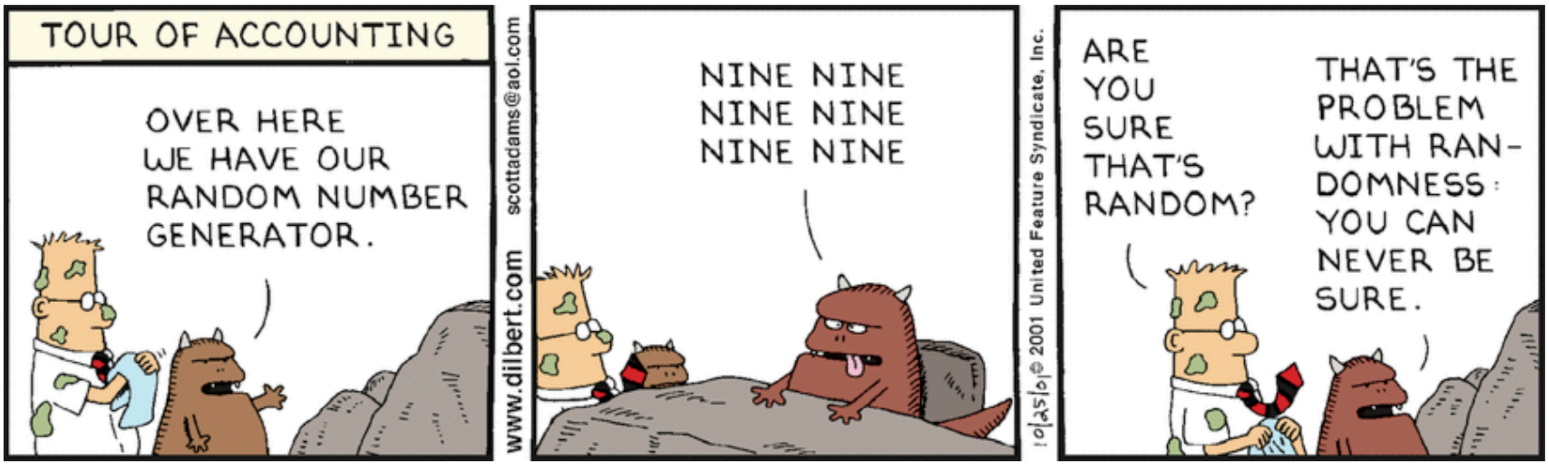
Stat 88: Probability & Mathematical Statistics in Data Science



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

Different ways of writing out distributions

X = # of H in 3 tosses of a fair coin

x	0	1	2	3
$f(x)=P(X=x)$	1/8	3/8	3/8	1/8
$F(x)=P(X\le x)$	1/8	4/8	7/8	8/8=1

w.p ⟺ "with probability"

← $F(x)$ defined for every real #

$$X=\begin{cases} 0 & \text{w.p. } 1/8 \\ 1 & \text{w.p. } 3/8 \\ 2 & \text{w.p } 3/8 \\ 3 & \text{w.p } 1/8 \end{cases}$$

$$f(x)=\begin{cases} 1/8, & x=0,3 \\ 3/8, & x=1,2 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x)=\begin{cases} 0, & x<0 \\ 1/8, & 0\le x<1 \\ 4/8, & 1\le x<2 \\ 7/8, & 2\le x<3 \\ 1, & 3\le x \end{cases}$$

Cdf and Exercise 4.5.2

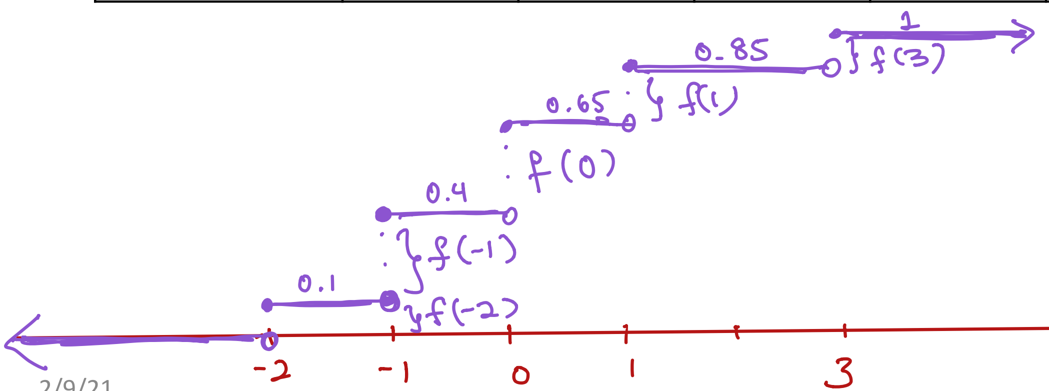
- Cumulative distribution function (**cdf = F(x)**) of a random variable X is another way of describing the distribution of the probability.
- $F(x) = P(X \le x)$
- $f(x) = P(X = x) = P(X \le x) - P(X \le x - 1) = F(x) - F(x-1)$

