Stat310

Introduction to inference

Hadley Wickham

1. Feedback
2. Motivation and introduction to inference
3. A survey
4. Estimators and their properties

Feedback

What I’m doing well to support your learning

What I’m doing well to support your learning

What I'm doing well to support your learning

--------------- extra credit ops

What I’m doing well to support your learning

--------------- extra credit ops

------------ clear/organised lectures

What I’m doing well to support your learning

-------------- extra credit ops

----------- clear/organised lectures

------- hw useful/effective

What I’m doing well to support your learning

-------------- extra credit ops

----------- clear/organised lectures

------- hw useful/effective

------ website

What I’m doing well to support your learning

-------------- extra credit ops

----------- clear/organised lectures

------- hw useful/effective

------ website

------ class involvement/your turn

What I’m doing well to support your learning

-------------- extra credit ops

----------- clear/organised lectures

------- hw useful/effective

------ website

------ class involvement/your turn

----- loud

What I’m doing well to support your learning

--------------- extra credit ops

------------ clear/organised lectures

------- hw useful/effective

------ website

------ class involvement/your turn

----- loud

----- lego

What I'm doing well to support your learning

--------------- extra credit ops

------------ clear/organised lectures

-------- hw useful/effective

------- website

------- class involvement/your turn

------ loud

------ lego

----- big writing

What I’m doing well to support your learning

-------------- extra credit ops

----------- clear/organised lectures

------- hw useful/effective

------ website

------ class involvement/your turn

----- loud

----- lego

---- big writing

---- available

What I’m doing well to support your learning

-------------- extra credit ops

----------- clear/organised lectures

------- hw useful/effective

------ website

------ class involvement/your turn

----- loud

----- lego

---- big writing

---- available

---- examples

What I’m doing well to support your learning

-------------- extra credit ops

----------- clear/organised lectures

------- hw useful/effective

------ website

------ class involvement/your turn

----- loud

----- lego

---- big writing

---- available

---- examples

---- help sessions more helpful

What I'm doing well to support your learning

--------------- extra credit ops

------------ clear/organised lectures

------- hw useful/effective

------ website

------ class involvement/your turn

----- loud

----- lego

---- big writing

---- available

---- examples

---- help sessions more helpful

--- stats in practice

What I'm doing well to support your learning

--------------- extra credit ops

------------ clear/organised lectures

-------- hw useful/effective

------- website

------- class involvement/your turn

------ loud

------ lego

----- big writing

----- available

----- examples

----- help sessions more helpful

---- stats in practice

Things that I’m doing that you like but I don’t really think actually supports your learning:

--- dogs

- accent
- cool dude
- nice clothes

What I could improve:

What I could improve:

What I could improve:

---------------- don’t make mistakes

What I could improve:

--------------- don’t make mistakes

--------- put slides online before class

What I could improve:

---------------- don’t make mistakes

--------- put slides online before class

------- more (fun) examples

What I could improve:

---------------- don’t make mistakes

---------- put slides online before class

-------- more (fun) examples

-------- write bigger

What I could improve:

---------------- don't make mistakes

---------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

What I could improve:

---------------- don’t make mistakes

--------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

What I could improve:

---------------- don’t make mistakes

---------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

---- make help sessions helpful

What I could improve:

--------------- don't make mistakes

--------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

---- make help sessions helpful

---- more intermediate steps

What I could improve:

--------------- don’t make mistakes

--------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

---- make help sessions helpful

---- more intermediate steps

--- more/longer review

What I could improve:

--------------- don’t make mistakes

--------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

---- make help sessions helpful

---- more intermediate steps

--- more/longer review

- gradebook on owlspace confusing

What I could improve:

--------------- don’t make mistakes

--------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

---- make help sessions helpful

---- more intermediate steps

--- more/longer review

- gradebook on owlspace confusing

- more classes

What I could improve:

--------------- don’t make mistakes

--------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

---- make help sessions helpful

---- more intermediate steps

--- more/longer review

- gradebook on owlspace confusing
- more classes

What I could improve:

