PROBABILITY & STATISTICS

Lecture 11 – Confidence Intervals

MLE FOR NORMAL DISTRIBUTIONS

estimating μ and σ

MLE FOR μ AND σ

- X1, X2, …, Xn – samples from the normal distribution.
- Parameters μ and σ are unknown.

MLE FOR μ AND σ

- X1, X2, …, Xn – samples from the normal distribution.
- Parameters μ and σ are unknown.
- $\hat{\mu} = ?$ $\hat{\sigma} = ?$

MLE FOR μ AND σ

- X1, X2, …, Xn – samples from the normal distribution.
- Parameters μ and σ are unknown.
- $\hat{\mu} =?$ $\hat{\sigma} =?$
- Maximum likelihood.

MLE FOR μ AND σ

- X_1, X_2, …, X_n – samples from the normal distribution.
- $\hat{\mu} =?$ $\hat{\sigma} =?$

$$L(\mu, \sigma) =$$

MLE FOR μ AND σ

- $X_1, X_2, \ldots, X_n$ – samples from the normal distribution.
- $\hat{\mu} = ?\ \hat{\sigma} = ?$

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

MLE FOR μ AND σ

- $X_1, X_2, \ldots, X_n$ – samples from the normal distribution.
- $\hat{\mu} = ?$ $\hat{\sigma} = ?$

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

$$\log L(\mu, \sigma) =$$

MLE FOR μ AND σ

- $X_1, X_2, \ldots, X_n$ – samples from the normal distribution.
- $\hat{\mu} =?$ $\hat{\sigma} =?$

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

$$\log L(\mu, \sigma) = -\sum_{i=1}^{n} \log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 =$$

MLE FOR μ AND σ

- $X_1, X_2, \ldots, X_n$ – samples from the normal distribution.
- $\hat{\mu} =?$ $\hat{\sigma} =?$

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2}$$

$$\log L(\mu, \sigma) = -\sum_{i=1}^{n} \log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 =$$

$$= -n \log \sigma - n \log \sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

MLE FOR μ AND σ

$$\log L(\mu, \sigma) = -n \log \sigma - n \log \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\frac{d}{d\mu} \log L(\mu, \sigma) =$$

$$\frac{d}{d\sigma} \log L(\mu, \sigma) =$$

MLE FOR μ AND σ

$$\log L(\mu, \sigma) = -n \log \sigma - n \log \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\frac{d}{d\mu} \log L(\mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) = 0$$

$$\frac{d}{d\sigma} \log L(\mu, \sigma) =$$

MLE FOR μ AND σ

$$\log L(\mu, \sigma) = -n \log \sigma - n \log \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\frac{d}{d\mu} \log L(\mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) = 0$$

$$\frac{d}{d\sigma} \log L(\mu, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (X_i - \mu)^2 = 0$$

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0 \qquad \frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

$$\frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

$$\frac{1}{\sigma^2} \neq 0 \Rightarrow$$

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0 \qquad \frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$$

$$\frac{1}{\sigma^2} \neq 0 \Rightarrow$$

$$\sum_{i=1}^{n}(X_i-\mu) = \sum_{i=1}^{n} X_i - n\mu = 0$$

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0$$

$$\frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$$

$$\frac{1}{\sigma^2} \neq 0 \Rightarrow$$

$$\sum_{i=1}^{n}(X_i-\mu) = \sum_{i=1}^{n}X_i - n\mu = 0$$

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

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

$$\frac{1}{\sigma^2} \neq 0 \Rightarrow$$

$$\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n} X_i - n\mu = 0$$

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

$$\frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

$$-n\sigma^2 + \sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

$$\frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

$$\frac{1}{\sigma^2} \neq 0 \Rightarrow$$

$$-n\sigma^2 + \sum_{i=1}^{n}(X_i - \boxed{\mu})^2 = 0$$

$$\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n} X_i - n\mu = 0$$

$$\boxed{\boldsymbol{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}}}$$

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

$$\frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

$$\frac{1}{\sigma^2} \neq 0 \Rightarrow$$

$$\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n} X_i - n\mu = 0$$

$$\boxed{\boldsymbol{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}}}$$

$$-n\sigma^2 + \sum_{i=1}^{n}(X_i - \boxed{\mu})^2 = 0$$

$$-n\sigma^2 + \sum_{i=1}^{n}(X_i - \bar{X})^2 = 0$$

MLE FOR μ AND σ

$$\frac{d}{d\mu}\log L(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

$$\frac{1}{\sigma^2} \neq 0 \Rightarrow$$

$$\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n} X_i - n\mu = 0$$

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

$$\frac{d}{d\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

$$-n\sigma^2 + \sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

$$-n\sigma^2 + \sum_{i=1}^{n}(X_i - \bar{X})^2 = 0$$

$$\boldsymbol{\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2}$$

MLE FOR μ AND σ

- $X_1, X_2, \ldots, X_n$ – samples from the normal distribution.
- Parameters μ and σ are unknown.
- $\hat{\mu} = ?$ $\hat{\sigma} = ?$
- Maximum likelihood:

