Sampling distribution

132 orange-spotted warblers. 1 indicates infected

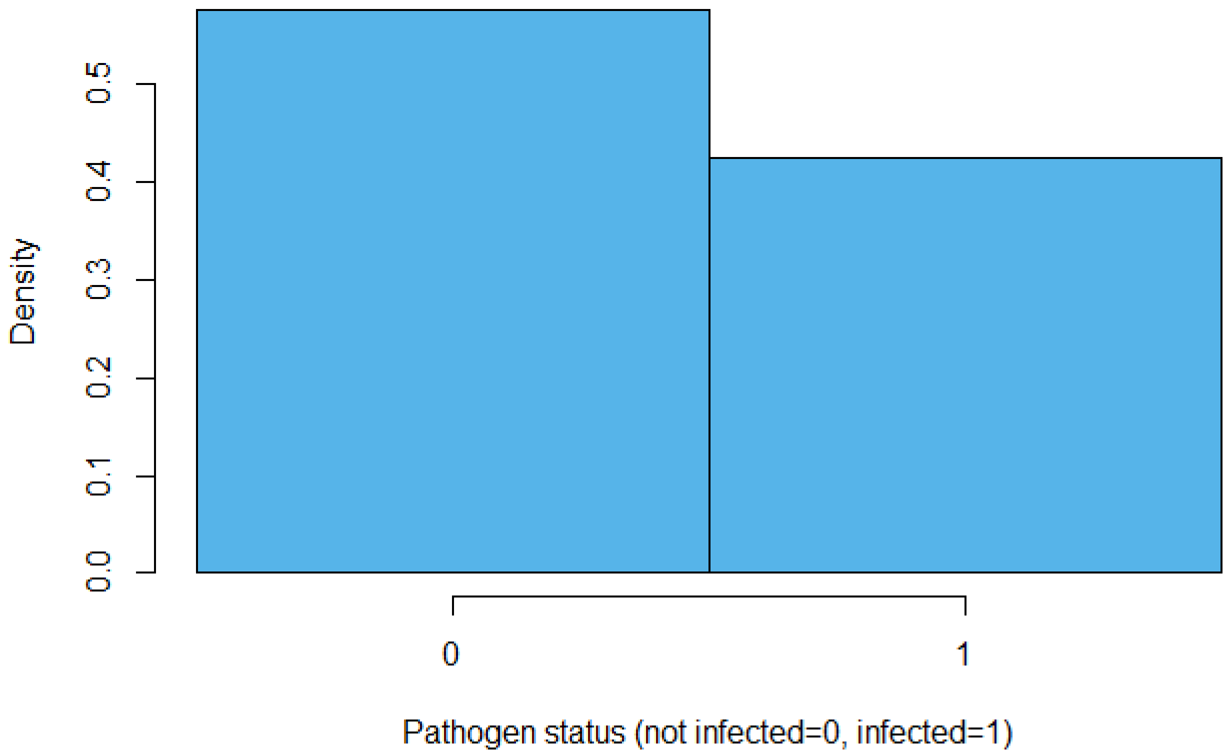
```
pathogen <-
    c(1,1,1,1,1,0,0,0,0,0,0,0,1,1,1,0,1,1,0,1,1,0,0,0,1,1,0,0,1,1,0,1,0,0,0,0,
      1,1,1,0,1,1,0,1,1,0,0,1,1,0,0,1,1,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,1,0,0,1,1,
      0,1,0,0,0,1,0,0,0,1,0,0,1,0,0,1,1,0,1,1,0,1,1,0,0,0,0,0,0,0,0,1,0,1,1,1,0,
      1,0,1,0,0,0,0,0,0,1,1,0,0,0,1,1,1,1,0,0,1)
```

