INTRODUCTION TO STATISTICS

LECTURE 7

LAST TIME

- Normal distribution
    - probability of being in an interval;
    - some properties;
    - MLE for μ and σ.

LAST TIME

- Normal distribution
  - probability of being in an interval;
  - some properties;
  - MLE for μ and σ.

- Central limit theorem – an introduction.

TODAY

- Central limit theorem
  - revision;
  - practical exercise.

TODAY

- Central limit theorem
  - revision;
  - practical exercise.

- Properties of point estimators
  - bias, variance and consistency.

TODAY

- Central limit theorem
  - revision;
  - practical exercise.
- Properties of point estimators
  - bias, variance and consistency.
- Confidence intervals

CENTRAL LIMIT THEOREM

Why normal distribution is so important

CENTRAL LIMIT THEOREM

In the video:

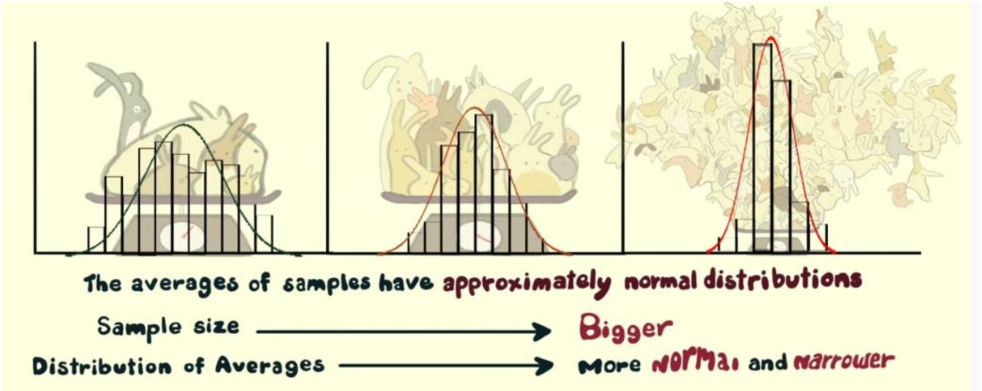
- Instead of measuring every single rabbit, weigh samples of size N and compute sample means.

CENTRAL LIMIT THEOREM

In the video:

- Instead of measuring every single rabbit, weigh samples of size N and compute sample means.

- ***Central limit theorem (informally):*** the larger the N, the more "normal" the distribution of the sample averages is.



CENTRAL LIMIT THEOREM

Samples $X_1, X_2, \ldots, X_n$:

- i.i.d.
- a *finite* mean μ and *finite* variance σ^2

CENTRAL LIMIT THEOREM

Samples $X_1, X_2, \ldots, X_n$:

- i.i.d.
- a *finite* mean μ and *finite* variance σ^2

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

CENTRAL LIMIT THEOREM

Samples $X_1, X_2, \ldots, X_n$:

- i.i.d.
- a *finite* mean μ and *finite* variance σ^2

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

CENTRAL LIMIT THEOREM

- $X \sim Po(5)$ - number of errors per computer program
  - $E(X) = Var(X) = 5$

- $X_1, X_2, \ldots, X_{125}$ - number of errors in the programs.

$$P(\bar{X}_n \leq 5.5) =$$

CENTRAL LIMIT THEOREM

- $X \sim Po(5)$ - number of errors per computer program
  - $E(X) = Var(X) = 5$
- $X_1, X_2, \dots, X_{125}$ - number of errors in the programs.

$$P(\bar{X}_n \leq 5.5) =$$

$$= P\left(\frac{\sqrt{125}(\bar{X}_n - 5)}{\sqrt{5}} \leq \qquad\qquad\qquad \right) \approx$$

CENTRAL LIMIT THEOREM

- $X \sim Po(5)$ - number of errors per computer program
  - $E(X) = Var(X) = 5$

- $X_1, X_2, \dots, X_{125}$ - number of errors in the programs.

$$P(\bar{X}_n \leq 5.5) =$$

$$= P\left(\frac{\sqrt{125}(\bar{X}_n - 5)}{\sqrt{5}} \leq \frac{\sqrt{125}(5.5 - 5)}{\sqrt{5}}\right) \approx$$

CENTRAL LIMIT THEOREM

- $X \sim Po(5)$ - number of errors per computer program
  - $E(X) = Var(X) = 5$

- $X_1, X_2, \ldots, X_{125}$ - number of errors in the programs.

$$P(\bar{X}_n \leq 5.5) =$$

$$= P\left(\frac{\sqrt{125}(\bar{X}_n - 5)}{\sqrt{5}} \leq \frac{\sqrt{125}(5.5 - 5)}{\sqrt{5}}\right) \approx$$

$$\approx P(Z \leq 2.5) =$$

CENTRAL LIMIT THEOREM

- $X \sim Po(5)$ - number of errors per computer program
  - $E(X) = Var(X) = 5$

- $X_1, X_2, \ldots, X_{125}$ - number of errors in the programs.

$$P(\bar{X}_n \leq 5.5) =$$

$$= P\left(\frac{\sqrt{125}(\bar{X}_n - 5)}{\sqrt{5}} \leq \frac{\sqrt{125}(5.5 - 5)}{\sqrt{5}}\right) \approx$$

$$\approx P(Z \leq 2.5) = 0.9938$$

CLT IN ACTION

Google Classroom -> Lecture 7 -> Mean of means

PROPERTIES OF ESTIMATORS

Bias, variance and consistency

PROPERTIES OF ESTIMATORS

- Parameter estimation:

  Given samples $X_1, X_2, \ldots X_n$

  Obtain estimate $\hat{\theta}$ of an unknown parameter θ.

PROPERTIES OF ESTIMATORS

- Parameter estimation:

  Given samples $X_1, X_2, \ldots X_n$
  Obtain estimate $\hat{\theta}$ of an unknown parameter θ.

- $\hat{\theta} = T(X_1, \ldots, X_n)$, $T(X)$ − estimator.

PROPERTIES OF ESTIMATORS

- Parameter estimation:
  Given samples $X_1, X_2, \ldots X_n$
  Obtain estimate $\hat{\theta}$ of an unknown parameter θ.

- $\hat{\theta} = T(X_1, \ldots, X_n)$, $\qquad T(X) -$ estimator.

- So far, we only constructed ML estimators.

PROPERTIES OF ESTIMATORS

- Parameter estimation:
  Given samples $X_1, X_2, \dots X_n$
  Obtain estimate $\hat{\theta}$ of an unknown parameter θ.

- $\hat{\theta} = T(X_1, \dots, X_n)$, $\quad T(X) -$ estimator.

- So far, we only constructed ML estimators.
  But *anything* can be an estimator:

PROPERTIES OF ESTIMATORS

- Parameter estimation:

  Given samples $X_1, X_2, \dots X_n$
  Obtain estimate $\hat{\theta}$ of an unknown parameter θ.

- $\hat{\theta} = T(X_1, \dots, X_n)$, $\qquad T(X) -$ estimator.

- So far, we only constructed ML estimators.
  But *anything* can be an estimator:

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \qquad \hat{\theta}_2 = \frac{X_1 + X_n}{2}, \qquad \hat{\theta}_3 = \max(X_1, \dots, X_n)\,, \qquad \dots$$

PROPERTIES OF ESTIMATORS

- Parameter estimation:

  Given samples $X_1, X_2, \ldots X_n$
  Obtain estimate $\hat{\theta}$ of an unknown parameter θ.

- $\hat{\theta} = T(X_1, \ldots, X_n)$, $\quad\quad T(X) -$ estimator.

- So far, we only constructed ML estimators.
  But *anything* can be an estimator:

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \qquad \hat{\theta}_2 = \frac{X_1 + X_n}{2}, \qquad \hat{\theta}_3 = \max(X_1, \ldots, X_n)\,, \qquad \ldots$$

- We need to compare different estimators.

PROPERTIES OF ESTIMATORS

- Bias
- Variance
- Consistency

BIAS

An estimator $T(X)$ is unbiased of θ if

$$E\big(T(X)\big) = \theta$$

BIAS

An estimator $T(X)$ is unbiased of θ if

$$E\big(T(X)\big) = \theta$$

Bias is defined as

$$bias\big(T(X)\big) = \theta - E\big(T(X)\big)$$

BIAS: EXAMPLE 1

$$X_1, \dots, X_n \sim Bernoulli(p)$$

$\hat{p}_{ML} = \frac{1}{n} \sum_{i=1}^{n} X_i$ − (un)biased?

