Estimation

Experimental Research

Professor Yamil Velez — 9/20/22

Last week

- Average treatment effect
- Directed acyclic graphs (DAGs)
  - Confounds
  - Mediators
  - Colliders
- DAG assignment

This week

- How do we estimate these quantities in practice?
  - Population
  - Expectation
  - Sample mean
  - Estimating ATE as a difference in sample means
  - Estimating uncertainty
    - Standard deviations
    - Standard errors

Populations vs. Samples

- Populations are the targets of our inferences
  - American public
  - Students
  - Companies
- To understand populations, we derive representative samples
- Much of applied statistics is geared toward understanding when sample-derived estimates match population estimates

Expectation

- Suppose we have a random variable *X*
- $E(X) = \mu = \sum xP(x)$
- Example:
  - Fair die (x ranges from 1-6; P(x) for each face is 1/6)
  - $1 \times 1/6 + 2 \times 1/6 + 3 \times 1/6 + 4 \times 1/6 + 5 \times 1/6 + 6 \times 1/6$
  - $E(X) = 3.5$
- With a finite population, *E(X)* equals the arithmetic mean
  - Try it out with the example above

Expected value

- Interpretations
  - Long-run average
    - E(X) captures the average over an infinite number of trials
  - Center of mass
    - Each random variable has a probability density function
    - E(X) is the center of mass

Deriving estimates from samples

- We rarely have population-level data, and even when we have a population, there exist other ways in which units may have been assigned to the treatment

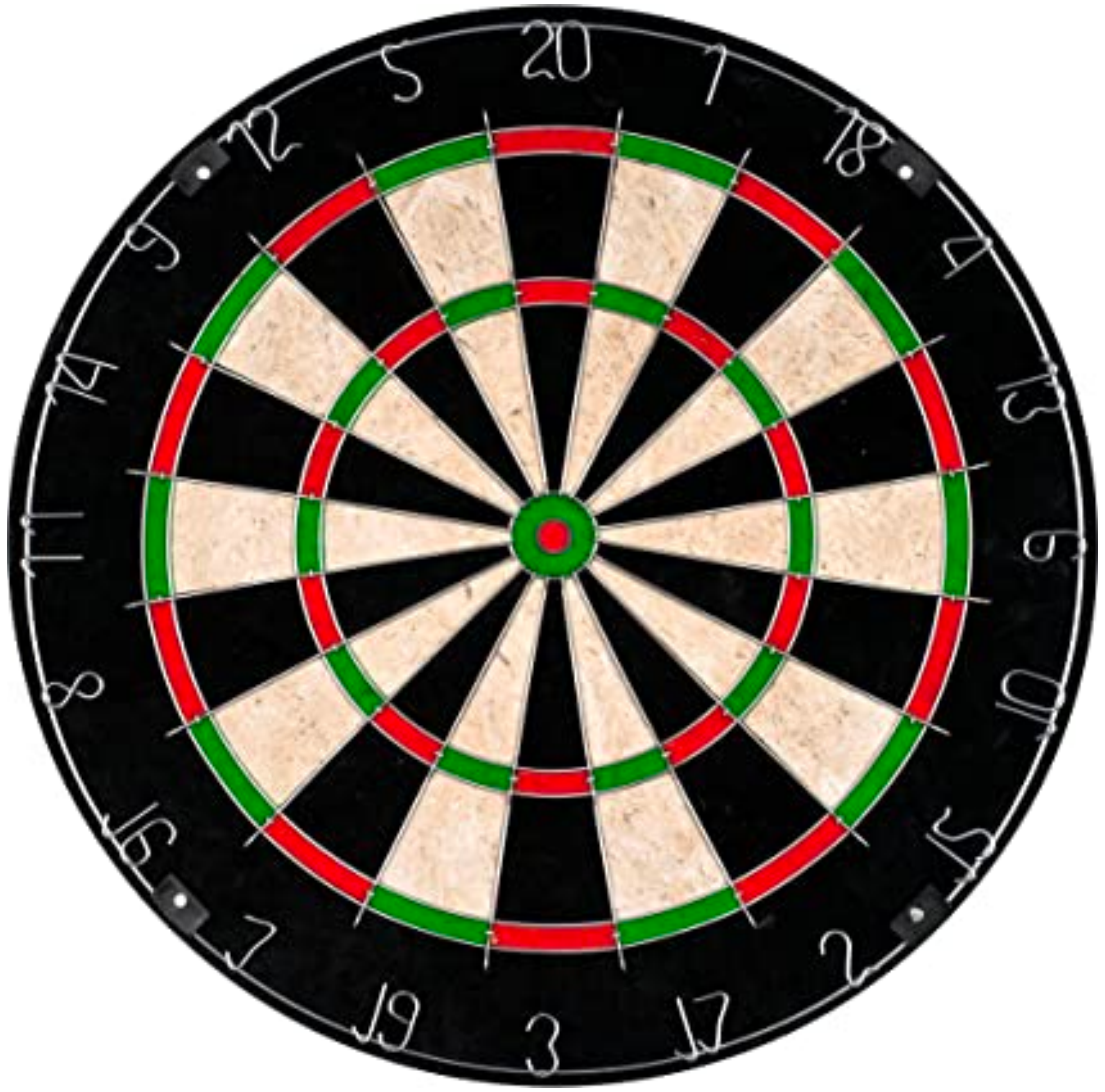
- How do generate estimates from samples that approximate the estimates we would hypothetically want at the population level?

