Stat 88 lec 9

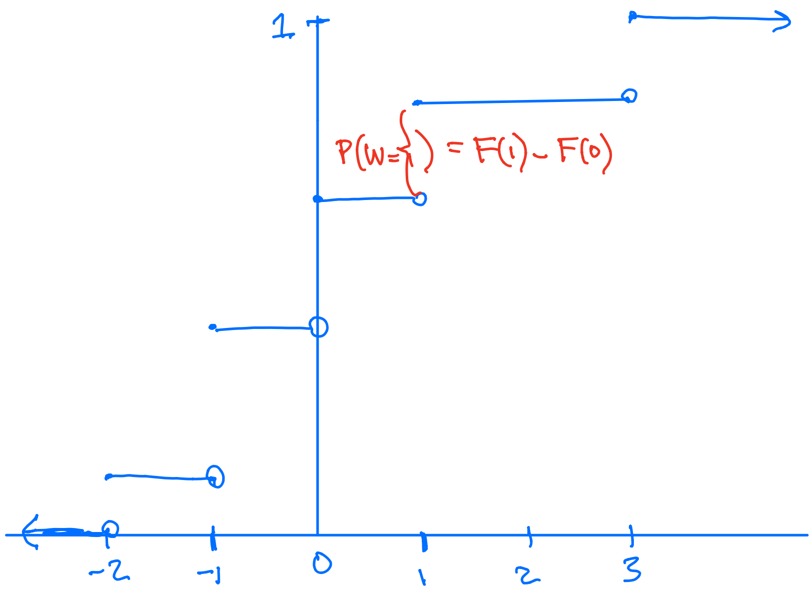
warm up 2:00–2:10

exercise 4.5.2

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

$F(w) = P(W \le w)$

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



Last time SLC drop in M–Th 11–5 pm

Sec 4.1 Cumulative distribution function (CDF)

The cdf of a RV X is $F(a) = P(X \le a)$

Purpose The cdf is an alternative way to specify a distribution.

ex given the graph of the cdf, in the warmup can you figure out what is $P(W=1)$

$F(1) - F(0)$

more generally $P(W=a) = F(a) - F(a-1)$

Use Solutions to many problems can be expressed in terms of CDF and Python has built in cdf function

ex Fisher Exact test result.

$$P(X \ge 50) = \sum_{g=50}^{60} \frac{\binom{80}{g}\binom{20}{60-g}}{\binom{100}{60}}$$

$$= 1 - P(X < 50)$$

$$= 1 - P(X \le 49) = 1 - F(49)$$

```
In [5]: from scipy import stats
        import numpy as np
```

```
In [15]: 1-stats.hypergeom.cdf(49, 100, 80, 60)
```

x, N, G, n

```
Out[15]: 0.22097998866696655
```

```
In [14]: sum(stats.hypergeom.pmf(np.arange(50,61), 100, 80, 60))
```

```
Out[14]: 0.22097998866696314
```

Today ① Sec 4.2 Waiting times

② Sec 4.3 Exponential Approximations.

tinyurl.com / feb7 -pt1

Stats 88

Friday February 7 2020

1. In a population, 30% of the individuals are green and the rest are blue. Suppose you draw individuals **with replacement** until you draw a blue. Is the binomial formula applicable to find the chance that you draw 10 times?

   binomial needs a fixed number of trials.

   a yes

   b no

   ggggggggg b $\quad (.3)^9 (.7)$

   waiting time until first success is the geometric distribution (sec 4.2).

① Sec 4.2 Waiting Times.

Waiting Time to the first success:

Consider a sequence of independent and identically distributed (iid) trials, each of which results in a success or a failure. Let p be the chance of success and q the chance of failure $(q = 1-p)$.

Let T_1 = # trials until the first success.

T_1 belongs to the geometric distribution

$T_1 \sim Geom(p)$

What is $P(T_1 = k) = ?$ $\quad q^{k-1}p$

What values does T_1 take? $\quad 1, 2, 3, \ldots$

What is the chance it takes at most 5 trials for 1st success?

