STAT 88: Lecture 9

Contents

Last time

Sec 4.1 Cummulative distribution function (CDF):

The CDF of a random variable X is $F(x) = P(X \leq x)$.

Purpose The CDF is an alternative way to specify a distribution:

$$P(X = x) = P(X \leq x) - P(X \leq x - 1) = F(x) - F(x - 1).$$

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

Warm up: (Exercise 4.5.2) A random variable W has the distribution shown in the table below. Sketch a graph of the cdf of W.

$P(W \leq w)$

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

Computation You can use the stats module of SciPy to calculate CDF.

```
from scipy import stats
import numpy as np
```

```
1 - stats.hypergeom.cdf(49, 100, 80, 60)
```

x N G n

```
0.22097998866696655
```

```
sum(stats.hypergeom.pmf(np.arange(50,61), 100, 80, 60))
```

```
0.22097998866696314
```

$$X \sim HG(100, 80, 60)$$

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

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

$$= 1 - P(X \leq 49)$$

$$= 1 - F(49)$$

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 follows a distribution called "Geometric" distribution,

$$T_1 \sim \text{Geom}(p).$$

What is $P(T_1 = k)$?

What values does T_1 take?

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

CDF for Geom(p)?

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

(a) Find $P(X = 39)$.

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

Waiting time till the rth success: 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 $P(X = 39)$.

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

Example: (Exercise 4.5.5) Cards are dealt one by one at random without replacement till the fourth ace appears. Let X be the number of cards dealt.

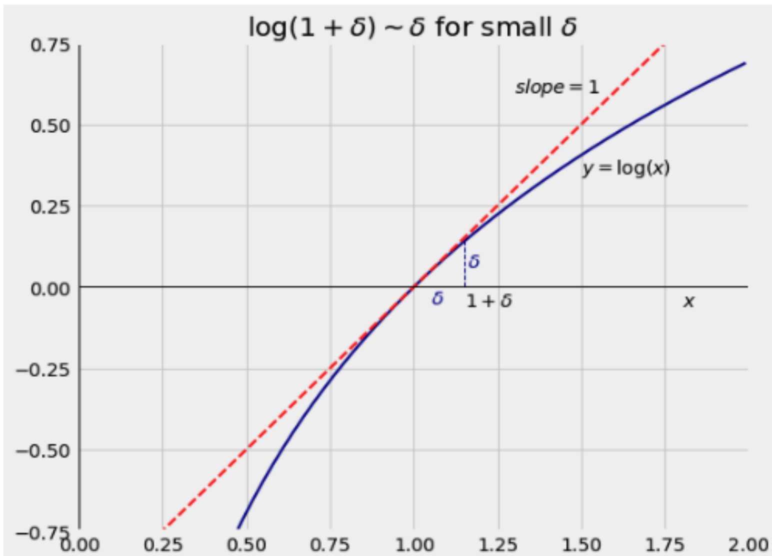
(a) Find $P(X = 39)$.

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

4.3. Exponential Approximations

A useful approximation from Calculus:

$$\log(1+\delta) \approx \quad \text{for small } \delta.$$



$f(x) = \log x$ 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 $\log(1+\delta) \approx \delta$.

Example: Approximate $x = \left(1 - \frac{3}{100}\right)^{100}$.

$$\log x = \log\left(1 - \frac{3}{100}\right)^{100} = 100 \cdot \log\left(1 - \frac{3}{100}\right) \approx 100\left(-\frac{3}{100}\right).$$

So $x \approx e^{-3}$.

Give exponential approximation for

(a) $x = \left(1 - \frac{2}{1000}\right)^{5000}$.

(b) $(1-p)^n$ for large n and small p.