- How do we guarantee better vs. worse estimates?

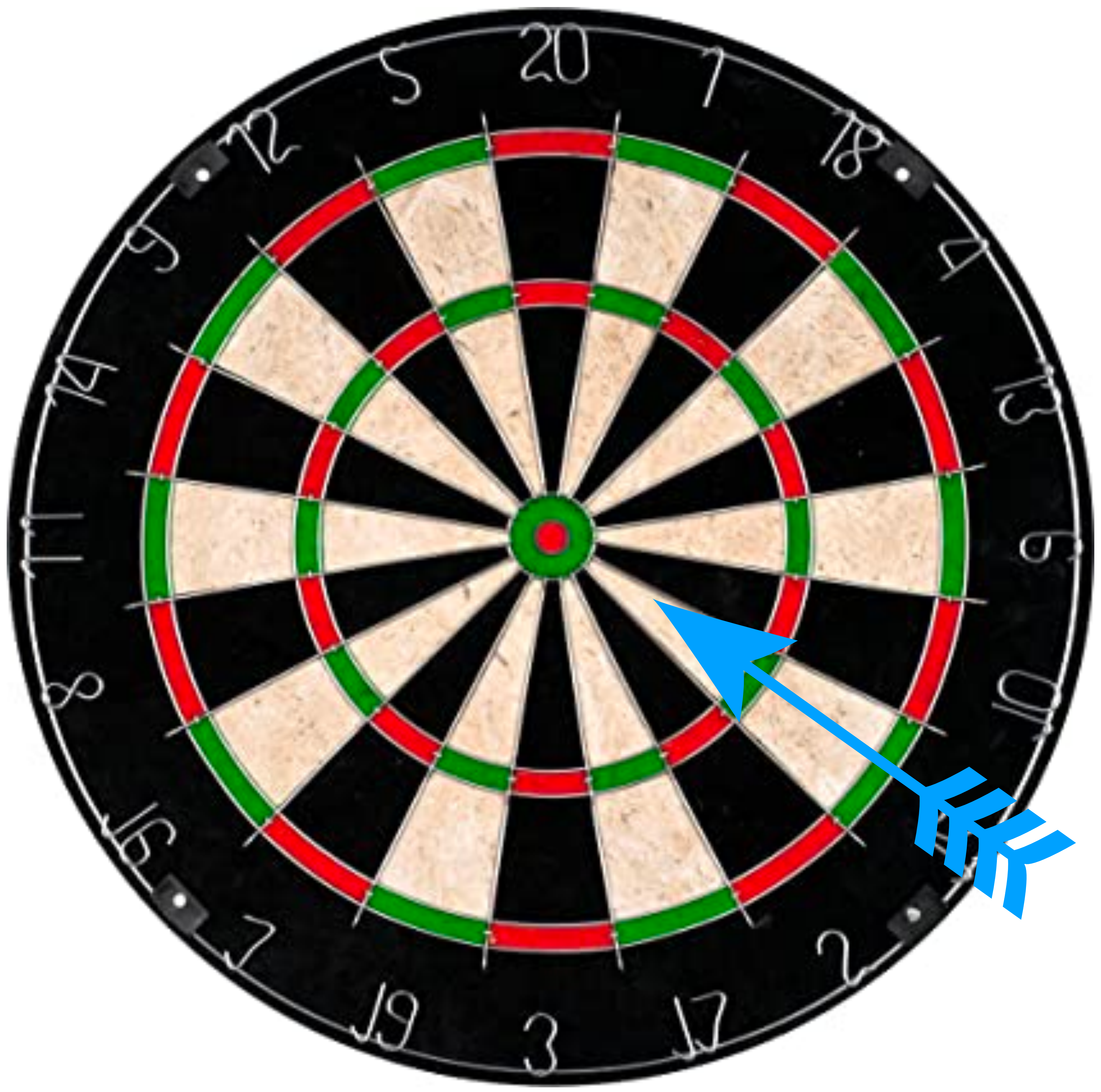
Metrics

- What does it mean for an estimate to be better or worse?
  - Bias
    - The degree to which sample estimates match population estimates on average
    - $\mathbb{E}[\theta - \hat{\theta}]$
    - If estimating the mean, this is the difference between the population mean and the sample mean. An unbiased estimator would be an estimator that exactly matches the population mean on average.
  - Consistency
    - Whether an estimator produces an estimate that converges in probability to the true estimate
    - As sample size increases, we should obtain a better population estimate
    - $P(|\theta - \hat{\theta}| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$
  - Efficiency
    - A good estimator minimizes variability in estimates across multiple trials

