Questions?

Probability Rules

Frequentist v Bayesian

Probability Mass Functions

P-values (homework)

From the Bernoulli to the Binomial

Bernoulli PMF

$$P(X = 1) = p$$

Binomial PMF

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

One Coin Flip

H

T

One Coin Flip

H	$$\binom{1}{1} = 1$$
T	$$\binom{1}{0} = 1$$

One Coin Flip

H	$\binom{1}{1} = 1$

Read as "one choose one"

One Coin Flip

H $$\binom{1}{1} = 1$$

$$P(X=1) = \binom{1}{1} 0.5^1 (1-0.5)^{1-1} = 1(0.5)1$$

One Coin Flip

Read as "one choose zero"

T $$\binom{1}{0} = 1$$

One Coin Flip

$$P(X=1) = \binom{1}{0} 0.5^0 (1-0.5)^{1-0} = 1(1)0.5$$

T $\binom{1}{0} = 1$

Two Coin Flips

HH

HT

TH

TT

Two Coin Flips

HH

$$\binom{2}{2} = 1$$

HT

TH

$$\binom{2}{1} = 2$$

TT

$$\binom{2}{0} = 1$$

Three Coin Flips

HHH
HHT
HTH
THH
HTT
THT
TTH
TTT

Three Coin Flips

HHH

HHT
HTH
THH

HTT
THT
TTH

TTT

$$\binom{3}{3} = 1$$

$$\binom{3}{2} = 3$$

$$\binom{3}{1} = 3$$

$$\binom{3}{0} = 1$$

dbinomial(2,0.5,3)