---------------- don't make mistakes

--------- put slides online before class

------- more (fun) examples

------- write bigger

------ slow down

----- put up solutions to your turns

---- make help sessions helpful

---- more intermediate steps

--- more/longer review

- gradebook on owlspace confusing

- more classes

- still not enough scooby snacks

Homework grade

- Convert each grade to a percentage
- Drop lowest two grades (excluding bonus homeworks)
- Adjust for half-point homeworks (0 and 6)
- Add all percentages
- Divide by number of homeworks - 3

What you're doing well to support your learning

-------------------- attending class

-------------- doing hw

----------- help sessions

----- taking notes

-------- starting hw earlier

------ hw: hard/better/more carefully

---- actively engaged in your turns

---- taking good notes

--- paying more attention/more focussed

What you could do better (non-specific)

------ study more

------ pay better attention

-- be more awake

- listen more intently
- not clever enough
- be more focussed
- engage more
- focus more

What you could do better (specific)

------------- review more

------ read textbook

----- start hw earlier

---- devote more time to class

-- more coffee

- create review sheet as we go

- ask more questions

Links

http://calnewport.com/blog/about/

http://calnewport.com/blog/category/tips-notetaking/

http://calnewport.com/blog/category/tips-studying/

http://calnewport.com/blog/category/tips-time-management-scheduling-productivity/

Inference

What proportion of Rice students smoke marijuana?

Data vs. Distributions

Random experiments produce data.

A repeatable random experiment has some underlying distribution.

We want to go from the data to say something about the underlying distribution. So far we have been doing the **opposite**.

Inference

Up to now: Given a sequence of random variables with distribution F, what can we say about the mean of a sample?

What we really want: Given the mean of a sample, what can we say about the underlying sequence of random variables?

Marijuana

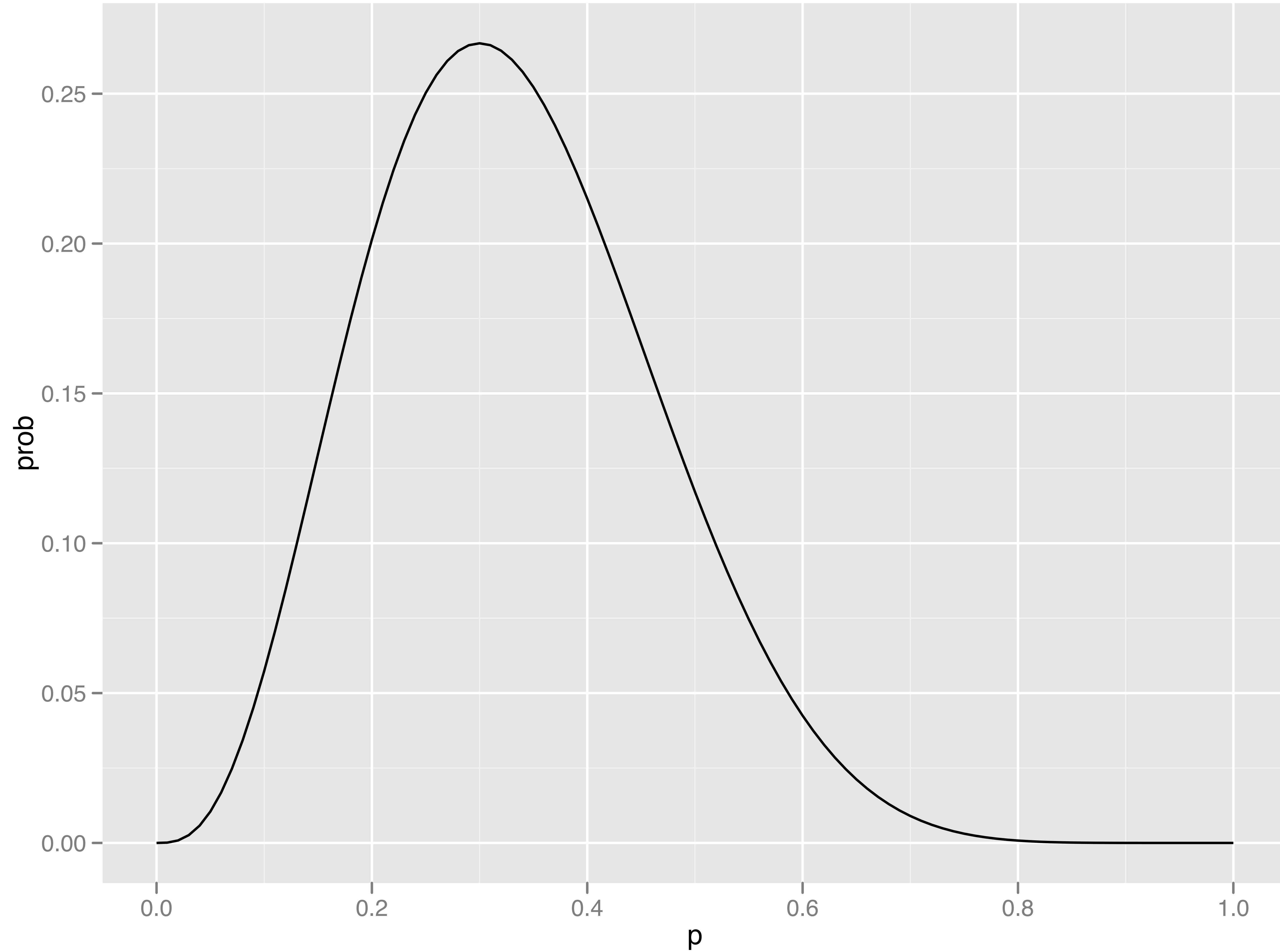
Up to now: If I told you that smoking pot was Binomially distributed with probability p, you could tell me what I would expect to see if I sampled n people at random.