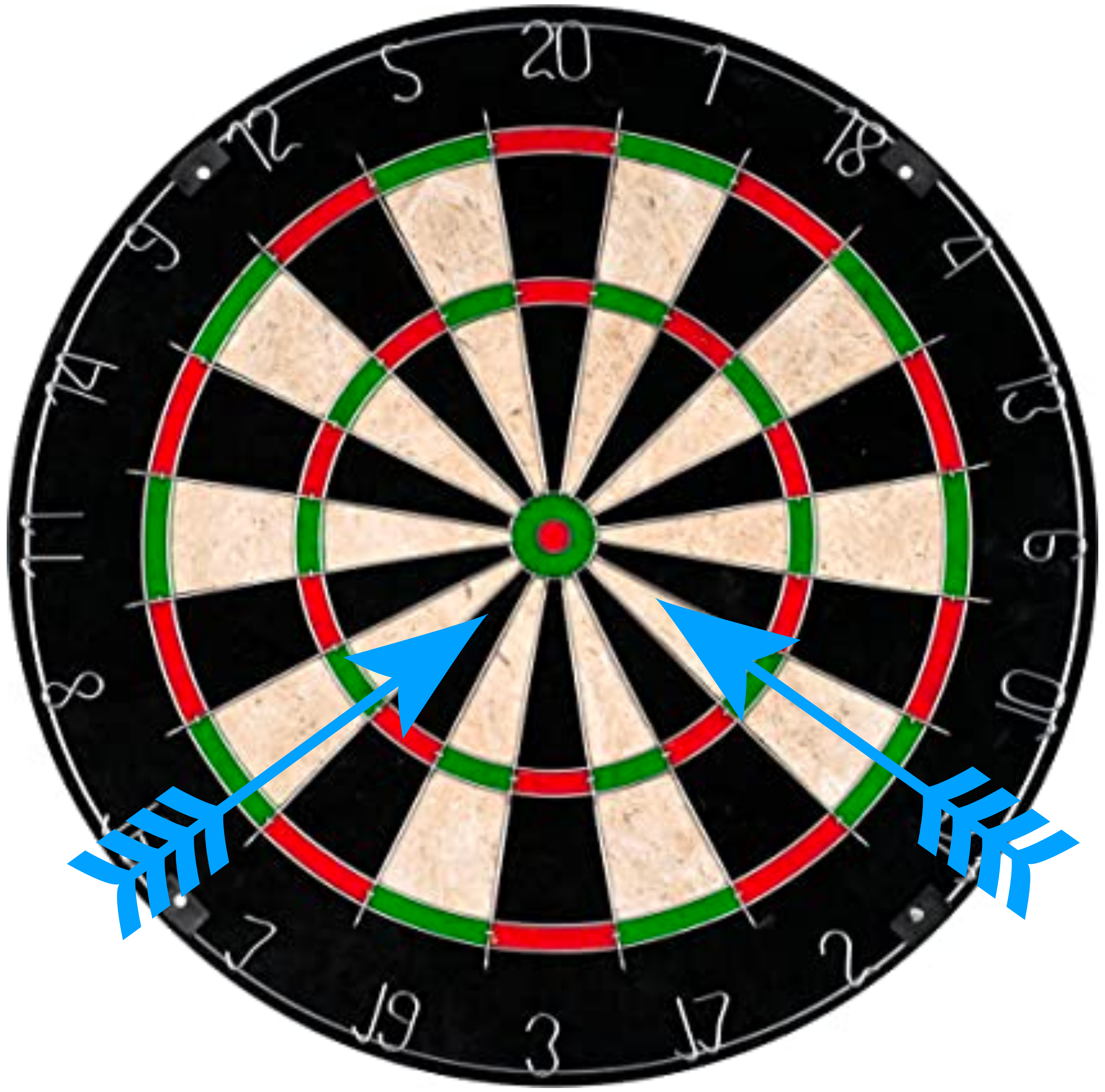
Dartboard



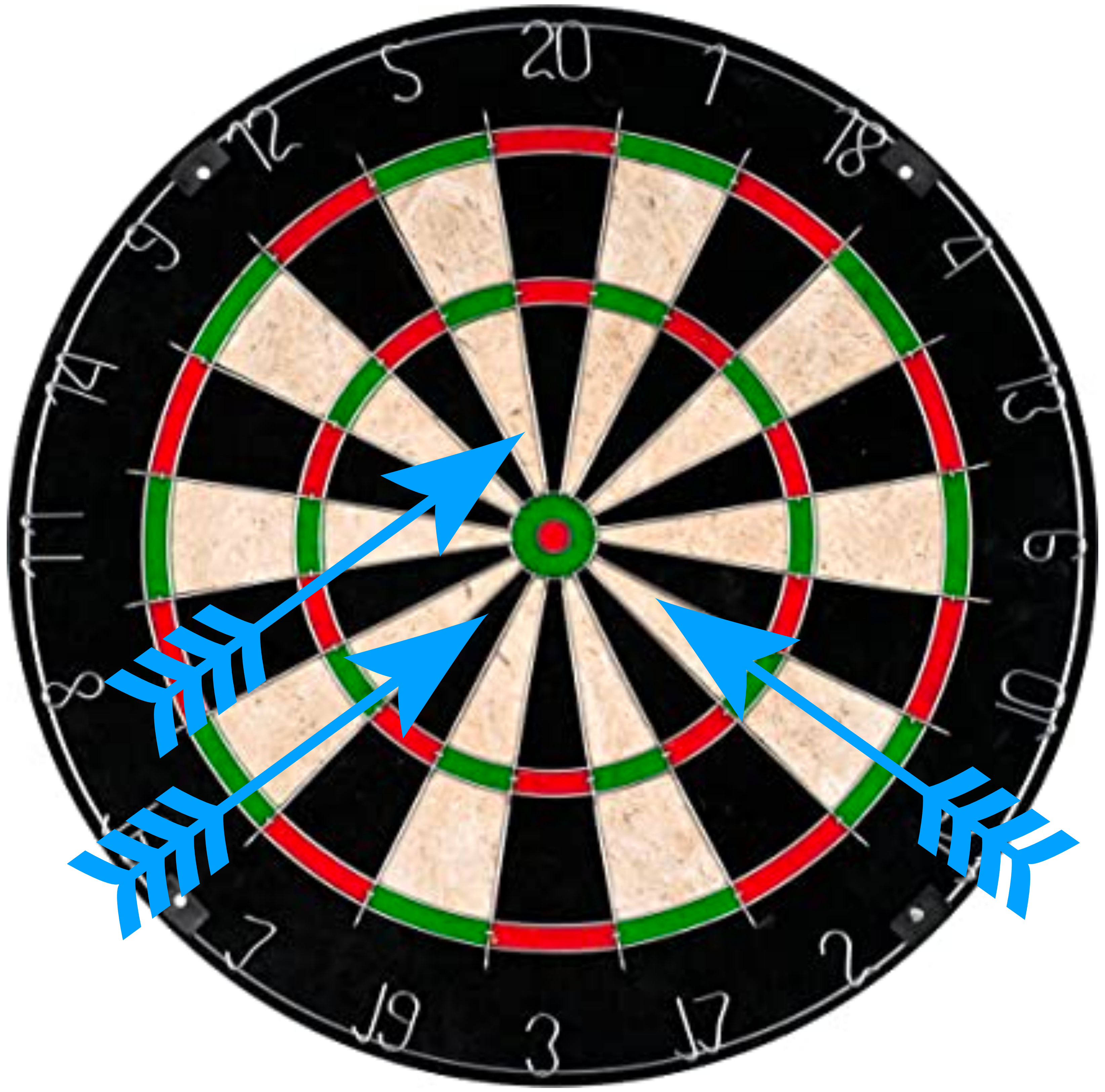
Dartboard



Dartboard



Dartboard

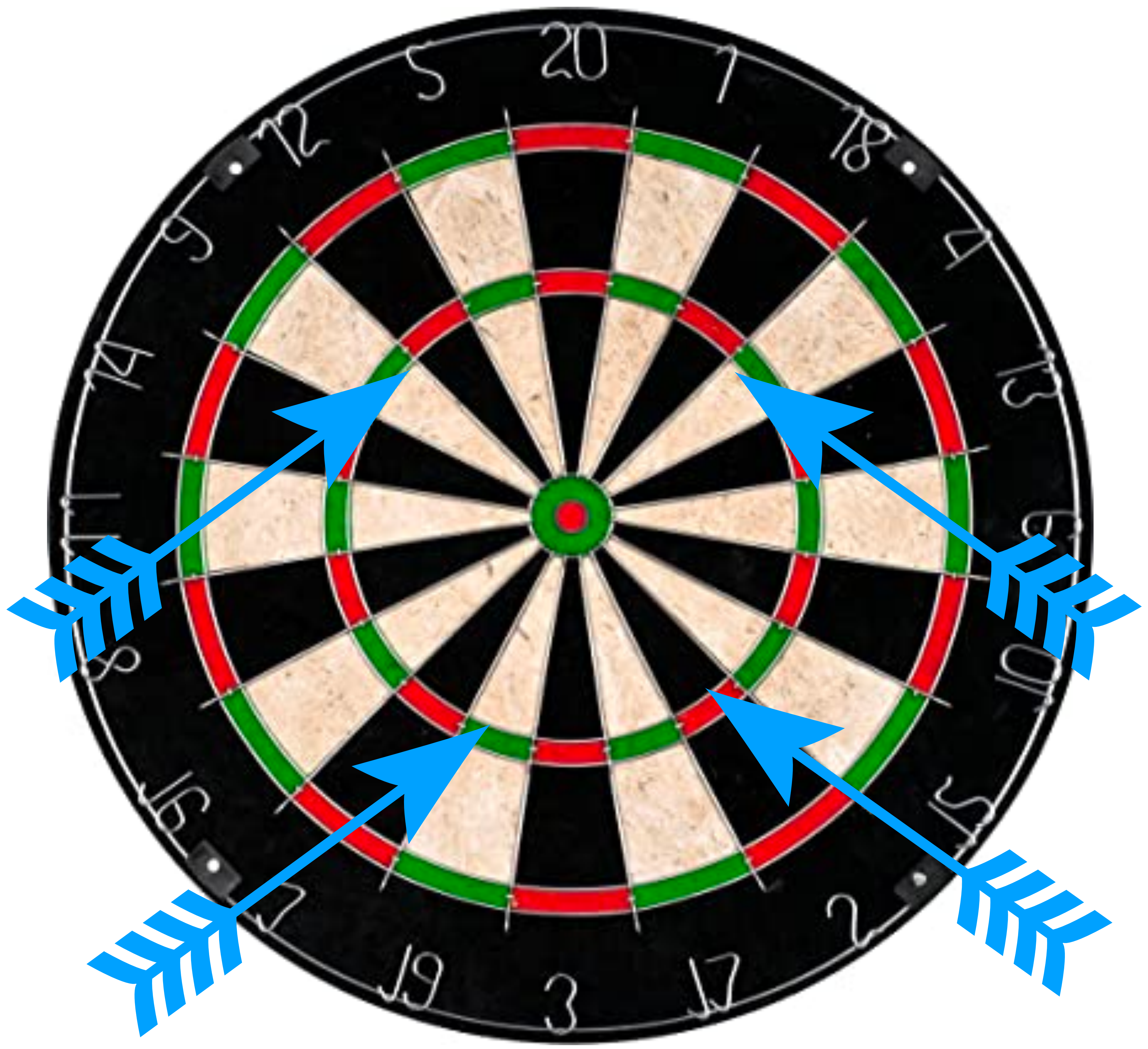


Dartboard

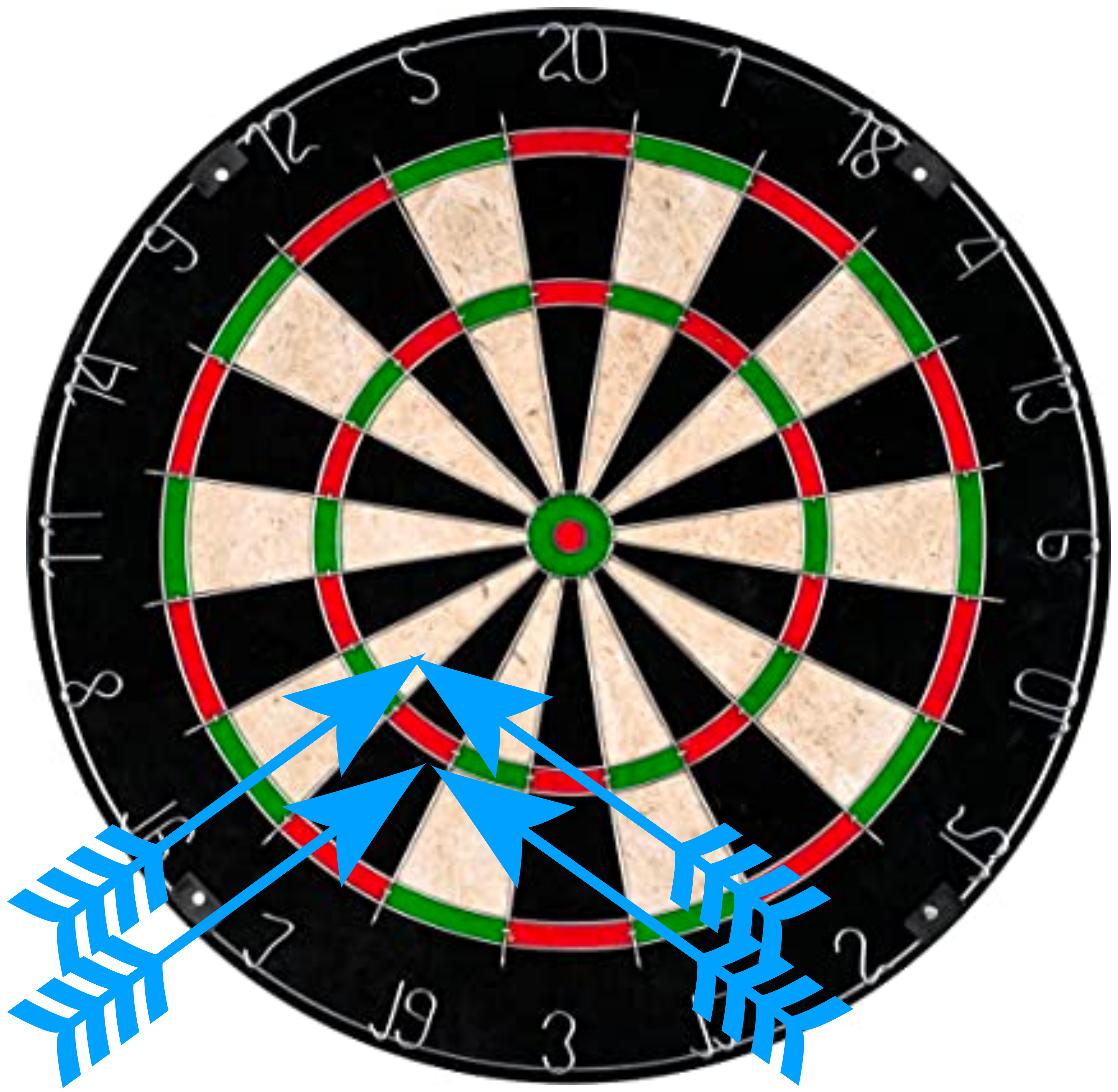
Dartboard

Unbiasedness

Dartboard