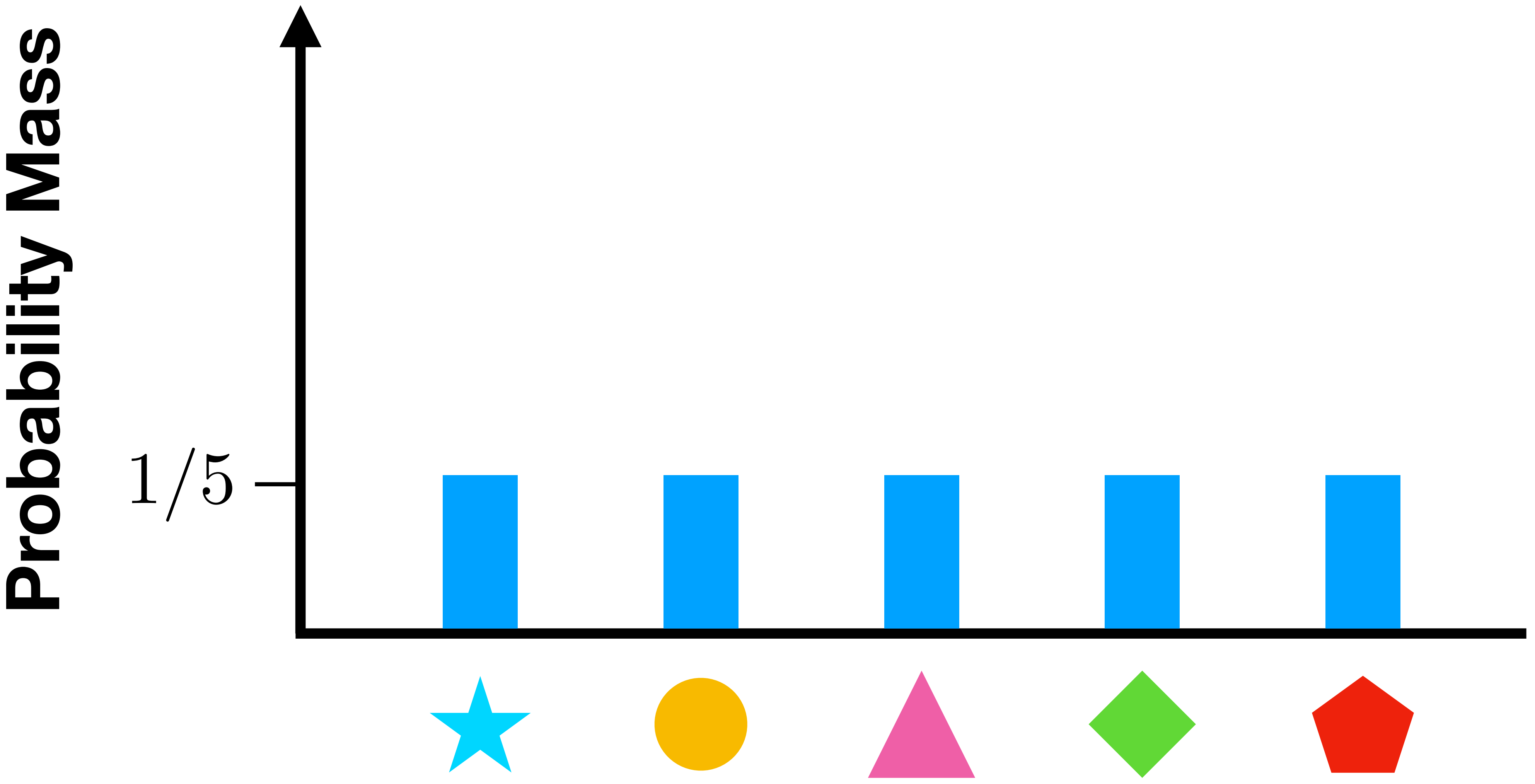
HHT
HTH
THH

$$\binom{3}{2} = 3$$