MLE FOR μ AND σ

- $X_1, X_2, \ldots, X_n$ – samples from the normal distribution.
- Parameters μ and σ are unknown.
- $\hat{\mu} = ?$ $\hat{\sigma} = ?$
- Maximum likelihood:
  - $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ **- MLE for μ is sample mean**

MLE FOR μ AND σ

- $X_1, X_2, \ldots, X_n$ – samples from the normal distribution.
- Parameters μ and σ are unknown.
- $\hat{\mu} = ?$ $\hat{\sigma} = ?$
- Maximum likelihood:
  - $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ - **MLE for μ is sample mean**
  - $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$ - **MLE for σ^2 is sample variance**

PROPERTIES OF ESTIMATORS

Bias, variance and consistency

PROPERTIES OF ESTIMATORS

- Parameter estimation:

  Given samples $X_1, X_2, \dots X_n$

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

PROPERTIES OF ESTIMATORS

- Parameter estimation:

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

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

PROPERTIES OF ESTIMATORS

- Parameter estimation:

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

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

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

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

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

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

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:

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

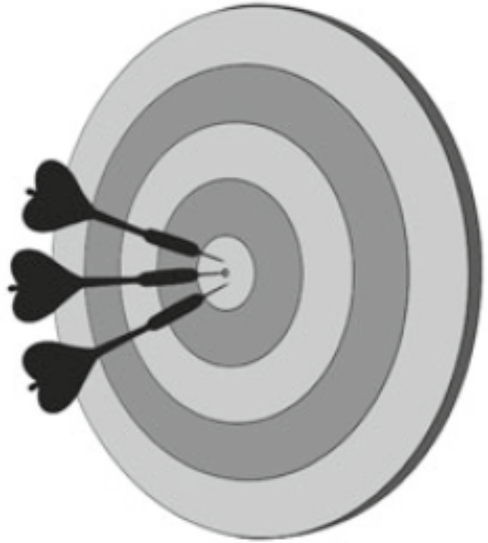
- We need to compare different estimators.

