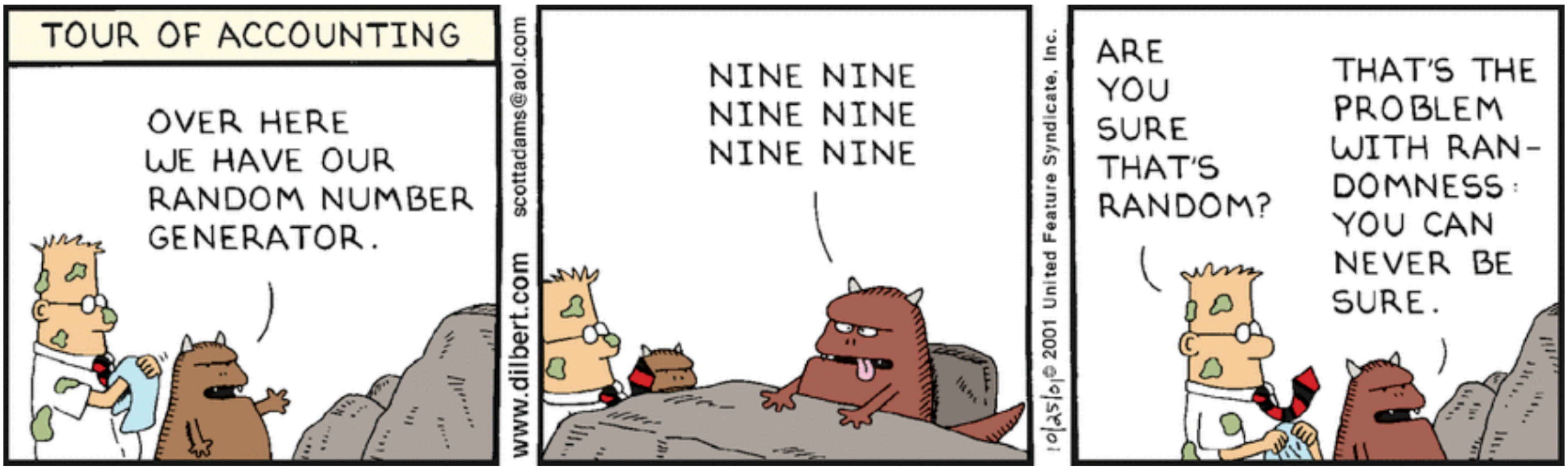
Stat 88: Probability & Mathematical Statistics in Data Science



Lecture 10: 2/10/2021

Waiting times, exponential approximations

Sections 4.2, 4.3

Agenda

- Warm up with a cdf problem
- 4.2 Waiting times
- 4.3 Exponential approximations

Cdf and Exercise 4.5.2

- Cumulative distribution function (**cdf = F(x)**) of a random variable X is another way of describing the distribution of the probability.
- $F(x) = P(X \leq x)$
- $f(x) = P(X = x) = P(X \leq x) - P(X \leq x - 1) = F(x) - F(x - 1)$

- A random variable W has the distribution shown in the table below. Sketch a graph of the cdf of W.

w	-2	-1	0	1	3
$P(W = w)$	0.1	0.3	0.25	0.2	0.15

Cdf and pmf

- A random variable W has the distribution shown in the table below. Sketch a graph of the pmf of W, and shade in F(1)

w	-2	-1	0	1	3
$P(W = w)$	0.1	0.3	0.25	0.2	0.15

Cdf is very useful because we often need sums of probabilities.

- Draw 12 balls w/o repl. from a box with 10 red and 15 blue balls.

P(at least 5 red balls in sample)?

4.2: Waiting times

- Say Ali keeps playing roulette, and betting on red each time. The ***waiting time*** of a red win is the number of spins until they see a red (so the number of spins until and including the time the ball lands on a red pocket).