BIAS: EXAMPLE 1

$$X_1, \dots, X_n \sim Bernoulli(p)$$

$\hat{p}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – (un)biased?

$$E(\hat{p}_{ML}) =$$

BIAS: EXAMPLE 1

$$X_1, \dots, X_n \sim Bernoulli(p)$$

$\hat{p}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – (un)biased?

$$E(\hat{p}_{ML}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) =$$

BIAS: EXAMPLE 1

$$X_1, \dots, X_n \sim Bernoulli(p)$$

$\hat{p}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – (un)biased?

$$E(\hat{p}_{ML}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) =$$

BIAS: EXAMPLE 1

$$X_1, \dots, X_n \sim Bernoulli(p)$$

$\hat{p}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – (un)biased?

$$E(\hat{p}_{ML}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot np = p$$

BIAS: EXAMPLE 2

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – (un)biased?

BIAS: EXAMPLE 2

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – (un)biased?

$$E(\hat{\mu}_{ML}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n\mu = \mu$$

BIAS: EXAMPLE 3

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$\hat{\sigma}^2_{ML} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 -$ (un)biased?

BIAS: EXAMPLE 3

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$\hat{\sigma}^2_{ML} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ – (un)biased?

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 =$$

BIAS: EXAMPLE 3

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$\hat{\sigma}^2_{ML} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ − (un)biased?

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^2 =$$

BIAS: EXAMPLE 3

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$$

$\hat{\sigma}^2_{ML} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 -$ (un)biased?

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

BIAS: EXAMPLE 3

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

BIAS: EXAMPLE 3

$$\hat{\sigma}_{ML}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\mathrm{E}(\hat{\sigma}_{ML}^2) =$$

BIAS: EXAMPLE 3

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 - \mathrm{E}\bar{X}^2$$

BIAS: EXAMPLE 3

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 - \mathrm{E}\bar{X}^2$$

$$EX_i^2 = ?$$

BIAS: EXAMPLE 3

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 - \mathrm{E}\bar{X}^2$$

$$EX_i^2 = ?$$

$$Var(X_i) = \sigma^2 =$$

BIAS: EXAMPLE 3

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 - \mathrm{E}\bar{X}^2$$

$$EX_i^2 = ?$$

$$Var(X_i) = \sigma^2 = EX_i^2 - (EX_i)^2 =$$

BIAS: EXAMPLE 3

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 - \mathrm{E}\bar{X}^2$$

$$EX_i^2 = ?$$

$$Var(X_i) = \sigma^2 = EX_i^2 - (EX_i)^2 = EX_i^2 - \mu^2$$

BIAS: EXAMPLE 3

$$\hat{\sigma}_{ML}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 - \mathrm{E}\bar{X}^2$$

$$EX_i^2 = ?$$

$$Var(X_i) = \sigma^2 = EX_i^2 - (EX_i)^2 = EX_i^2 - \mu^2$$

$$E\left(X_i^2\right) = \sigma^2 + \mu^2$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

$$Var(\bar{X}) =$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

$$Var(\bar{X}) = \mathrm{E}\bar{X}^2 - (\mathrm{E}\bar{X})^2 =$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

$$Var(\bar{X}) = \mathrm{E}\bar{X}^2 - (\mathrm{E}\bar{X})^2 = \mathrm{E}\bar{X}^2 - \mu^2$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

$$Var(\bar{X}) = \mathrm{E}\bar{X}^2 - (\mathrm{E}\bar{X})^2 = \mathrm{E}\bar{X}^2 - \mu^2$$

$$Var(\bar{X}) =$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

$$Var(\bar{X}) = \mathrm{E}\bar{X}^2 - (\mathrm{E}\bar{X})^2 = \mathrm{E}\bar{X}^2 - \mu^2$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) =$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

$$Var(\bar{X}) = \mathrm{E}\bar{X}^2 - (\mathrm{E}\bar{X})^2 = \mathrm{E}\bar{X}^2 - \mu^2$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \mathrm{E}\bar{X}^2$$

$$\mathrm{E}\bar{X}^2 = ?$$

$$Var(\bar{X}) = \mathrm{E}\bar{X}^2 - (\mathrm{E}\bar{X})^2 = \mathrm{E}\bar{X}^2 - \mu^2$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\mathrm{E}\bar{X}^2 = \frac{\sigma^2}{n} + \mu^2$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 =$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{(n-1)\sigma^2}{n}$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{(n-1)\sigma^2}{n}$$

Sample variance $\mathrm{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a **biased** estimator for the parameter σ^2 of the Normal distribution.

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}_{ML}^2) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{(n-1)\sigma^2}{n}$$

Sample variance $\mathrm{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a **biased** estimator for the parameter σ^2 of the Normal distribution.

- How to make it **unbiased**?

$\mathrm{S}_{unbiased}^2 =$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{(n-1)\sigma^2}{n}$$

Sample variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a **biased** estimator for the parameter σ^2 of the Normal distribution.

- How to make it **unbiased**?

$$S^2_{unbiased} = \frac{\boldsymbol{n}}{(\boldsymbol{n}-\mathbf{1})} \cdot S^2 =$$

BIAS: EXAMPLE 3

$$\mathrm{E}(\hat{\sigma}^2_{ML}) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{(n-1)\sigma^2}{n}$$

Sample variance $\mathrm{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a **biased** estimator for the parameter σ^2 of the Normal distribution.

- How to make it **unbiased**?

$$\mathrm{S}^2_{unbiased} = \frac{\boldsymbol{n}}{(\boldsymbol{n}-\mathbf{1})} \cdot S^2 = \frac{\mathbf{1}}{\boldsymbol{n}-\mathbf{1}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} (\boldsymbol{X_i} - \bar{\boldsymbol{X}})^{\mathbf{2}}$$

VARIANCE

Variance of an estimator $T(X)$ is defined as

$$Var\big(T(X)\big) = E\big\{T(X) - E\big(T(X)\big)\big\}^2$$

VARIANCE: EXAMPLE

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

VARIANCE: EXAMPLE

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i \,, = \bar{X}, \qquad Var(\hat{\mu}_{ML}) = ?$$

VARIANCE: EXAMPLE

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$$

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i \,, = \bar{X}, \qquad Var(\hat{\mu}_{ML}) = ?$$

$$Var(\bar{X}) =$$

VARIANCE: EXAMPLE

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i \, , = \bar{X}, \qquad Var(\hat{\mu}_{ML}) = ?$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) =$$

VARIANCE: EXAMPLE

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i \,, = \bar{X}, \qquad Var(\hat{\mu}_{ML}) = ?$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

BIAS AND VARIANCE EXAMPLE

- Can a biased and an unbiased estimator have the same variance?

BIAS AND VARIANCE EXAMPLE

- Can a biased and an unbiased estimator have the same variance?
- Example:

$$\hat{\mu}_{ML} = \bar{X} - \text{unbiased}$$

$$\hat{\mu}_2 = \hat{\mu}_{ML} + 1 - \text{biased}$$

BIAS AND VARIANCE EXAMPLE

- Can a biased and an unbiased estimator have the same variance?
- Example:

$$\hat{\mu}_{ML} = \bar{X} - \text{unbiased}$$

$$Var(\hat{\mu}_{ML}) = \frac{\sigma^2}{n}$$

$$\hat{\mu}_2 = \hat{\mu}_{ML} + 1 - \text{biased}$$

BIAS AND VARIANCE EXAMPLE

- Can a biased and an unbiased estimator have the same variance?
- Example:

$$\hat{\mu}_{ML} = \bar{X} - \text{unbiased}$$

$$Var(\hat{\mu}_{ML}) = \frac{\sigma^2}{n}$$

$$\hat{\mu}_2 = \hat{\mu}_{ML} + 1 - \text{biased}$$

but

$$Var\ \hat{\mu}_2\ =$$

BIAS AND VARIANCE EXAMPLE

- Can a biased and an unbiased estimator have the same variance?
- Example:

