

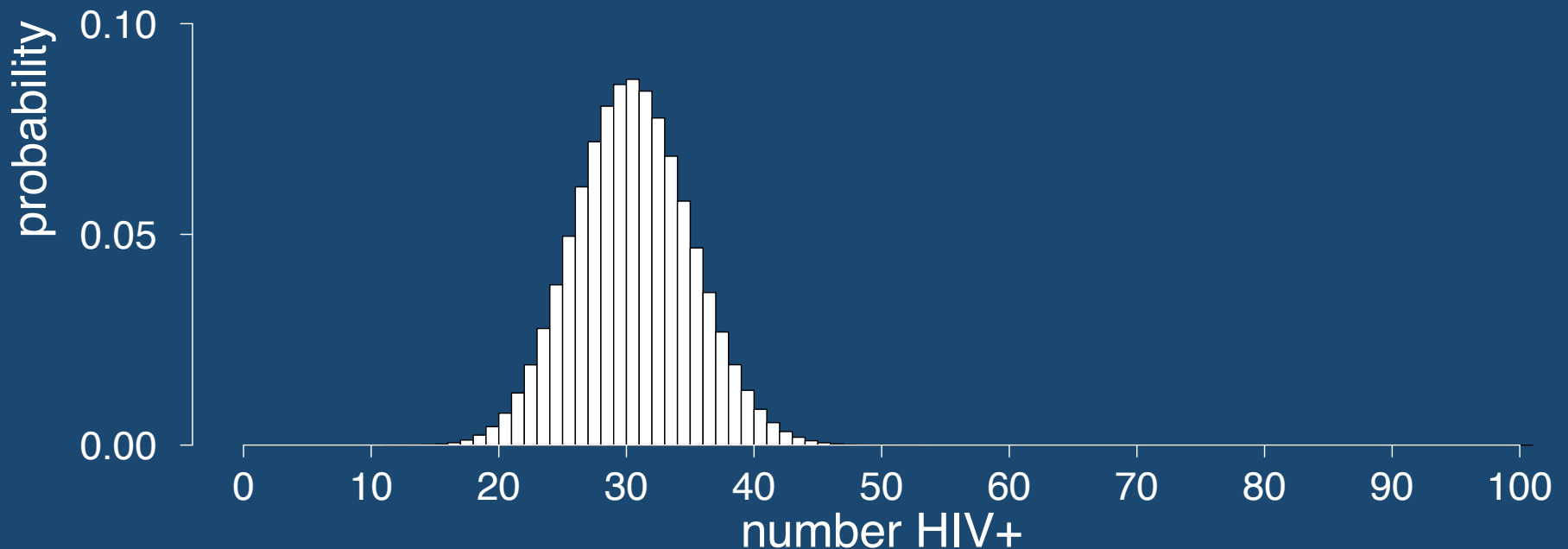
Introduction to Likelihood

Meaningful Modeling of Epidemiologic Data, 2015
AIMS, Muizenberg, South Africa

Steve Bellan, PhD, MPH
University of Texas at Austin

In a population of 1,000,000 people with a true prevalence of 30%, the probability distribution of number of positive individuals if 100 are sampled:

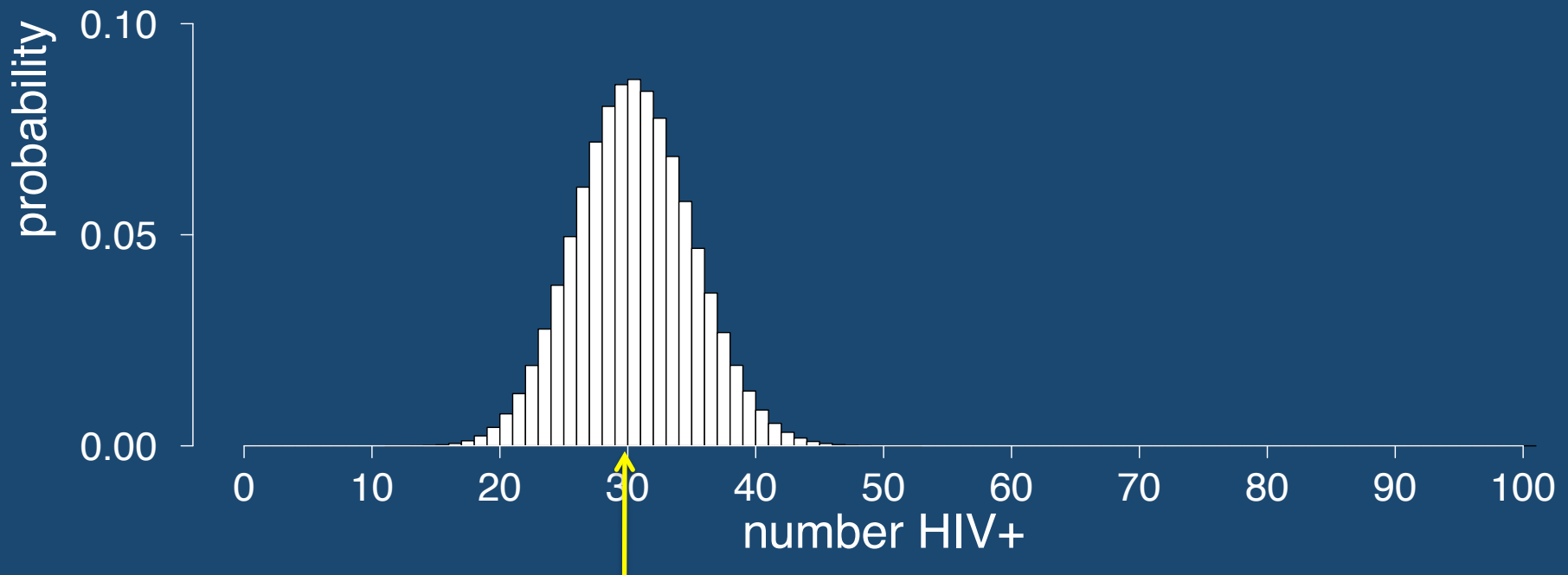
$$f(x) = \binom{100}{x} (0.3)^x (0.7)^{100-x}$$



```
barplot(dbinom(x = 0:100, size = 100, prob = .3), names.arg = 0:size)
```

In a population of 1,000,000 people with a true prevalence of 30%, the probability distribution of number of positive individuals if 100 are sampled:

$$f(x) = \binom{100}{x} (0.3)^x (0.7)^{100-x}$$

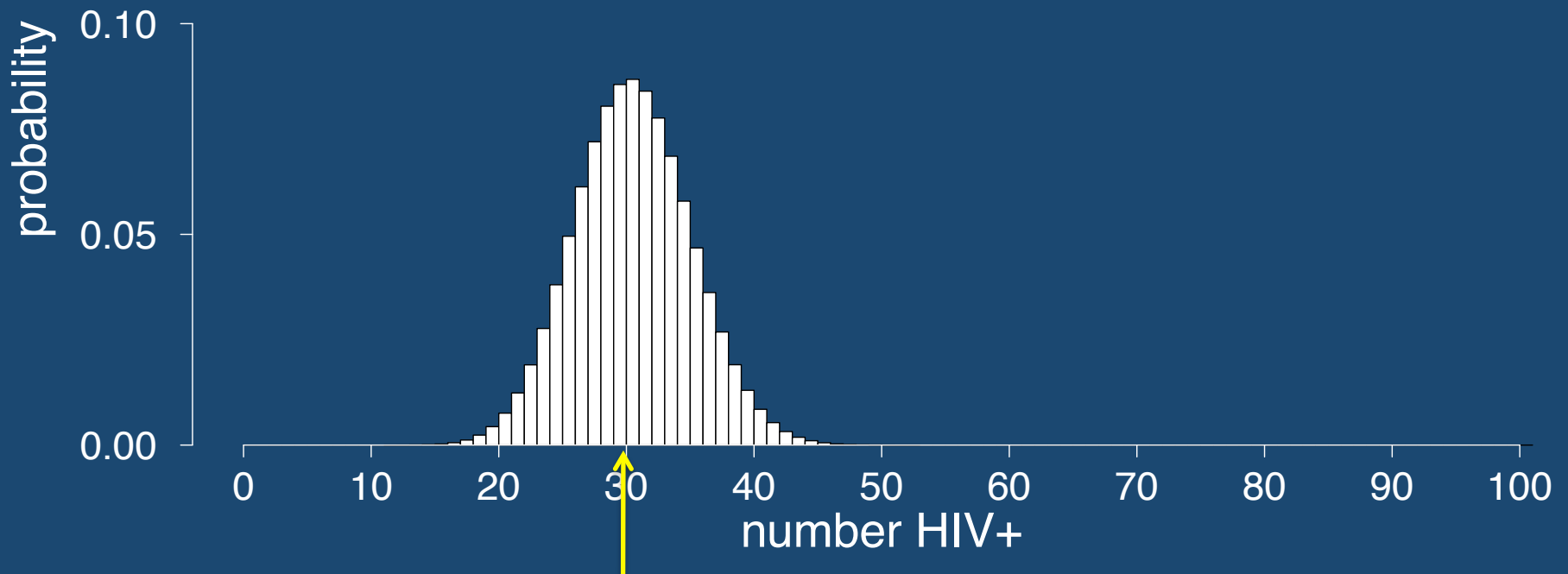


We sample 100 people once and 28 are positive:

```
> rbinom(n = 1, size = 100, prob = .3)
[1] 28
```

We don't know the true prevalence!

But we can calculate the probability of 28 or a more extreme value occurring for a given prevalence.



We sample 100 people once and 28 are positive:

```
> rbinom(n = 1, size = 100, prob = .3)  
[1] 28
```

Cumulative Probability & P Values

If true prevalence were 30%, then $p(28 \text{ or more extreme})$ is

$$2 * \text{pbinom}(28, 100, 0.3, \text{lower.tail} = \text{TRUE})$$

$$p = 0.754$$

for 30% prevalence:

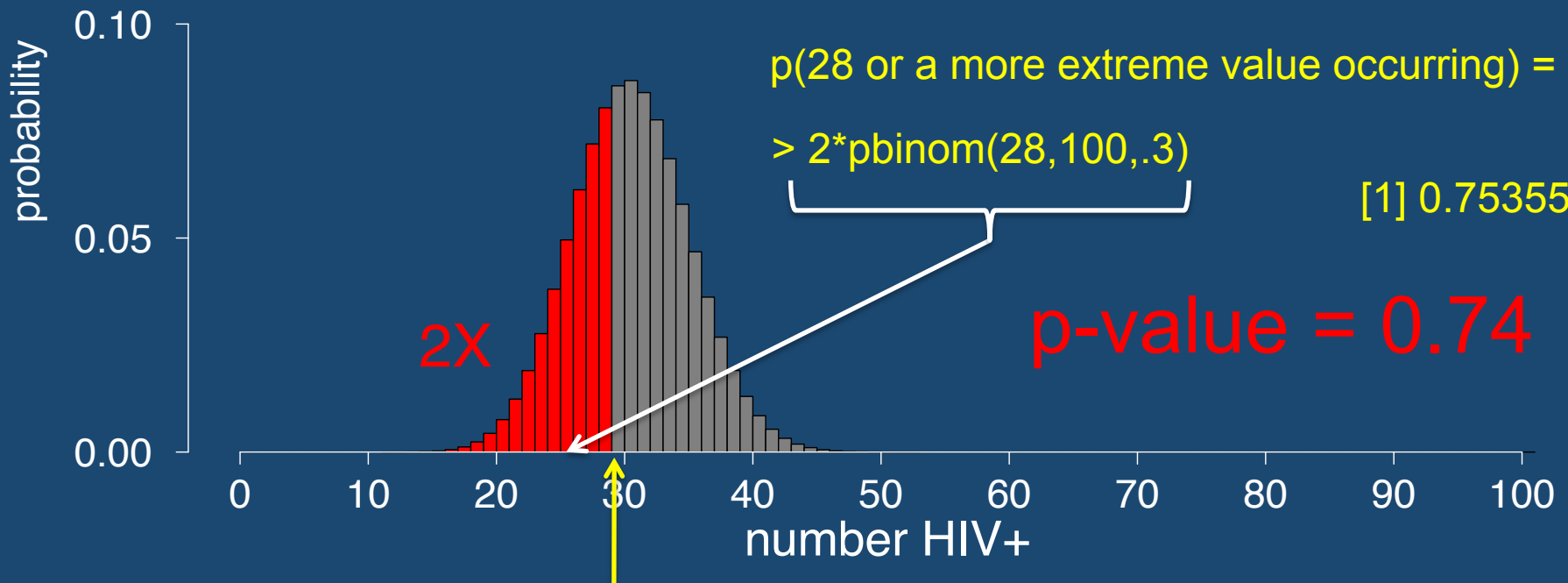
$$p(28 \text{ or a more extreme value occurring}) =$$