Take a sample:

```
sample(pathogen,10)
```

0 1 0 0 0 0 0 1 0 0
pathogen prevalence = 0.2

Our scientific observation



True prevalence is 0.424

Sampling distribution algorithm

repeat very many times   law of large numbers

    sample n units from the population

    calculate the sample statistic

plot sampling distribution (histogram) of the sample statistic

for pathogen prevalence

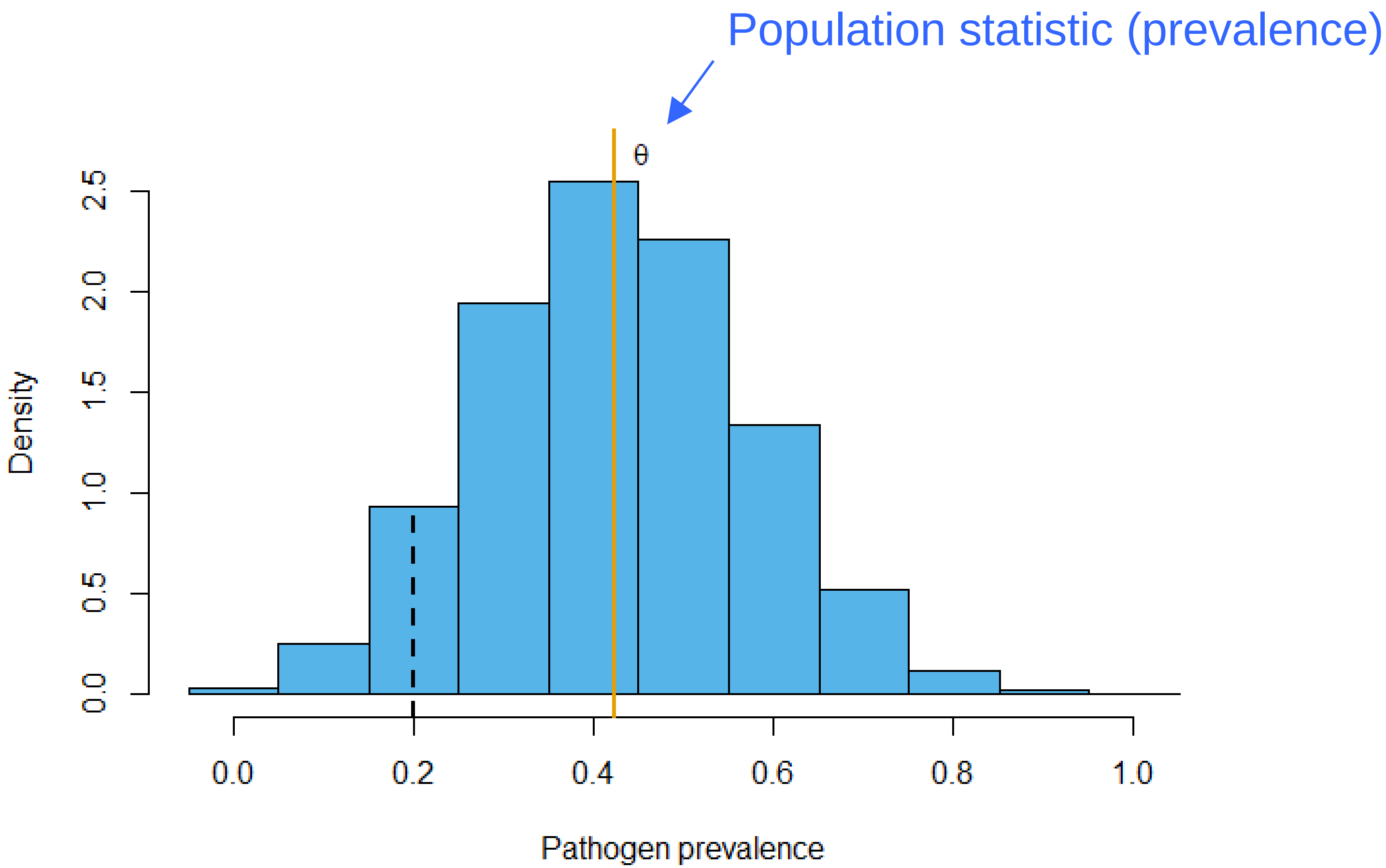
for a large number of repeated samples

    randomly sample 10 birds from the population

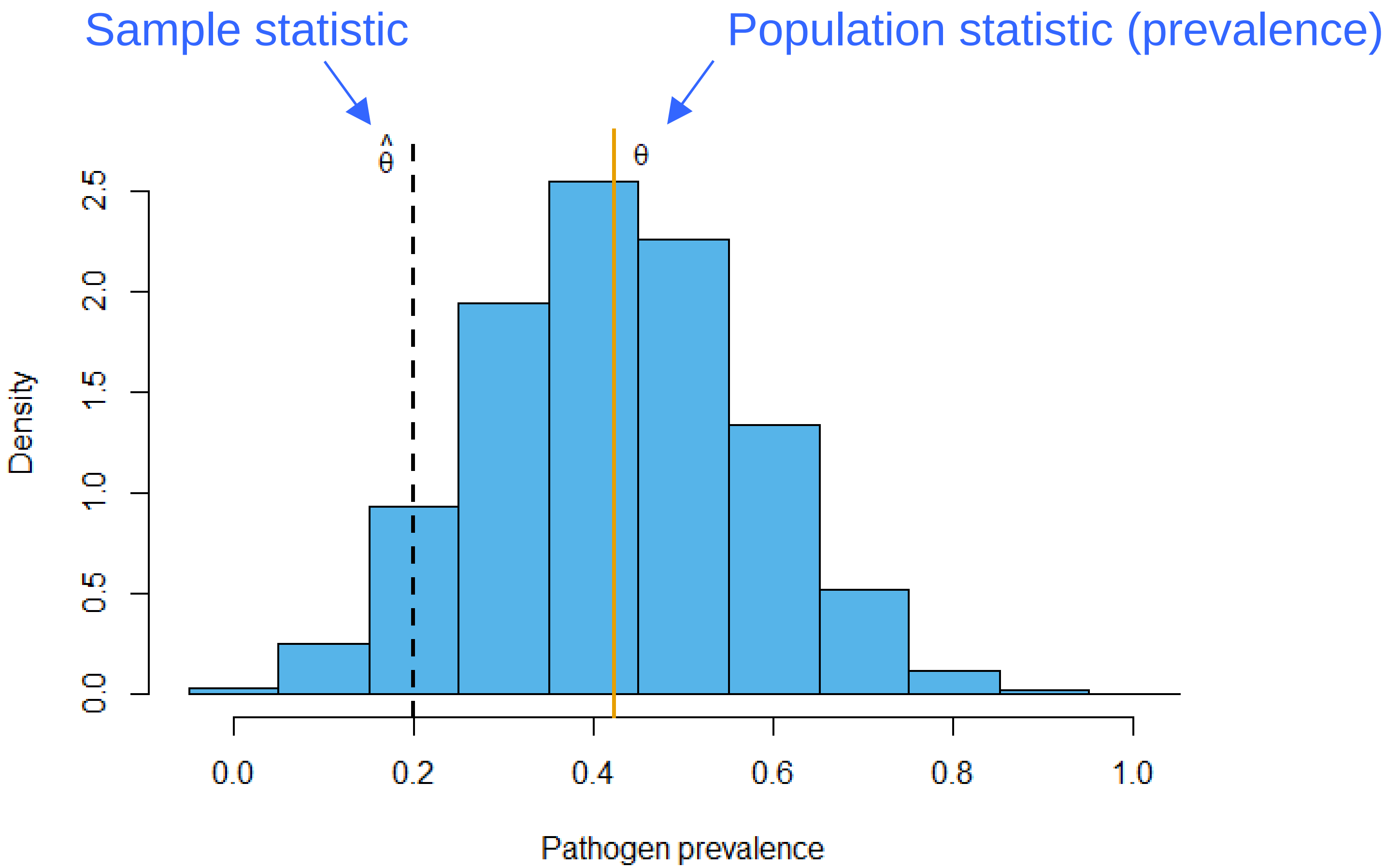
    calculate the prevalence in the sample