$$\hat{\mu}_{ML} = \bar{X} - \text{unbiased}$$

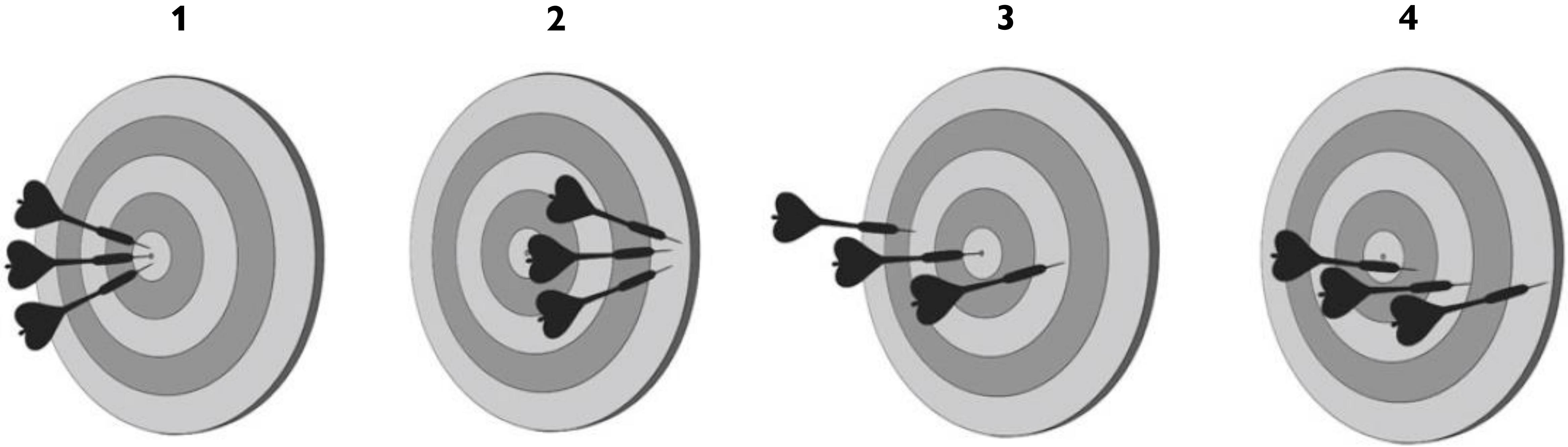
$$Var(\hat{\mu}_{ML}) = \frac{\sigma^2}{n}$$

$$\hat{\mu}_2 = \hat{\mu}_{ML} + 1 - \text{biased}$$

but

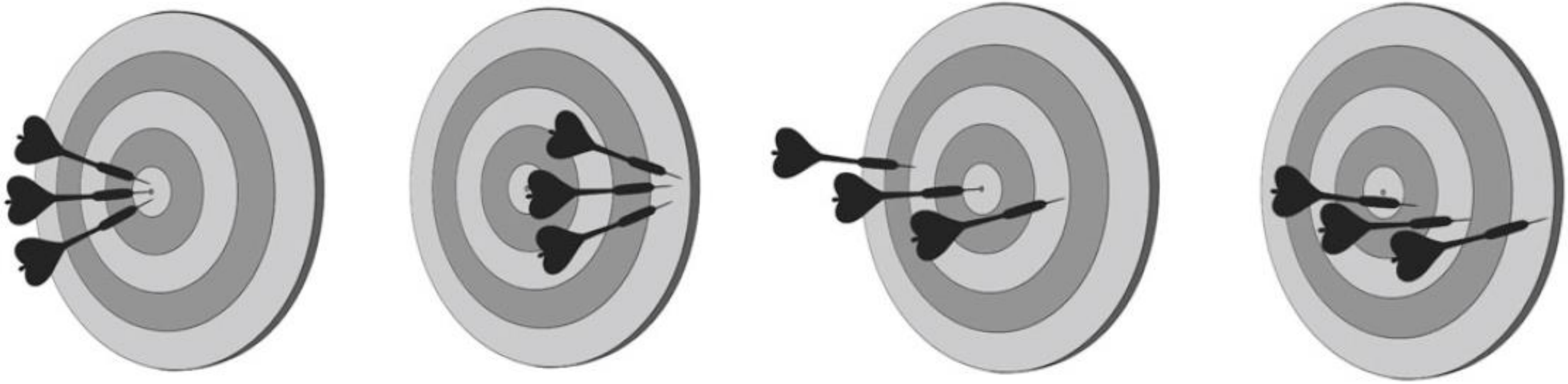
$$Var(\hat{\mu}_2) = \frac{\sigma^2}{n}$$