What we want: If I sample n people at random and find out m of them smoke pot, what does that tell me about p?

Your turn

Let's I selected 10 people at random from the Rice campus and asked them if they'd smoked pot in the last month. Three said yes.

Let Y_i be 1 if person i smoked pot and 0 otherwise. What's a good guess for distribution of Y_i? What assumptions did you make?


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

Plug-in principle

A good first guess at the true mean, variance, or any other parameter of a distribution is simply to estimate it from the data.

We'll learn more sophisticated ways in the days to come.

Problems

If I picked someone at random out of the Rice campus directory, and asked them if they'd smoked pot in the last month, how likely do you think I am to get an honest answer?

How can we get around this?

Your turn

Compare your survey to someone with the other colour. What is the difference? How could you use the answers to estimate how many people smoke pot?

Questions

1. Have you smoked pot in the last month?
2. Flip a coin. If it's heads, write yes. Otherwise write yes if you've smoked pot in the last month, and no otherwise.
3. In the last month, how many of the following activities have you done? One list contains smoking pot, the other doesn't.

Results

1. The proportion of yeses in Q1
2. The proportion of yeses in (Q2 - 0.5) * 2
3. Average Q3a - average Q3b

Your turn

Fill in the survey. Make sure no one else can see your answers!

Your turn

While we tally the results, brainstorm as many different properties that you could use to decide which estimator is best.

Remember that there is some true unknown proportion that we are trying to estimate.

Results

Definitions

Parameter space: set of all possible parameter values

Estimator: process/function which takes data and gives best guess for parameter. Usually written with a hat. Is a random variable and so has a distribution.

Point estimate: single best guess.