plot sampling distribution (histogram) of prevalence

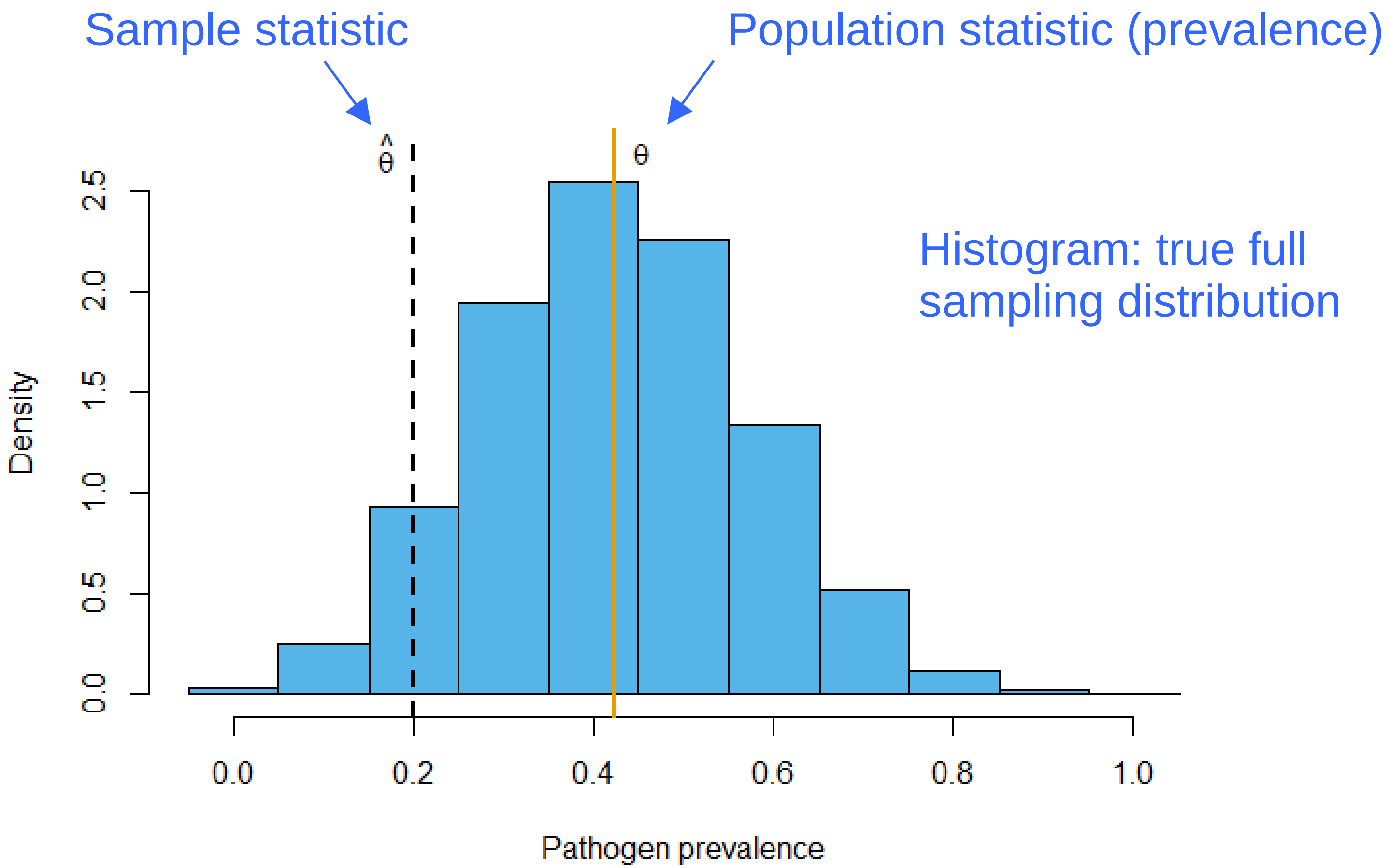
The sampling distribution for prevalence