BIAS VS VARIANCE



BIAS VS VARIANCE

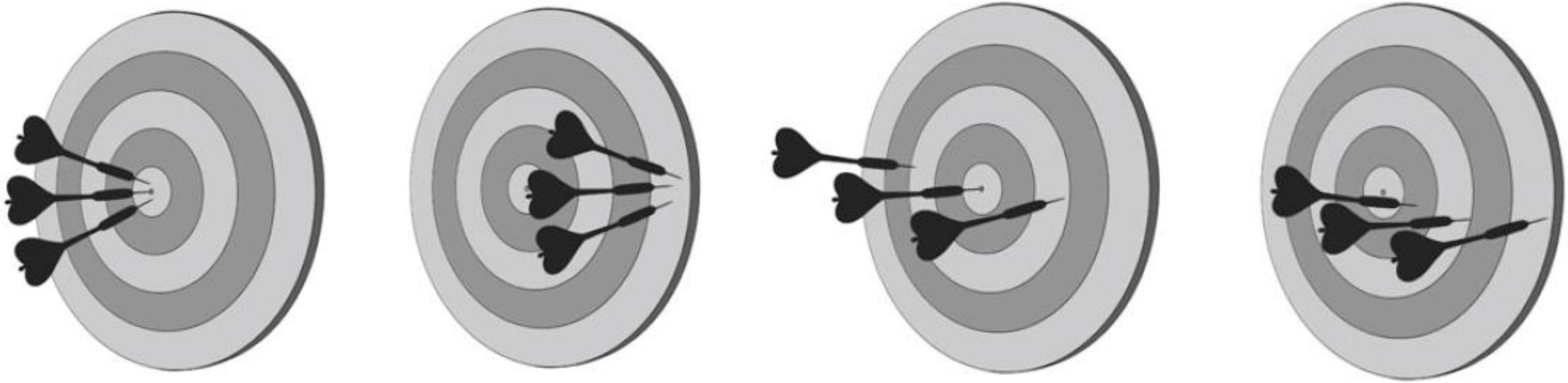
1 **2** **3** **4**



LOW BIAS
LOW VARIANCE

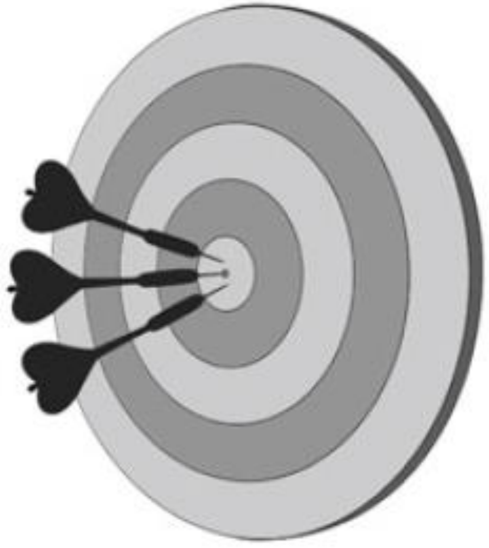
BIAS VS VARIANCE

1 2 3 4



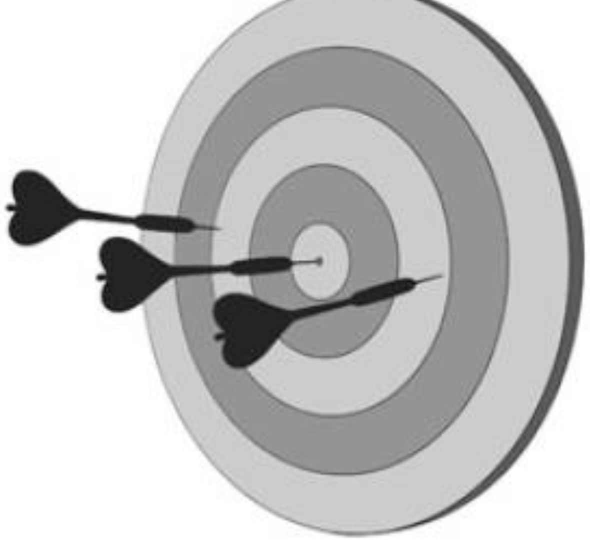
BIAS VS VARIANCE

1



2

3

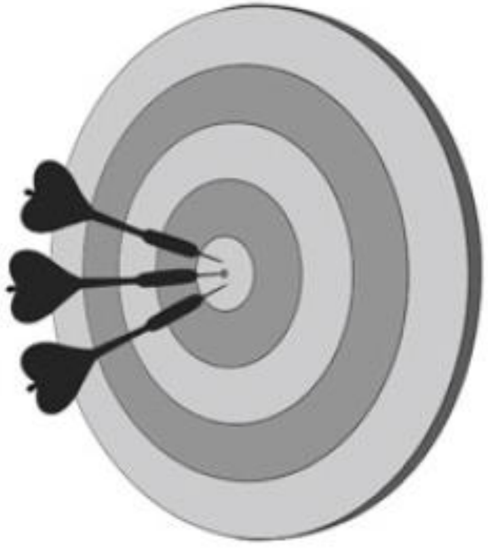


LOW BIAS
HIGH VARIANCE

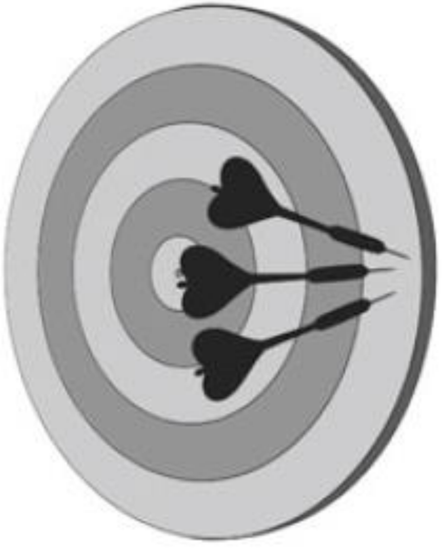
4

BIAS VS VARIANCE

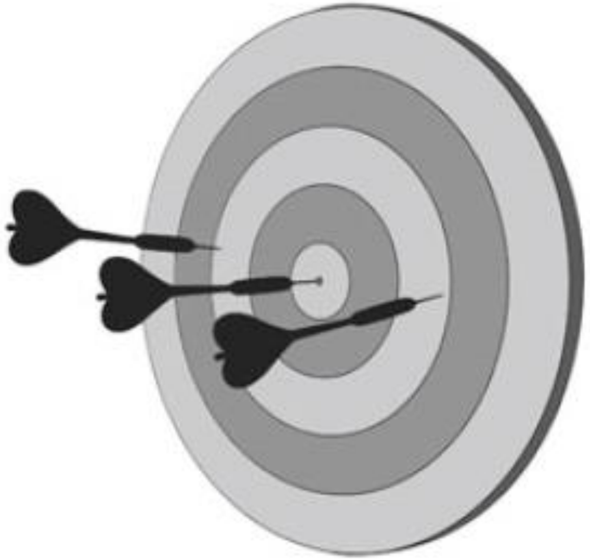
1



2



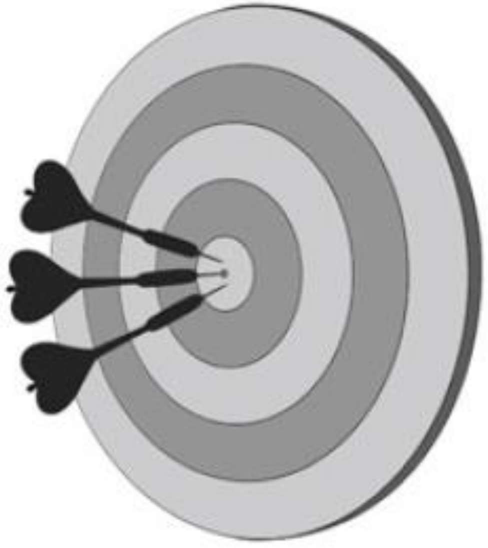
3



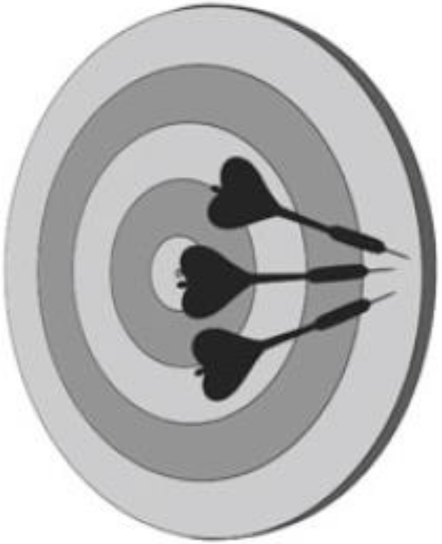
4

BIAS VS VARIANCE

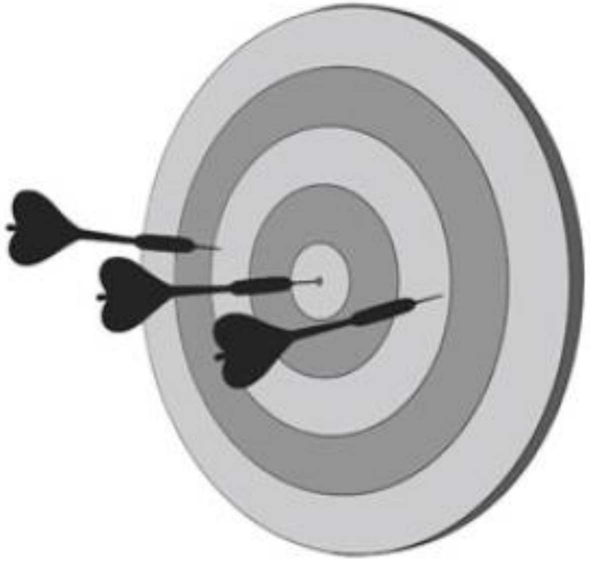
1



2



3

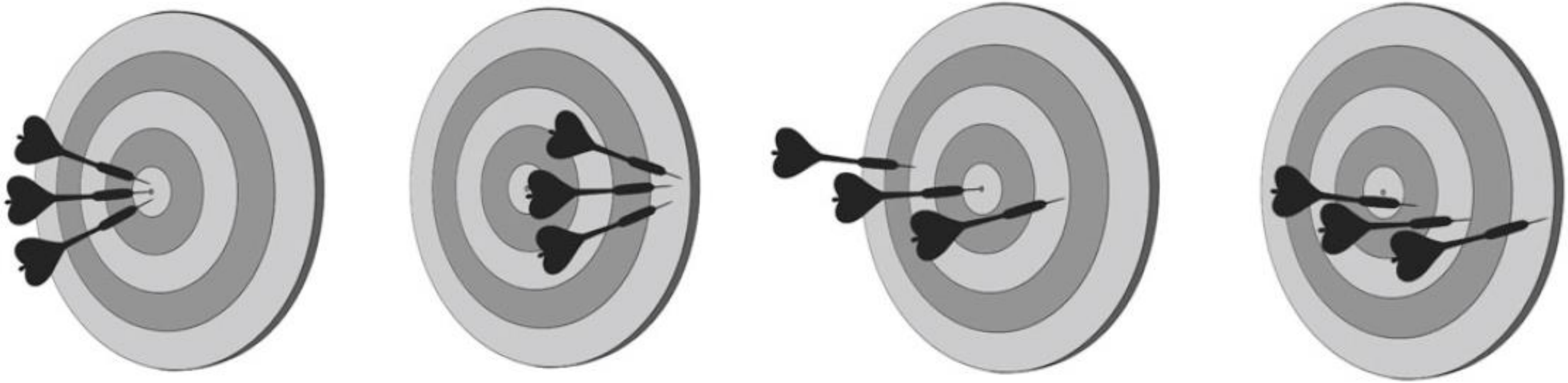


4

HIGH BIAS
LOW VARIANCE

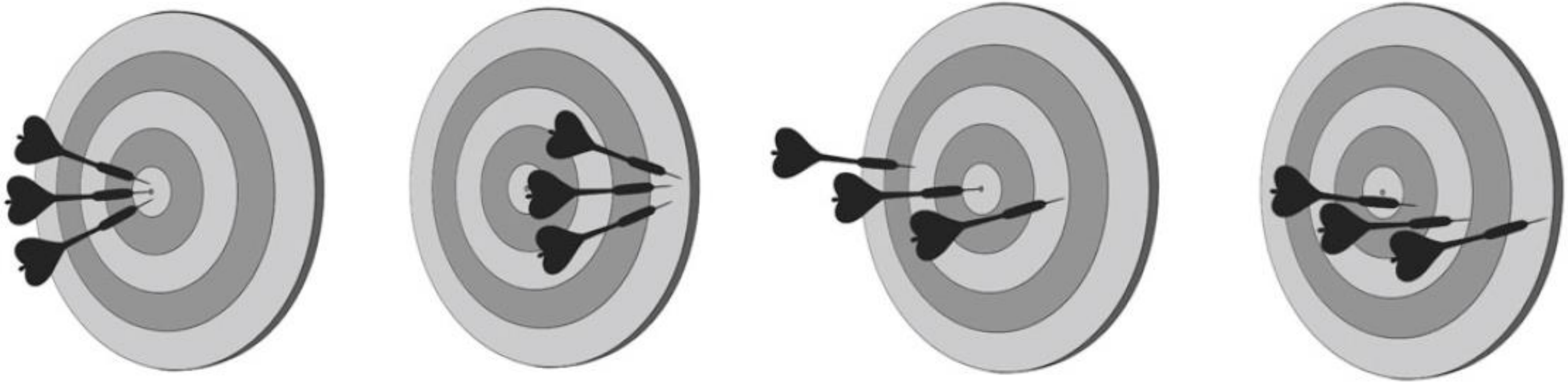
BIAS VS VARIANCE

1 2 3 4



BIAS VS VARIANCE

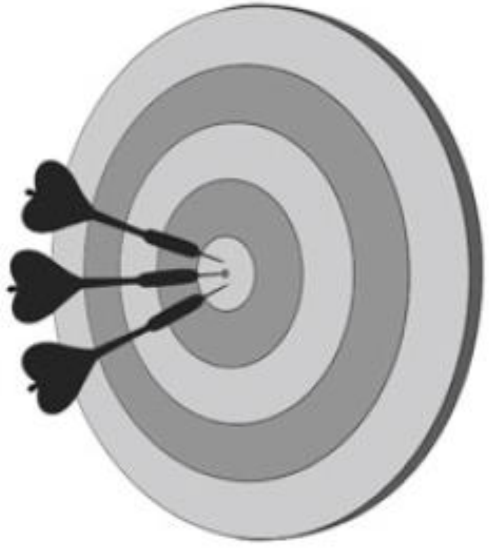
1 **2** **3** **4**



HIGH BIAS
HIGH VARIANCE

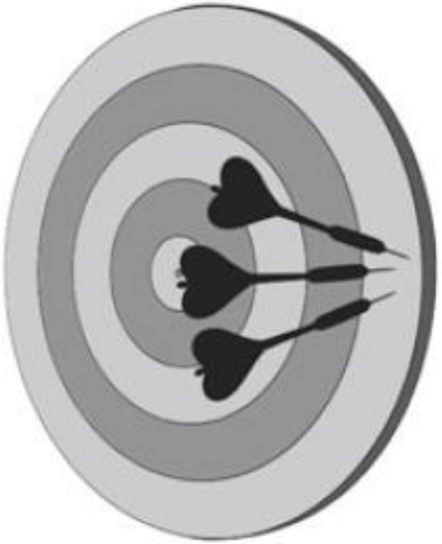
BIAS VS VARIANCE

1



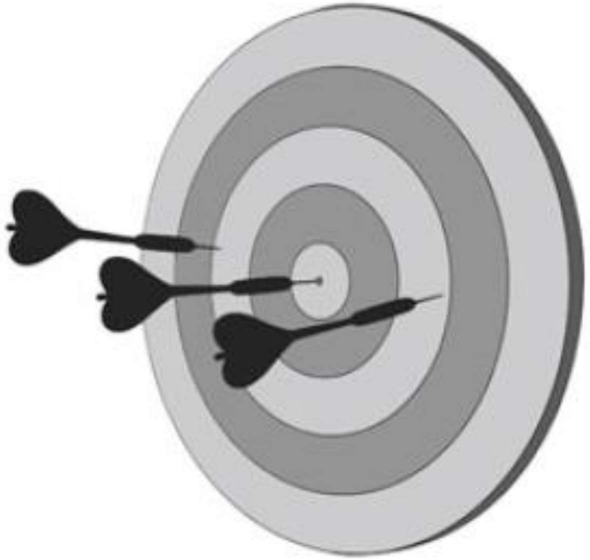
LOW BIAS
LOW VARIANCE

2



HIGH BIAS
LOW VARIANCE

3



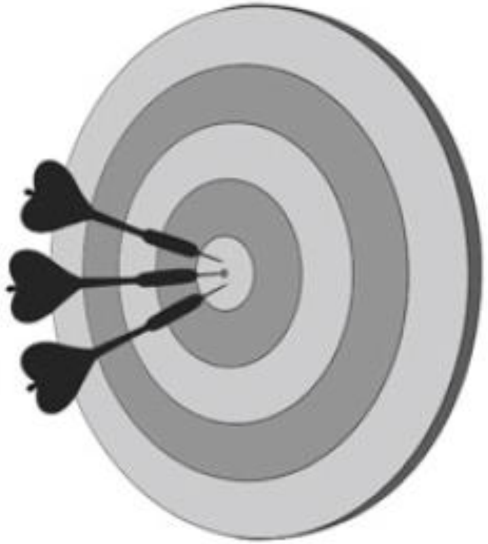
LOW BIAS
HIGH VARIANCE

4

HIGH BIAS
HIGH VARIANCE

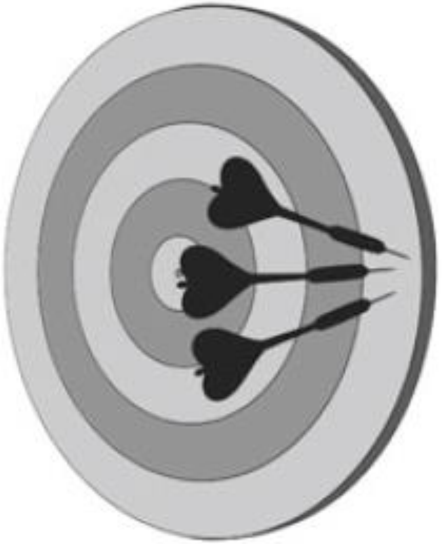
BIAS VS VARIANCE

1



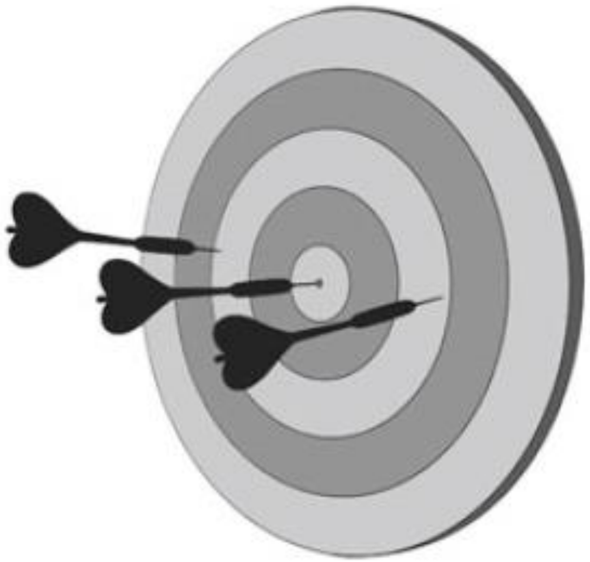
LOW BIAS
LOW VARIANCE

2



HIGH BIAS
LOW VARIANCE

3



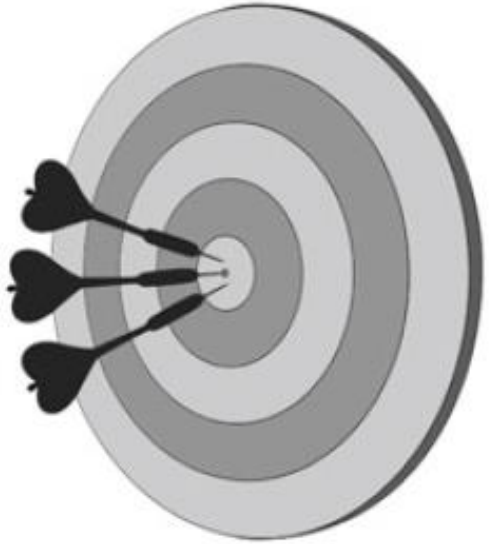
LOW BIAS
HIGH VARIANCE

4

HIGH BIAS
HIGH VARIANCE

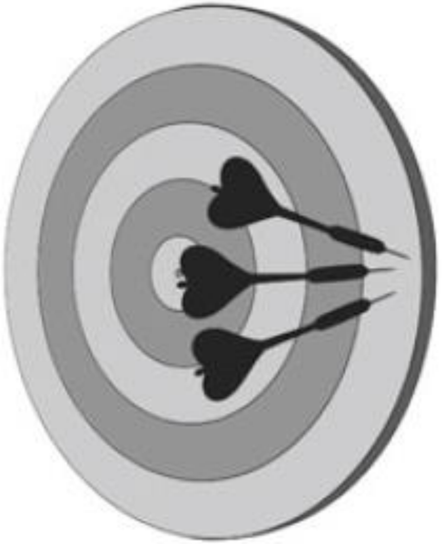
BIAS VS VARIANCE

1



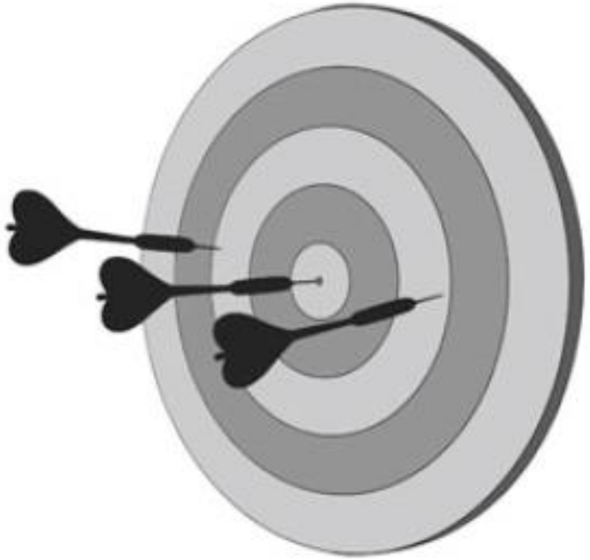
LOW BIAS
LOW VARIANCE

2



HIGH BIAS
LOW VARIANCE

3



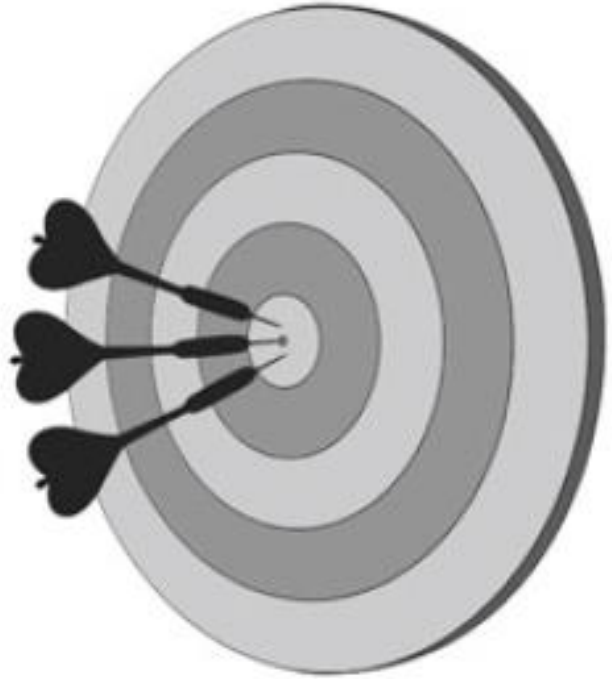
LOW BIAS
HIGH VARIANCE

4

HIGH BIAS
HIGH VARIANCE

BIAS-VARIANCE TRADE-OFF

- Impossible to simultaneously optimize bias and variance.
- Related to *under*- and *overfitting* in Machine Learning.



LOW BIAS
LOW VARIANCE

CONSISTENCY

For a good estimator, as the sample size increases, the values of the estimator should get closer to the parameter being estimated.