Other estimators

Important properties of an estimator

Unbiased

Small variance

Consistent

Sufficient

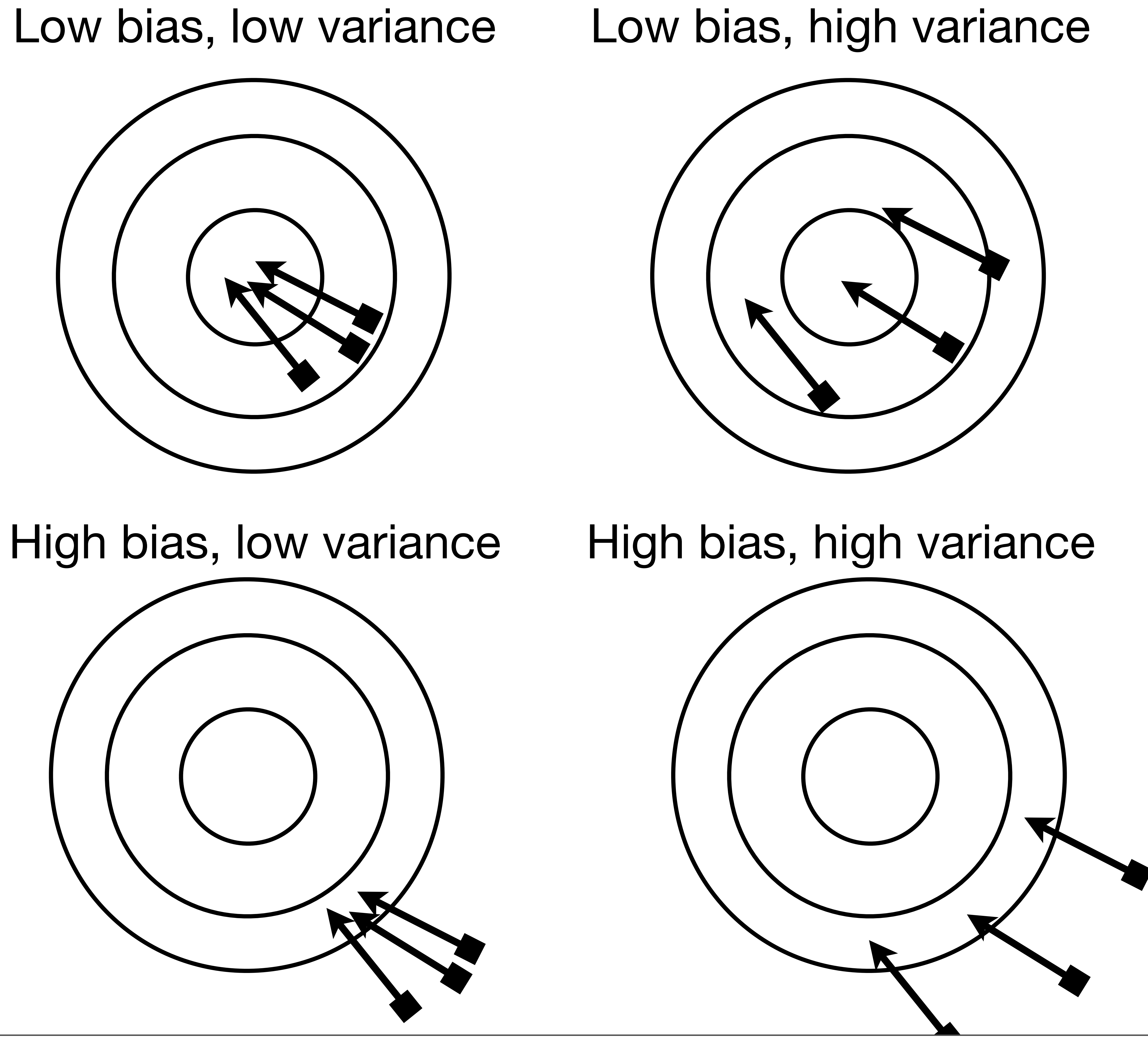