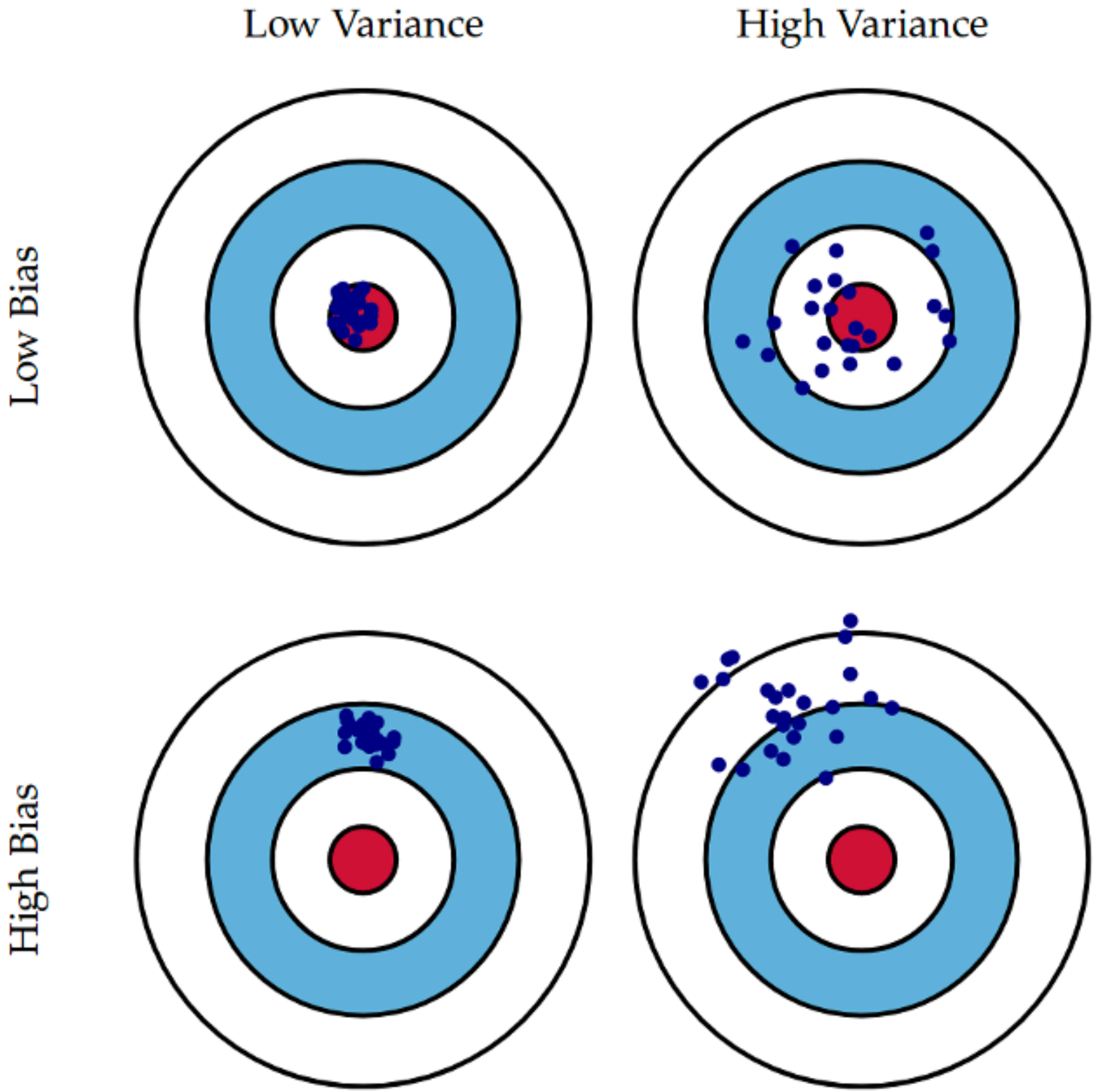


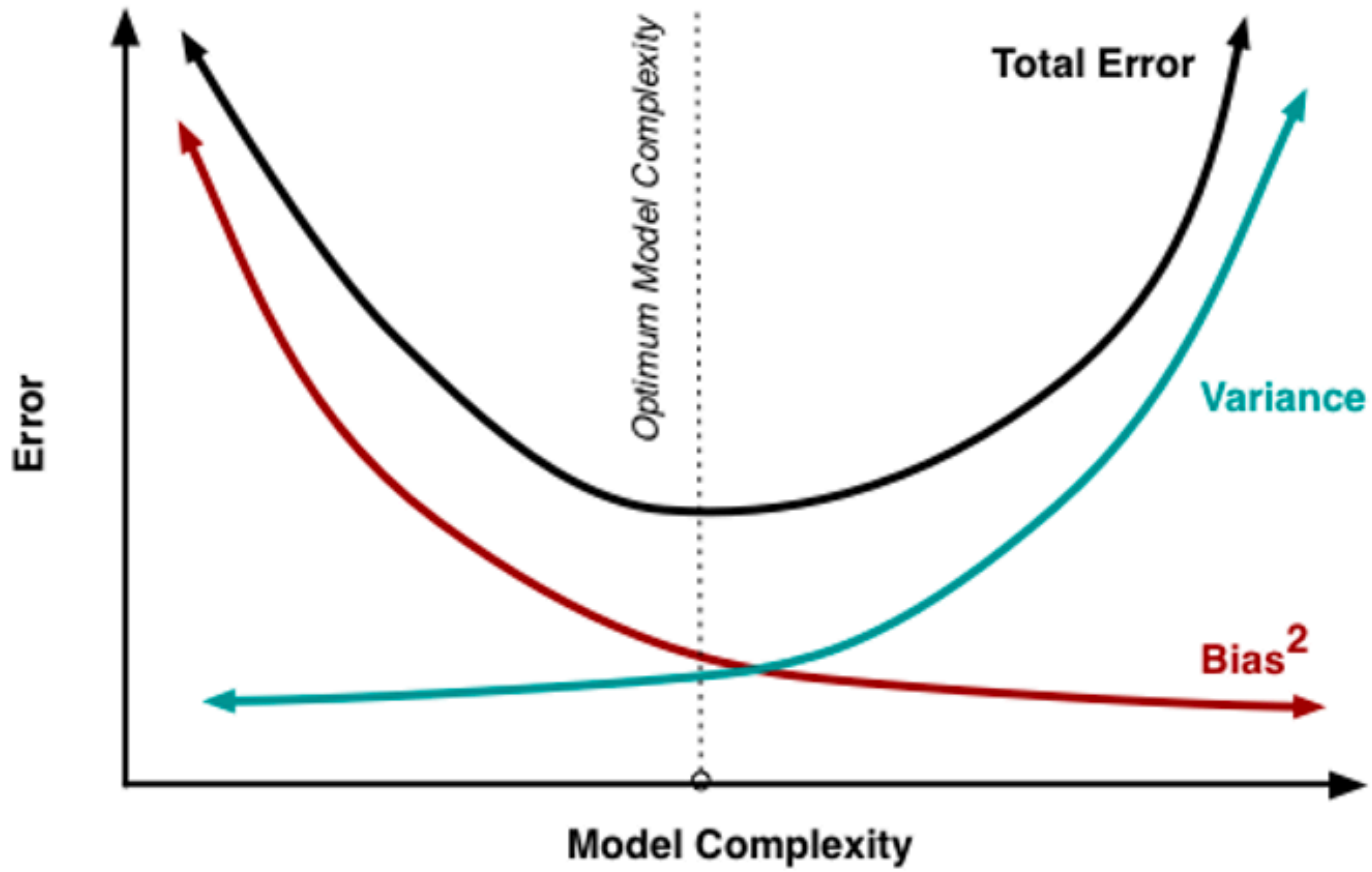
Dartboard



Biased but efficient

Dartboard


Low Variance
High Variance
Low Bias
High Bias



Optimum Model Complexity
Total Error
Variance
Error
Bias^2
Model Complexity

The sample mean is an unbiased estimator of the population mean