$$> 2 * \text{pbinom}(28, 100, .3)$$

[1] 0.7535564

2X

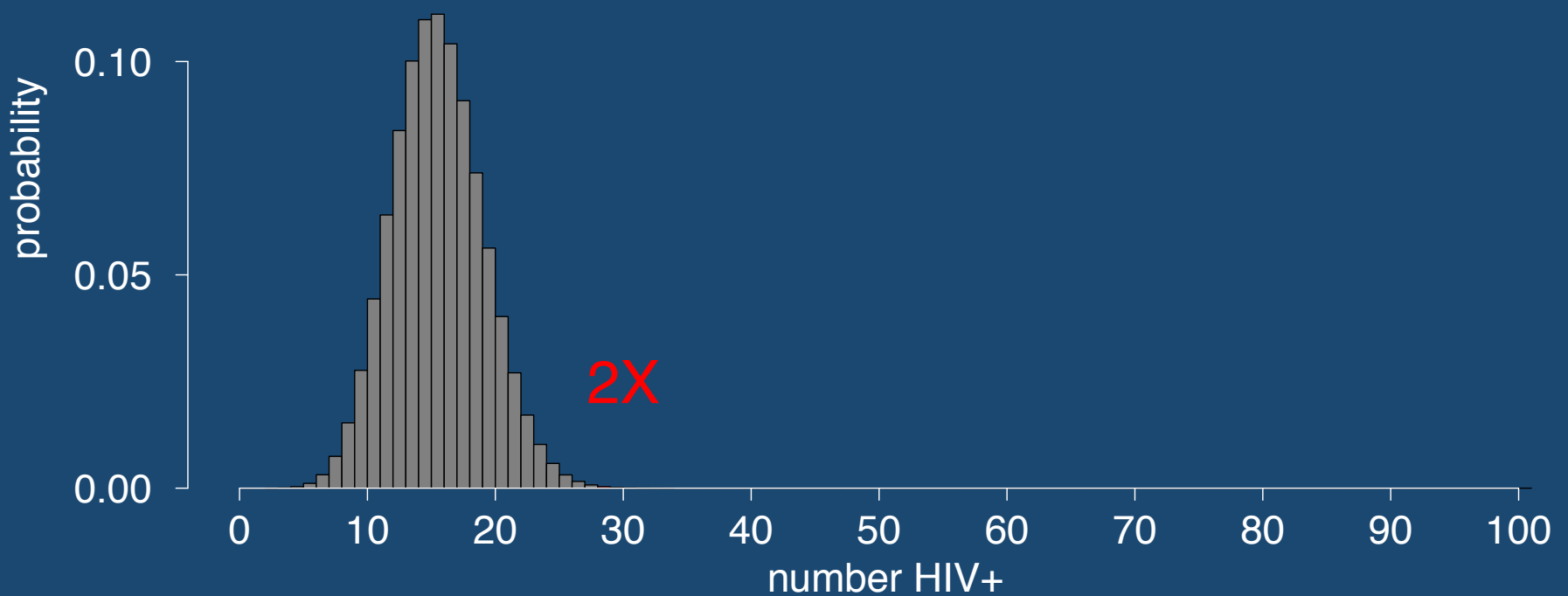
p-value = 0.74



We sample 100 people once and 28 are positive.

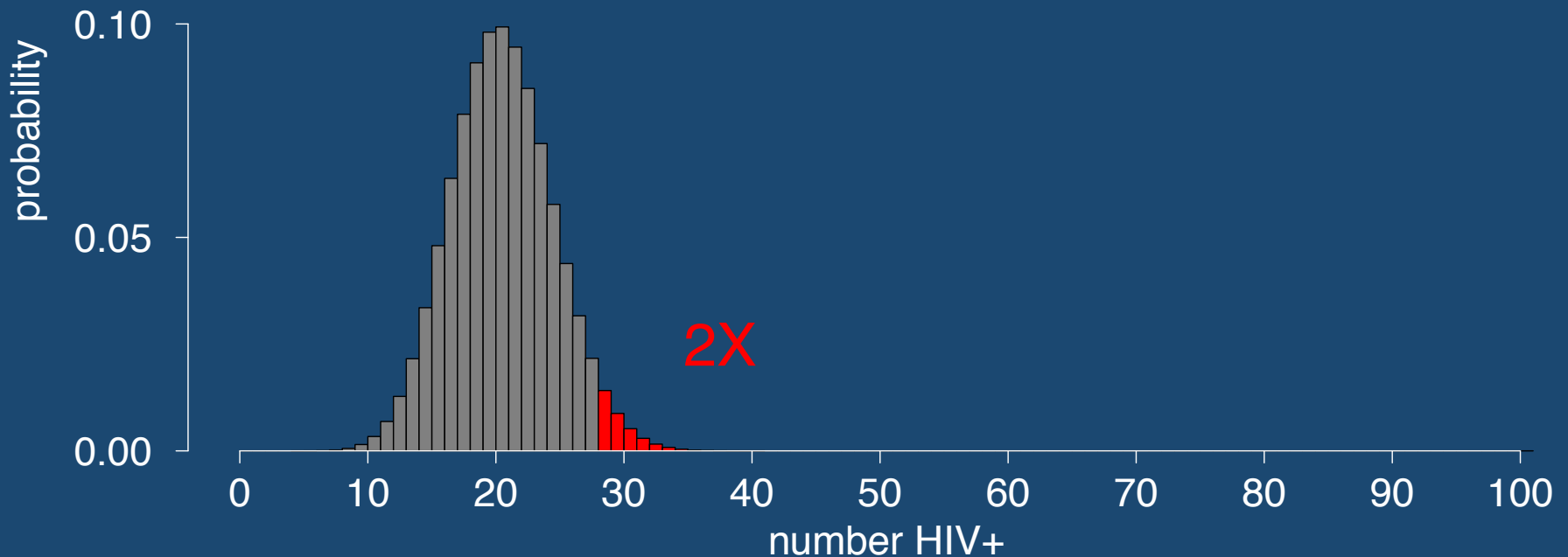
If true prevalence were 15%, then $p(28 \text{ or more extreme})$ is
 $2 * \text{pbinom}(28-1, 100, 0.15, \text{lower.tail} = \text{FALSE})$

$p = 0.00123$



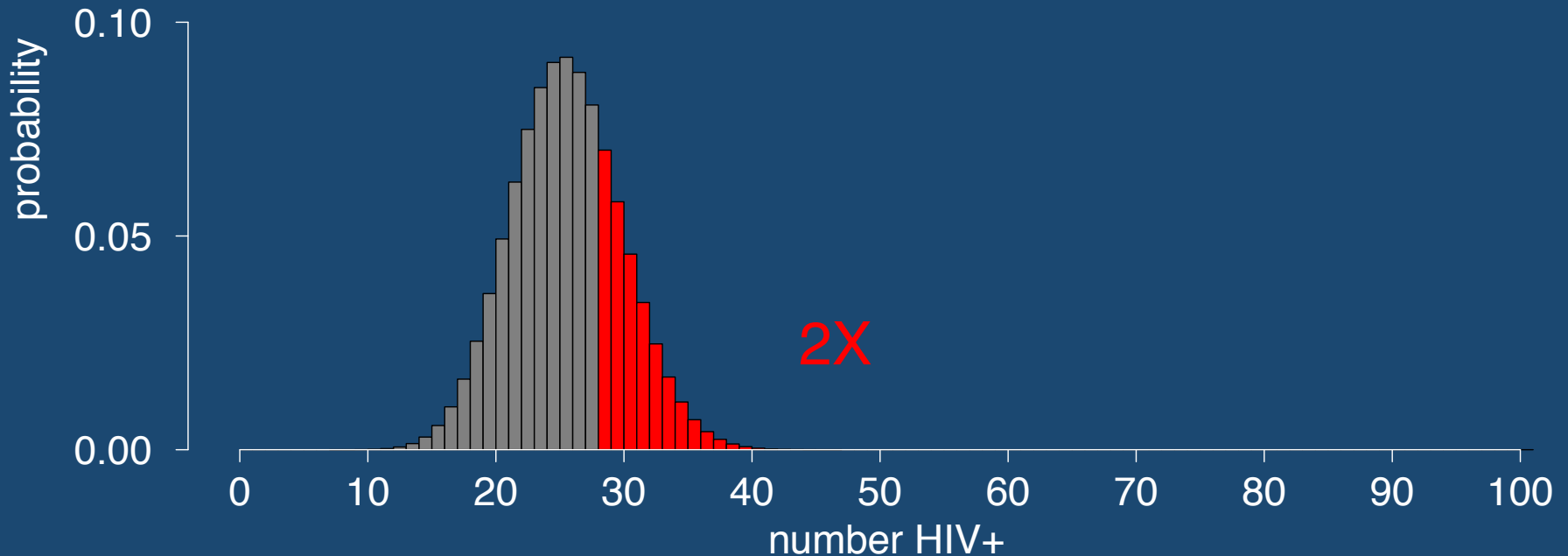
If true prevalence were 20%, then $p(28 \text{ or more extreme})$ is
 $2 * \text{pbinom}(28-1, 100, 0.2, \text{lower.tail} = \text{FALSE})$

$p = 0.0683$



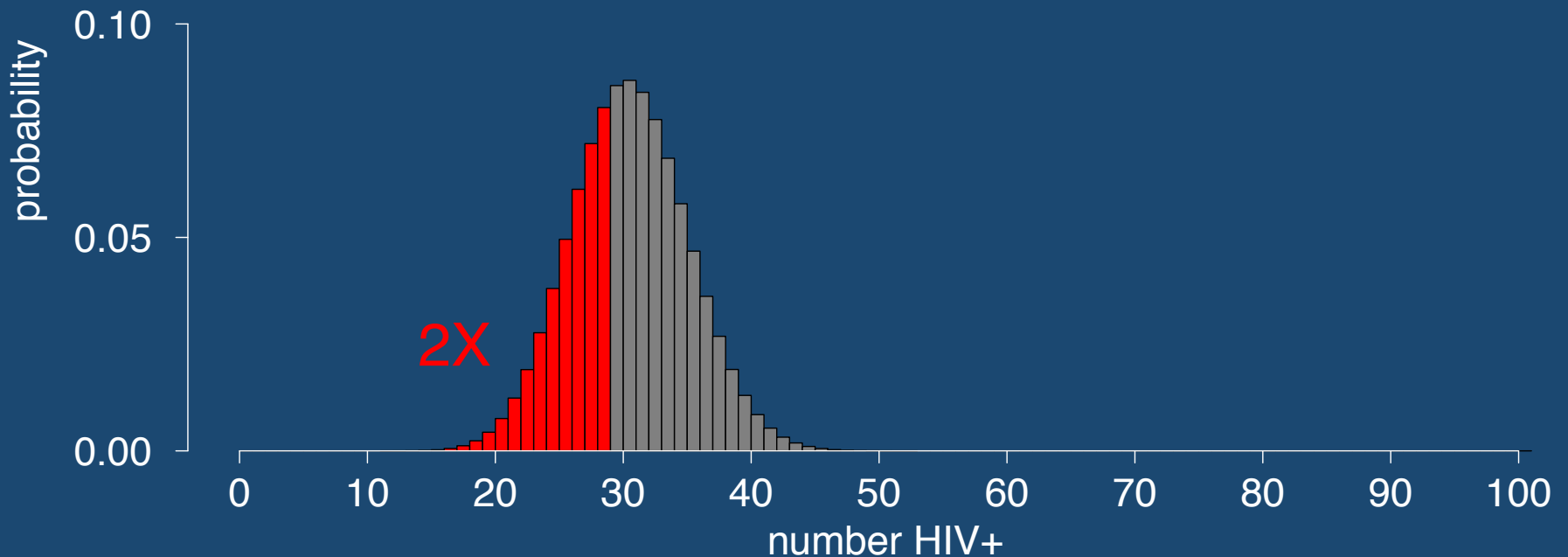
If true prevalence were 25%, then $p(28 \text{ or more extreme})$ is
 $2 * \text{pbinom}(28-1, 100, 0.25, \text{lower.tail} = \text{FALSE})$

$p = 0.555$



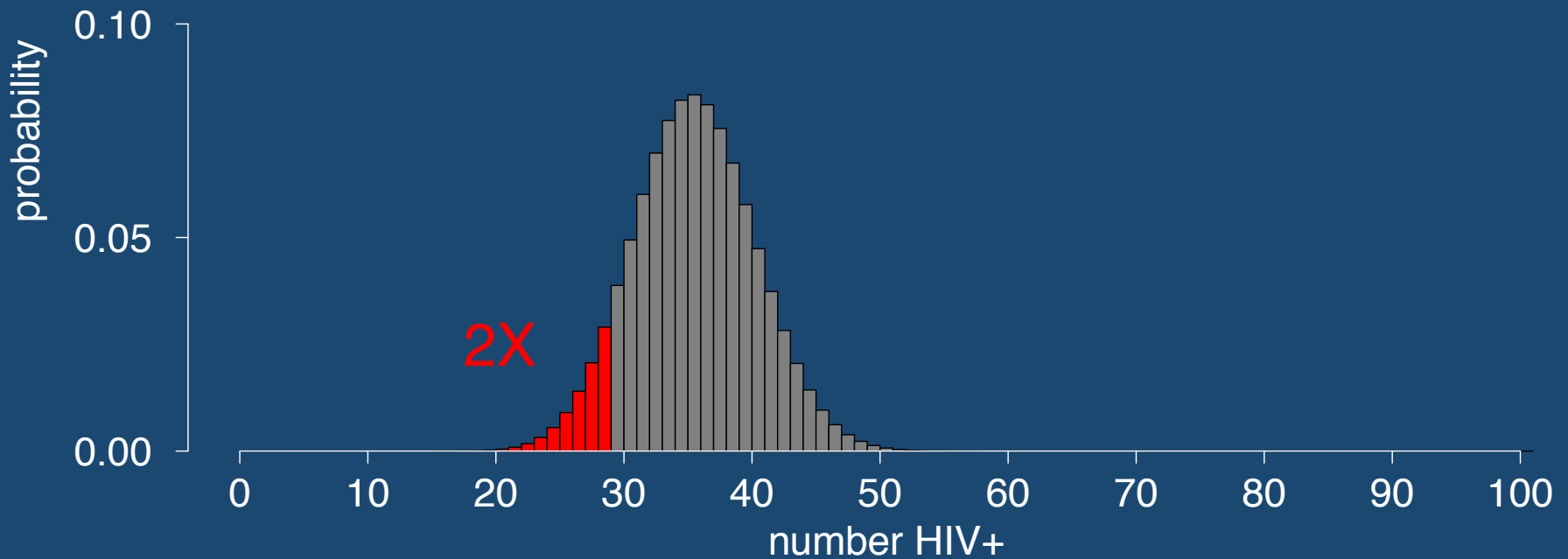
If true prevalence were 30%, then $p(28 \text{ or more extreme})$ is
 $2 * \text{pbinom}(28, 100, 0.3, \text{lower.tail} = \text{TRUE})$