Geometric sum from algebra

$$P(T_1 \le 5) = p + qp + q^2p + q^3p + q^4p = \cancel{p}(1 + q + q^2 + q^3 + q^4) = 1 - q^5$$

$$1 + q + q^2 + q^3 + q^4 = \frac{1-q^5}{1-q}, \quad 1-q = \cancel{p}$$

formula for a finite geometric sum.

$$P(T_1 \le 5) = 1 - P(T_1 > 5) = 1 - q^5, \quad P(T_1 > 5) = q^5$$

Note $F(5) = P(T_1 \le 5) = 1 - q^5$

CDF for $Geom(p)$? ⟶ $F(x) = 1 - q^x$ for $x = 1, 2, 3, \ldots$

ex

Cards are dealt one by one at random with replacement till the first ace appears. Let X be the number of cards dealt.

$X \sim \text{Geom}(1/13)$, with $p = 1/13$

a) Find $P(X = 39)$. $\quad \left(\frac{12}{13}\right)^{38}\left(\frac{1}{13}\right)$

b) Find $P(X > 20)$. $\quad q^{20} = \left(\frac{12}{13}\right)^{20}$.

Waiting time till the r^{th} success:

Lets do a related problem!

Cards are dealt one by one at random with replacement till the fourth ace appears. Let X be the number of cards dealt.

a) Find $P(X = 39)$. $= P(\underline{3 \text{ aces out of } 38} \text{ and } \underline{39^{th} \text{ card is ace}})$

$$= \binom{38}{3}\left(\frac{1}{13}\right)^3\left(\frac{12}{13}\right)^{35} \cdot \frac{1}{13} = \binom{38}{3}\left(\frac{1}{13}\right)^4\left(\frac{12}{13}\right)^{35}$$

b) Find $P(X > 20)$.

$$= P(\text{fewer than 4 aces in 20 cards})$$

$$= \sum_{k=0}^{3}\binom{20}{k}\left(\frac{1}{13}\right)^k\left(\frac{12}{13}\right)^{20-k}$$

In Python this is

```
stats.binom.cdf(3, 20, 1/13)
```

Picture

$$\underbrace{\overset{1}{\_}\ \overset{2}{\_}\ \overset{3}{\_}\ \cdots\ \overset{38}{\_}}_{\text{3 aces out of 38 cards}}\ \overset{39}{\underline{S}}$$

$\uparrow$ 4^{th} ace.

One more variation:

exercise 4.5.5

5. Cards are dealt one by one at random without replacement till the fourth ace appears. Let X be the number of cards dealt.

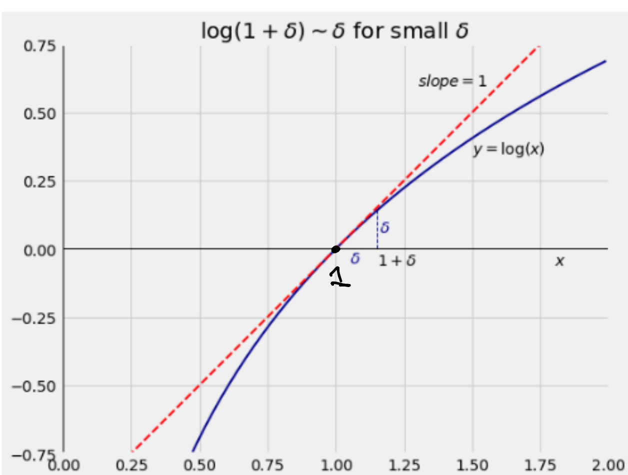
a) Find $P(X = 39)$. $\dfrac{\binom{4}{3}\binom{48}{35}}{\binom{52}{38}} \cdot \dfrac{1}{14}$

b) Find $P(X > 20)$. $\displaystyle\sum_{g=0}^{3} \dfrac{\binom{4}{g}\binom{48}{20-g}}{\binom{52}{20}}$

② Sec 4.3 <u>Exponential approximations</u>

A useful approximation from Calculus

$\log(1+\delta) \approx \delta$ for small δ



$f(x) = \log(x)$ is locally flat at $x=1$ with slope 1

Since $f'(x) = \frac{1}{x}$ so $f'(1) = 1$,

So starting at $x=1$ if run by δ you rise by δ.

So $\log(1+\delta) \approx \delta$.