PROPERTIES OF ESTIMATORS

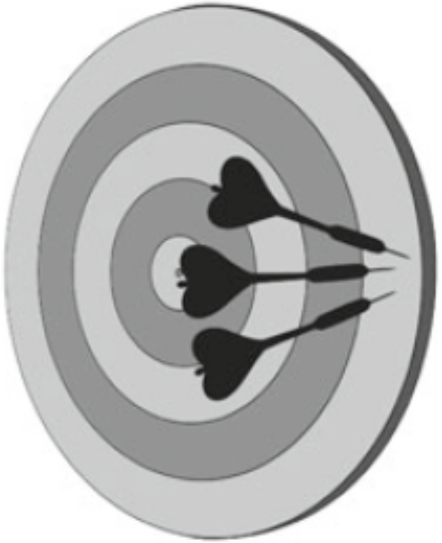
- Bias
- Variance
- Consistency

BIAS VS VARIANCE

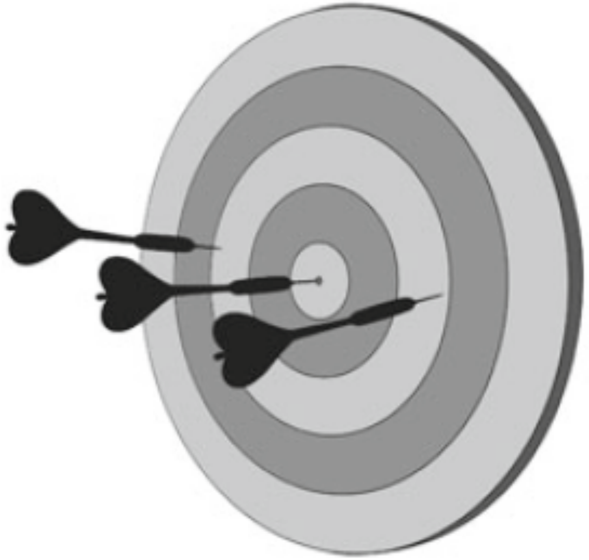
1



2

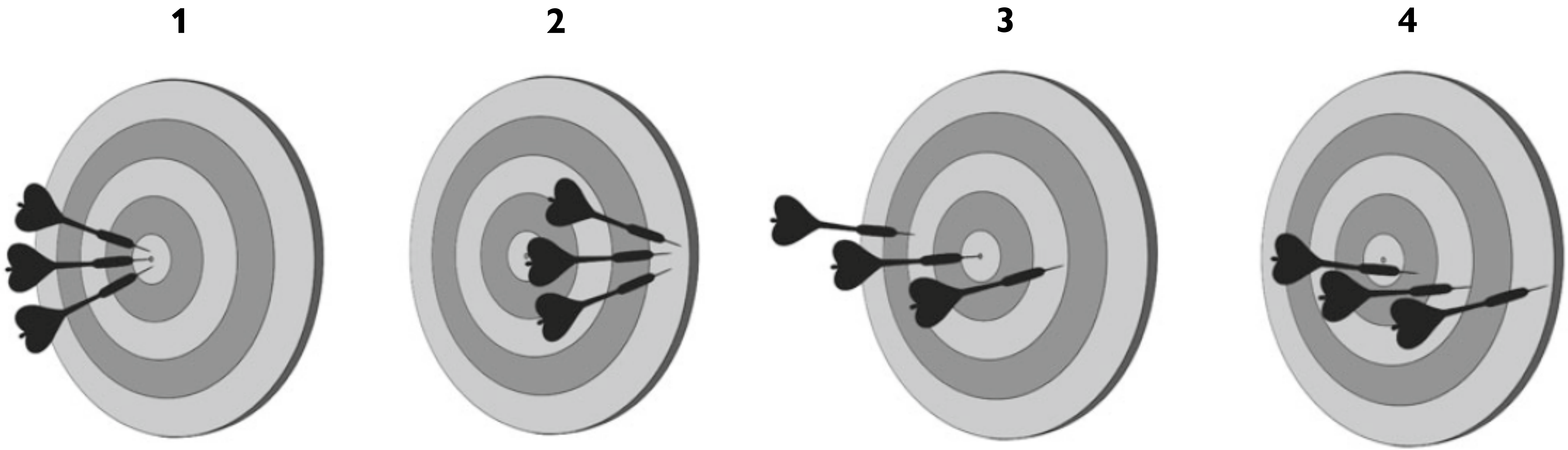


3



4

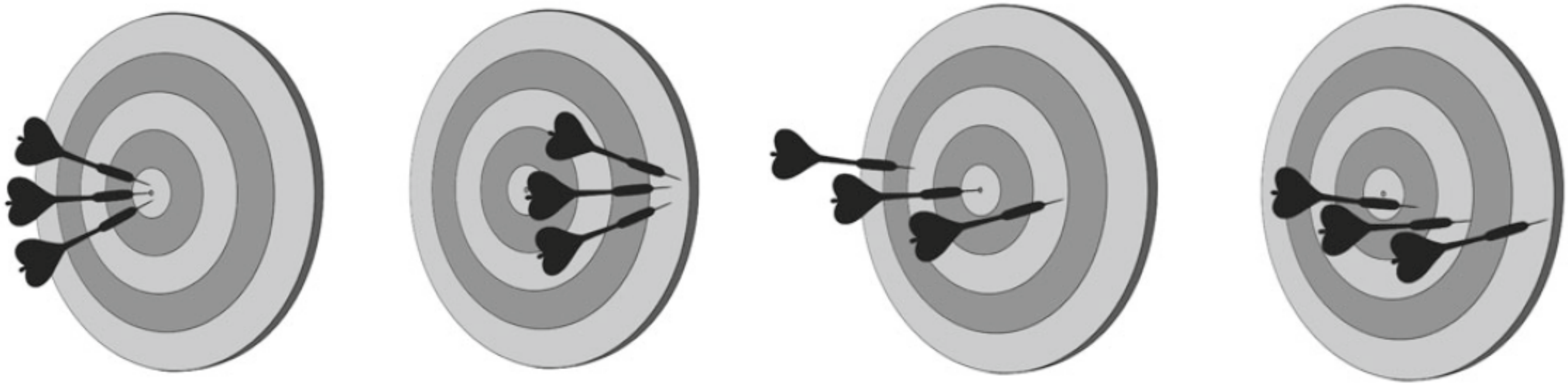
BIAS VS VARIANCE



LOW BIAS
LOW VARIANCE

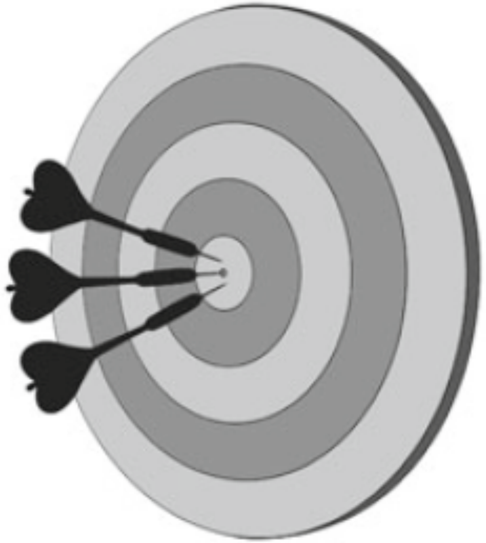
BIAS VS VARIANCE

1 2 3 4

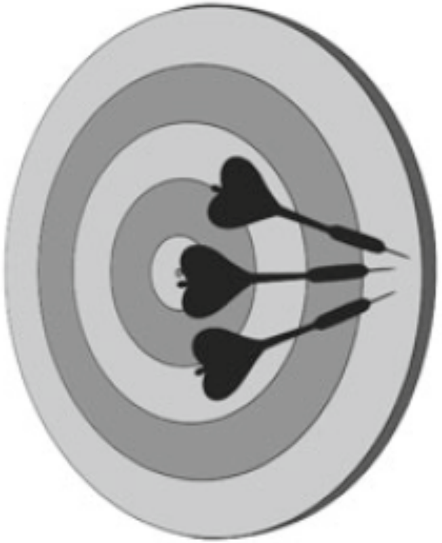


BIAS VS VARIANCE

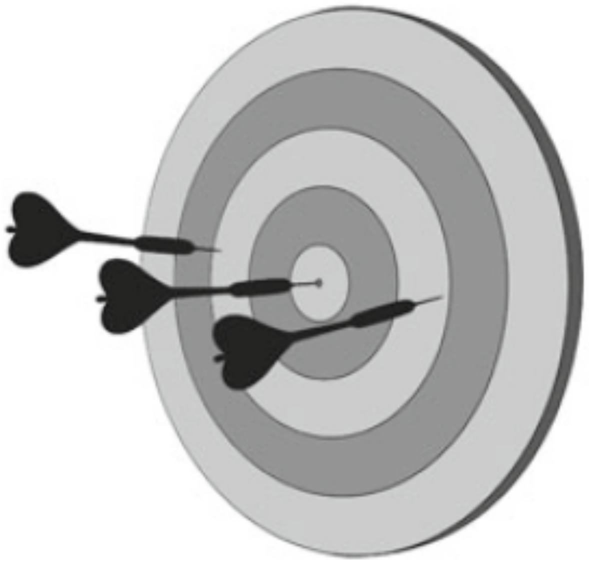
1



2



3

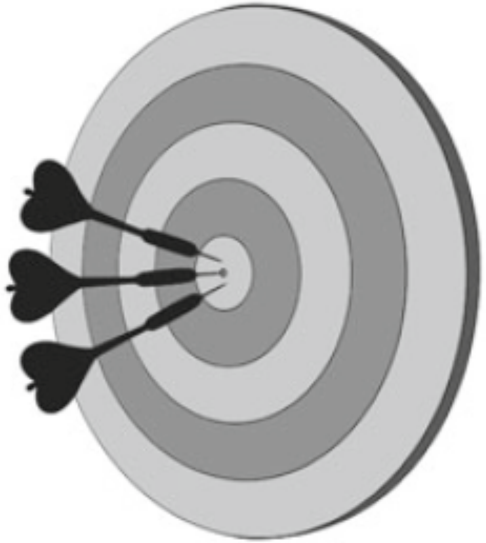


LOW BIAS
HIGH VARIANCE

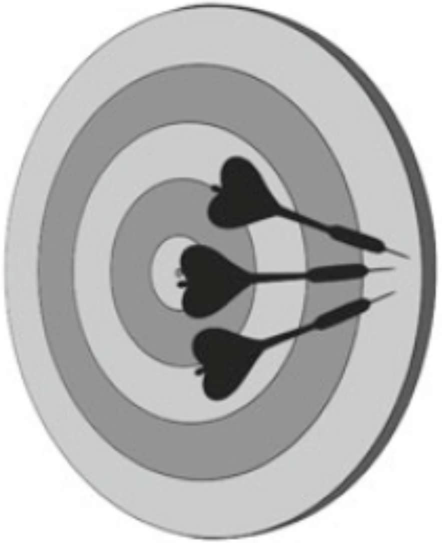
4

BIAS VS VARIANCE

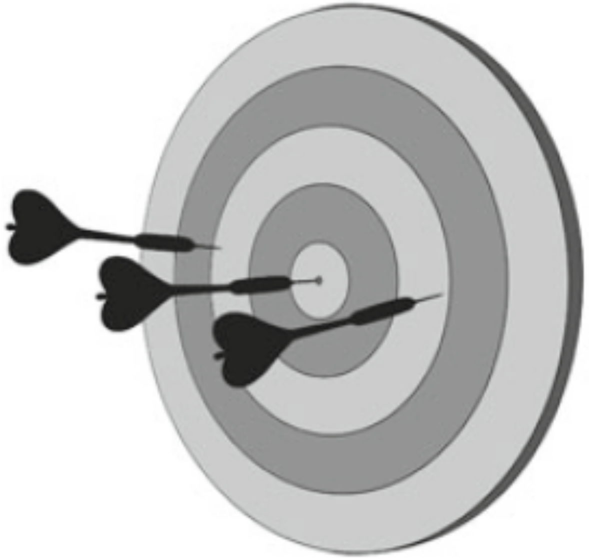
1



2



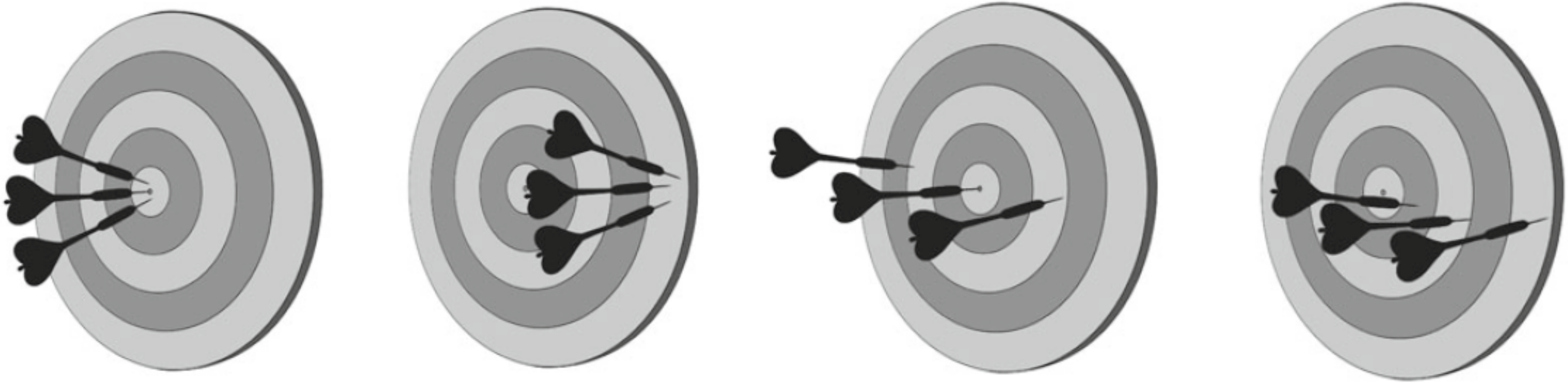
3



4

BIAS VS VARIANCE

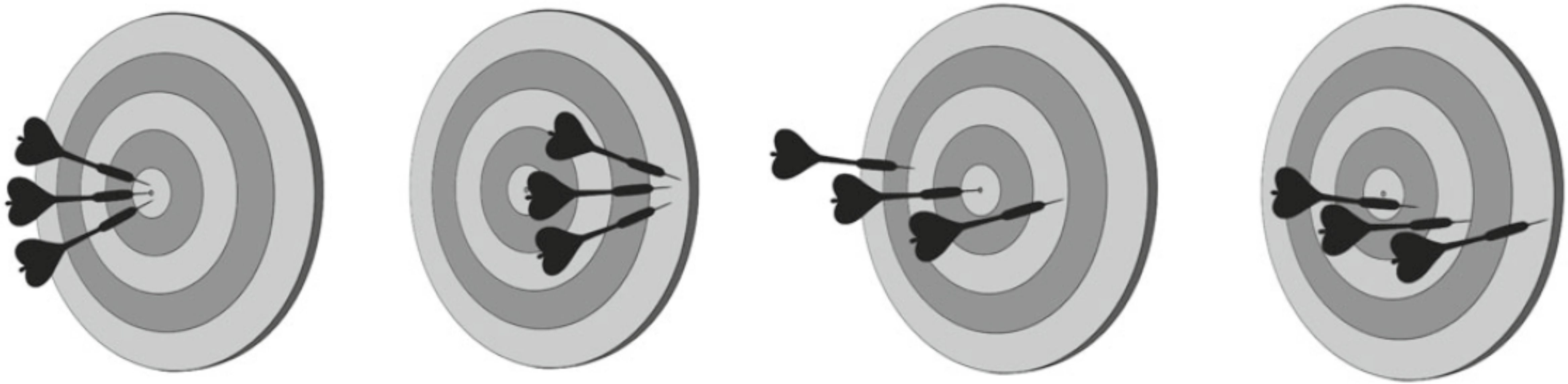
1 2 3 4



HIGH BIAS
LOW VARIANCE

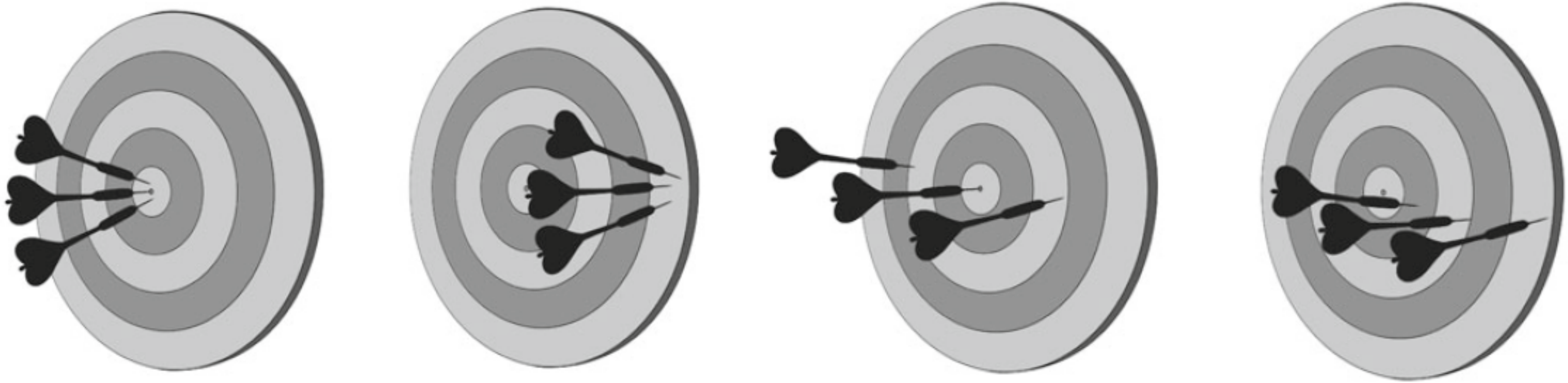
BIAS VS VARIANCE

1 2 3 4



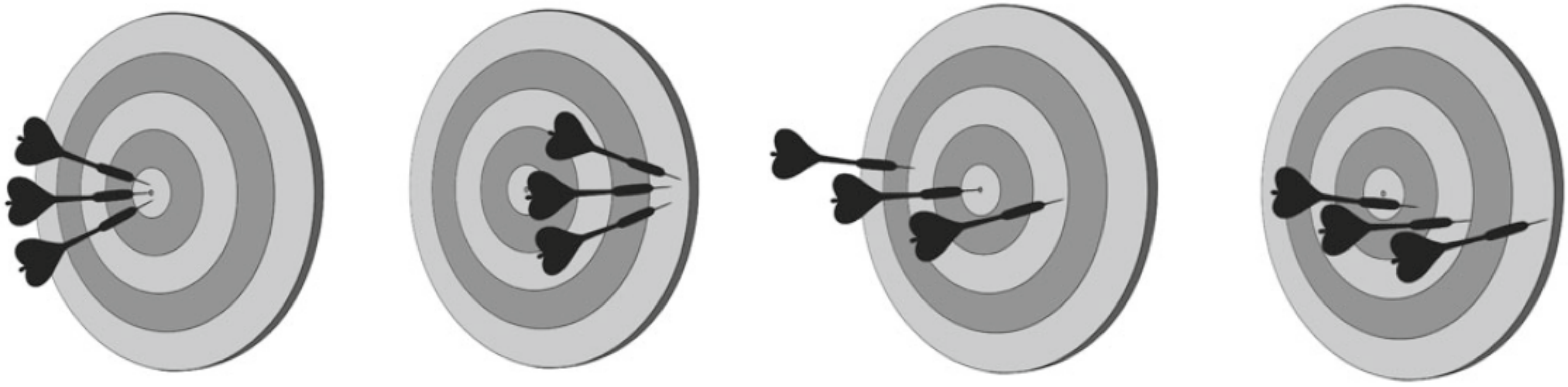