$p = 0.754$



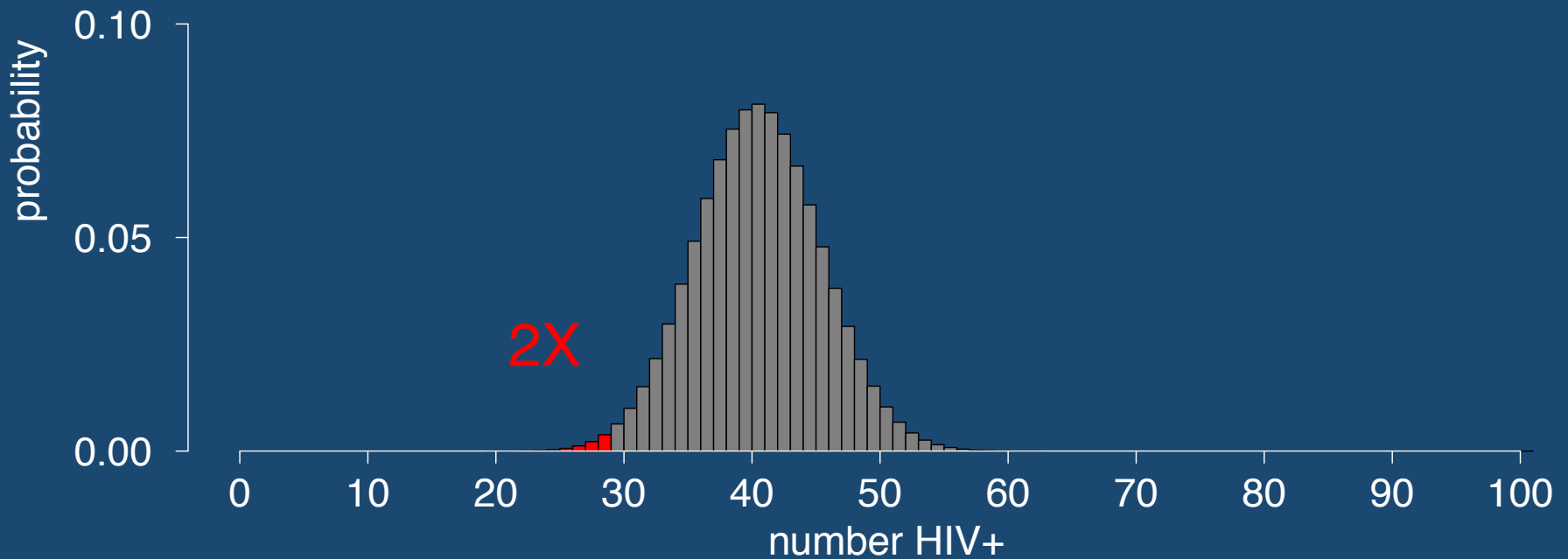
If true prevalence were 35%, then $p(28 \text{ or more extreme})$ is
 $2 * pbinom(28, 100, 0.35, \text{lower.tail} = \text{TRUE})$

$p = 0.17$



If true prevalence were 40%, then $p(28 \text{ or more extreme})$ is
 $2 * \text{pbinom}(28, 100, 0.4, \text{lower.tail} = \text{TRUE})$

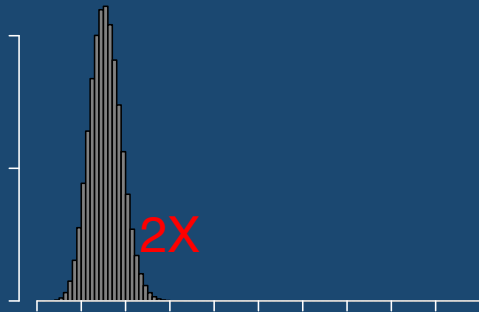
$p = 0.0169$



Summary

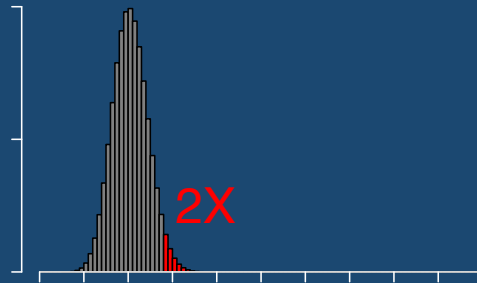
hypothetical
prevalence: 15 %

$p = 0.00123$



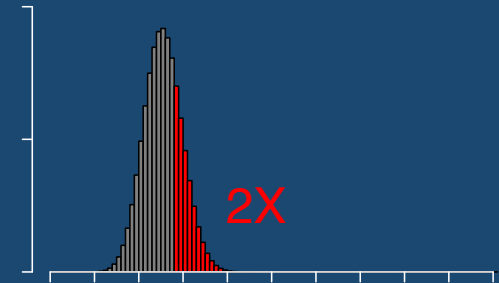
hypothetical
prevalence: 20 %

$p = 0.0683$



hypothetical
prevalence: 25 %

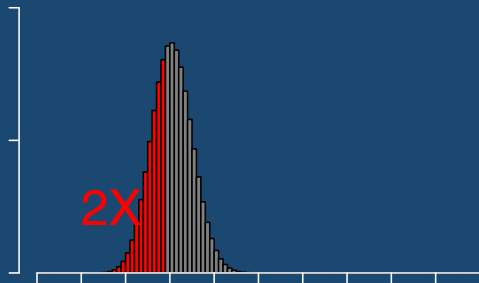
$p = 0.555$



probability

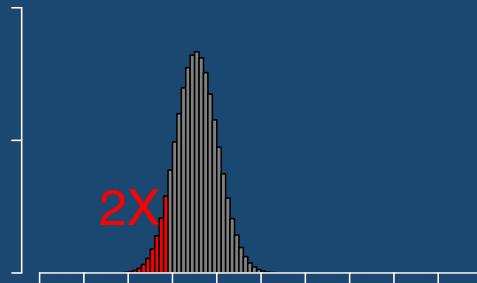
hypothetical
prevalence: 30 %

$p = 0.754$



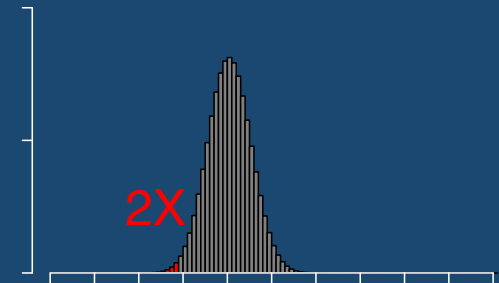
hypothetical
prevalence: 35 %

$p = 0.17$



hypothetical
prevalence: 40 %

$p = 0.0169$



number HIV+

Which hypotheses do we reject?

IF GIVEN THE HYPOTHESIS

$p \text{ value} < \text{cutoff}$

THEN REJECT HYPOTHESIS

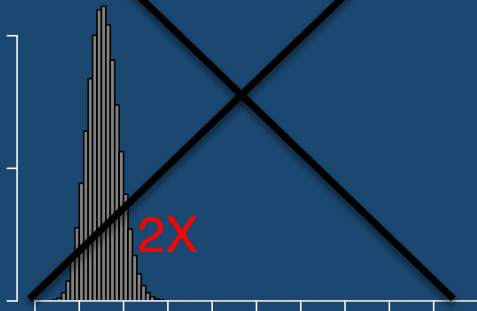
Cutoff usually chosen as $\alpha = 0.05$

Which hypotheses do we reject?

