# Stat310

## Estimation

Hadley Wickham

1. Recap
2. Method of moments
3. Maximum likelihood

# Recap

# Definitions

**Parameter space**: set of all possible parameter values

**Estimator**: function which takes data and gives best guess for parameter. Usually written with a hat. Is a **random variable**

**Point estimate**: single best guess.

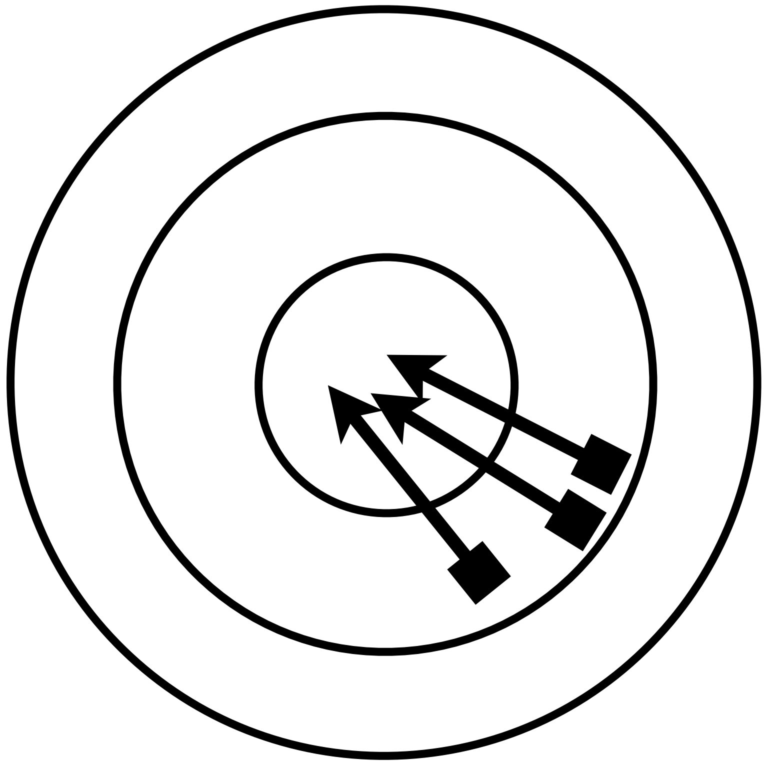
# Estimating marijuana usage of Rice students

Used three different estimators.

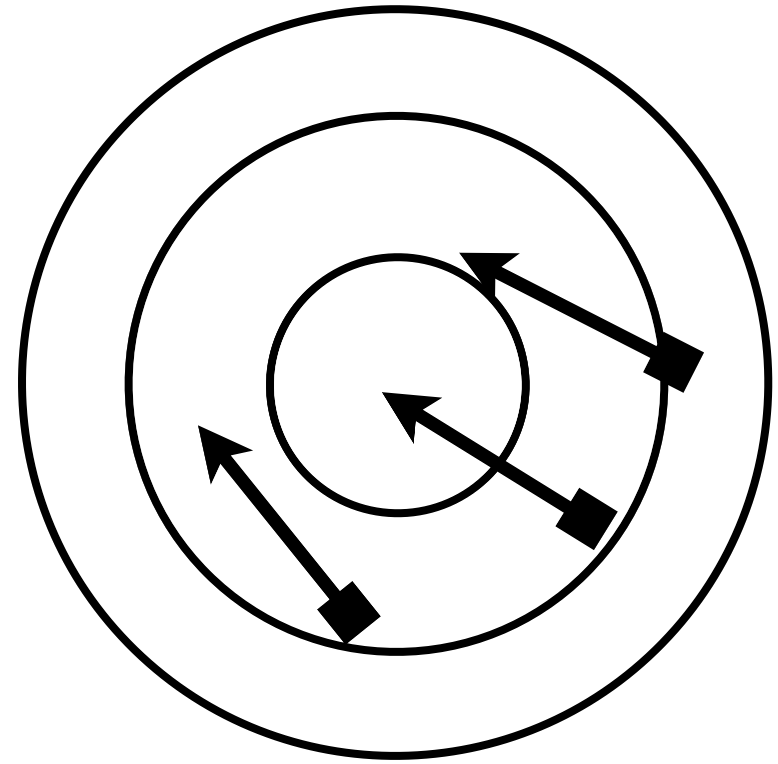
All gave different results, and would give different results if I repeated the experiment.

How do we know which one is best?