$$\mathrm{E}(\hat{\theta}) = \theta$$

Unbiased

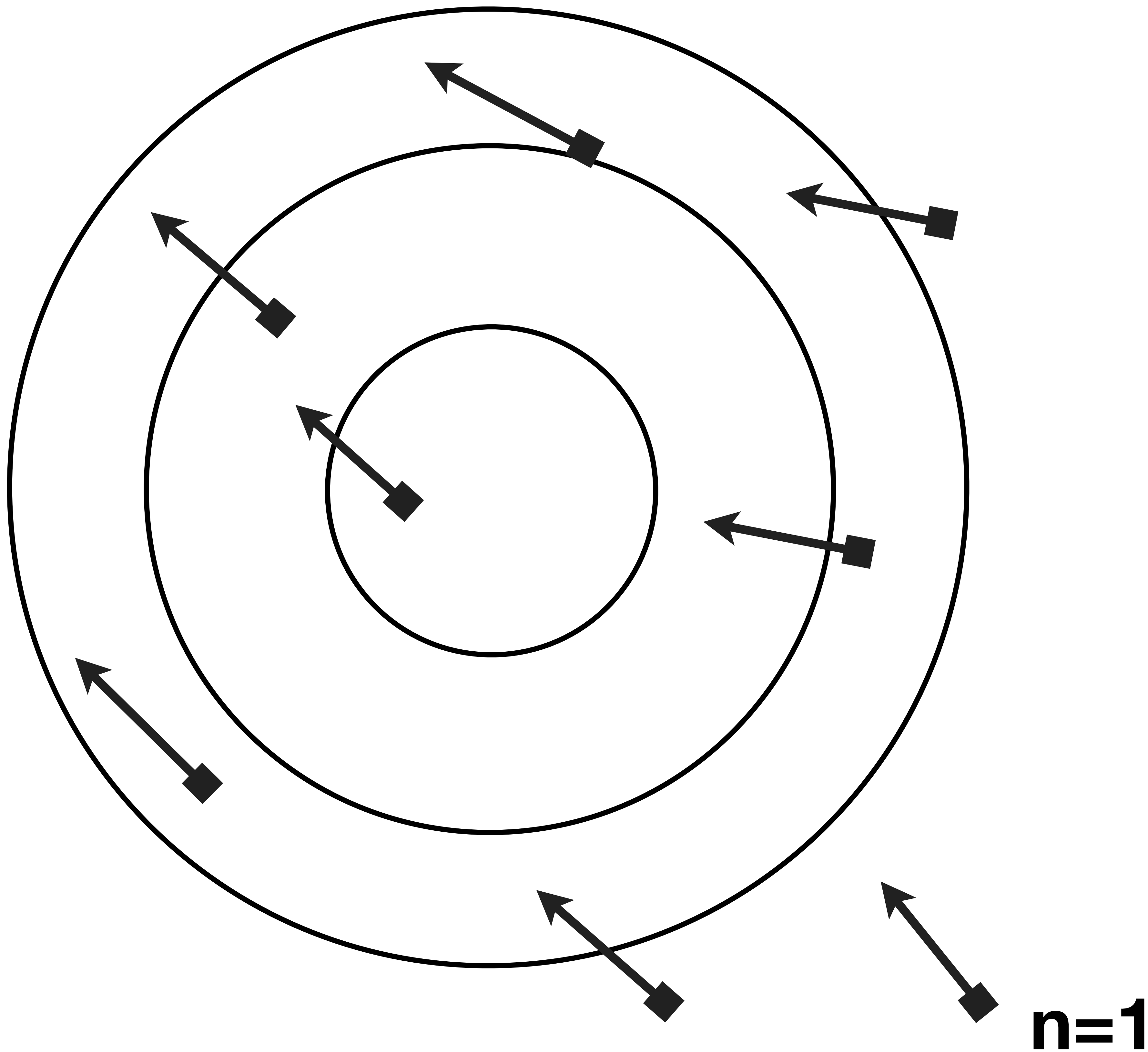
$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