probability

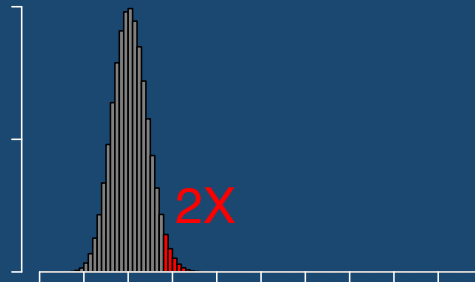
hypothetical prevalence: 15 %

$p = 0.00123$



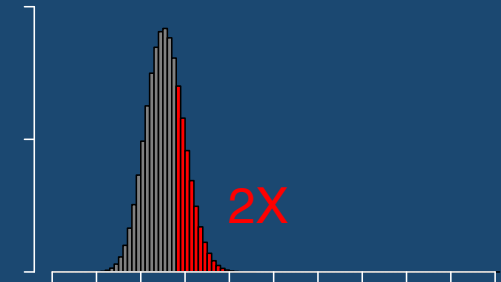
hypothetical prevalence: 20 %

$p = 0.0683$



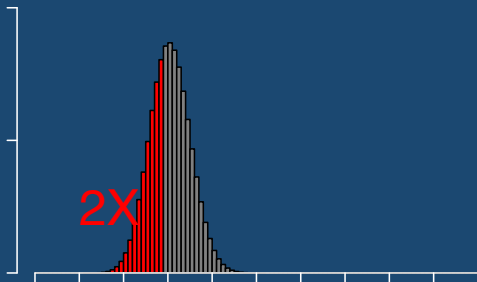
hypothetical prevalence: 25 %

$p = 0.555$



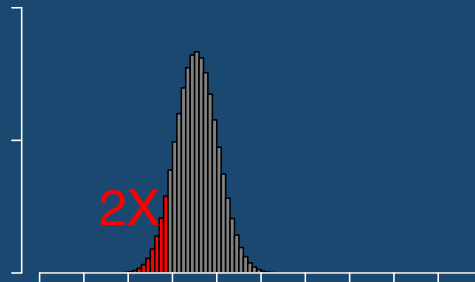
hypothetical prevalence: 30 %

$p = 0.754$



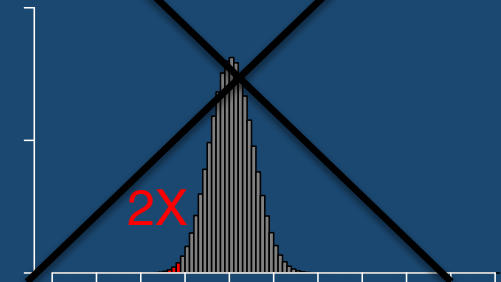
hypothetical prevalence: 35 %

$p = 0.17$



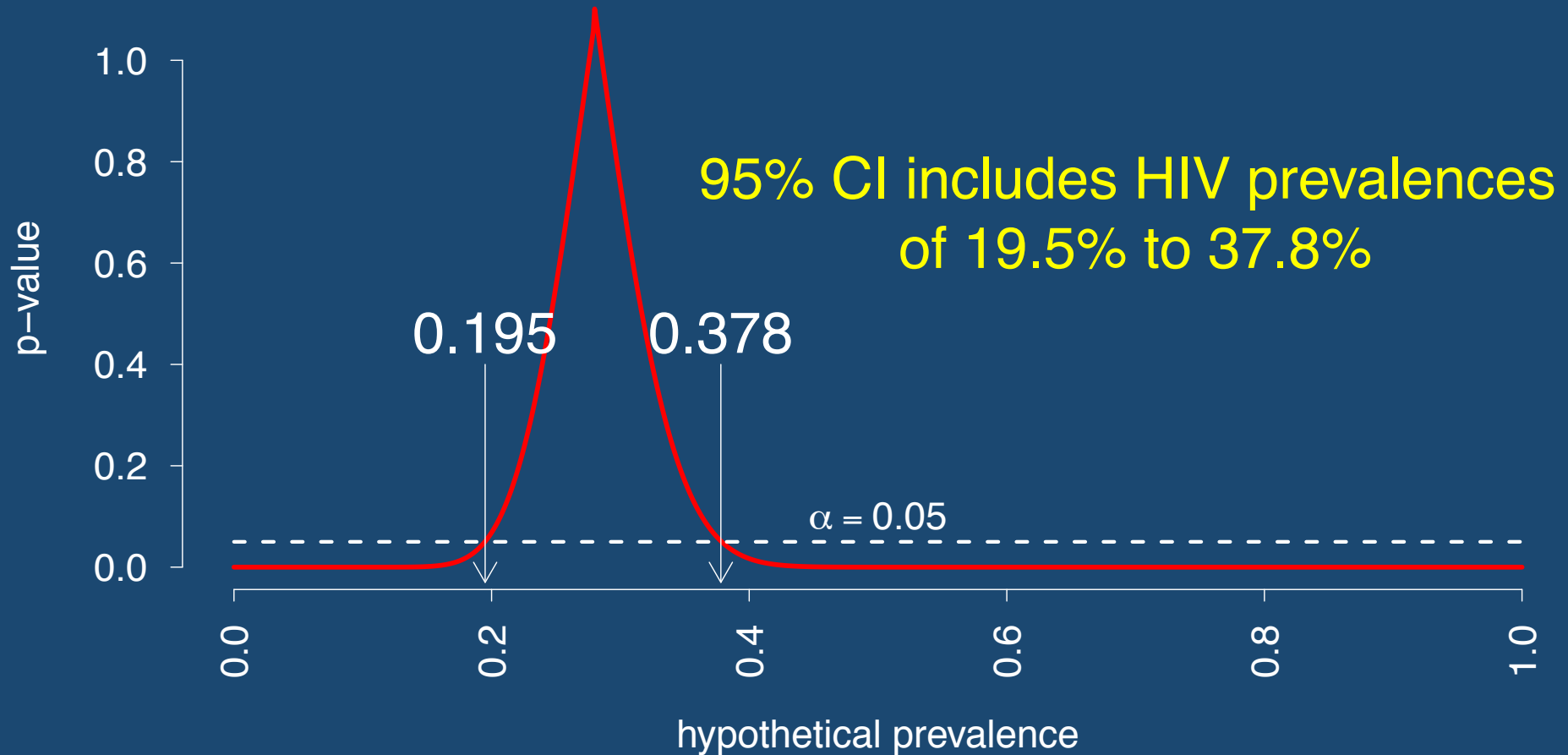
hypothetical prevalence: 40 %

$p = 0.0169$



number HIV+

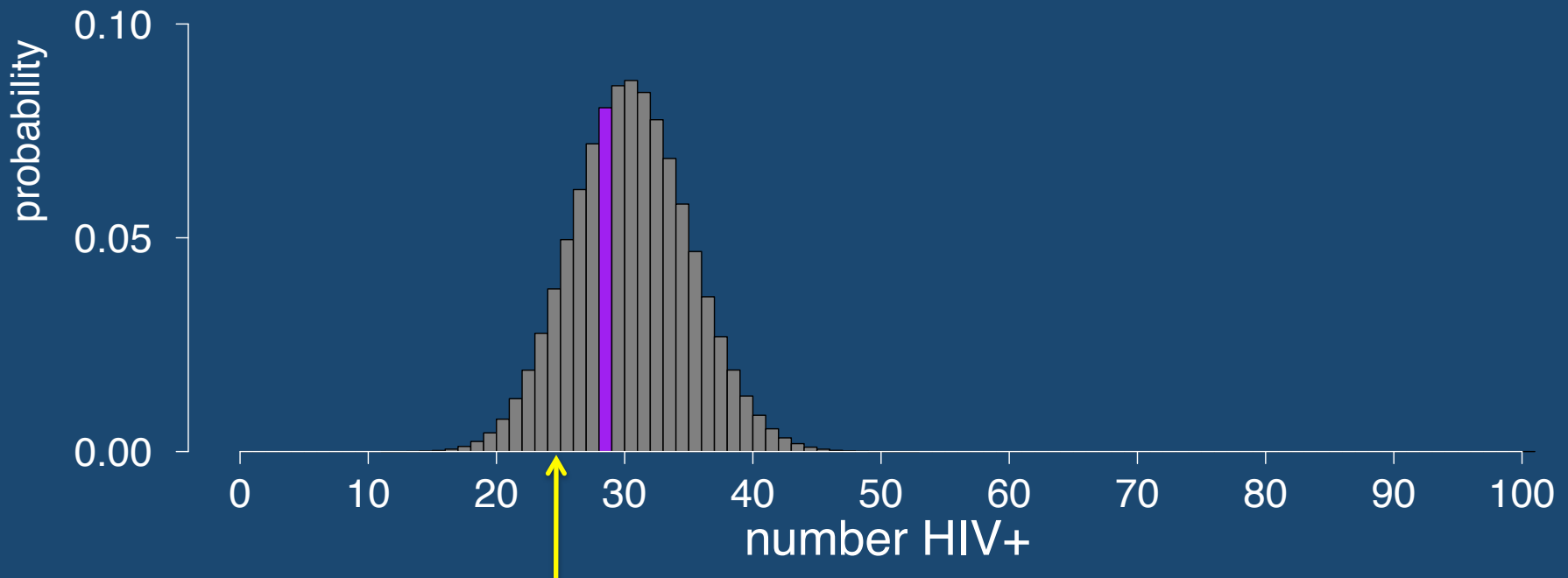
Which hypotheses do we NOT reject: **CONFIDENCE INTERVAL**



Let's take another approach

hypothetical prevalence: 30 %

$$\text{dbinom}(28, 100, 0.3) = 0.0804$$

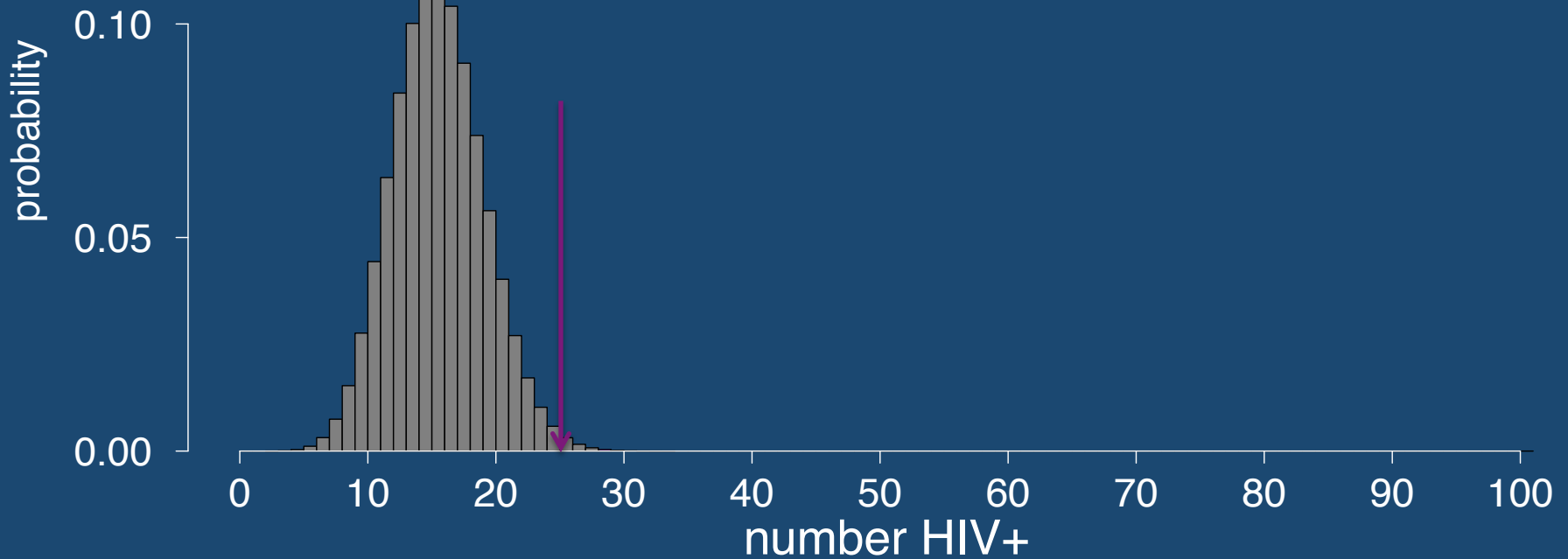


We sample 100 people once and 28 are positive:

```
> rbinom(n = 1, size = 100, prob = .3)
[1] 28
```

