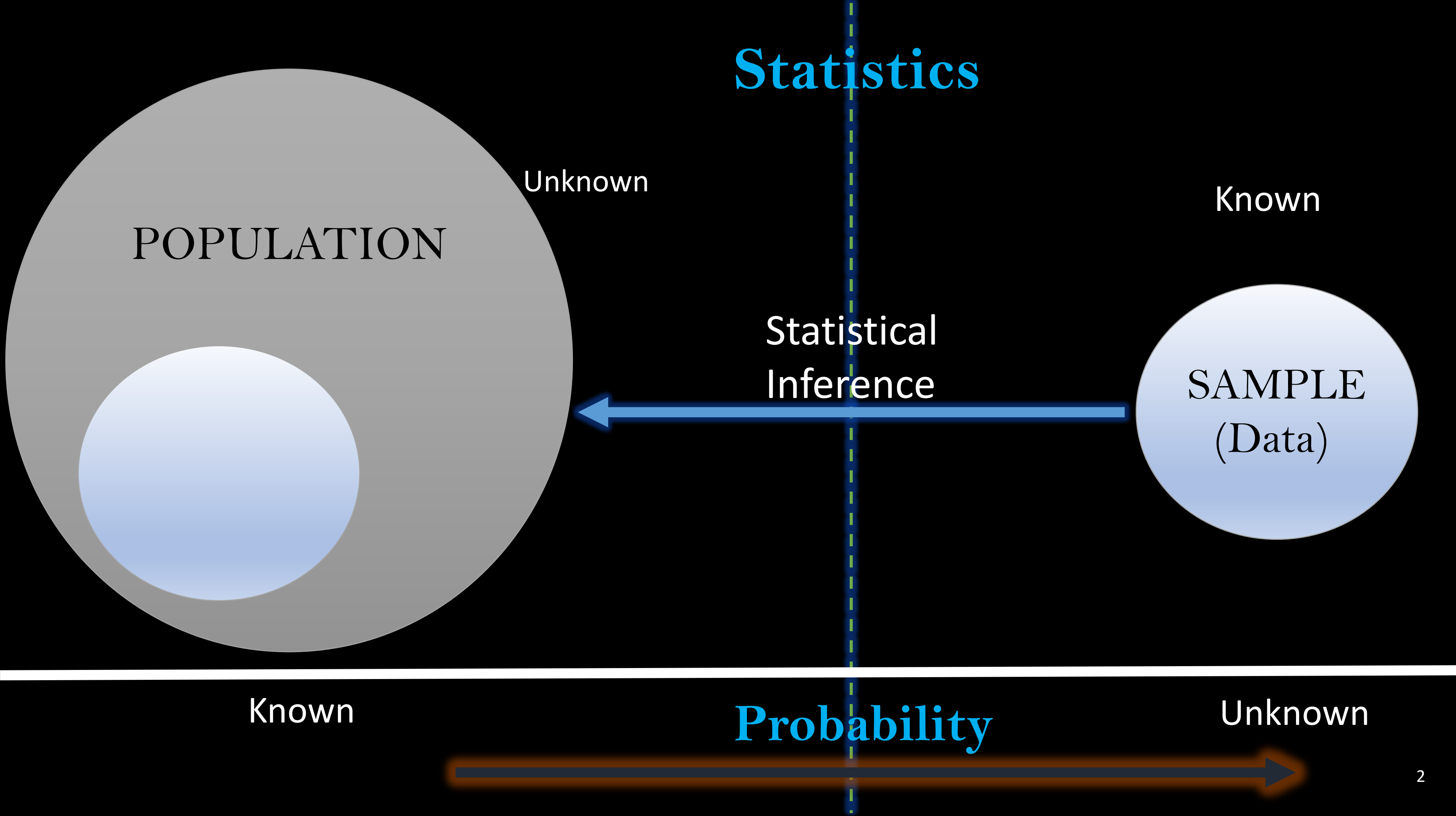
2.3 Variance and Standard Deviation


Statistics
Unknown
POPULATION
Known
Statistical
Inference
SAMPLE
(Data)
Known
Probability
Unknown

Notation review:

f(x): pmf

F(x): cdf

X: Random Variable

x: value that the random variable takes

Note: the following applies to (discrete) pmfs: *Assuming S = {1,2,3,...}*

$$\boxed{f(x) = P[X = x] = P[X \leq x] - P[X \leq (x-1)]}$$

Variance

One way to characterize a random variable is by its location (**mean, median**).