Minimum variance

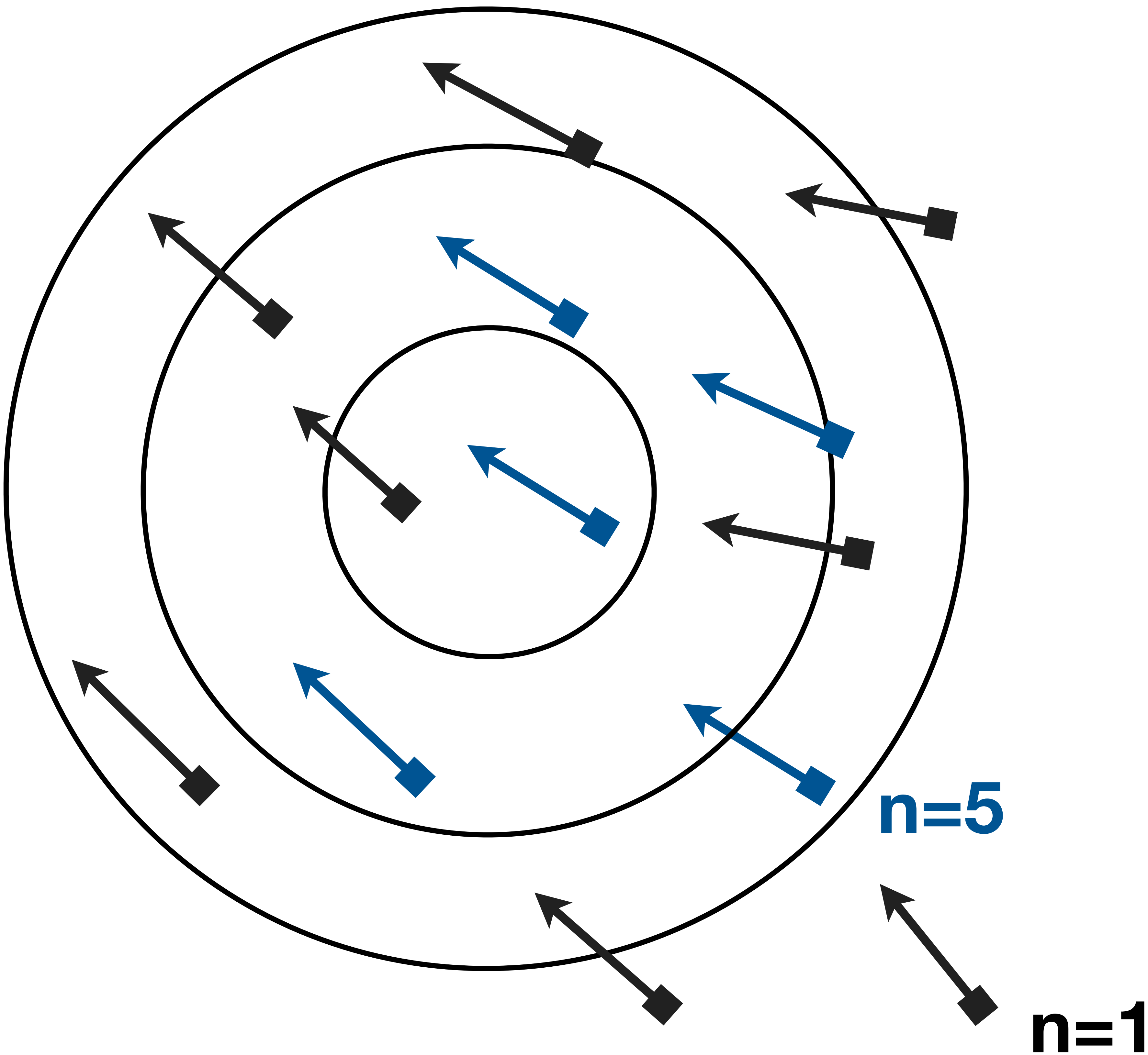
Common problem is to find UMVE (unbiased minimum variance estimator) across all possible estimators


Low bias, low variance
Low bias, high variance
High bias, low variance
High bias, high variance