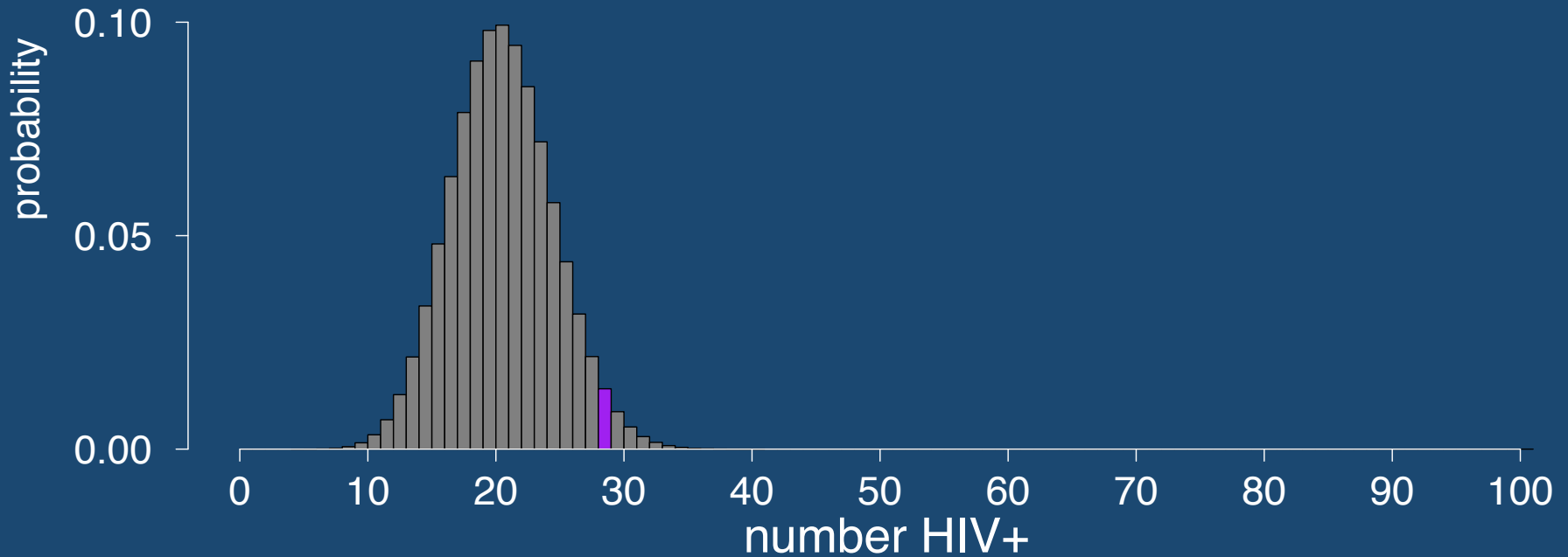

hypothetical prevalence: 15 %

$\text{dbinom}(28, 100, 0.15) = 0.000353$



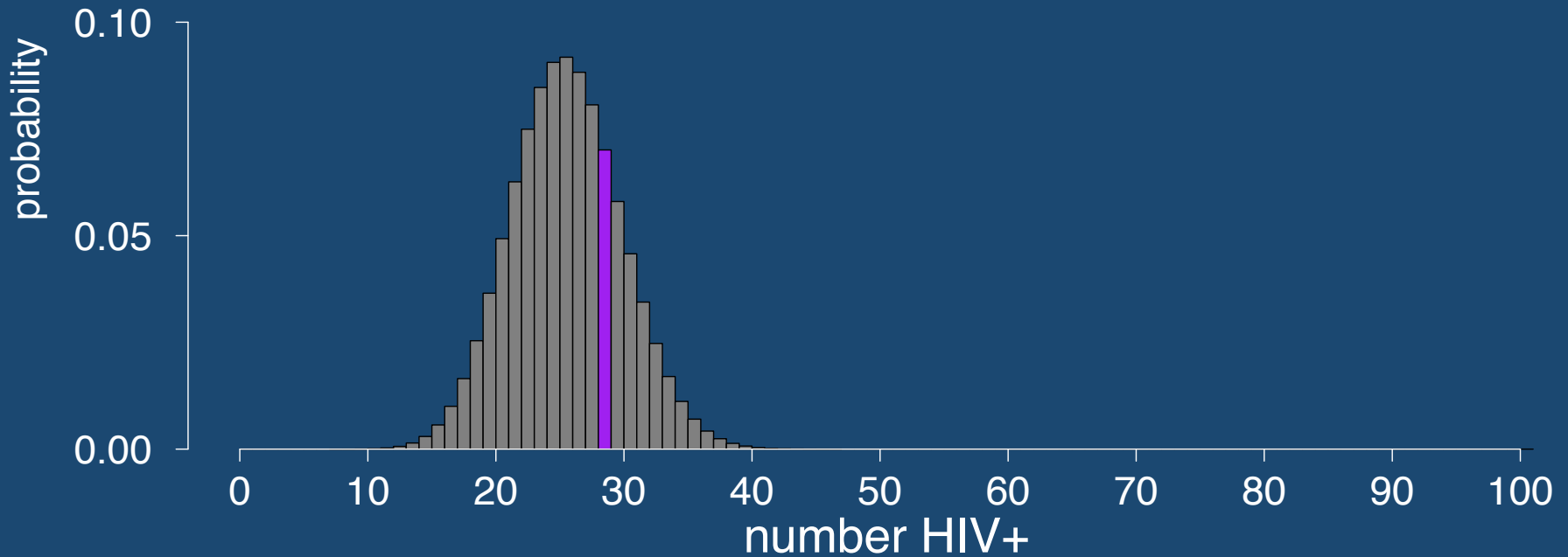
hypothetical prevalence: 20 %

$$\text{dbinom}(28, 100, 0.2) = 0.0141$$



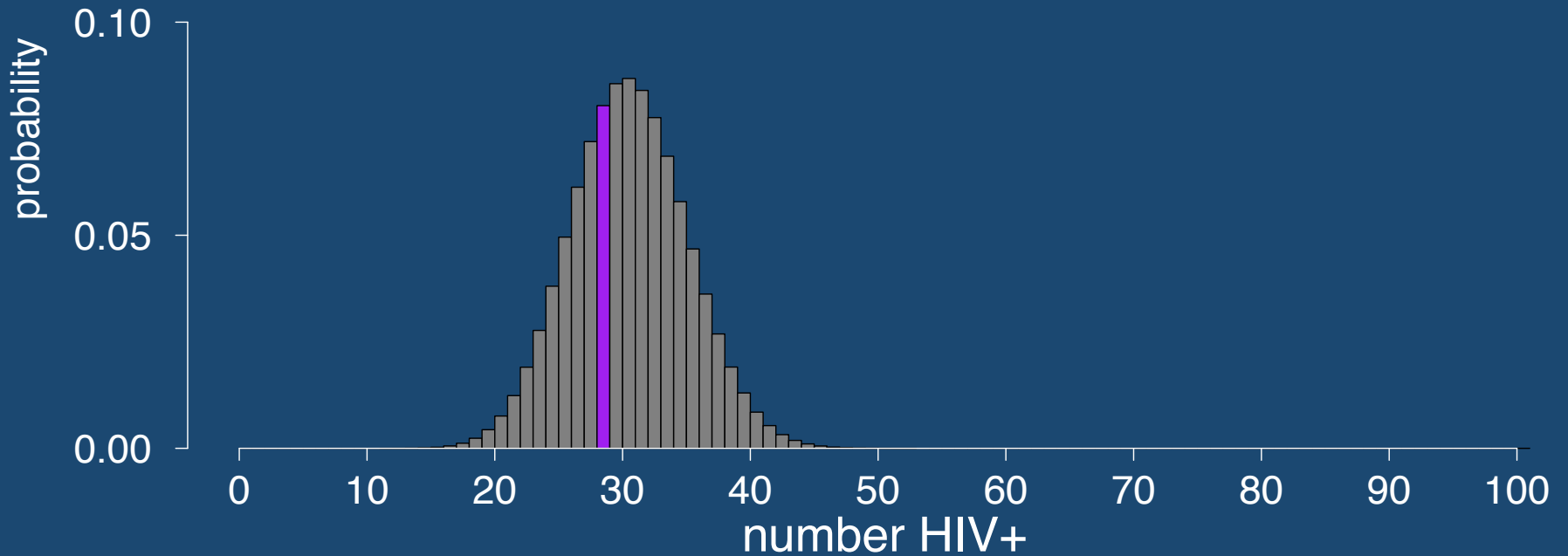
hypothetical prevalence: 25 %

$\text{dbinom}(28, 100, 0.25) = 0.0701$



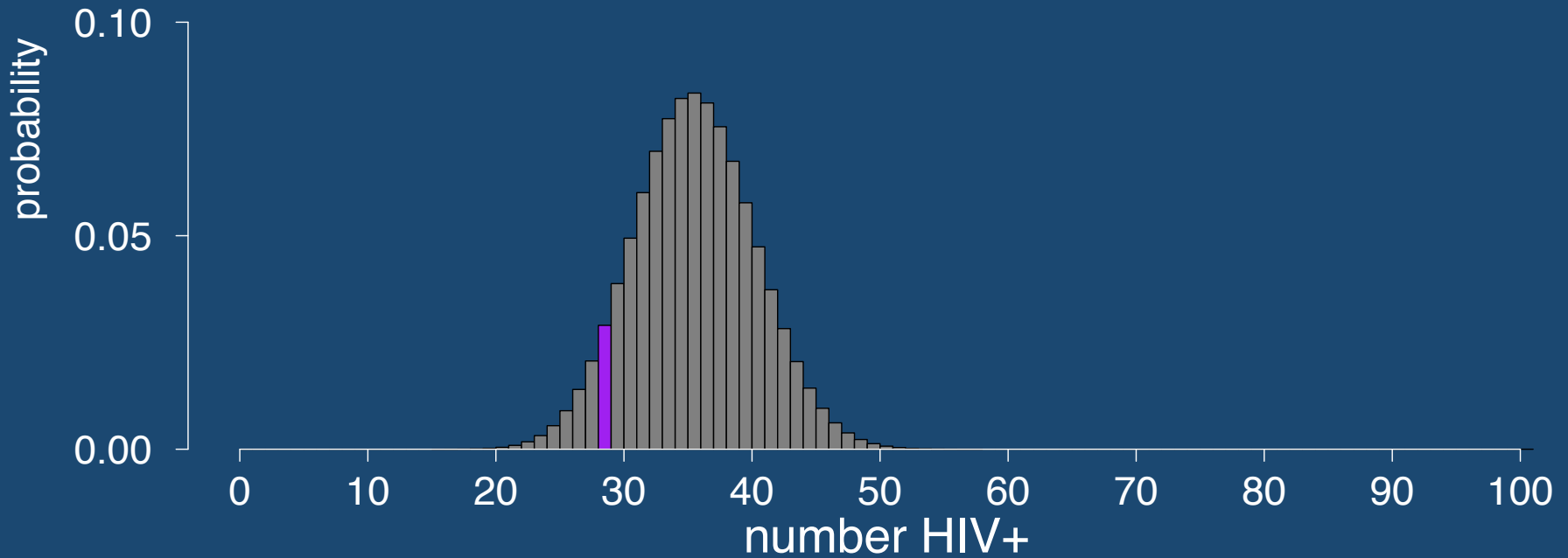
hypothetical prevalence: 30 %

$\text{dbinom}(28, 100, 0.3) = 0.0804$



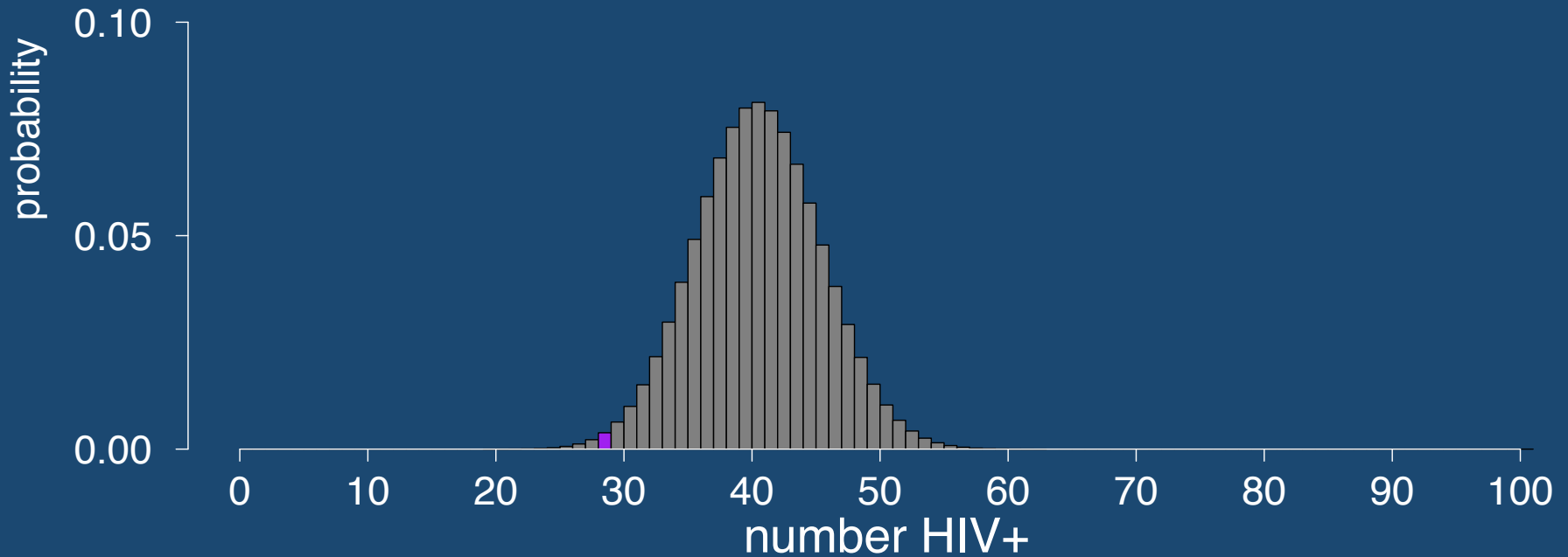
hypothetical prevalence: 35 %

$$\text{dbinom}(28, 100, 0.35) = 0.029$$



hypothetical prevalence: 40 %

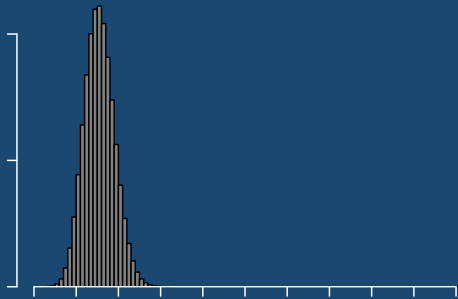
$\text{dbinom}(28, 100, 0.4) = 0.00383$



Which prevalence gives the greatest probability of observing **exactly** 28/100?

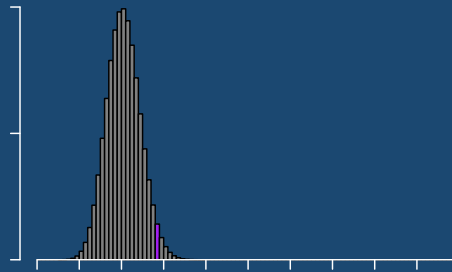
hypothetical prevalence: 15 %

0.000353



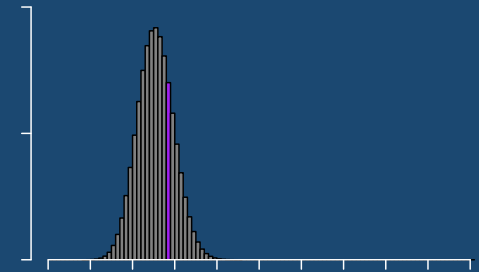
hypothetical prevalence: 20 %

0.0141



hypothetical prevalence: 25 %

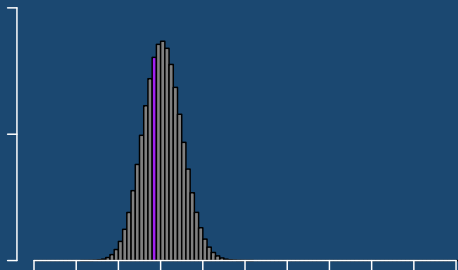
0.0701



probability

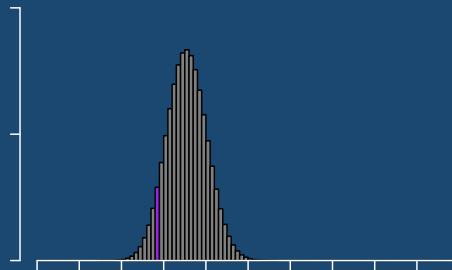
hypothetical prevalence: 30 %

0.0804



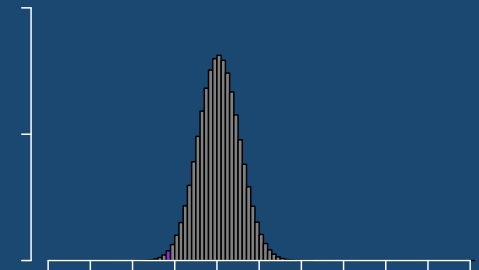
hypothetical prevalence: 35 %

0.029



hypothetical prevalence: 40 %

0.00383

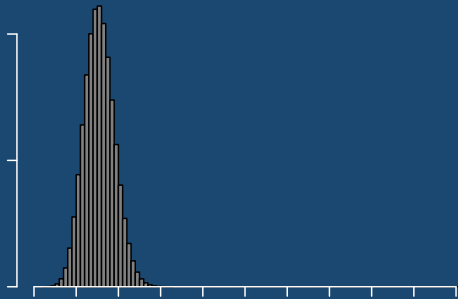


number HIV+

Which of these prevalence values is most likely given our data?