The sampling distribution for prevalence



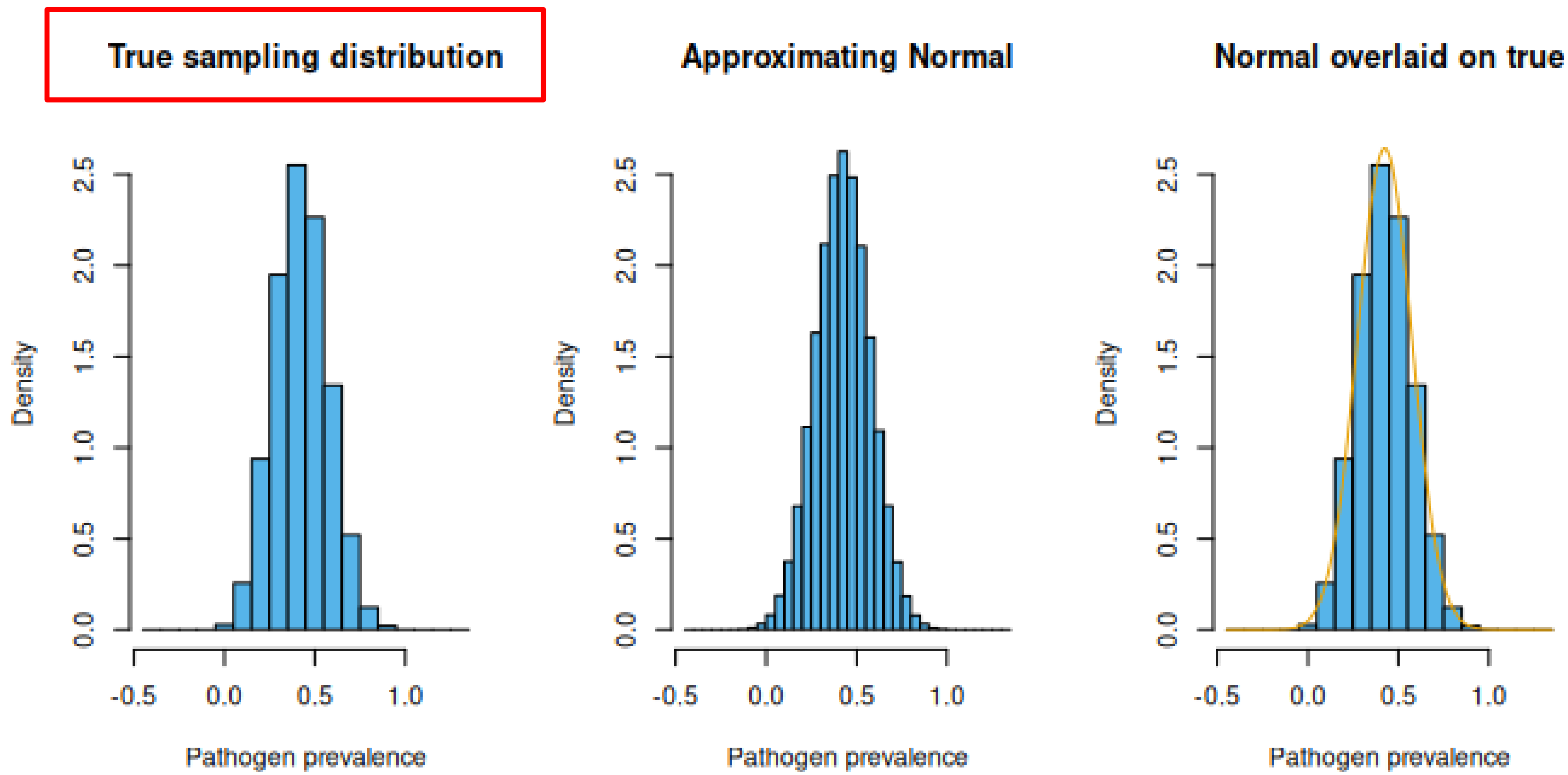
The sampling distribution for prevalence