- What is the probability that Ali will wait for 4 spins before their first win? (That is, the first time the ball lands in red is the 4th spin or trial)

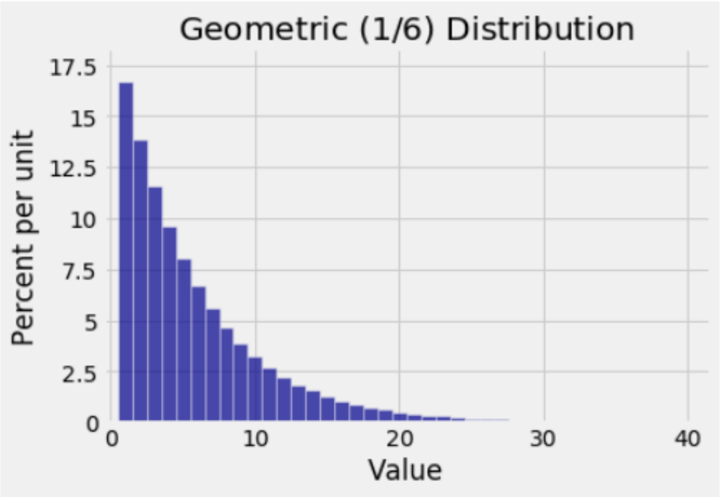
- Say we have a sequence of ***independent*** trials (roulette spins, coin tosses, die rolls etc) each of which has outcomes of success or failure, and $\boldsymbol{P(S) = p}$ on each trial.
- Let T_1 be the number of trials up to and including the first success. Then T_1 is the ***waiting time until the first success***.
- What are the values T_1 takes? What is its pmf $f(x)$?

Geometric distribution

- Say T_1 has the **geometric distribution**, denoted $\boldsymbol{T_1 \sim Geom(p)}$ on {1, 2, 3,…}

- $f(k) \ = \ P(T_1 = k) \ =$

- Check that it sums to 1. What is the cdf for this distribution? Can you think of an easy way to write down the cdf?



Waiting time until r^{th} success

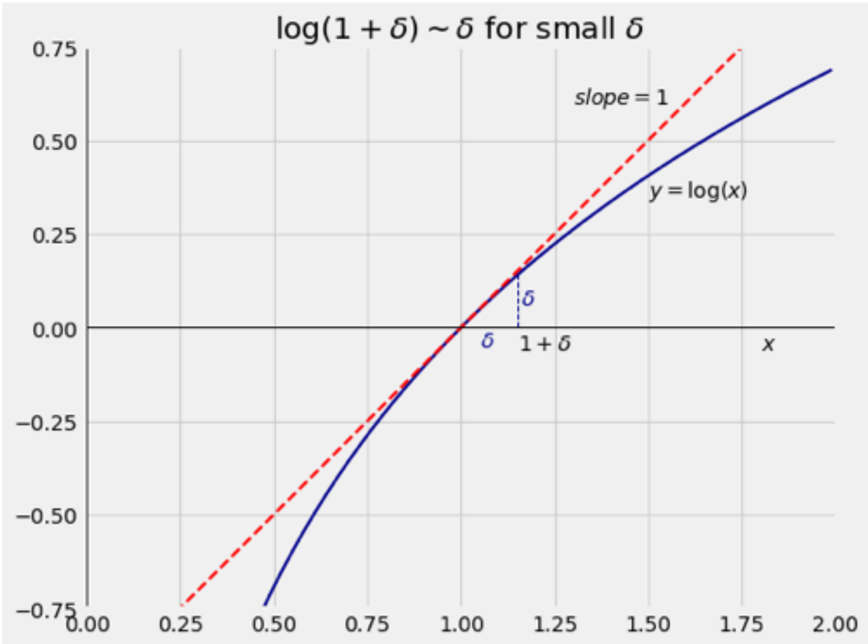
- Say we roll a 8 sided die.
- What is the chance that the first time we roll an eight is on the 11^{th} try?

- What is the chance that it takes us 15 times until the 4^{th} time we roll eight? (That is, the waiting time until the 4^{th} time we roll an eight is 15)

- 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.

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. (Niote: 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, …

- 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