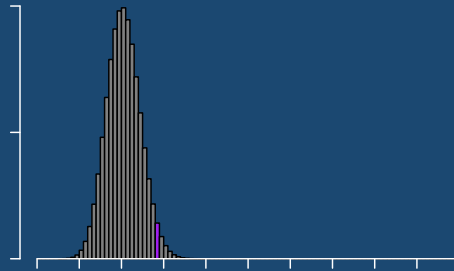
hypothetical prevalence: 15 %

0.000353



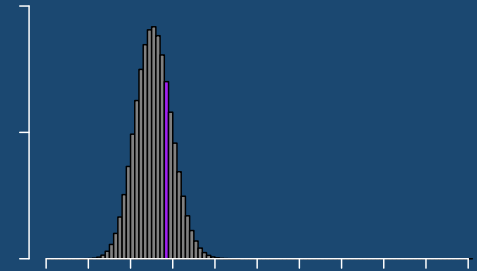
hypothetical prevalence: 20 %

0.0141



hypothetical prevalence: 25 %

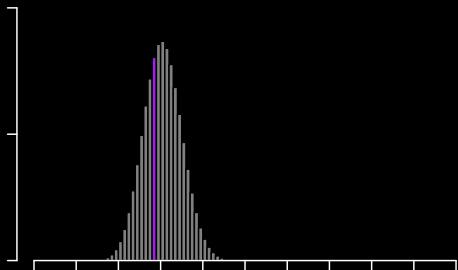
0.0701



probability

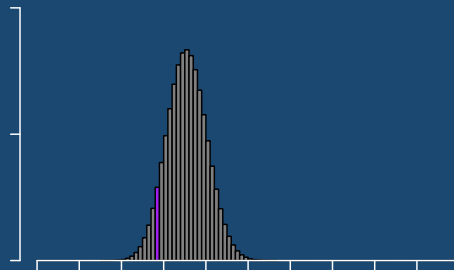
hypothetical prevalence: 30 %

0.0804



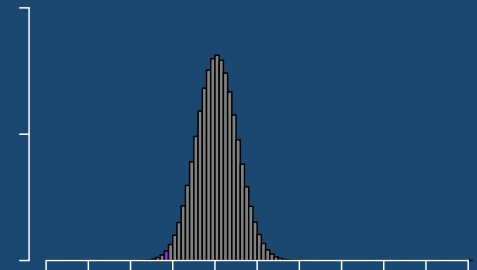
hypothetical prevalence: 35 %

0.029



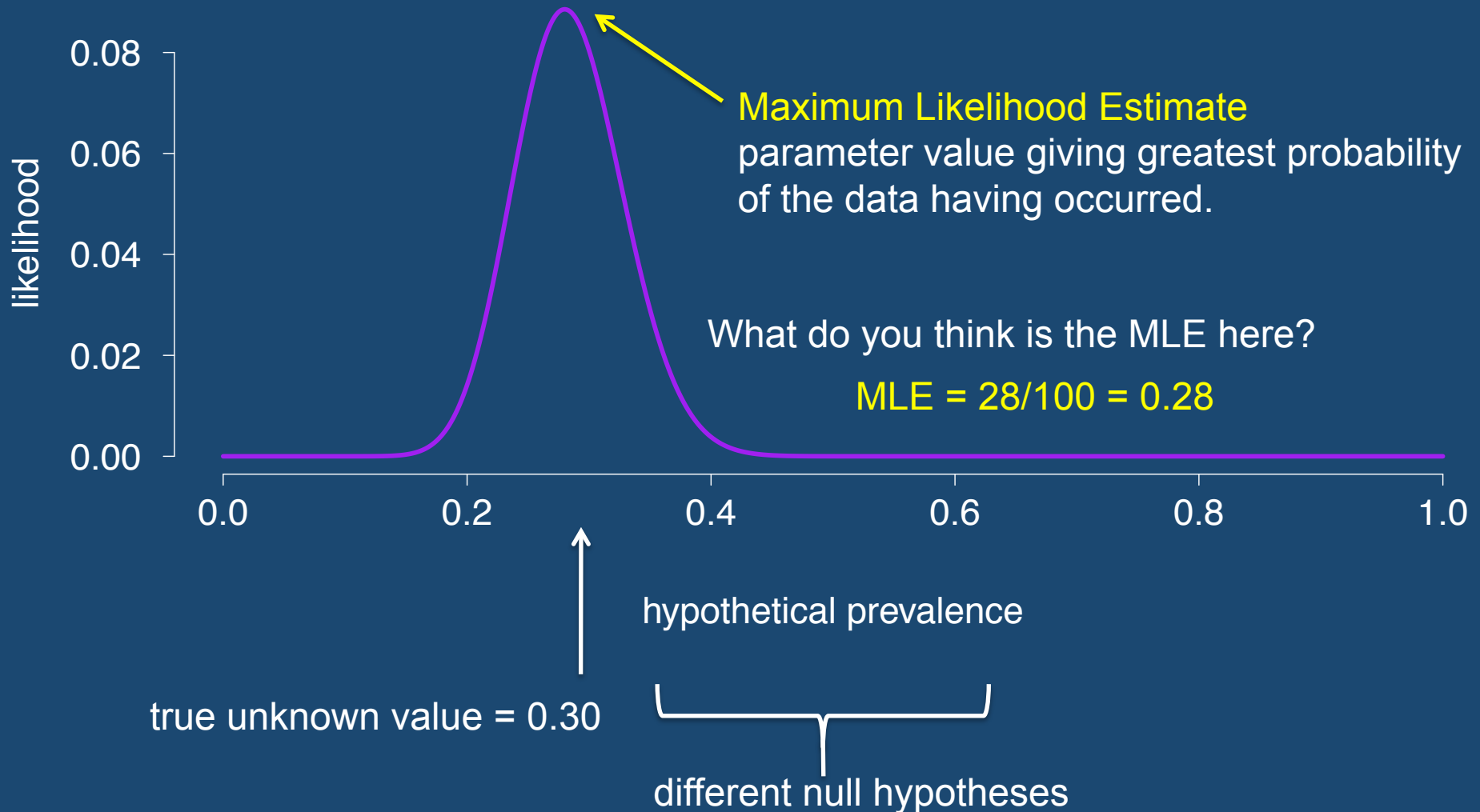
hypothetical prevalence: 40 %

0.00383



number HIV+

$p(\text{our data given prevalence}) = \text{LIKELIHOOD}$



Defining Likelihood

- $L(\text{parameter} \mid \text{data}) = p(\text{data} \mid \text{parameter})$

- Not a probability distribution.

function of x



PDF: $f(x|p) = \binom{n}{x} (p)^x (1 - p)^{n-x}$

- Probabilities taken from many different distributions.

LIKELIHOOD: $L(p|x) = \binom{n}{x} (p)^x (1 - p)^{n-x}$

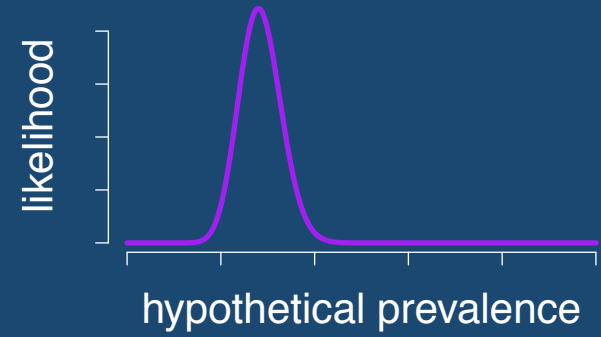


function of p

Deriving the Maximum Likelihood Estimate

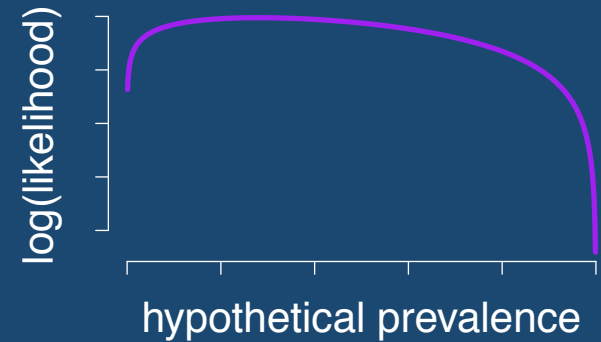
maximize

$$L(p) = \binom{n}{x} (p)^x (1-p)^{n-x}$$



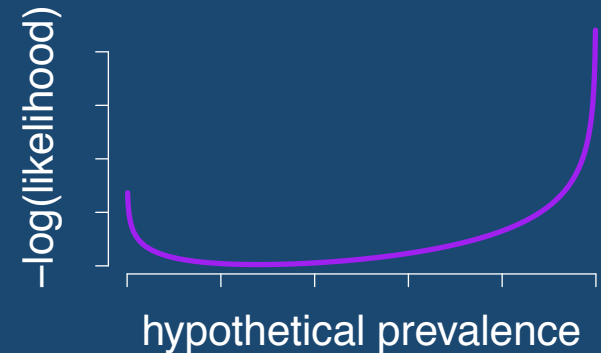
maximize

$$\log(L(p)) = \log \left[\binom{n}{x} (p)^x (1-p)^{n-x} \right]$$



minimize

$$l(p) = -\log \left[\binom{n}{x} (p)^x (1-p)^{n-x} \right]$$



Deriving the **Maximum Likelihood Estimate**

$$l(p) = -\log(L(p)) = -\log \left[\binom{n}{x} (p)^x (1-p)^{n-x} \right]$$

Deriving the Maximum Likelihood Estimate

$$l(p) = -\log \binom{n}{x} - x \log(p) - (n - x) \log(1 - p)$$

$$\frac{dl(p)}{dp} = 0 \quad -\frac{x}{p} - \frac{-(n - x)}{(1 - p)}$$

$$0 = -x + \hat{p}n$$

$$0 = -\frac{x}{\hat{p}} + \frac{(n - x)}{(1 - \hat{p})}$$

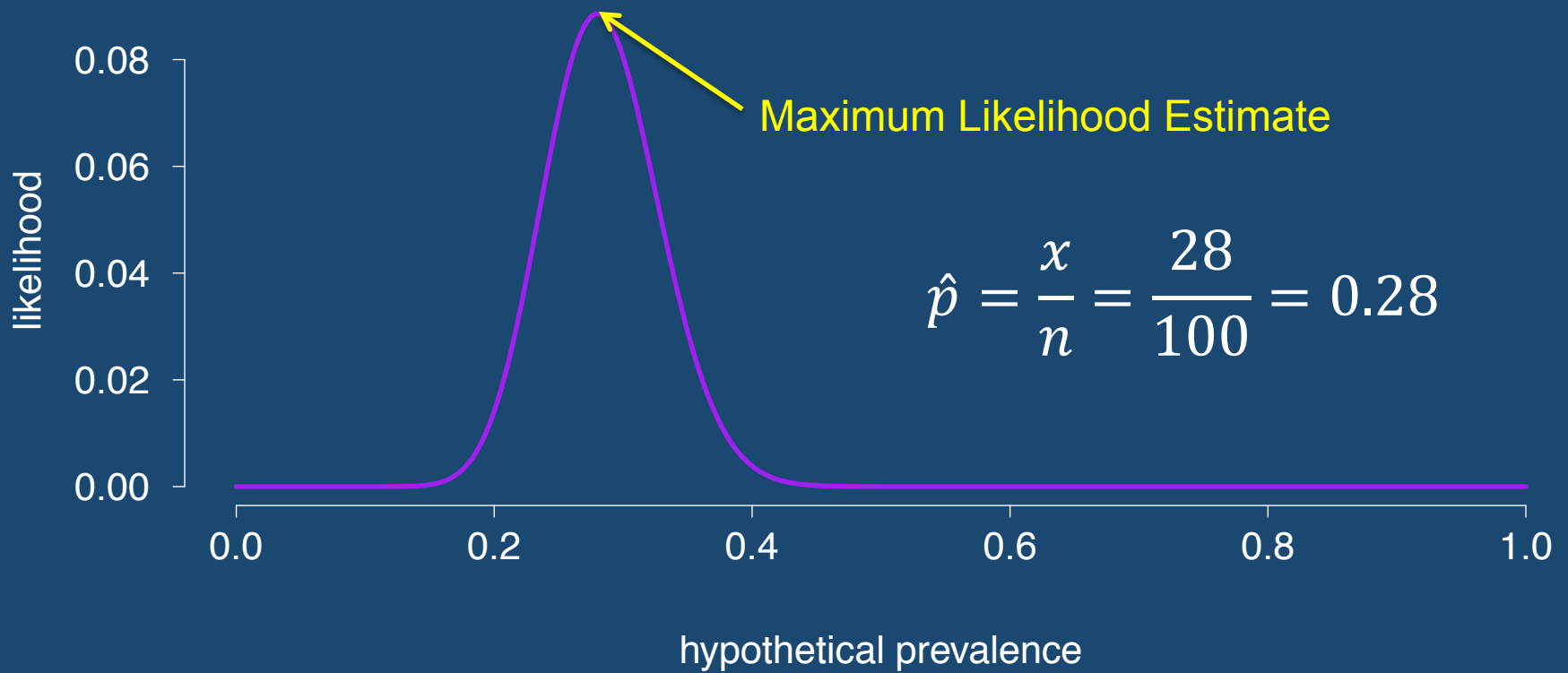
$$0 = \frac{-(1 - \hat{p})x + \hat{p}(n - x)}{(1 - \hat{p})}$$