- What is an unbiased estimator of the population mean?
- $\bar{X} = 1/n \sum_{i}^{n} X_i$
- We want to show that $E(\bar{X}) = E(X)$
    - Step 1: Substitute $\bar{X}$ with equation above – $\mathbb{E}\left(1/n \sum_{i}^{n} X_i\right)$
    - Step 2: Extract non-random numbers out of the expectation – $1/n \sum E(X_i)$
    - Step 3: Substitute $E(X_i)$ with μ – $1/n \sum \mu$
    - Step 4: The expected value of $\bar{X}$ is μ

Estimating the Average Treatment Effect

- $ATE = \mathbb{E}(Y_i(1) - Y_i(0))$
- $E(Y_i(1)) - E(Y_i(0))$
- $E[Y_i \mid T_i = 1] - E[Y_i \mid T_i = 0]$
- $\widehat{ATE} = \frac{1}{n} \sum_{i \in T_i = 1} Y_i - \frac{1}{n} \sum_{i \in T_i = 0} Y_i$
  - In plain language, difference in mean outcomes between those in the treatment and those in the control

ATE in Practice

i	Yi	Ti
1	1	1
2	3	0
3	0	1
4	1	0

ATE in Practice

i	Yi	Ti
1	1	1
2	3	0
3	0	1
4	1	0

Treated mean: .5

ATE in Practice

i	Yi	Ti
1	1	1
2	3	0
3	0	1
4	1	0

Treated mean: .5

Untreated mean: 2

ATE in Practice

i	Yi	Ti
1	1	1
2	3	0
3	0	1
4	1	0

Treated mean: .5

Untreated mean: 2

ATE = -1.5