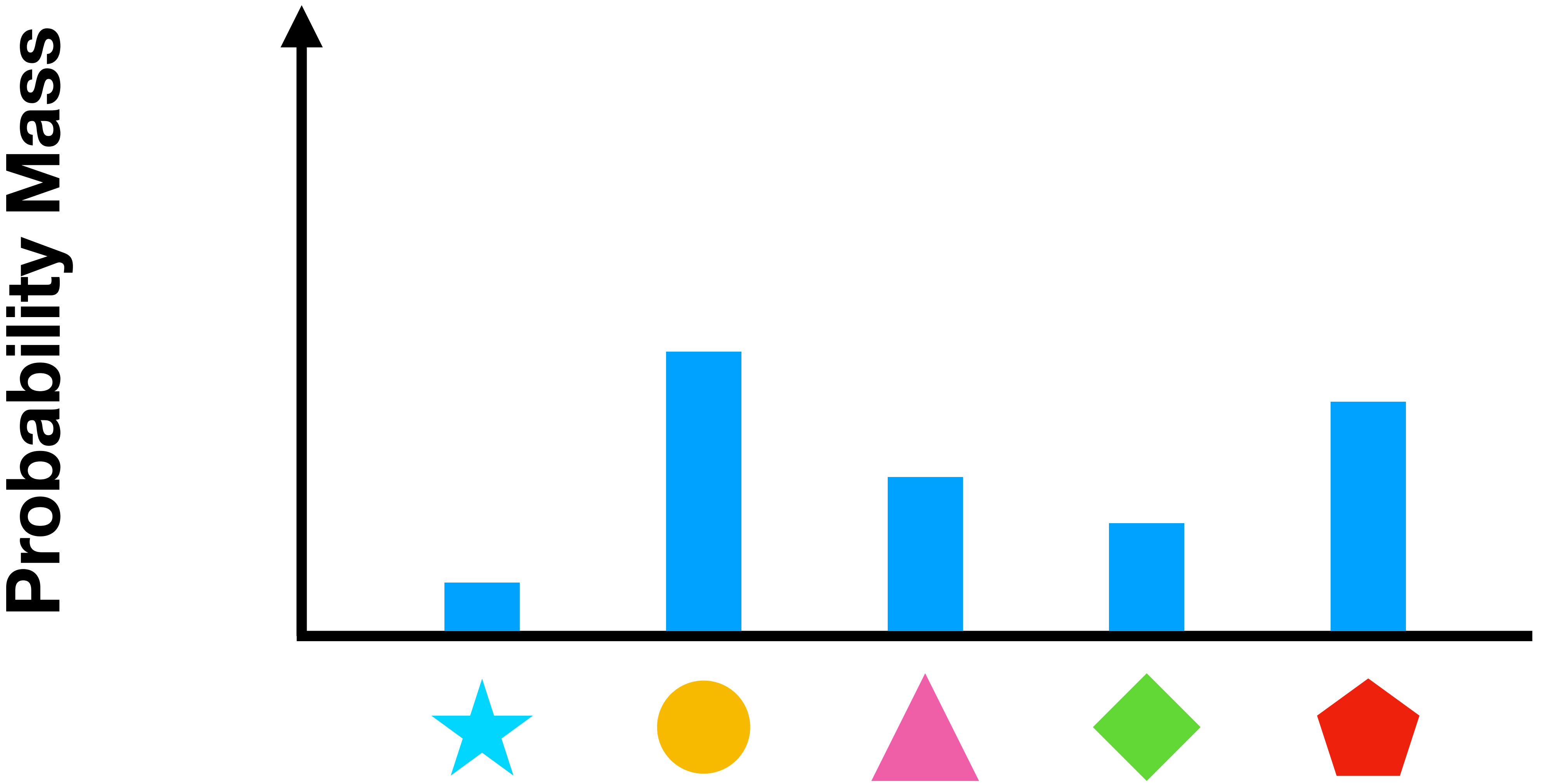
$$P(X=2) = \binom{3}{2} p^k (1-p)^{n-k} = 3(0.5)^2(1-0.5)^1 = 0.375$$