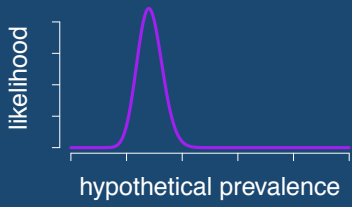
$$\hat{p} = \frac{x}{n}$$

$$0 = -x + \cancel{\hat{p}x} + \hat{p}n - \cancel{\hat{p}x}$$

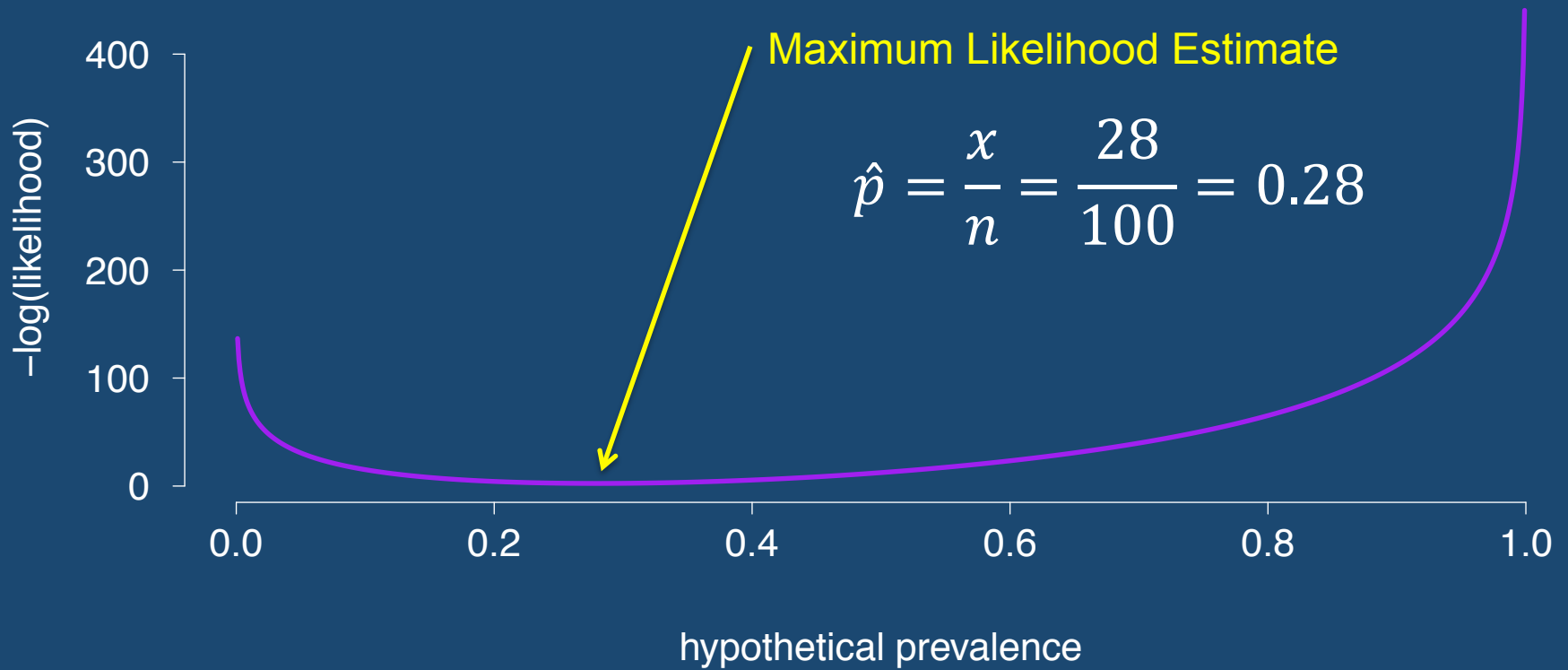
The proportion of positives!

Likelihood





we usually minimize the $-\log(\text{likelihood})$



Building Confidence Intervals

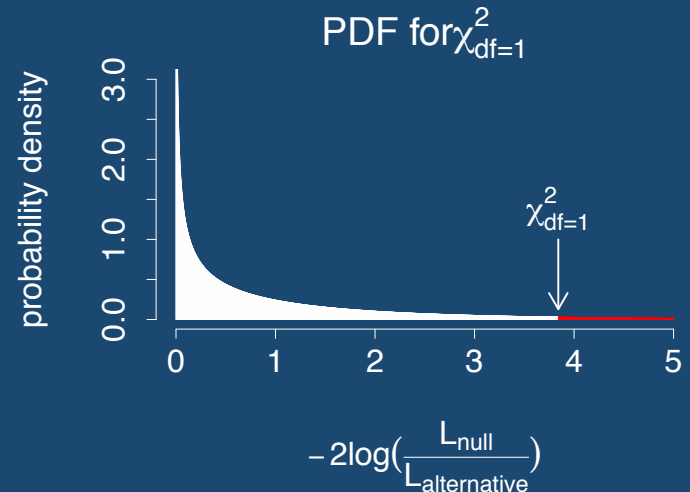
Likelihood Ratio Test

If the null hypothesis were true then

$$2 \log \left(\frac{L(\text{alternative hypothesis})}{L(\text{null hypothesis})} \right) \sim \chi_{df=1}^2$$

Why does this work?

- Adding irrelevant parameters *always* improves the fit.
- How much should fit improve due to chance alone by adding an irrelevant parameter?
- Fit improvement, as measured above, is approximately χ_{df}^2 distributed with $df =$ to the difference in parameters used to fit.



Building Confidence Intervals

Likelihood Ratio Test

If the null hypothesis were true then

$$2 \log \left(\frac{L(\text{alternative hypothesis})}{L(\text{null hypothesis})} \right) \sim \chi_{df=1}^2$$

$$2 \log(L_{MLE}) - 2 \log(L_{null}) \sim \chi_{df=1}^2$$

$$-2l_{MLE} + 2l_{null} \sim \chi_{df=1}^2$$

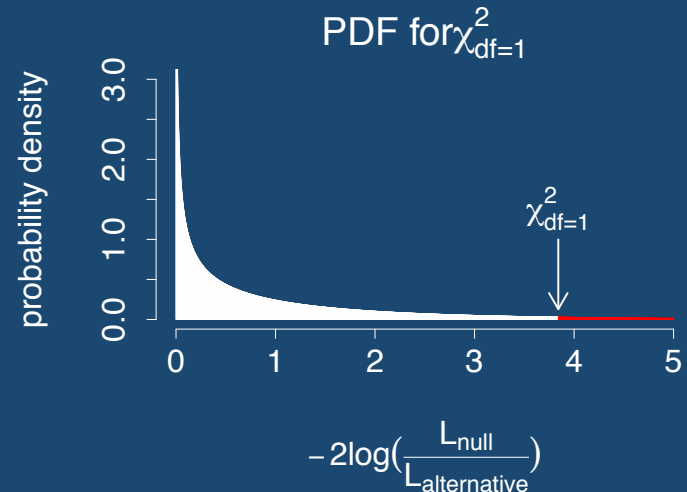
So if our $\alpha = .05$, then we reject any null hypothesis for which

$$-2l_{MLE} + 2l_{null} > \chi_{df=1, \alpha=.05}^2 = 3.84 \quad > \text{qchisq}(p = .95, df = 1) \\ [1] 3.841459$$

$$l_{null} - l_{MLE} > 1.92$$

If $\log(L_{MLE}) - \log(L_{null}) > 1.92$,

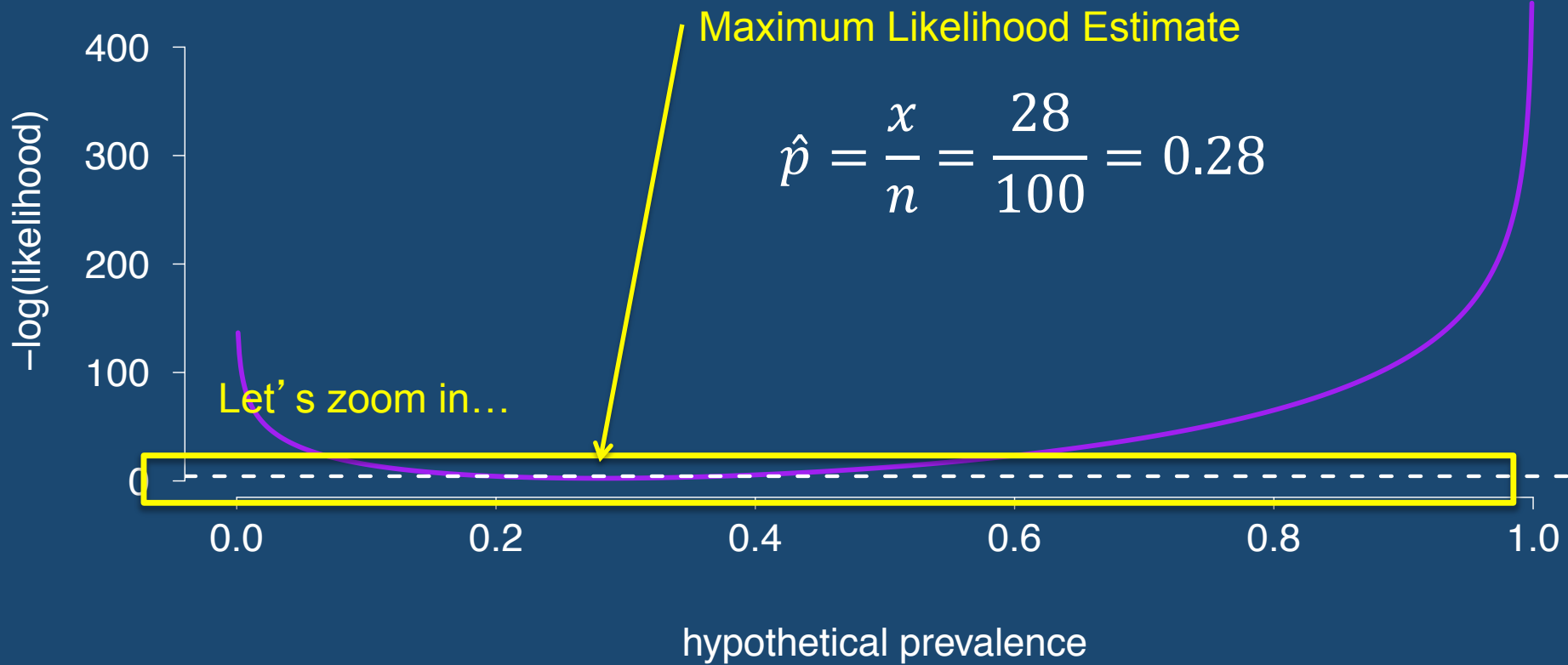
we reject that null hypothesis.