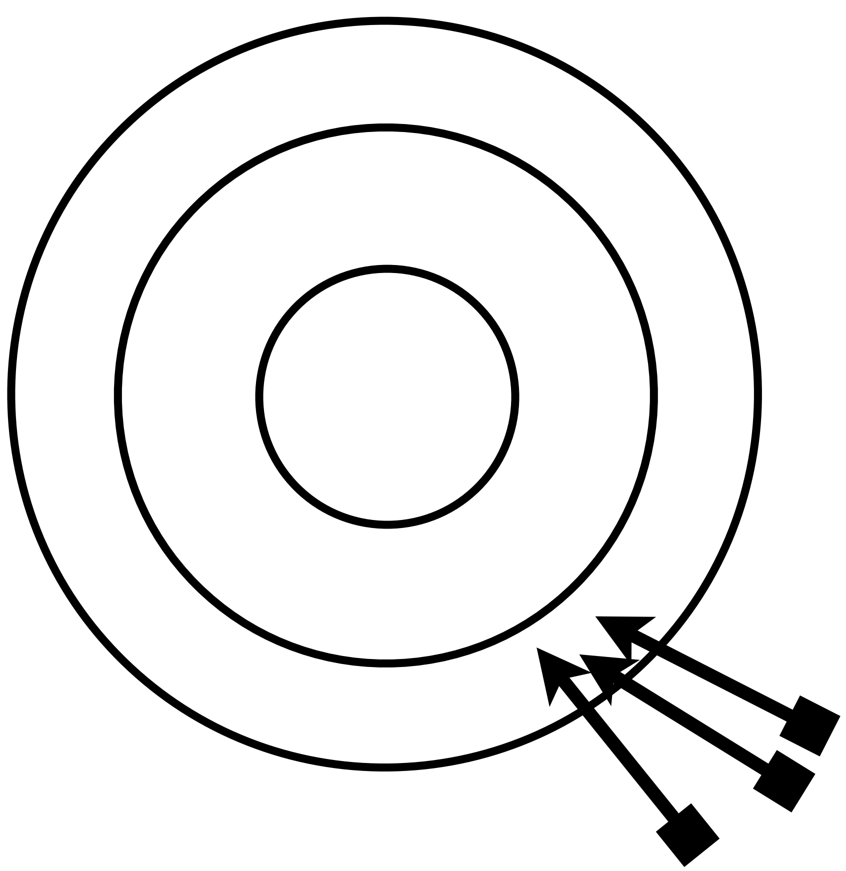
## Low bias, low variance



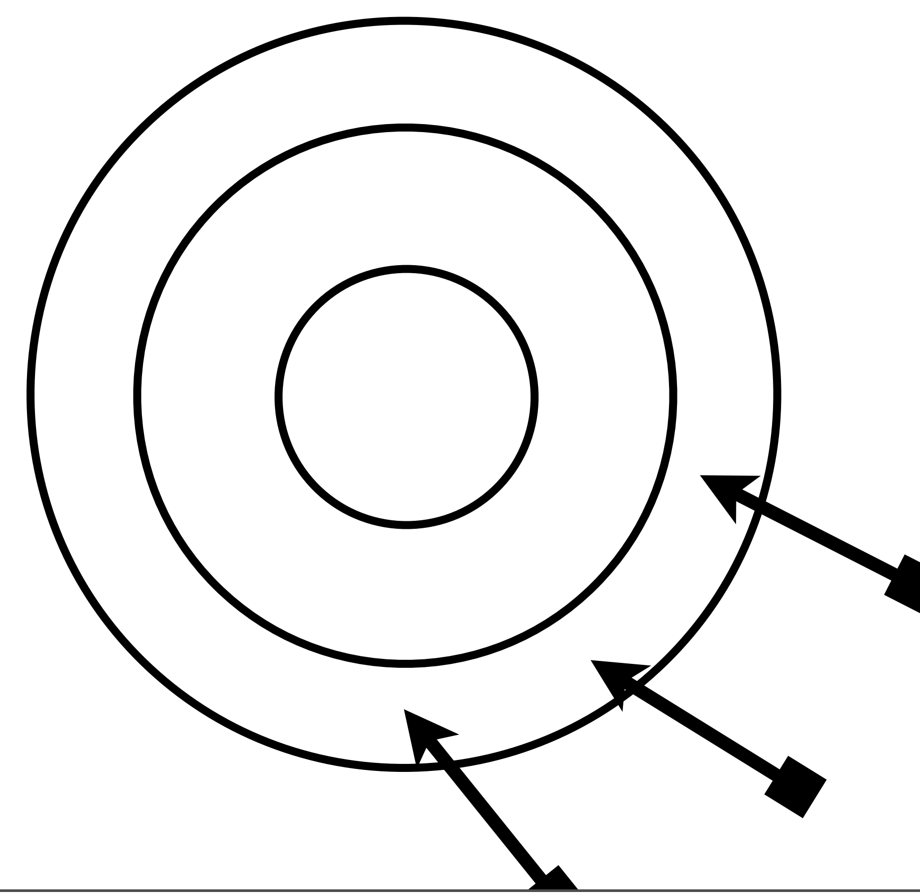
## Low bias, high variance



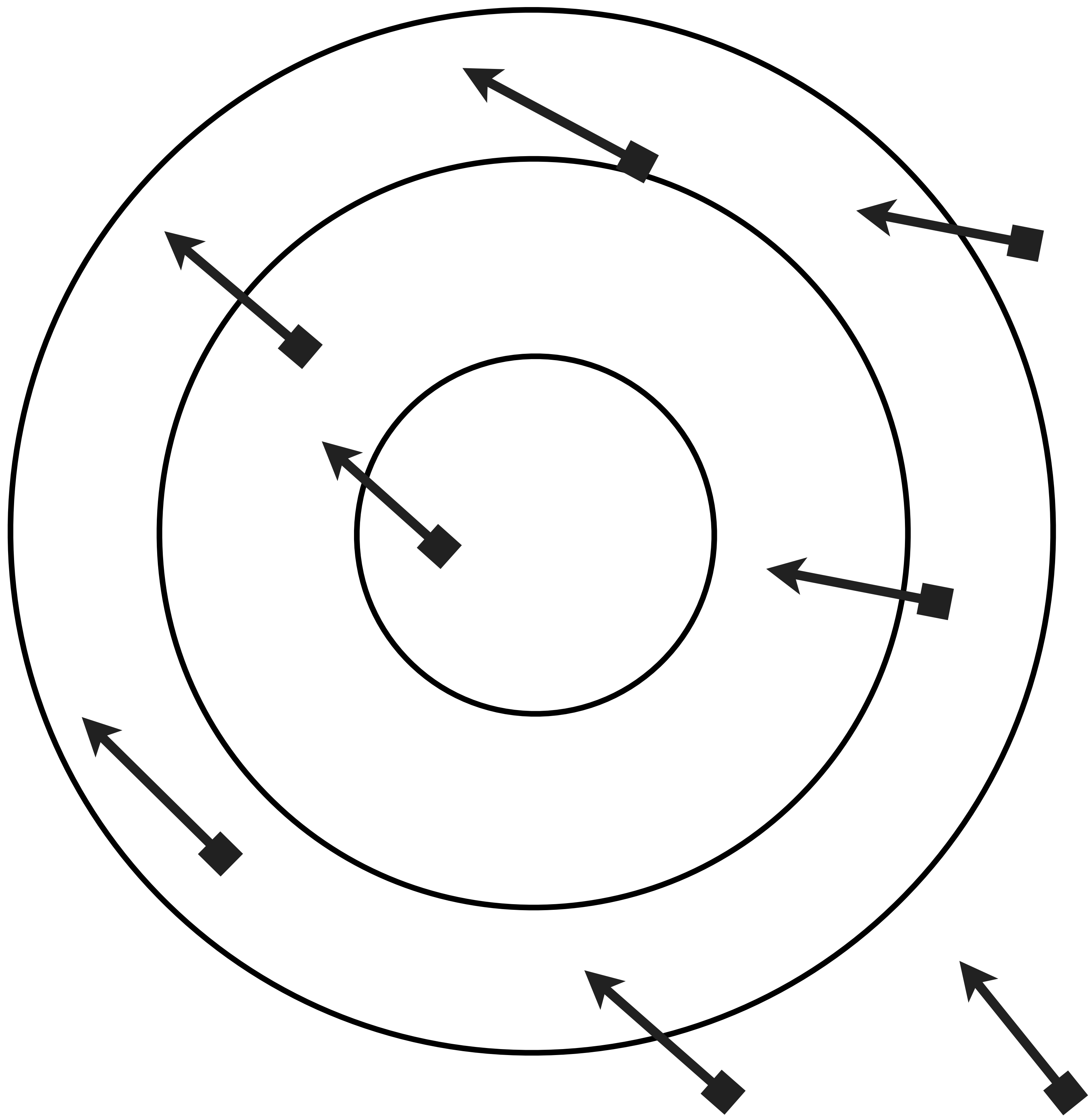
## High bias, low variance



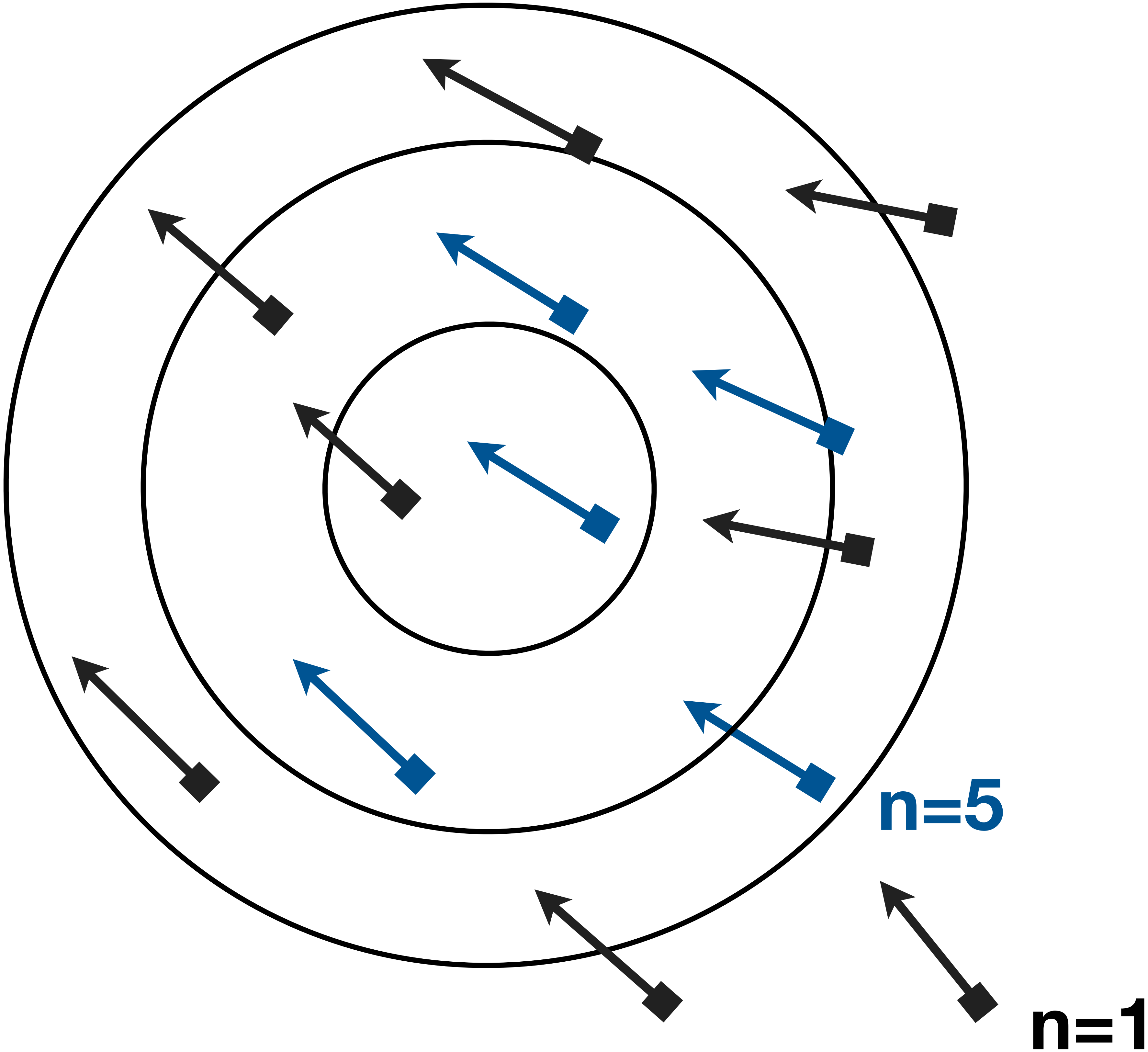
## High bias, high variance



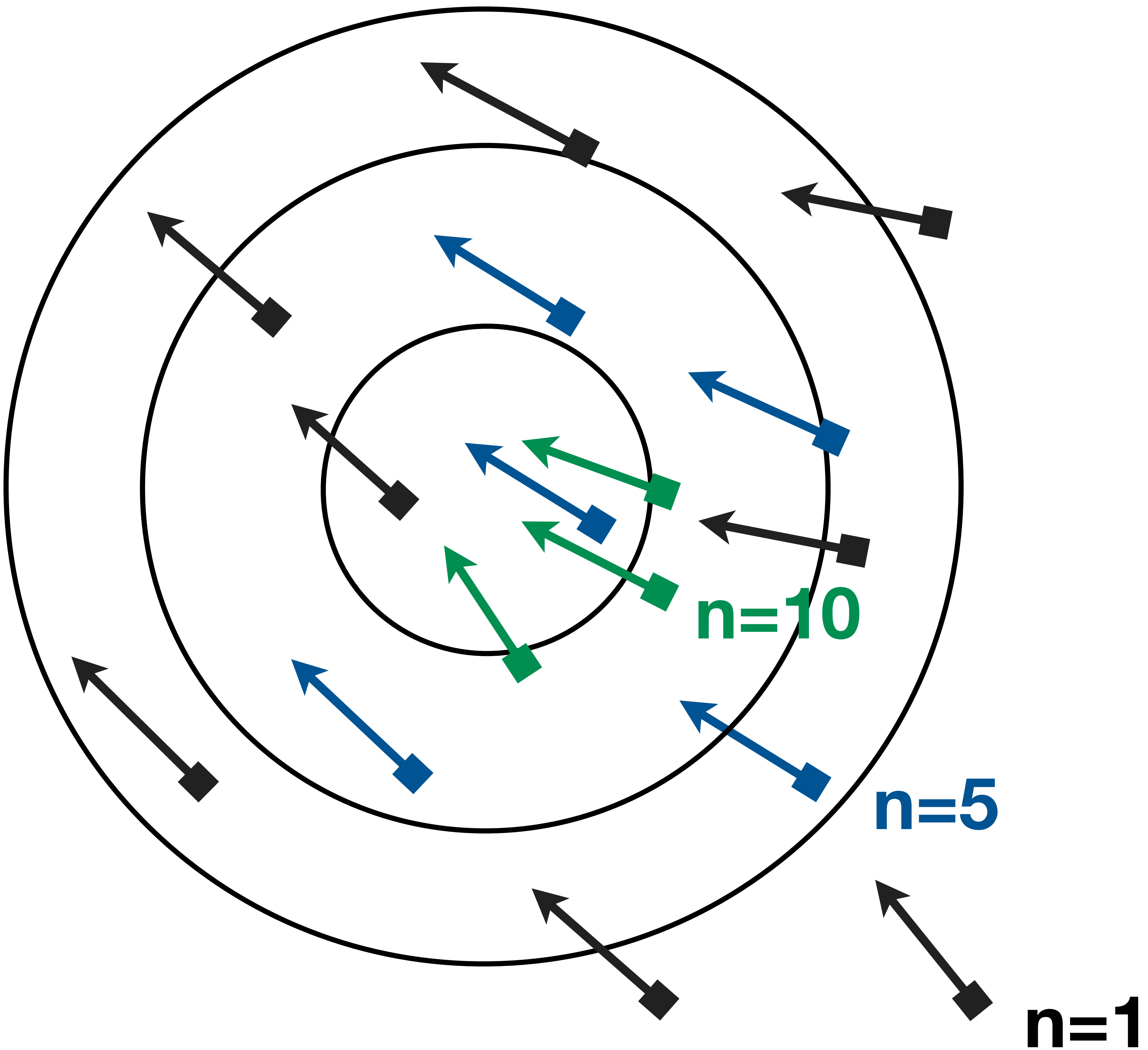
Consistent



# Consistent



# Consistent



# Goal

Develop a systematic way of creating new estimators, and deriving their distribution.

# Method of moments

# Method of moments

We know how to calculate sample moments (e.g. mean and variance of data)

We know what the moments of the distribution are in terms of the parameters.

Why not just match them up?

| | Distribution | Data |
|---|---|---|
| Mean | | |
| Variance | | |

# Gamma distribution

$X \sim \text{Gamma}(\alpha, \theta)$

$E(X) = \alpha\beta$  $Var(X) = \alpha\theta^2$

# Steps

- Write down formulas for mean and variance.
- Solve for the parameters
- Rewrite to use sample estimates
- (If there are more parameters, you'll need to use more moments, but that won't come up since we haven't learned any distributions with more than two parameters)

# Your turn

Find a method of moments estimator for $\lambda$ when the iid sequence is Poisson distributed.

Find another method of moments estimator.

# But...

Let X ~ Unif(0, θ)

What is a method of moments estimator for θ?

For one experiment we got the values 3, 5, 6, 18. What is the method of moments estimate for θ? Is it a good estimator?

# Pros/cons

Pro: Simple!

Con: Doesn’t always work.

Con: Often not best estimator

Con: Don’t know anything about the error in the estimate

# Maximum likelihood

# Basic idea

For a specific value of the parameter, what's the probability it generated this data?

We will only look at the case of one unknown parameter, but the method generalises to more than one.