Discrete Uniform Distribution

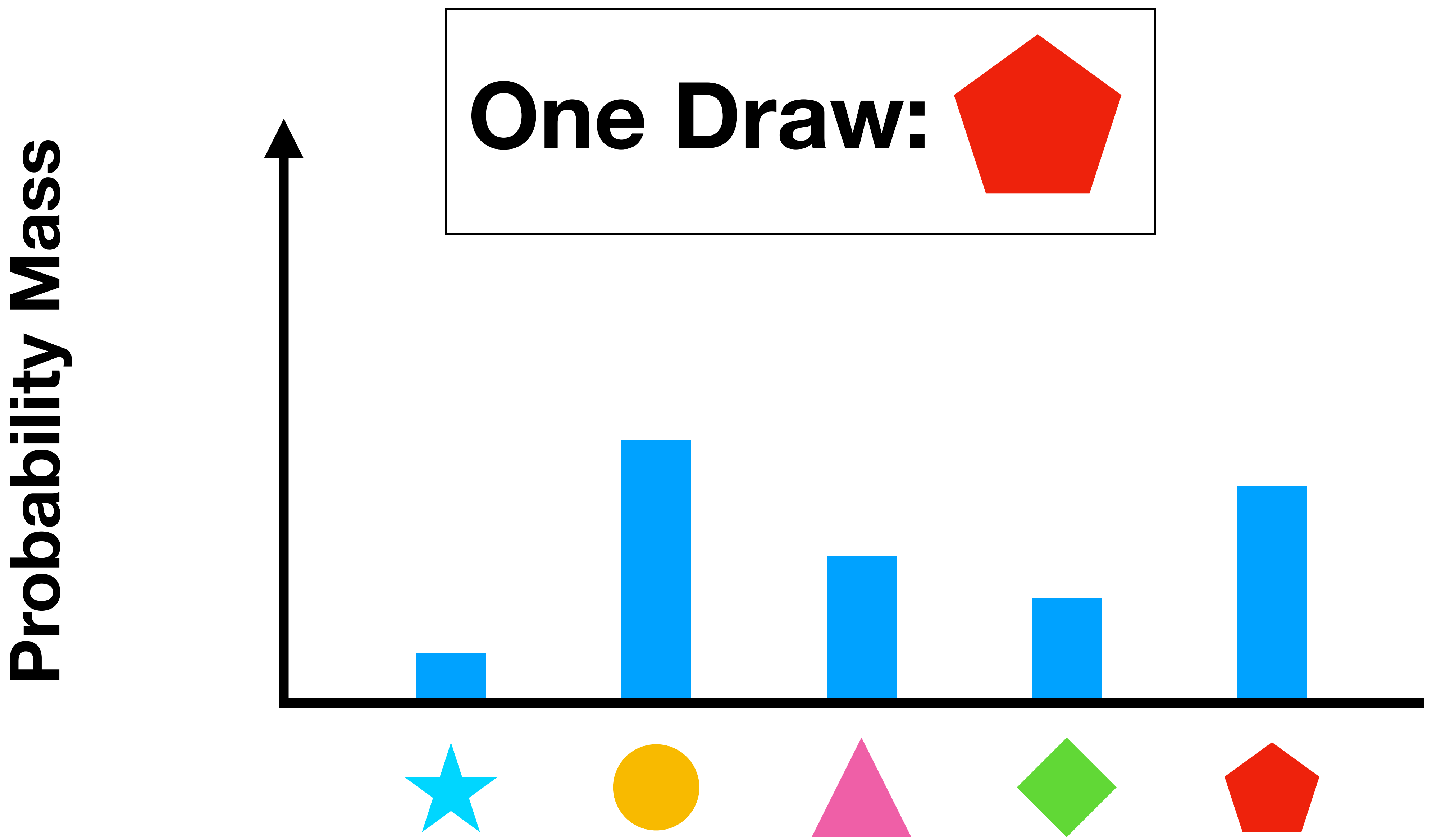
("choosing at random")