CONSISTENCY

For a good estimator, as the sample size increases, the values of the estimator should get closer to the parameter being estimated.

Let $T_1, T_2, \ldots T_n$ be a sequence of estimators for $\theta, T_k = T(X_1, \ldots, X_k)$.

CONSISTENCY

For a good estimator, as the sample size increases, the values of the estimator should get closer to the parameter being estimated.

Let $T_1, T_2, \ldots T_n$ be a sequence of estimators for $\theta, T_k = T(X_1, \ldots, X_k)$.

Then $\{T_n\}$ is consistent if $\forall \epsilon > 0$

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

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu,$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

Let's show that $\forall \epsilon > 0 \ \lim_{n\to\infty} P(|T_n - \mu| \geq \epsilon) = 0$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

Let's show that $\forall \epsilon > 0 \; \lim_{n\to\infty} P(|T_n - \mu| \geq \epsilon) = 0$

Let's chose $\epsilon = \frac{c\sigma}{\sqrt{n}}$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

Let's show that $\forall \epsilon > 0 \ \lim_{n\to\infty} P(|T_n - \mu| \geq \epsilon) = 0$

Let's chose $\epsilon = \frac{c\sigma}{\sqrt{n}}$

$$P\left(|T_n - \mu| \geq \frac{c\sigma}{\sqrt{n}}\right) \leq$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

Let's show that $\forall \epsilon > 0 \ \lim_{n\to\infty} P(|T_n - \mu| \geq \epsilon) = 0$

Let's chose $\epsilon = \frac{c\sigma}{\sqrt{n}}$

Chebyshev's inequality:
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P\left(|T_n - \mu| \geq \frac{c\sigma}{\sqrt{n}}\right) \leq$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

Let's show that $\forall \epsilon > 0 \ \lim_{n \to \infty} P(|T_n - \mu| \geq \epsilon) = 0$

Let's chose $\epsilon = \frac{c\sigma}{\sqrt{n}}$

Chebyshev's inequality:

$$P \quad X - \mu \quad \geq k\sigma \quad \leq \frac{1}{k^2}$$

$$P\left(|T_n - \mu| \geq \frac{c\sigma}{\sqrt{n}}\right) \leq \frac{1}{c^2} =$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

Let's show that $\forall \epsilon > 0 \ \lim_{n\to\infty} P(|T_n - \mu| \geq \epsilon) = 0$

Let's chose $\epsilon = \frac{c\sigma}{\sqrt{n}}$

Chebyshev's inequality:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P\left(|T_n - \mu| \geq \frac{c\sigma}{\sqrt{n}}\right) \leq \frac{1}{c^2} = \frac{\sigma^2}{n\epsilon^2}$$

CONSISTENCY: EXAMPLE

$$\hat{\mu} = \bar{X}, \qquad E\hat{\mu} = \mu, \qquad Var(\hat{\mu}) = \sigma^2/n$$

Let's show that $\forall \epsilon > 0 \ \lim_{n\to\infty} P(|T_n - \mu| \geq \epsilon) = 0$

Let's chose $\epsilon = \frac{c\sigma}{\sqrt{n}}$

Chebyshev's inequality:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P\left(|T_n - \mu| \geq \frac{c\sigma}{\sqrt{n}}\right) \leq \frac{1}{c^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \qquad n \to \infty$$

PROPERTIES OF MAX. LIKELIHOOD EST.

PROPERTIES OF MAX. LIKELIHOOD EST.

1. The MLE is **consistent**:

PROPERTIES OF MAX. LIKELIHOOD EST.

1. The MLE is **consistent**:

$$\hat{\theta} \xrightarrow{P} \theta$$

PROPERTIES OF MAX. LIKELIHOOD EST.

1. The MLE is **consistent**:

$$\hat{\theta} \xrightarrow{P} \theta$$

2. The MLE is **equivariant**:

PROPERTIES OF MAX. LIKELIHOOD EST.

1. The MLE is **consistent**:

$$\hat{\theta} \xrightarrow{P} \theta$$

2. The MLE is **equivariant**:

If $\hat{\theta}$ is MLE for θ than $g(\hat{\theta})$ is MLE for $g(\theta)$

PROPERTIES OF MAX. LIKELIHOOD EST.

1. The MLE is **consistent**:

$$\hat{\theta} \xrightarrow{P} \theta$$

2. The MLE is **equivariant**:

If $\hat{\theta}$ is MLE for θ than $g(\hat{\theta})$ is MLE for $g(\theta)$

3. MLE is **asymptotically efficient**:

PROPERTIES OF MAX. LIKELIHOOD EST.