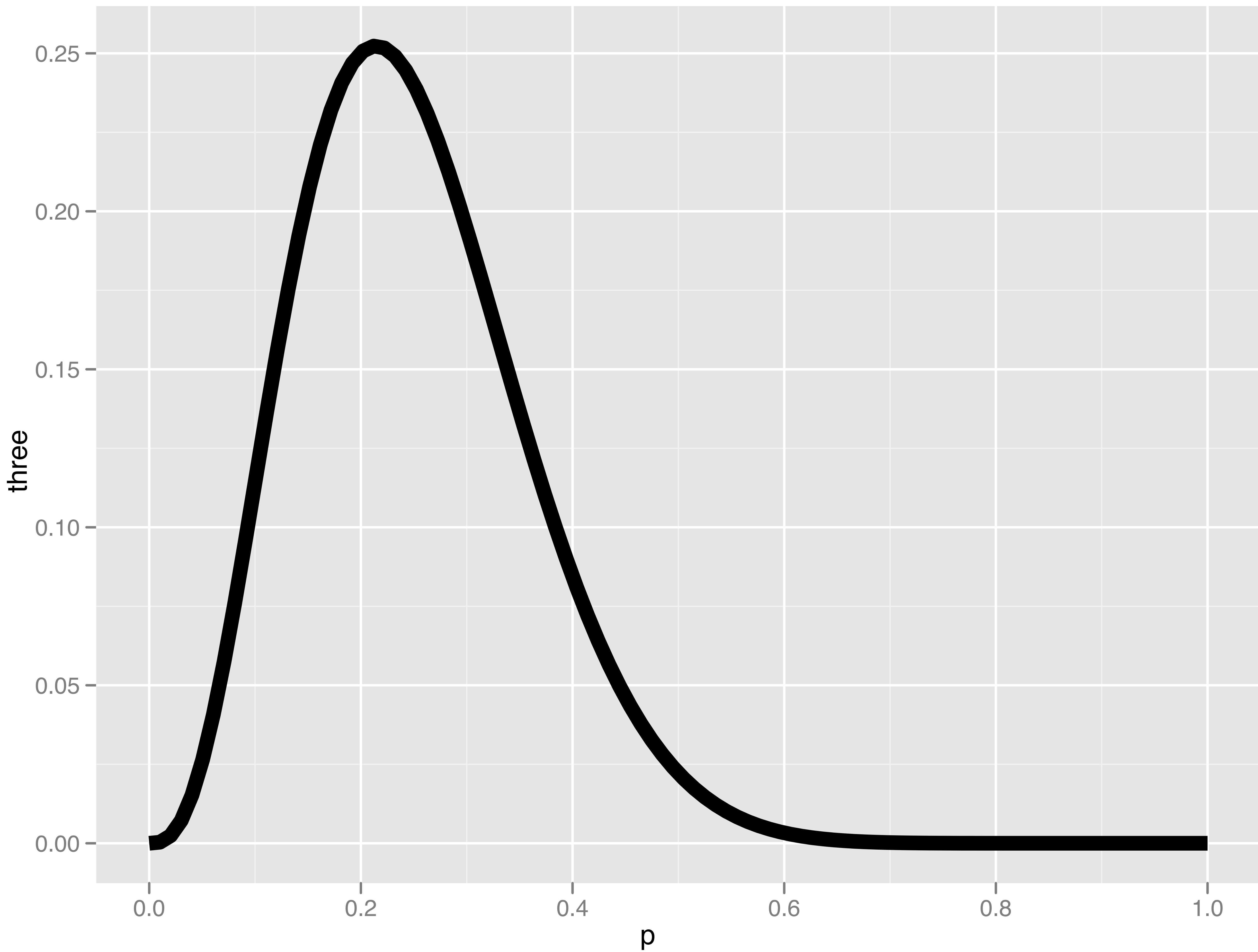
Amazing thing is that maximum likelihood estimators are approximately normally distributed (more on that next week)

# Example

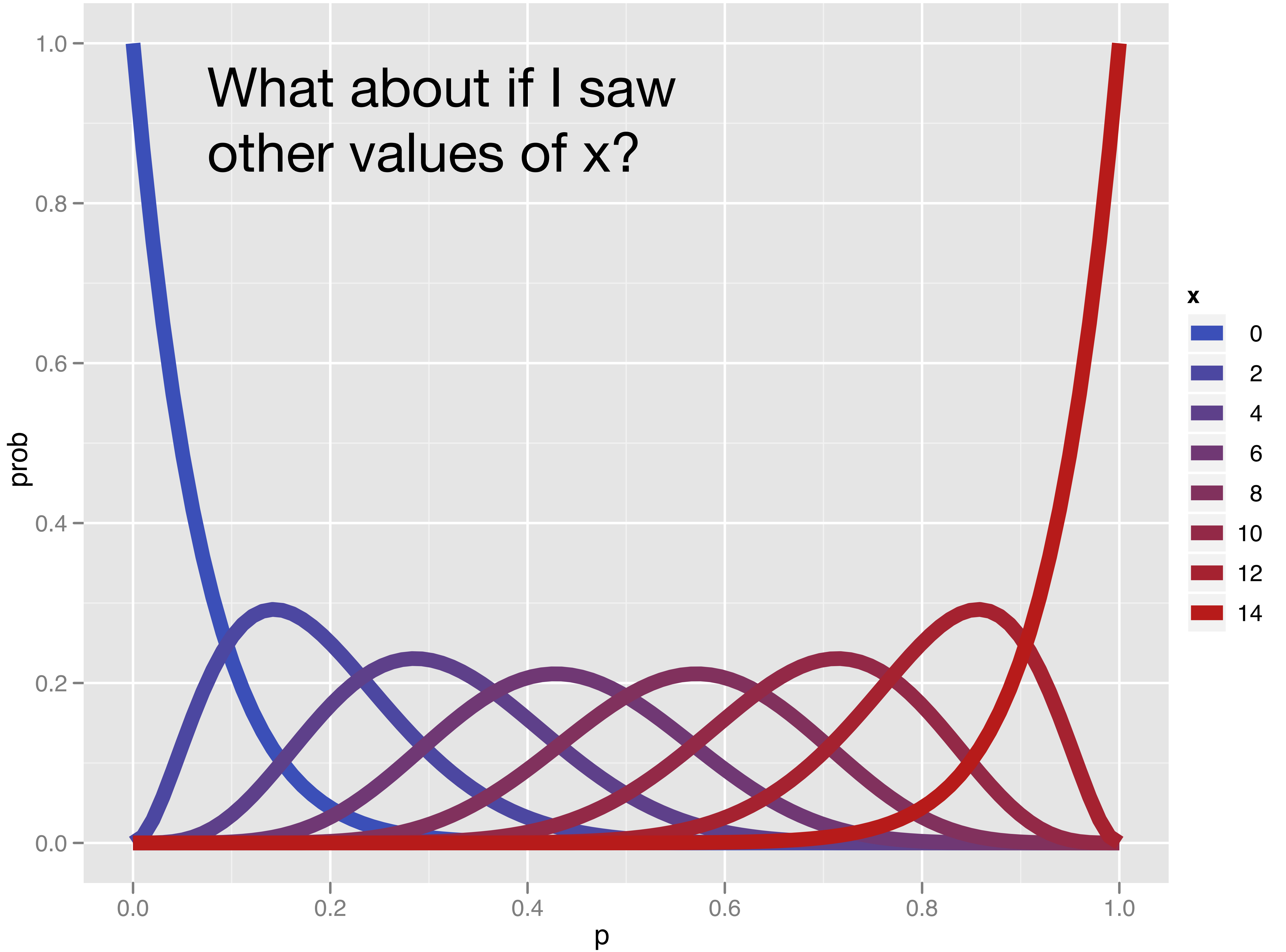
X ~ Binomial(14, p)

I do the experiment and get 3 successes.