BIAS VS VARIANCE

1 2 3 4



HIGH BIAS
HIGH VARIANCE

BIAS VS VARIANCE



1 — LOW BIAS, LOW VARIANCE

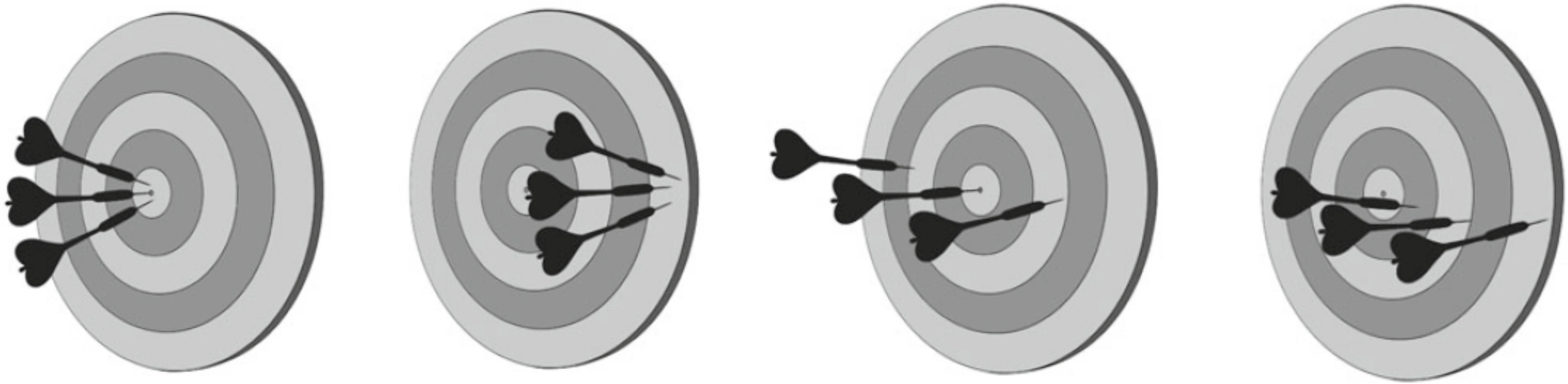
2 — HIGH BIAS, LOW VARIANCE

3 — LOW BIAS, HIGH VARIANCE

4 — HIGH BIAS, HIGH VARIANCE

BIAS VS VARIANCE

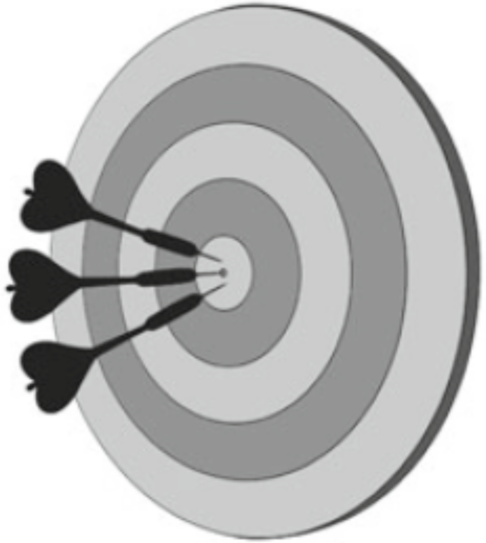
1	2	3	4



LOW BIAS LOW VARIANCE	HIGH BIAS LOW VARIANCE	LOW BIAS HIGH VARIANCE	HIGH BIAS HIGH VARIANCE

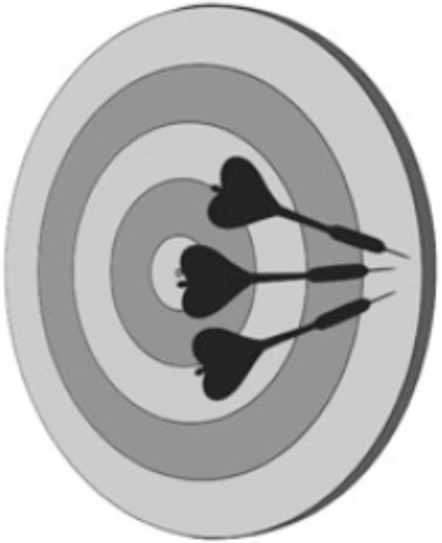
BIAS VS VARIANCE

1



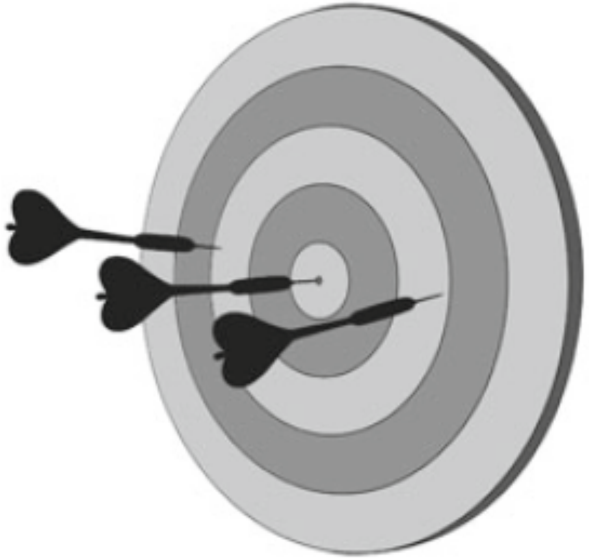
LOW BIAS
LOW VARIANCE

2



HIGH BIAS
LOW VARIANCE

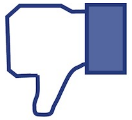
3



LOW BIAS
HIGH VARIANCE

4

HIGH BIAS
HIGH VARIANCE



BIAS

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

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

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

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