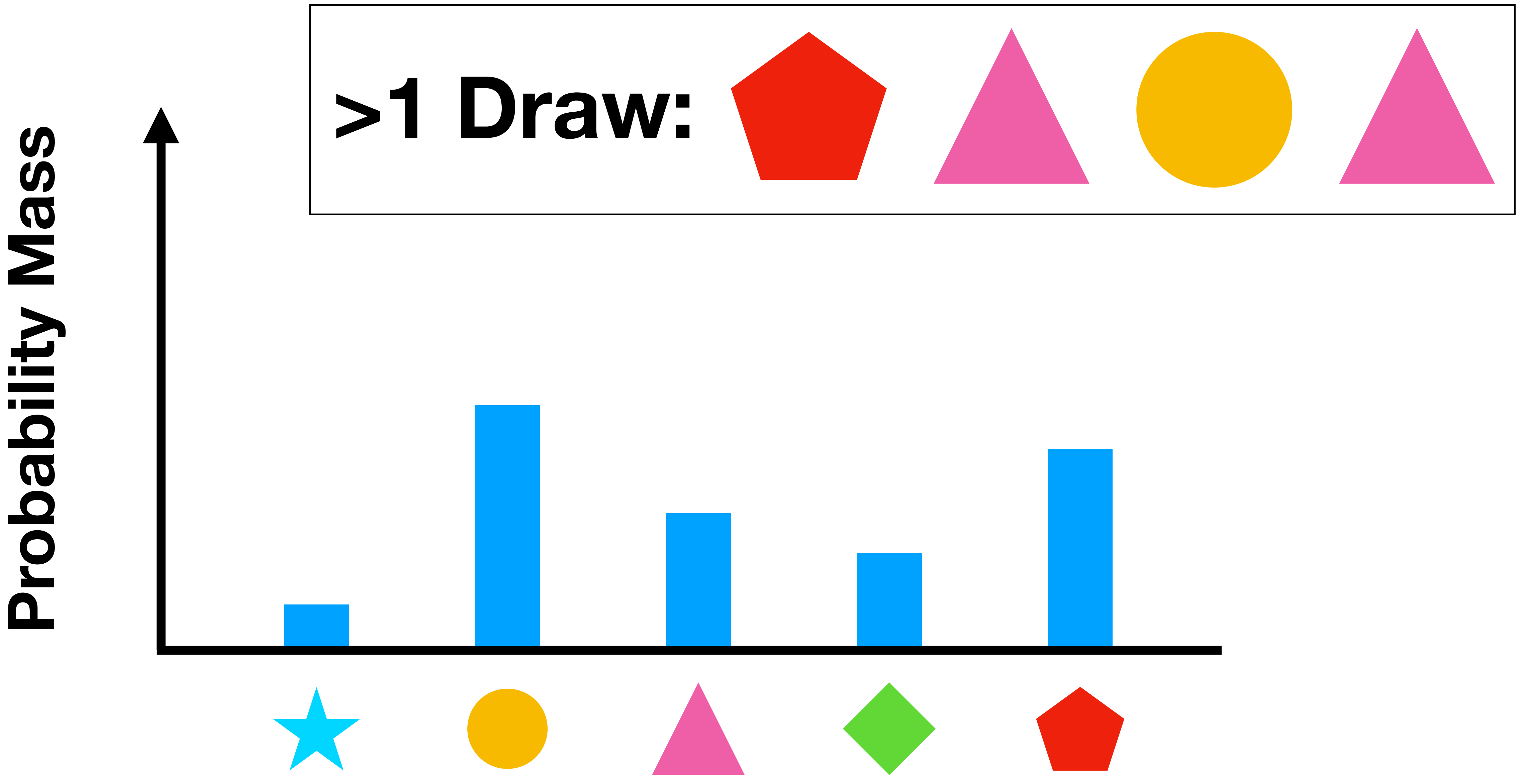
Categorical Distribution



Categorical Distribution



Multinomial Distribution

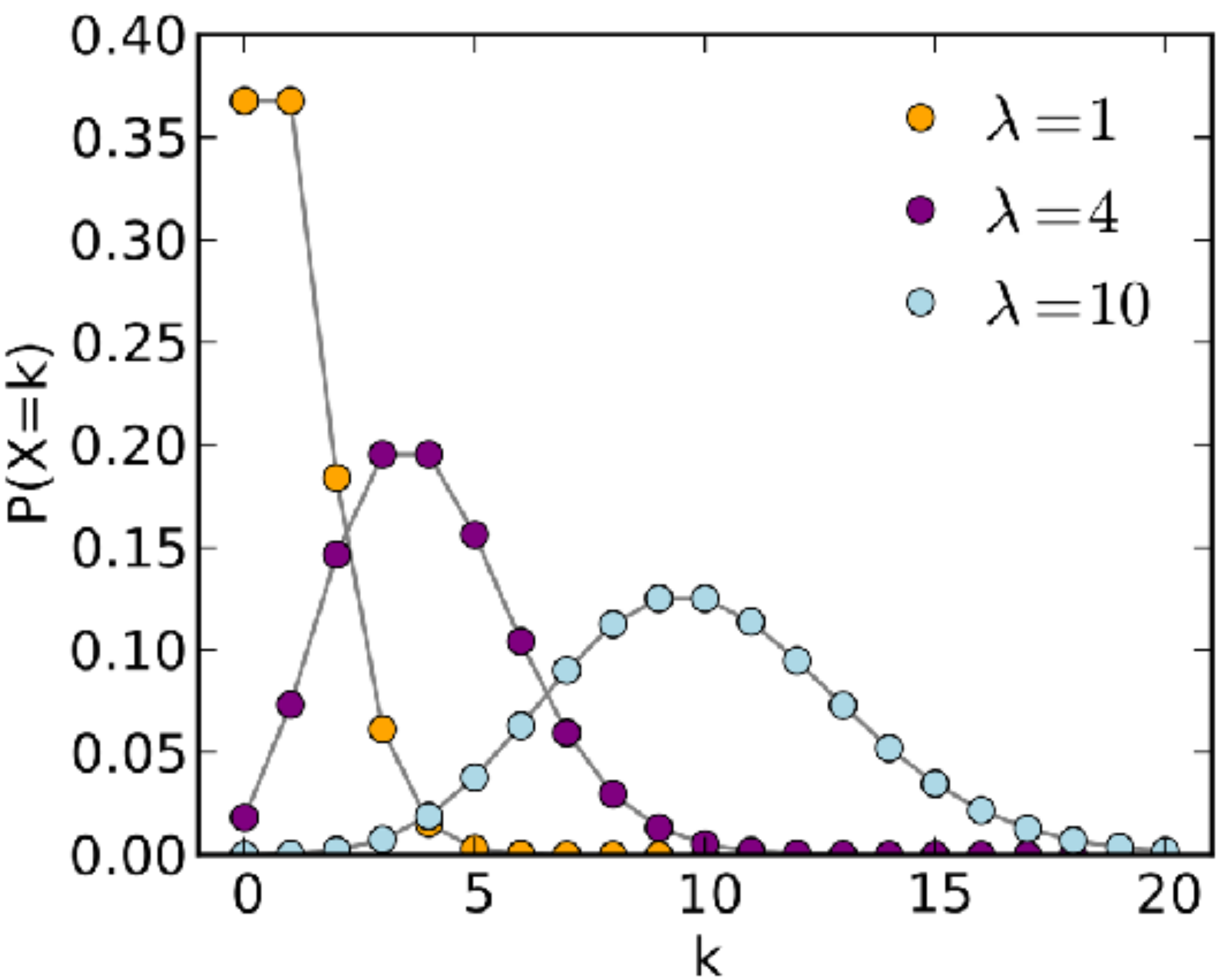


Exercise 5 in RevBayes

Custom functions and sampling from a multinomial

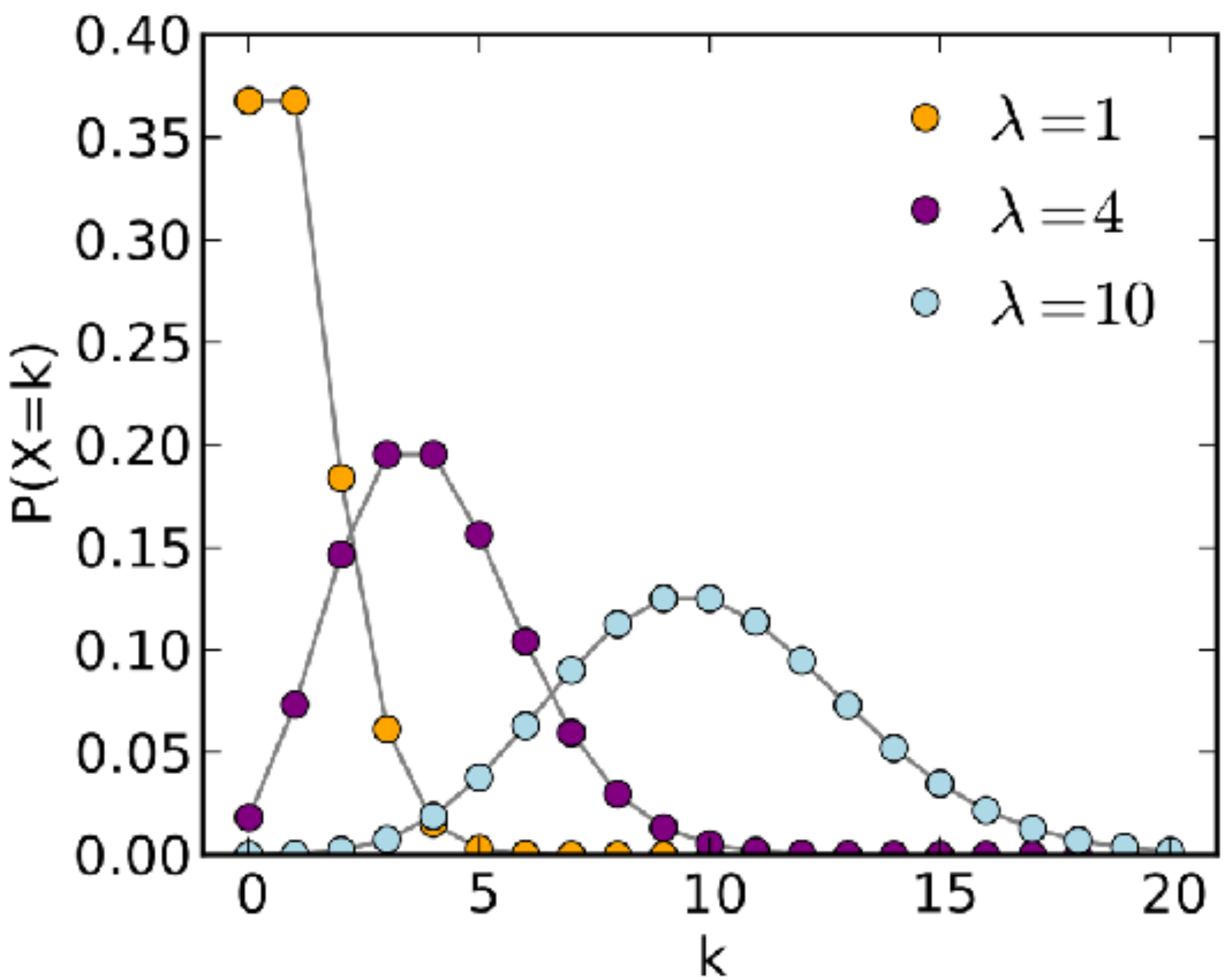
Poisson Distribution

This distribution models the number of events that occur in a given amount of time if there's a constant probability of an event occurring (for instance, the number of messages you receive if the senders are independent of one another). As with the binomial, this distribution can take any value greater than or equal to 0.



Poisson Distribution

The Poisson can be derived by starting with a binomial and increasing the number of trials, while keeping the expected number of successes constant (decreasing the probability of success for each individual trial). The Poisson is controlled by a rate parameter (the expected number of successes per unit time, space, etc.).