For each value of p, what's the probability that I get three successes?



0.25
0.20
0.15
0.10
0.05
0.00
three
0.0
0.2
0.4
0.6
0.8
1.0
p


What about if I saw
other values of x?
prob
p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
x
0
2
4
6
8
10
12
14

# Likelihood

To emphasise the focus on a fixed parameter value, we normally call the probabilities **likelihoods**: how likely is it that this data was generated by that parameter value?

(Likelihoods are usually very very small)

# Your turn

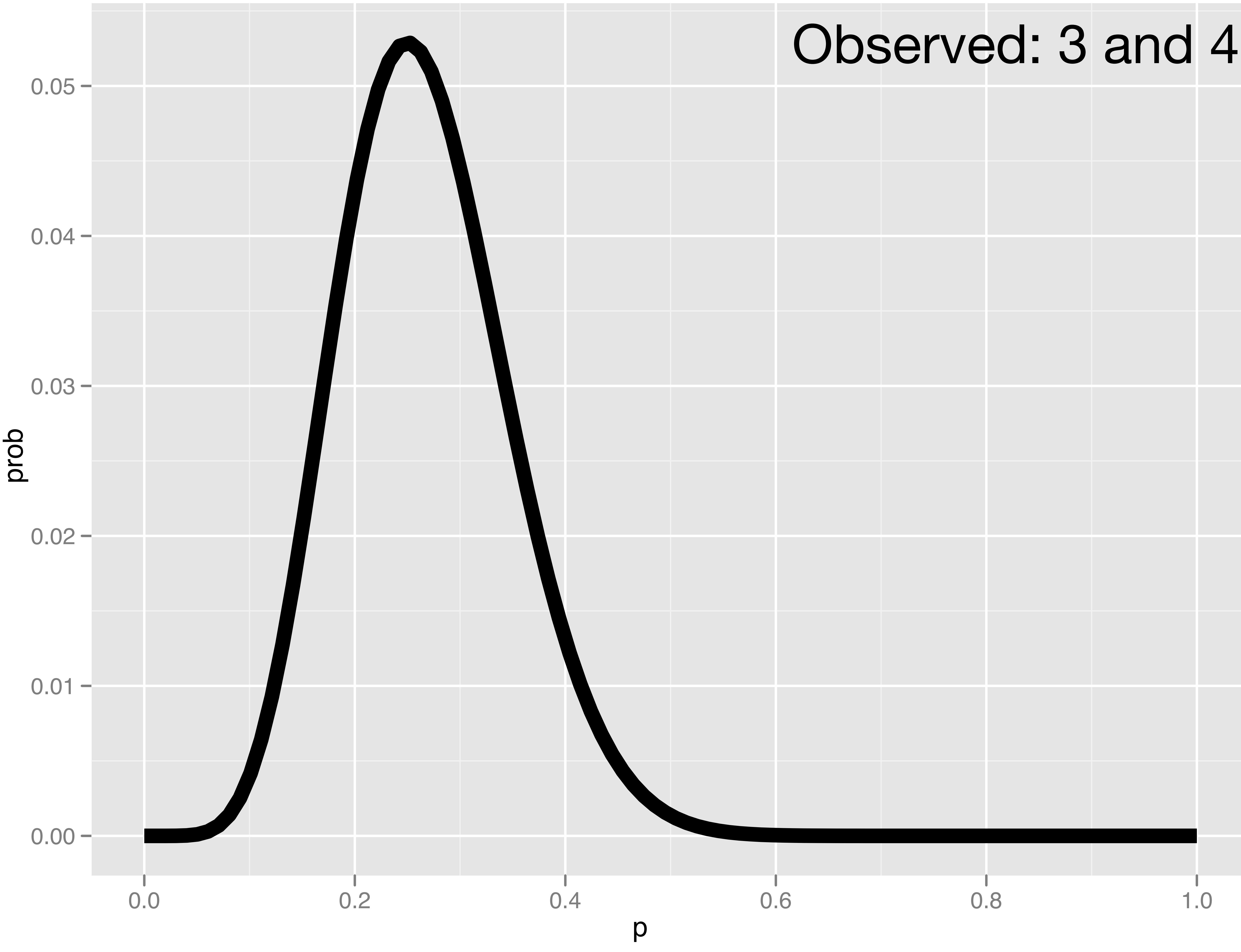
What happens if I repeat the experiment? (So I have two numbers) Can I still draw these plots? What would change?