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

BIAS: EXAMPLE 3

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

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

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

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

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

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

$$EX_i^2 = ?$$

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

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

BIAS: EXAMPLE 3

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

BIAS: EXAMPLE 3

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

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

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

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

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

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

BIAS: EXAMPLE 3

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

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

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

BIAS: EXAMPLE 3

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

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

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

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

BIAS: EXAMPLE 3

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

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

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

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

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

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

BIAS: EXAMPLE 3

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

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

BIAS: EXAMPLE 3

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

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**?

$S^2_{unbiased} =$

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

- How to make it **unbiased**?

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

VARIANCE

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

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

VARIANCE: EXAMPLE

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

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

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

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

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

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:

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

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

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

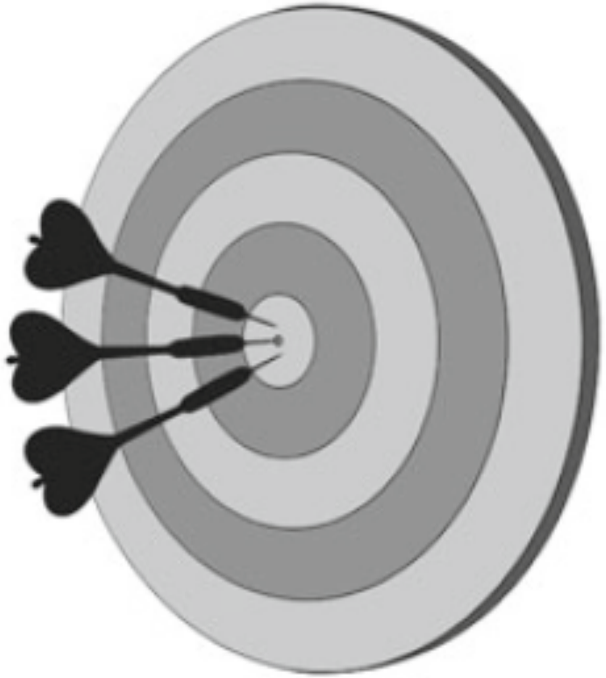
$$but$$

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

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

CONSISTENCY

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

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

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

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

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:

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

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

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 $\epsilon = \frac{c\sigma}{\sqrt{n}}$

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

Chebyshev's inequality:

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

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.