Building Confidence Intervals

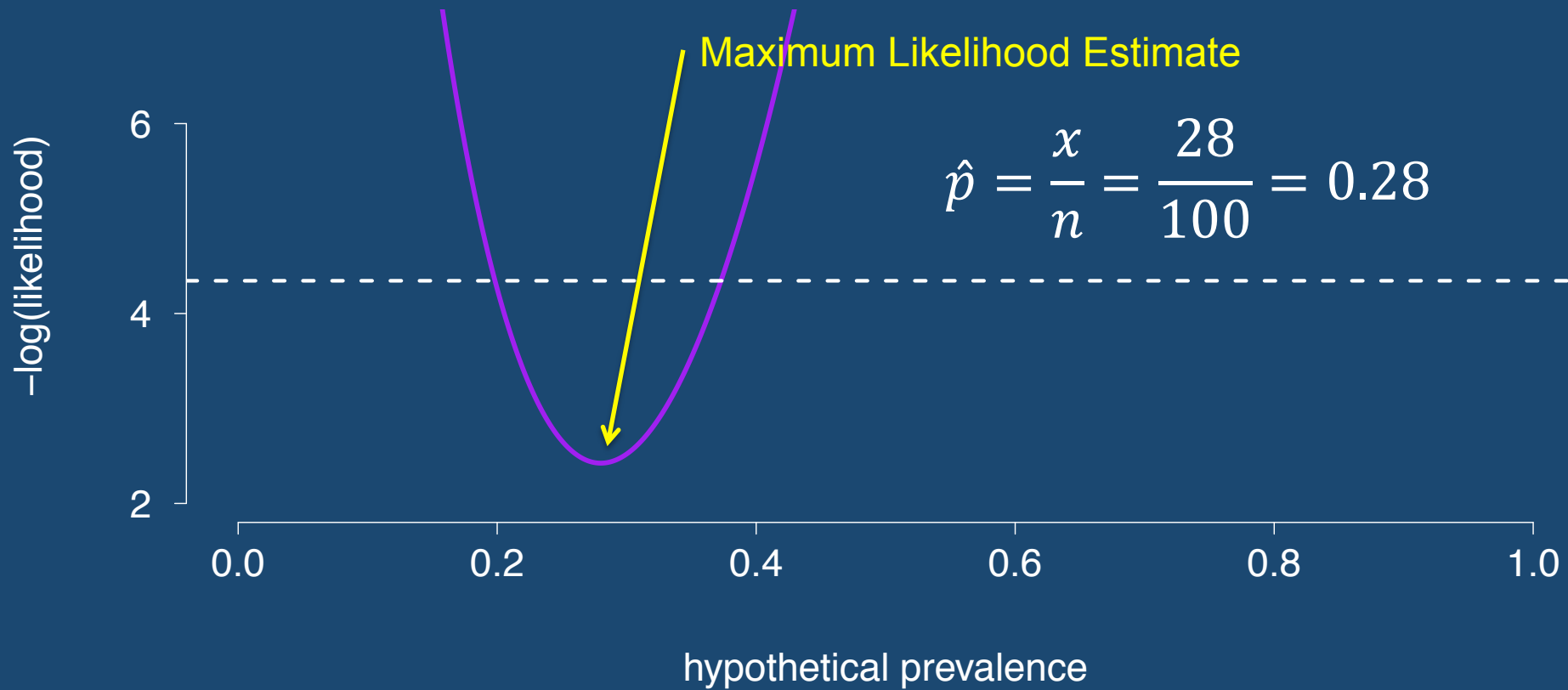
Likelihood Ratio Test

we usually minimize the $-\log(\text{likelihood})$



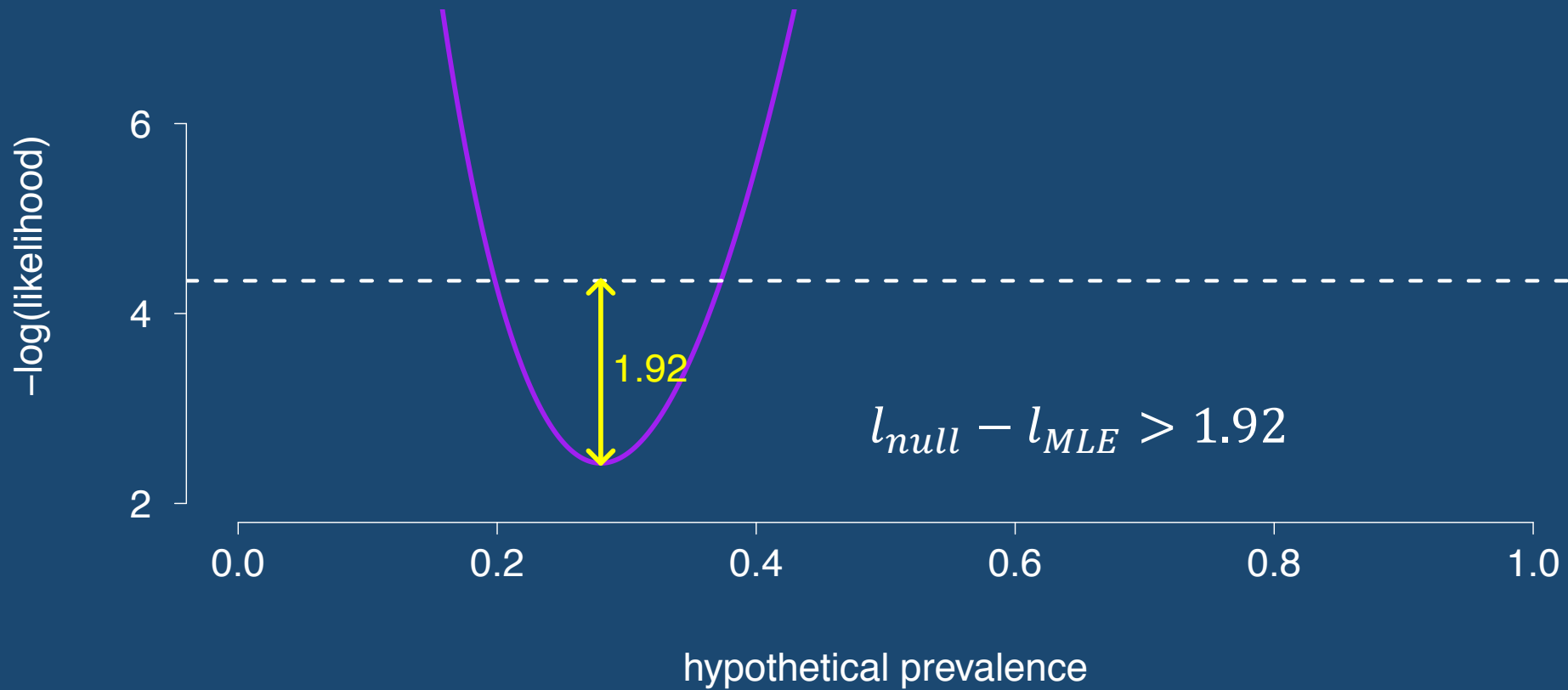
Building Confidence Intervals

Likelihood Ratio Test



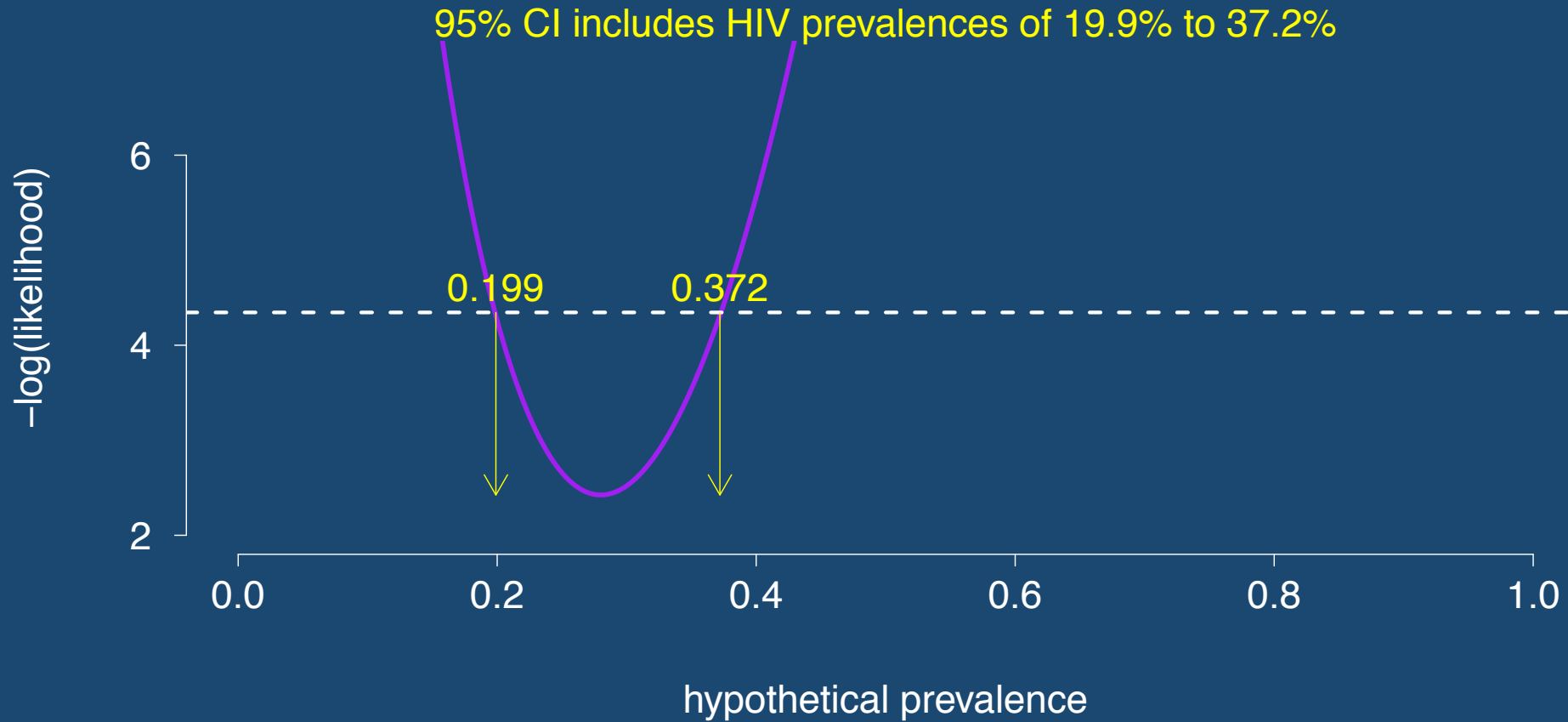
Building Confidence Intervals

Likelihood Ratio Test

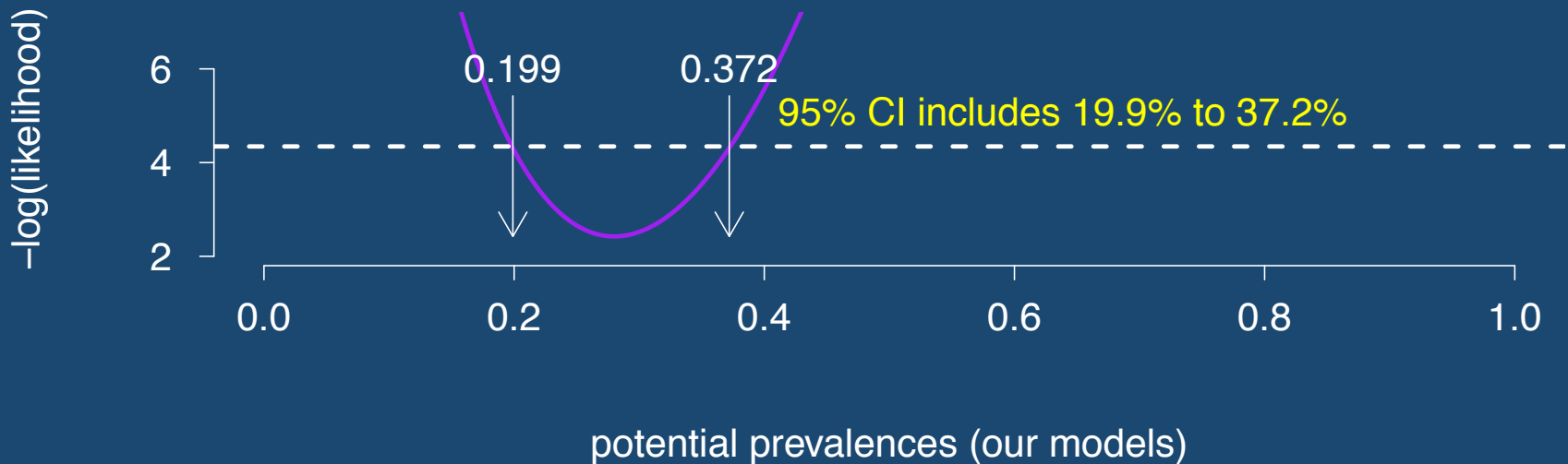
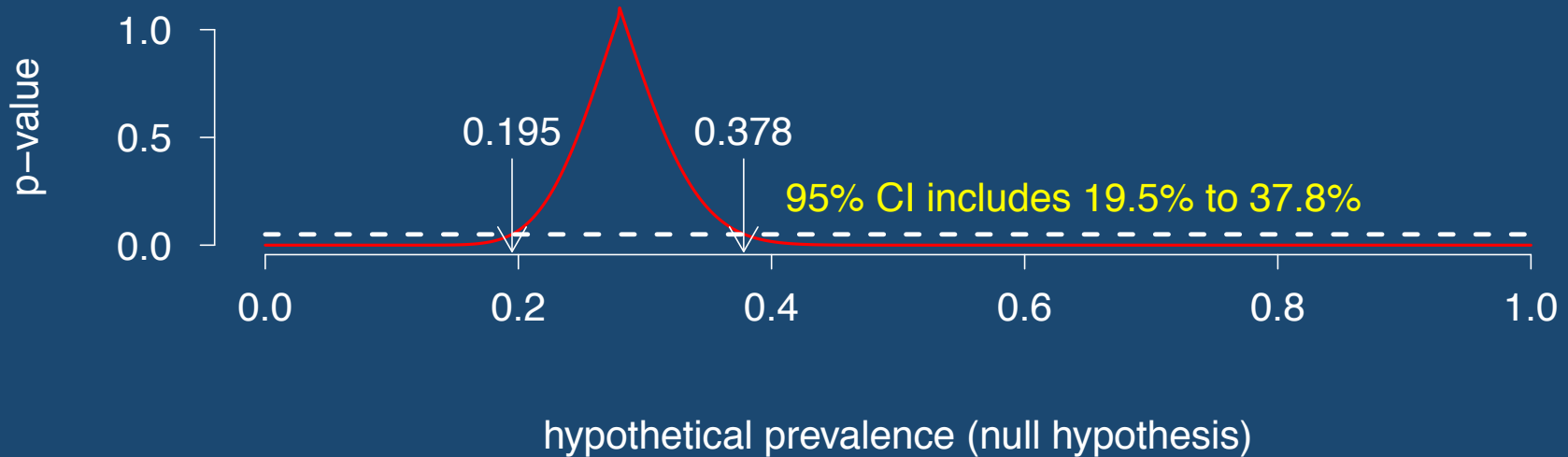


Building Confidence Intervals

Likelihood Ratio Test



Comparing Confidence Intervals



Advantages of Likelihood

- Practical method for
estimating parameters
estimating variance of our estimates
- Easily adaptable to different probability
distributions & dynamic models

Likelihood Profile Confidence Intervals

- Profiles one parameter,
with all others flexibly fit.



This presentation is made available through a Creative Commons Attribution-Noncommercial license. Details of the license and permitted uses are available at <http://creativecommons.org/licenses/by-nc/3.0/>



© 2010 Steve Bellan and the Meaningful Modeling of Epidemiological Data Clinic

Title: Introduction to Likelihood

Attribution: Steve Bellan, Clinic on the Meaningful Modeling of Epidemiological Data

Source URL: <http://lalashan.mcmaster.ca/theobio/mmed/index.php/>

For further information please contact Steve Bellan (sbellan@berkeley.edu).