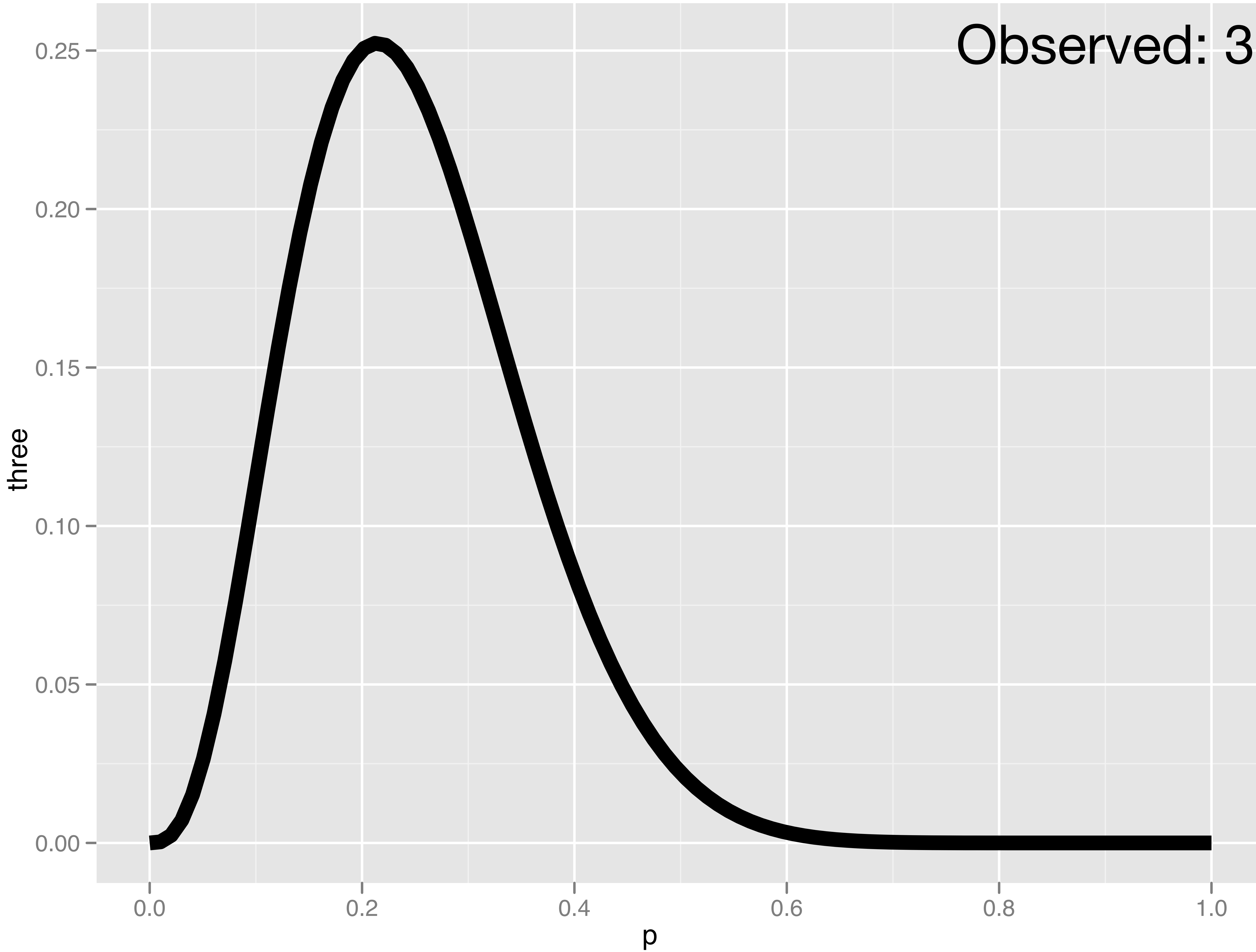
# Repeatability

Now have a bivariate pmf/pdf - but can assume they are independent, so it's still easy to calculate the probability for each value of p.

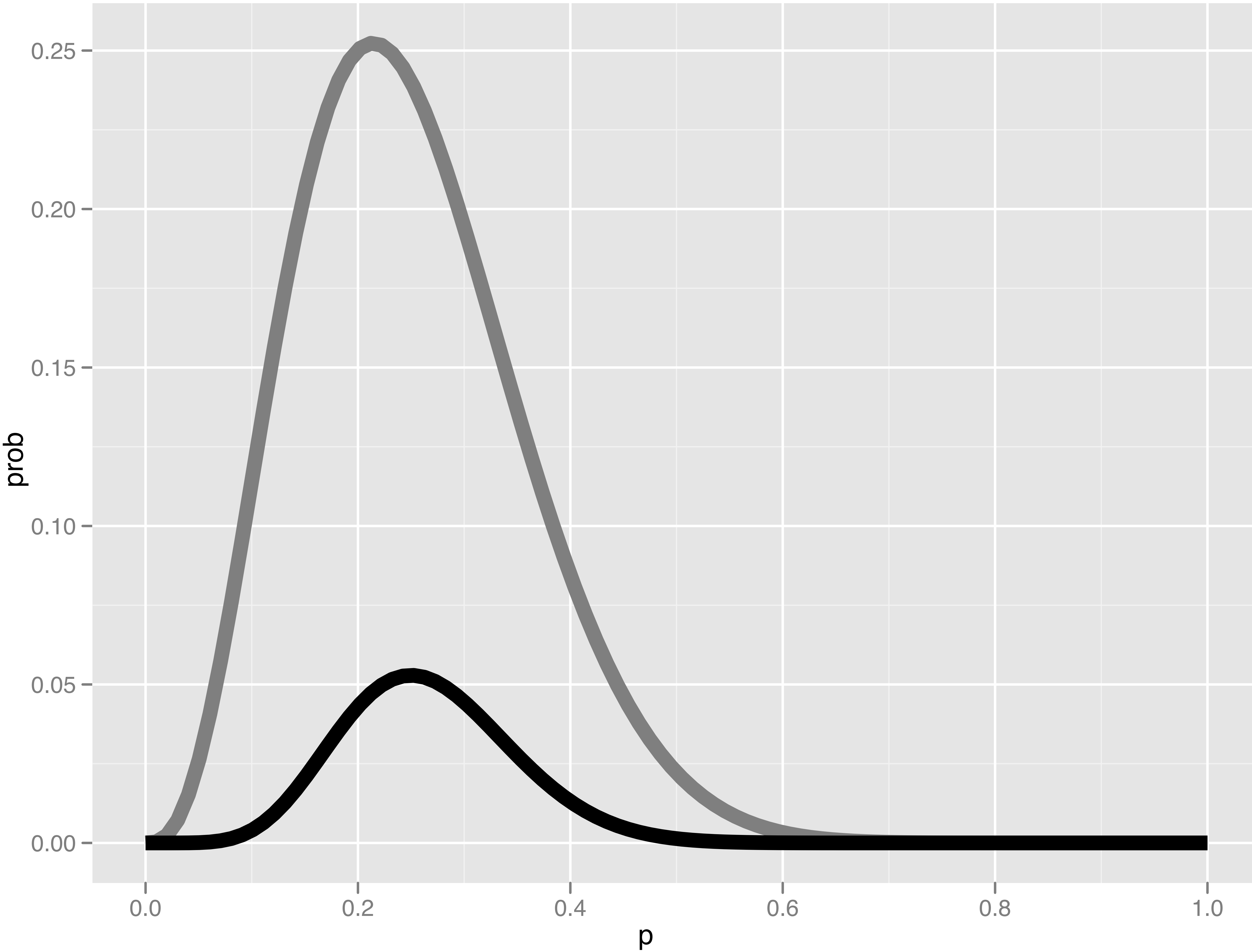
Let's say I repeated and I got a 4.  What does the plot look like now?