Another way is to describe how spread out it is (**variance**).

For a random variable, X, Variance can be written:

$$Var[X], \quad \sigma^2{}_{\mathrm{X}}, \quad or \quad \sigma^2$$

Variance

$$\boxed{\sigma^2 = E[(X-\mu)^2] = \sum_{all\ x}(x-\mu)^2 f(x)}$$

Also,

$$\begin{aligned}\sigma^2 = E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2.\end{aligned}$$

, or $\boxed{\sigma^2 = E[X^2] - (E[X])^2}$

Linear Transformation of a Random Variable – Basic Properties

$$\boxed{E[aX + b] = a \cdot E[X] + b}$$

$remember\colon \sigma^2 = E[(X-\mu)^2]$

Linear Transformation of a Random Variable – Basic Properties

$$\boxed{Var[aX + b] = a^2 \cdot Var[X]}$$

$$\boxed{SD[aX + b] = |a| \cdot SD[X]}$$

1. A pocket contains 5 billiard balls numbered 1 to 5. Suppose you pull out two of them at random.

a) How many different subsets of 2 billiards are there in this pocket?

b) Let X be the larger of the two numbers drawn. What is the pmf of X?

c) What is the cdf of X?

d) What is E[X]?

e) What is Var[X]?

Outcome	X
1,2	2
1,3	3
1,4	4
1,5	5
2,3	3
2,4	4
2,5	5
3,4	4
3,5	5
4,5	5

d) What is E[X]?

e) What is Var[X]?

2. Suppose a fair die is tossed 3 times. Let X be the largest number that shows up.

a) Find an **expression** for F(x).

b) Find an expression for f(x).

3. A fair coin is tossed three times. Let X be $\# \ of \ heads - \# \ of \ tails$ in the three tosses.

a) What is the space of X?

b) What is the pmf of X?

c) Sketch the CDF of X.

d) What is $E[X]$?

e) What is $Var[X]$?

3. A fair coin is tossed three times. Let X be $\# \; of \; heads \; - \; \# \; of \; tails$ in the three tosses.

d) What is $E[X]$?

e) What is $Var[X]$?

4. Suppose $E[X] = 20,\ SD[X] = 2$

$Let\ Y = 3X + 1.$

$Let\ Z = 3 - X$

Find $E[Y]$ and $Var[Y]$.

Find $E[Z]$ and $SD[Z]$.

5. Using R, take a random sample of size n `= 500` from the following pmf. Calculate the empirical mean and variance. Compare with the true mean and variance.

x	f(x)
1	0.1
2	0.4
3	0.2
4	0.3

$E[X] = 1(0.1) + 2(0.4)+ 3(0.2) + 4(0.3) \quad = 2.7$

$E[X^2] = 1^2(0.1) + 2^2(0.4) + 3^2(0.2) + 4^2(0.3) = 8.3$

$Var[X] = \quad = 1.01$

```r
n = 500
mydata <- sample(x = c(1,2,3,4), n, replace = T, prob = c(0.1, 0.4, 0.2, 0.3))
windows()  #if using a mac, type quartz()
hist(mydata, breaks = 0:4, freq = F)
mean(mydata)
sd(mydata)
var(mydata)
```

Additional Examples: Hw2 Ex. 4

Consider a random variable X with the probability mass function:

$$f(x) = \frac{\frac{1}{3}}{(3/2)^x}, \quad x = 0, 1, 2, 3, \ldots$$

(1) Calculate $E[X]$.

Hint:

The summation of this infinite sequence is known as an **arithmetico–geometric series**, and its most basic form has been called **Gabriel's staircase**:[2][3][4]

$$\boxed{\sum_{k=1}^{\infty} kr^k = \frac{r}{(1-r)^2}, \quad \text{for } 0 < r < 1}$$

https://en.wikipedia.org/wiki/Arithmetico%E2%80%93geometric_sequence