1. The MLE is **consistent**:

$$\hat{\theta} \xrightarrow{P} \theta$$

2. The MLE is **equivariant**:

If $\hat{\theta}$ is MLE for θ than $g(\hat{\theta})$ is MLE for $g(\theta)$

3. MLE is **asymptotically efficient**:
   roughly speaking, among well-behaved estimators, it has the smallest variance, at least for large samples.

CONFIDENCE INTERVALS

MOTIVATION

- Point estimation provides a single “best guess”.

MOTIVATION

- Point estimation provides a single “best guess”.

- How confident can we be in our estimation?

MOTIVATION

- Point estimation provides a single “best guess”.
- How confident can we be in our estimation?
- CI proposes a range of plausible values.

DEFINITION

A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (a, b)$ such that $\mathrm{T}_1 = t_1(X_1, \dots, X_n)$, $T_2 = t_2(X_1, \dots, X_n)$ and

$$P(T_1 < \theta < T_2) \geq 1 - \alpha$$

DEFINITION

A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (T_1, T_2)$ such that $\mathrm{T}_1 = t_1(X_1, \dots, X_n)$, $T_2 = t_2(X_1, \dots, X_n)$ and

$$P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Random** intervals: T_1 and T_2 are functions of random samples.

DEFINITION

A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (T_1, T_2)$ such that $\mathrm{T}_1 = t_1(X_1, \dots, X_n)$, $T_2 = t_2(X_1, \dots, X_n)$ and

$$P(T_1 < \theta < T_2) \geq 1-\alpha$$

- **Random** intervals: T_1 and T_2 are functions of random samples.
- θ is unknown, but fixed
  T_1 and T_2 are random

INTERPRETATION

$$C_n = (T_1, T_2):\ \ P(T_1 < \theta < T_2) \geq 1 - \alpha$$

INTERPRETATION

$$C_n = (T_1, T_2):\;\; P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Not** a probability statement about θ since it's fixed.

INTERPRETATION

$$C_n = (T_1, T_2): \;\; P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Not** a probability statement about θ since it's fixed.
- Common interpretation:

INTERPRETATION

$$C_n = (T_1, T_2):\ \ P(T_1 < \theta < T_2) \geq 1 - \alpha$$

- **Not** a probability statement about θ since it's fixed.

- Common interpretation:

If I repeat the experiment many times, the interval will contain the true value of θ 95% of the time (α=0.05).

CI FOR MEAN (VARIANCE IS KNOWN)

CI FOR μ (σ IS KNOWN)

- Suppose that $X_1, X_2, \dots, X_n$ - i.i.d. samples from a distribution with unknown mean μ and known variance σ^2.

CI FOR μ (σ IS KNOWN)

- Suppose that $X_1, X_2, \dots, X_n$ - i.i.d. samples from a distribution with unknown mean μ and known variance σ^2.

- Point estimate for μ:

$$\mu = \bar{X} = \sum_{i=1}^{n} X_i$$

CI FOR μ (σ IS KNOWN)

- Suppose that $X_1, X_2, \ldots, X_n$ - i.i.d. samples from a distribution with unknown mean μ and <u>known</u> variance σ^2.

- Point estimate for μ:

$$\mu = \bar{X} = \sum_{i=1}^{n} X_i$$

- How to compute a confidence interval?

CI FOR μ (σ IS KNOWN)

Central Limit Theorem:

CI FOR μ (σ IS KNOWN)

Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$

CI FOR μ (σ IS KNOWN)

Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0, 1)$

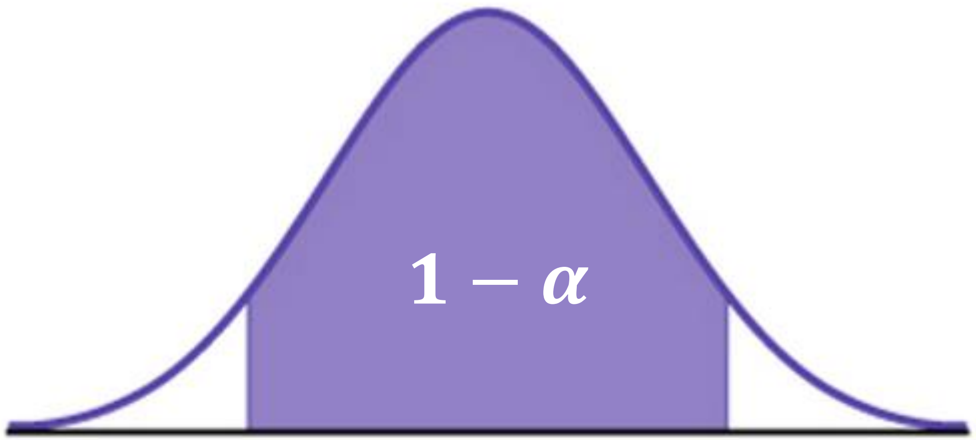
CI FOR μ (σ IS KNOWN)

Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0,1)$

$$P\left(\qquad < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < \qquad \right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

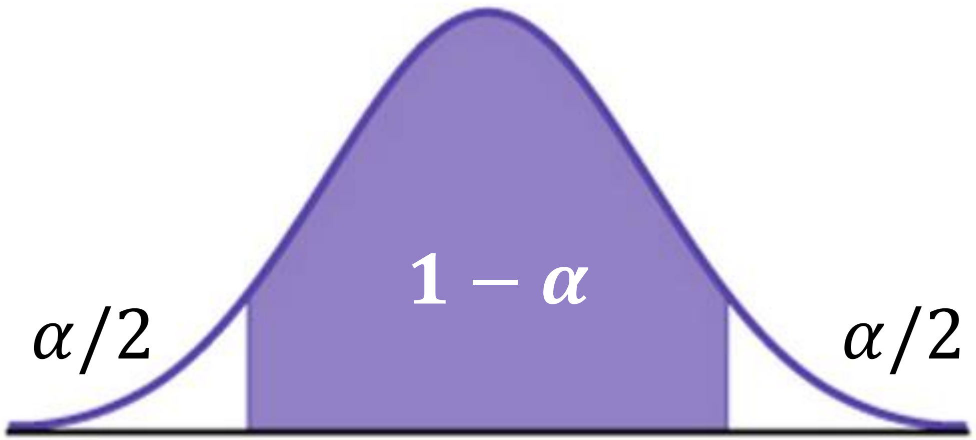
Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0,1)$



$$P\left(\qquad < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < \qquad \right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

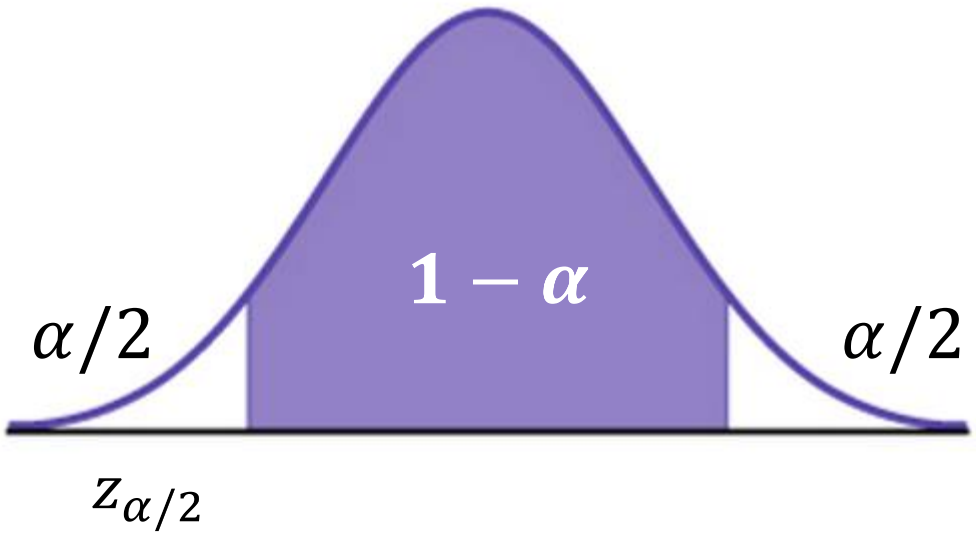
Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0,1)$



$$P\left(\qquad < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < \qquad \right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