**likelihood** = joint pdf


Observed: 3 and 4
0.05
0.04
0.03
0.02
0.01
0.00
prob
0.0
0.2
0.4
0.6
0.8
1.0
p



Observed: 3
three
p
0.25
0.20
0.15
0.10
0.05
0.00
0.0
0.2
0.4
0.6
0.8
1.0



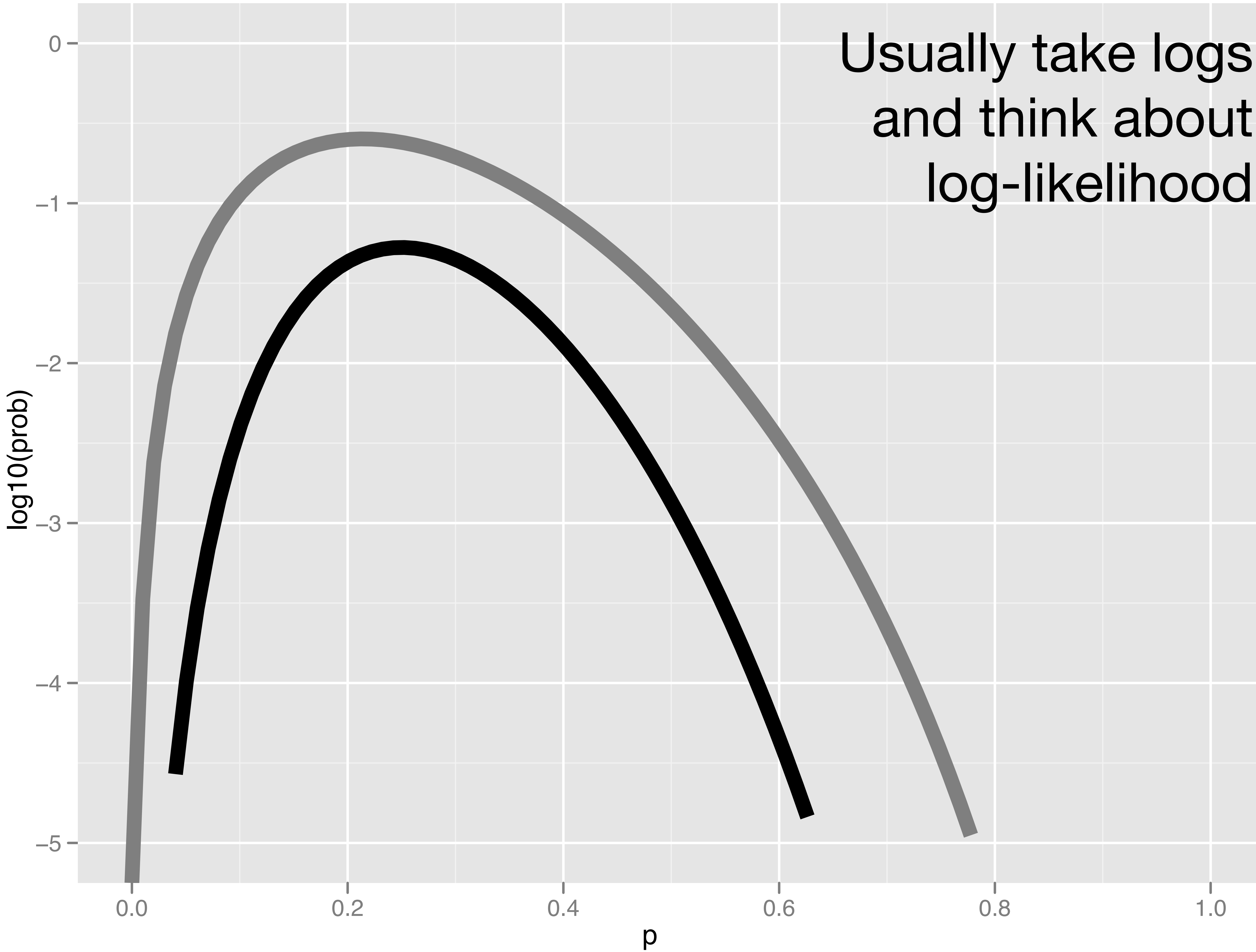
0.25
0.20
0.15
0.10
0.05
0.00
prob
0.0
0.2
0.4
0.6
0.8
1.0
p

# Three differences

Middle has shifted

Spread has decreased

Likelihood decreased by a lot



Usually take logs
and think about
log-likelihood
0
−1
−2
−3
−4
−5
log10(prob)
0.0
0.2
0.4
0.6
0.8
1.0
p

# General strategy

Flip the pmf/pdf around: given a set of x values, what's the probability that they were seen given a specified parameter value?

# Maximum likelihood

Write down the joint pdf
(pdf of iid random variables is?)

We have the data, so we can work out how likely each possible value of p is.

Then pick the p that is most likely.

Can do this **numerically** or **analytically**

# Steps

Write out likelihood (=joint pdf)

Write out log-likelihood

(Discard constants)

Find maximum:

- Differentiate and set to 0
- (Check second derivative is positive)
- (Check end points)

# Binomial

1. Joint pmf - be careful about x’s and indices

2. Log-likelihood - why?

3. Differentiate and set to zero

(We can use calculus because we’re thinking about the continuous p, not the discrete x’s)

# Your turn

Repeat the process to find an estimator of $\lambda$ when the sequence is iid Poisson($\lambda$).

1. Joint pmf

2. Log-likelihood

3. Differentiate and set to zero