- 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

$[-2, -1)$

$$F(x) = \begin{cases} 0, & x < -2 \\ 0.1, & -2 \le x < -1 \\ 0.4, & -1 \le x < 0 \\ 0.65, & 0 \le x < 1 \\ 0.85, & 1 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$



Cdf and pmf

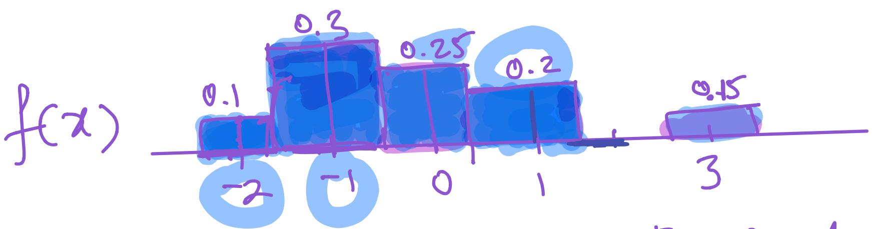
prob MASS function

- 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

$F(1) = P(X \le 1) = \sum_{x \le 1} f(x)$

area of bar = prob.



$F(2) = F(1)$

$F(2) = \sum_{x \le 2} f(x)$

$= 0.1 + 0.3 + 0.25$
$+ 0.2 + 0$
$= 0.85 = F(1$

$F(27) = 1$

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)?

$X \sim \text{Hypergeom}(N, G, n)$

$HG(25, 10, 12)$

$P(X \ge 5) = \sum^{10} \frac{\binom{10}{k}\binom{15}{12-k}}{\binom{25}{}} = 1 - P(X < 5)$

F(Jeff Bezos total wealth) = 1

F(# of stars in known universe) = 1

$P(X \leq 5) = F(5)$ $K=5$ (12) $= 1 - P(X \leq 4)$ $= 1 - F(4) = 1 - \text{stats.hypergeom.cdf}(4, 25, 10, 12)$

4.2: Waiting times

38

- 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).

$P(S)$ $P(R) = \frac{18}{38}$

$P(F)$ $P(NR) = \frac{20}{38}$

- 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)

X = # of spins up to & including 1st R.

$$P(X=4) = P(FFFS) = \left(\frac{20}{38}\right)^3 \cdot \left(\frac{18}{38}\right)$$

- 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 $f(x)$?

$P(S) = p$

$P(F) = 1-p$

$F F \ldots \ldots F \underset{k^{th}}{S}$

$$T_1 = 1, 2, 3, 4, \ldots$$

$$f(k) = P(T_1 = k) = (1-p)^{k-1} \cdot p$$

Geometric series $\rightarrow$ $1-p=q$, $\sum_{k=1}^{\infty} q^{k-1} = \frac{1}{1-q} = \frac{1}{p}$

Geometric distribution

$\sum_{k=1}^{\infty} f(k) = p \cdot \frac{1}{1-q} = p \cdot \frac{1}{p} = 1.$

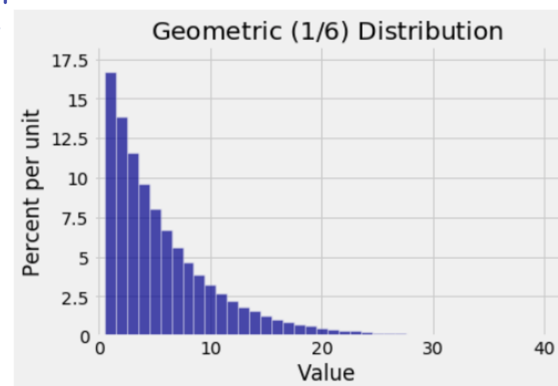
- Say T_1 has the **geometric distribution**, denoted $\boldsymbol{T_1 \sim Geom(p)}$ on $\{1, 2, 3, \ldots\}$
- $f(k) = P(T_1 = k) =$ $(1-p)^{k-1} \cdot p$

$= \sum_{k=1}^{\infty} f(k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = \underbrace{p \sum_{k=1}^{\infty} q^{k-1}}_{\text{pull out } p}$

- 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?

$F(x) = P(X \leq x) = \sum_{k=1}^{x} (1-p)^{k-1} p$

$= \sum_{k \leq 1}^{x} f(x)$



$P(T_1 \leq x) = 1 - \underbrace{P(T_1 > x)}_{} = 1 - q^x$

$P(T_1 > x)$

$P(T_1 > 5) = q^5 \leftarrow$ prob of first 5 trials being Failures.

Waiting time until r^{th} success

$\underbrace{- - - - - - - \; . \; . \; . \; .}_{} \underset{\uparrow 17}{S}$

Back to roulette & Ali

Say Ali spins until their 5^{th} win

$P(T_5 = 17) = \binom{16}{4}\left(\frac{18}{38}\right)^4\left(\frac{20}{38}\right)^{12} \cdot \left(\frac{18}{38}\right)$

(T_5 waiting time until 5^{th} win)

- 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.