Confidence interval

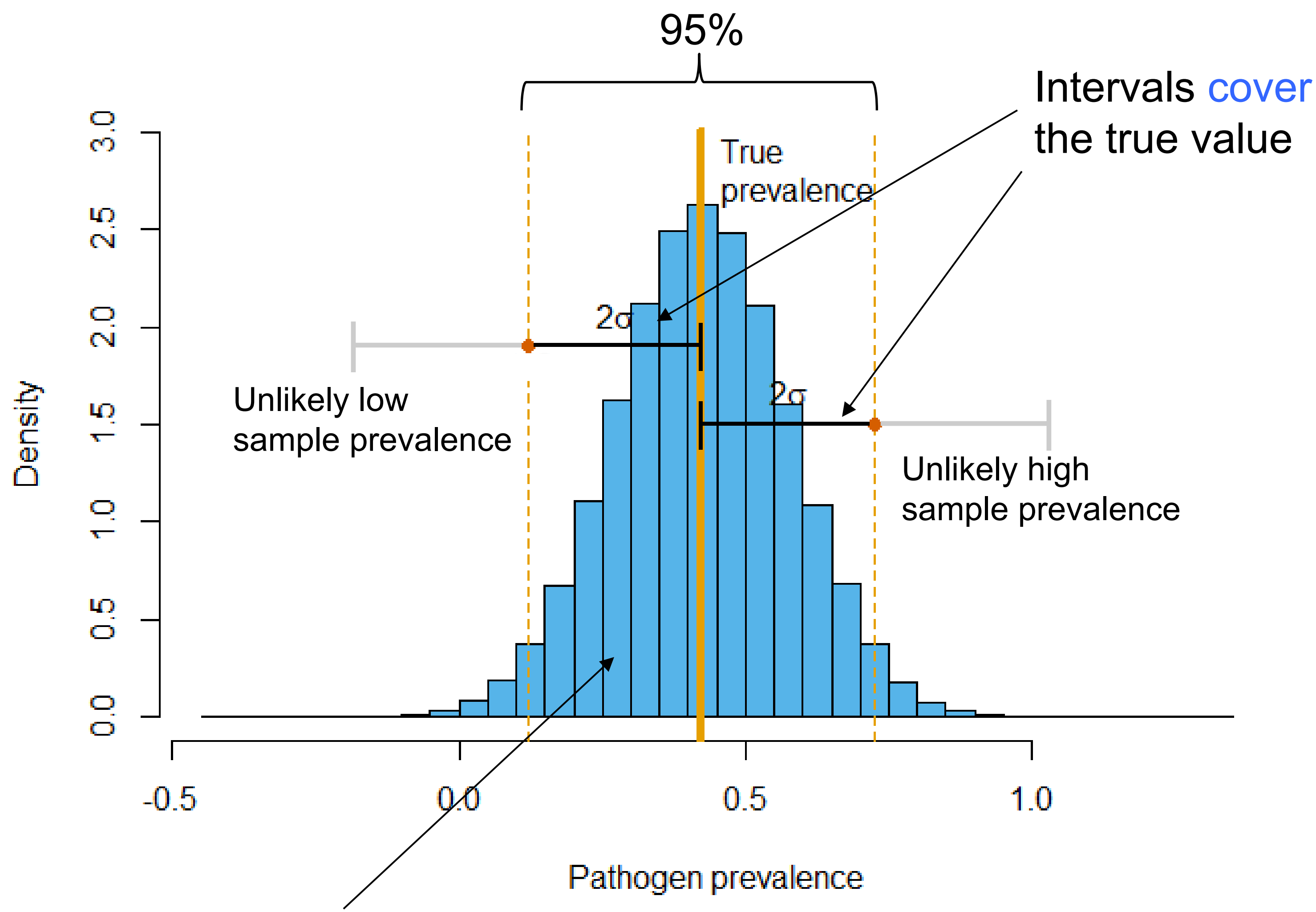
- An interval calculated by some procedure that would **contain** (or **cover**) the true population value 95% of the time, **if sampling and calculating an interval were repeated a very large number of times**

Confidence = reliability of the procedure



True sampling distribution
Approximating Normal
Normal overlaid on true
Density
0.0
0.5
1.0
1.5
2.0
2.5
-0.5
0.0
0.5
1.0
Pathogen prevalence

Construct an interval to cover true value



Normal distribution approximating the true sampling distribution

Plug in principle

- We don't know the true sampling distribution or its parameters
- Plug in the sample instead as an estimate
  - in this example we can use the standard error of the sample as an estimate of the standard deviation of the sampling distribution

Coverage algorithm

repeat very many times
    sample n units from the population
    calculate the sample statistic
    calculate the interval for the sample statistic
calculate frequency true value is in the interval

Calibrates the degree of confidence in the procedure

Calibration: 95.6% confidence intervals

In first 100, 6 do not cover the true value
(we expect about 5/100)