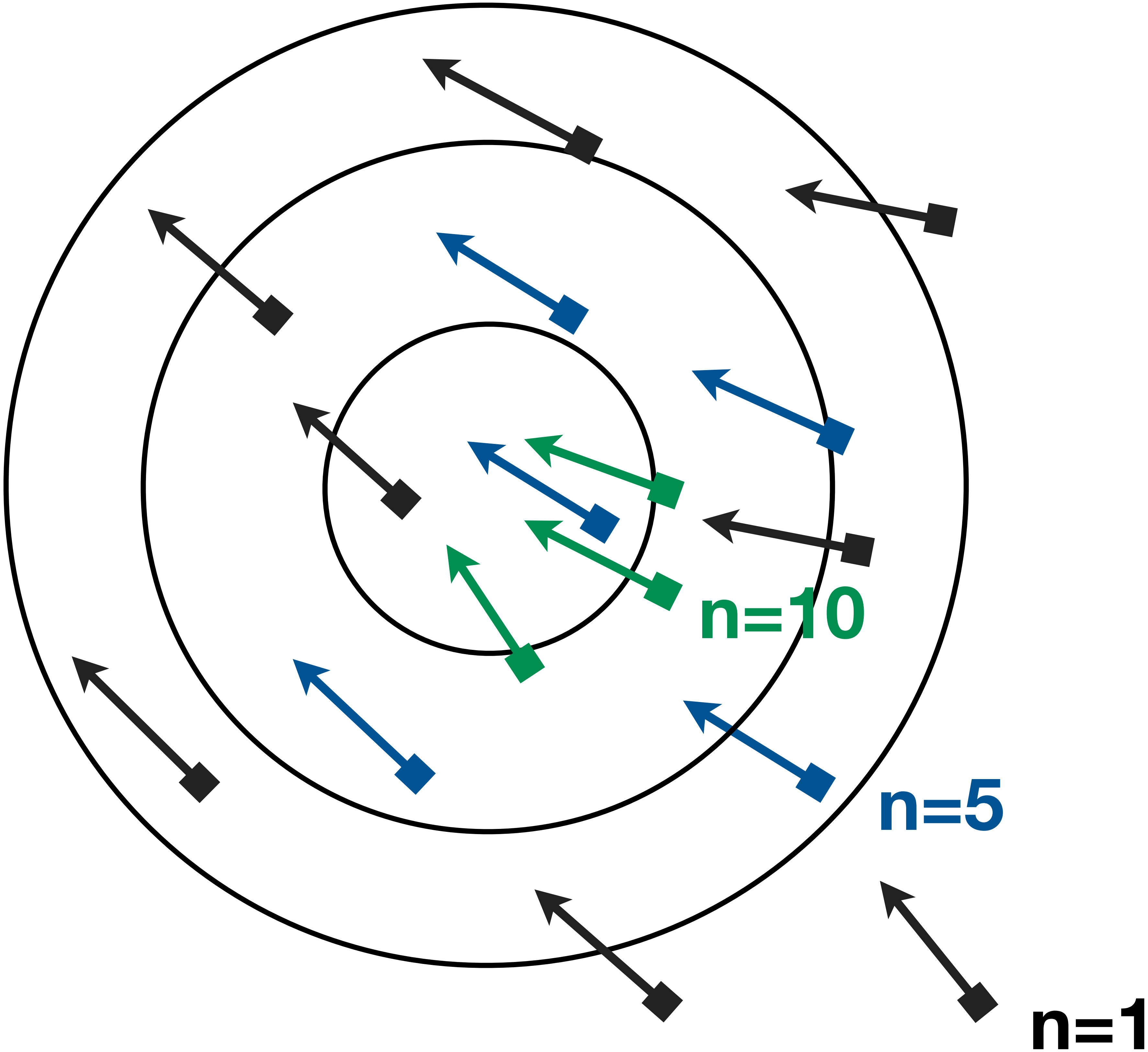
Consistent



Consistent



Consistent



$$\hat{\theta}_n \to \theta$$

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1$$

Your turn

How can we describe this as a random experiment? Do you have any concerns about applying the results of this survey to the entire Rice campus?

Experiments

Remember, the context is always the random experiment. For this case the randomness comes from picking the person at random.

Repeating the experiment does not mean asking each person again (the answer would be the same) but asking a different set of people.

Next time

Method of moments

Maximum likelihood