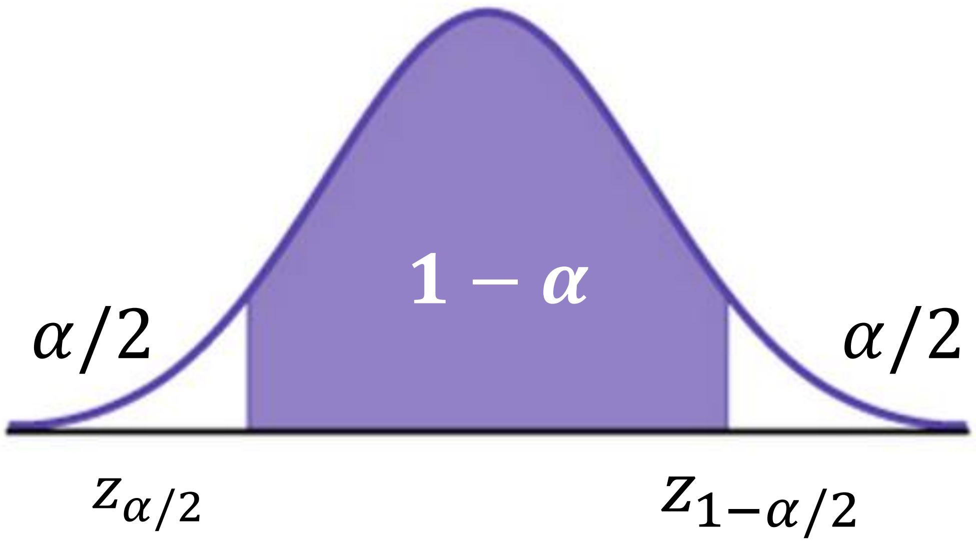
Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0,1)$



$$P\left(\qquad < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < \qquad \right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

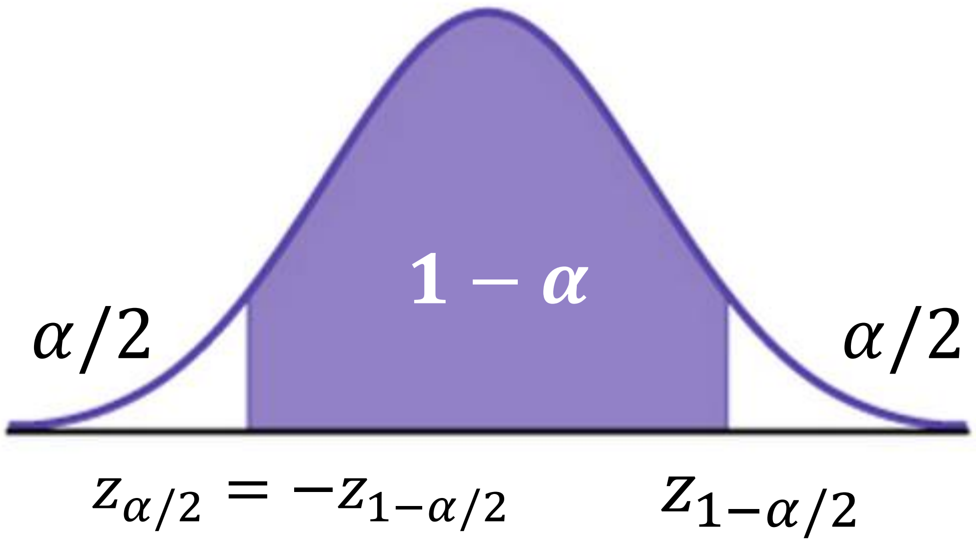
Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0,1)$



$$P\left(\qquad < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < \qquad \right) = 1-\alpha$$

CI FOR μ (σ IS KNOWN)

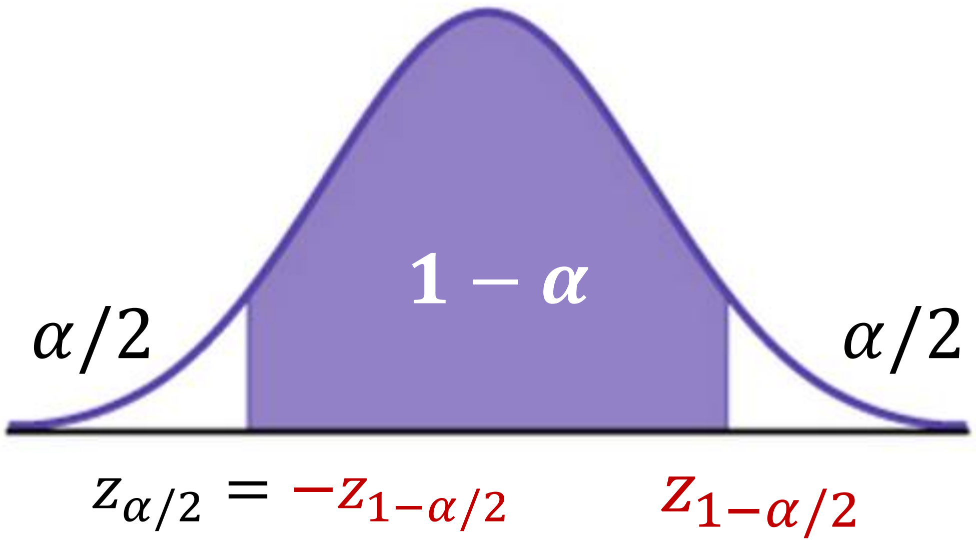
Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0,1)$



$$P\left(\qquad < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < \qquad \right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

Central Limit Theorem: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \approx N(0,1)$



$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

CI FOR μ (σ IS KNOWN)

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

$$P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.95$

$$P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.95$

$$P\left(5 - \frac{1}{10} z_{1-\frac{0.05}{2}} < \mu < 5 + \frac{1}{10} z_{1-\frac{0.05}{2}}\right) = 0.95$$

CI FOR μ (σ IS KNOWN)

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.95$

$$P\left(5 - \frac{1}{10} z_{1-\frac{0.05}{2}} < \mu < 5 + \frac{1}{10} z_{1-\frac{0.05}{2}}\right) = 0.95$$

CI FOR μ (σ IS KNOWN)

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.95$

$$P\left(5 - \frac{1}{10} z_{1-\frac{0.05}{2}} < \mu < 5 + \frac{1}{10} z_{1-\frac{0.05}{2}}\right) = 0.95$$

$$P(5 - 0.196 < \mu < 5 + 0.196) = 0.95$$

CI FOR μ (σ IS KNOWN)

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.95$

$$P\left(5 - \frac{1}{10} z_{1-\frac{0.05}{2}} < \mu < 5 + \frac{1}{10} z_{1-\frac{0.05}{2}}\right) = 0.95$$

$$P(5 - 0.196 < \mu < 5 + 0.196) = 0.95$$

$$(\mathbf{5 - 0.196\,; 5 + 0.196})$$

CI FOR μ (σ IS KNOWN)

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.90$

$$P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

CI FOR μ (σ IS KNOWN)

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.90$

$$P\left(5 - \frac{1}{10} z_{1-\frac{0.1}{2}} < \mu < 5 + \frac{1}{10} z_{1-\frac{0.1}{2}}\right) = 0.90$$

CI FOR μ (σ IS KNOWN)

Quantile (p)	$\Phi^{-1}(p, 0, 1)$
0.995	2.58
0.99	2.33
0.975	1.96
0.95	1.64
0.9	1.28

- Example:

$$n = 100, \qquad \bar{X} = 5, \qquad \sigma = 1$$

- Construct CI for μ at the level $1 - \alpha = 0.90$

$$P\left(5 - \frac{1}{10} z_{0.95} < \mu < 5 + \frac{1}{10} z_{0.95}\right) = 0.90$$

$$P(5 - 0.164 < \mu < 5 + 0.164) = 0.95$$

$$(\mathbf{5 - 0.164\, ;\, 5 + 0.164})$$

TO SUM SUP

- Central Limit Theorem
- Properties of estimators
  - bias, variance, consistency;
  - properties of ML estimates.
- Confidence intervals

MID-TERM

- Tomorrow, Wednesday, December 9
- 09:00 – 12:00 (no class)
- Assignment will become available on Google Classroom
- You should submit by 12:00

Topics:

- Descriptive statistics
- Discrete distributions
- Continuous random variables (CDFs, PDFs, probabilities)
- Maximum likelihood (discrete and continuous)

MID-TERM

- Tomorrow, Wednesday, December 9

- 09:00 – 12:00 (no class)

- Assignment will become available on Google Classroom

- You should submit by 12:00