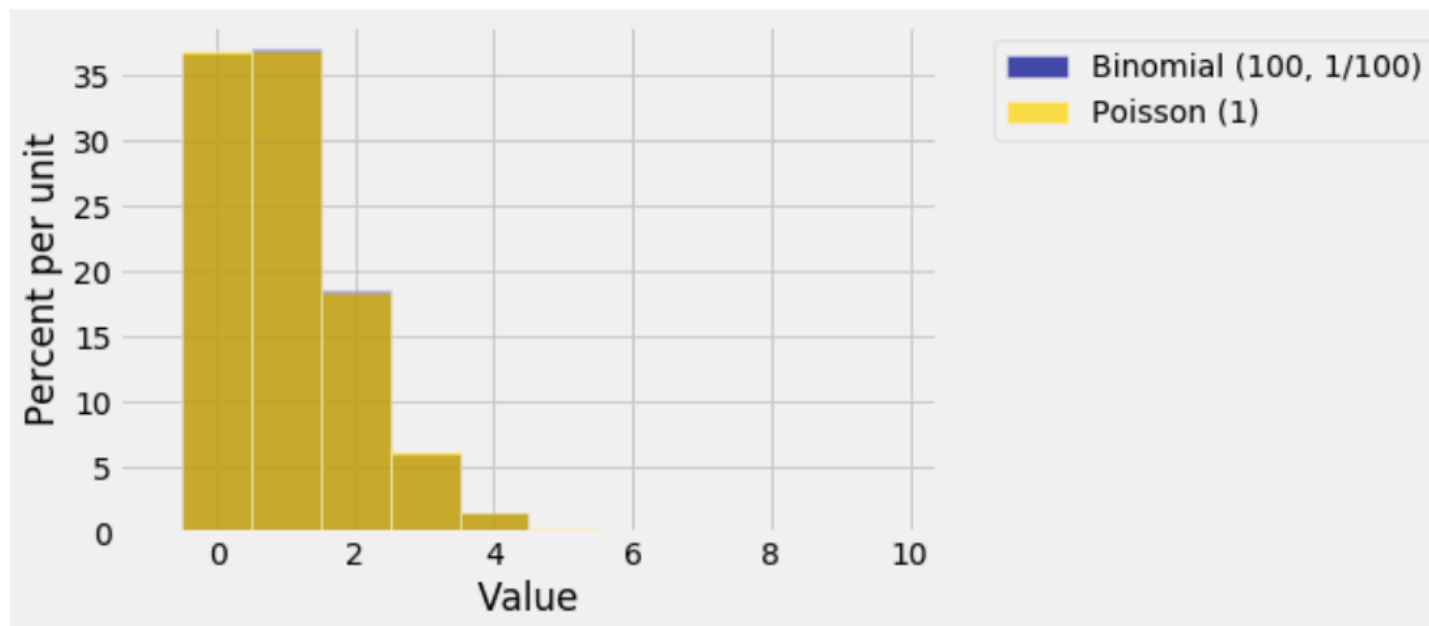
- We say that a random variable X has the Poisson distribution if

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

- The parameter of the distribution is μ

Relationship between Poisson and Binomial distributions

- **The Law of Small Numbers:** when n is large and p is small, the binomial (n,p) distribution is *well approximated* by the Poisson(μ) distribution where $\mu=np$.



Exercise 4.5.7

A book has 20 chapters. In each chapter the number of misprints has the Poisson distribution with parameter 2, independently of the misprints in other chapters.

- a) Find the chance that Chapter 1 has more than two misprints.
- b) Find the chance that the book has no misprints.
- c) Find the chance that two of the chapters have three misprints each.

Sums of independent Poisson random variables

- If X and Y are random variables such that
- X and Y are independent,
- X has the Poisson(μ) distribution, and
- Y has the Poisson(λ) distribution,
- then the sum $S=X+Y$ has the Poisson ($\mu+\lambda$) distribution.

Exercise 4.5.8

In the first hour that a bank opens, the customers who enter are of **three** kinds: those who only require teller service, those who only want to use the ATM, and those who only require special services (neither the tellers nor the ATM). Assume that the numbers of customers of the three kinds are independent of each other, and also that:

- the number that only require teller service has the Poisson (6) distribution,
- the number that only want to use the ATM has the Poisson (2) distribution, and
- the number that only require special services has the Poisson (1) distribution.

Suppose you observe the bank in the first hour that it opens. In each part below, find the chance of the event described.

- 12 customers enter the bank
- more than 12 customers enter the bank
- customers do enter but none requires special services