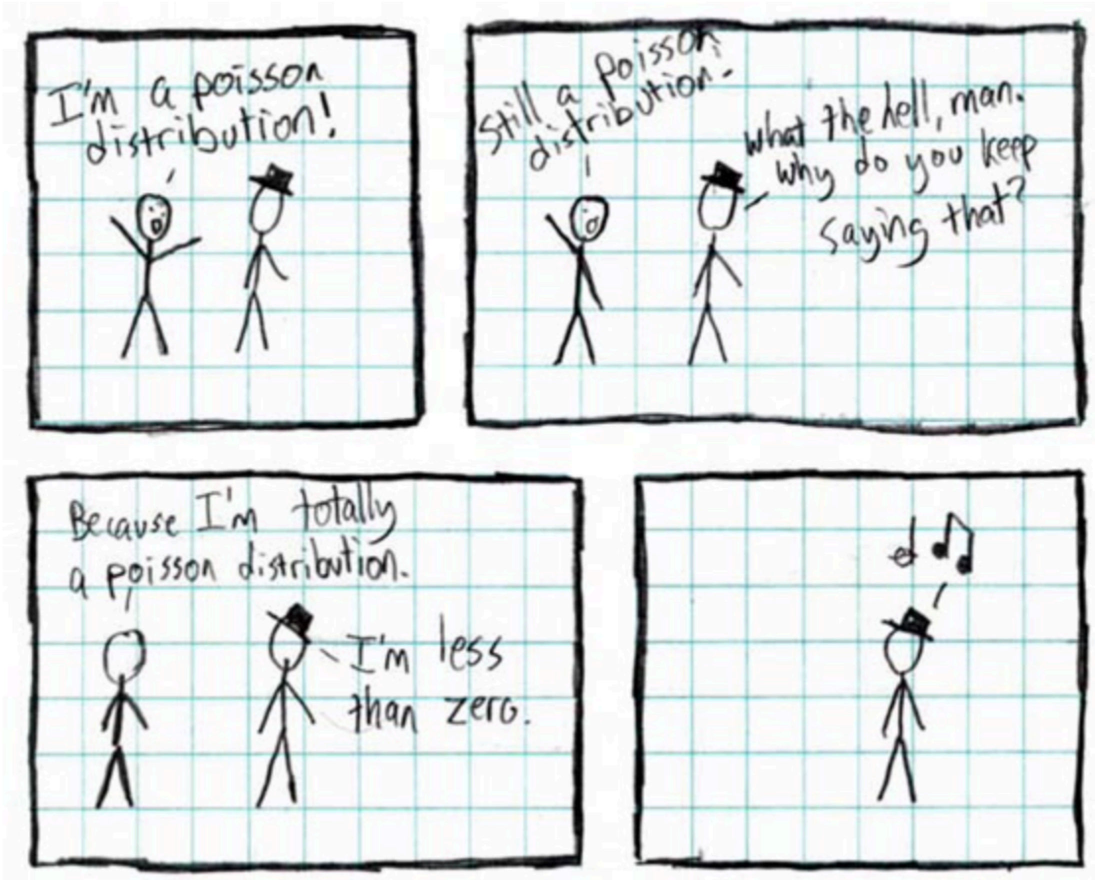
Stat 88: Probability & Math. Stat in Data Science



https://xkcd.com/12/

Lecture 15: 2/21/2024

Rare events, the Poisson distribution, Expected value

4.2, 4.3, 4.4

Shobhana Stoyanov

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 that 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, $X = 0, 1, 2, \ldots$

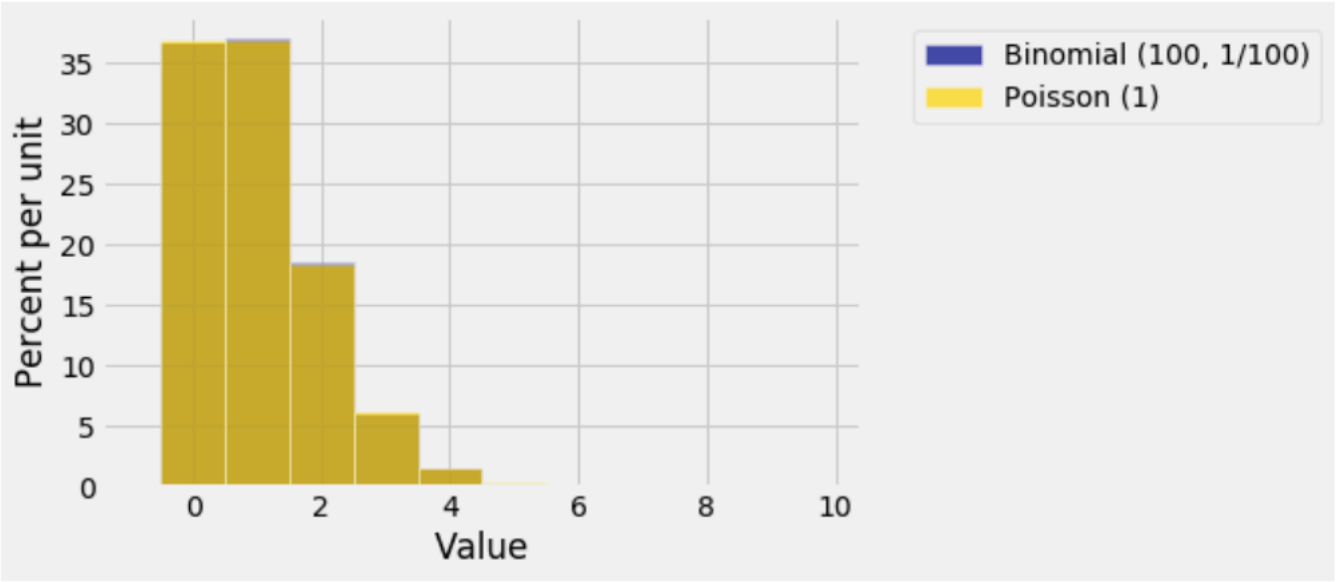
- We say that a random variable X has the ***Poisson*** distribution if