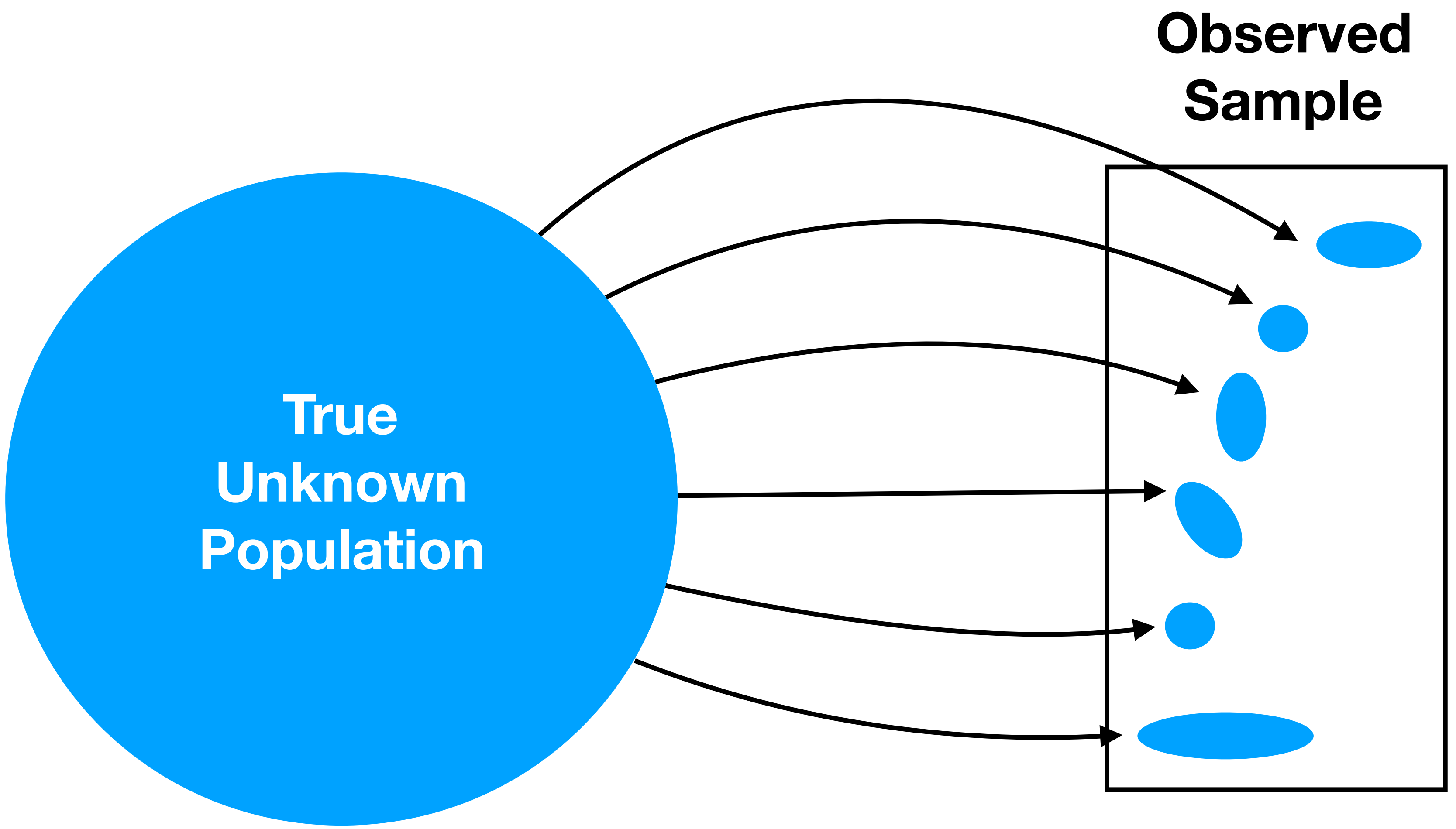


Exercise 6 in RevBayes

Exploring the Poisson

Switching gears a bit…

Samples v Populations



Trying to learn about this . . . . but have observed this.

Bootstrapping

Let's say we have a sample of values…

`[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]`

…with mean…

`3.5151`

How confident are we that this value is a good representation of the mean value for the entire population from which we've sampled?

Bootstrapping

If we know the distribution from which the data were sampled, we sometimes have a formula available to tells us about the 95% confidence interval on our estimate.

[ASIDE: What is a 95% confidence interval?]

But what if we don’t know what kind of distribution our data were drawn from?

Bootstrapping

In that situation, we need to "pull ourselves up by our bootstraps", without a formula.

Brad Efron, a Stanford statistician, gave us a procedure to do just that.

If we generate a series of pseudo-replicate datasets, by sampling with replacement from the original dataset, we can get an estimate of our confidence in the empirical estimate.

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

[2.148, 2.668, 3.647, 3.141, 2.148, 3.141, 2.051, 3.318, 3.346, 3.141]

Example Pseudoreplicate