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

- 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.

a) 12 customers enter the bank

b) more than 12 customers enter the bank

c) customers do enter but none requires special services

5.1: Expected Value of a random variable

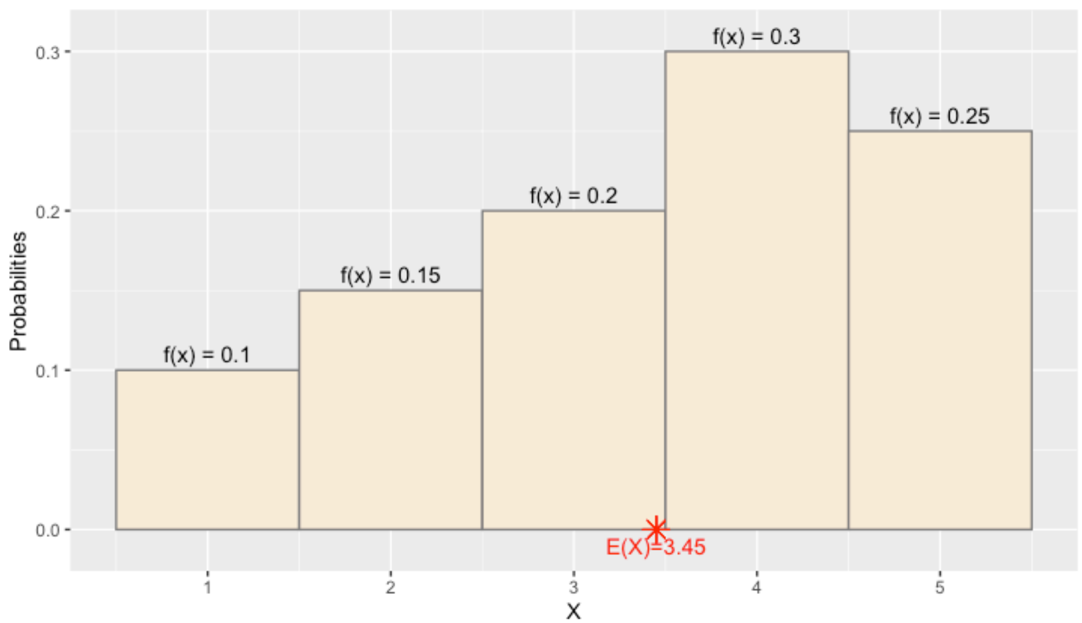
- The ***Expectation*** or ***Expected Value*** of a random variable X is defined to be the sum of all the products $(\boldsymbol{x} \times \boldsymbol{f}(\boldsymbol{x}))$ over all possible values x of the random variable X: that is, the expectation of a random variable is a ***weighted average*** of all the possible values that the rv can take, weighted by the probability of each value.
- $\boldsymbol{E(X) = \sum x \bullet P(X = x) = \sum x \cdot f(x)}$

- For example, toss a coin 3 times, let X = # of heads. Write down $f(x)$, and then use the formula to compute the expectation of X.

- Note: If X takes finitely many values, no problem. If it takes infinitely (countable, but not finite) many values (such as Poisson, or Geometric), then we have to be more careful when we take the sums.

Example (from text)

x	1	2	3	4	5
f(x)	0.1	0.15	0.2	0.3	0.25



Notes about E(X)

- Same units as X
- Not necessarily an attainable value (for example, in In 2020, there was an average of 1.93 children under 18 per family in the United States)
- Expectation is a long-run average value of X
- Center of mass or center of gravity (balancing point) for the distribution
- Expectation of a constant is that constant $E(c) = c$
- Linearity: $E(aX + b) = aE(X) + b$
- Example: for the X defined in the previous slide, find E(2X-1)

Special examples

- Bernoulli (Indicators)

- Uniform

- Poisson