Unbiased Estimators

Dirk Grunwald
University of Colorado, Boulder

Unbiased Estimators

- Estimates
- Estimators
- Bias *vs.* Unbiased
- Estimator of the Mean
- Estimator of Variance

Estimates

Given a dataset $x_1, x_2, \ldots, x_n$ from a set of random variables $X_1, X_2, \ldots, X_n$,

Estimates

Given a dataset $x_1, x_2, \ldots, x_n$ from a set of random variables $X_1, X_2, \ldots, X_n$,

Estimates

An *estimate* is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$; *i.e.* t is some function of the dataset only:

$$t = h(x_1, x_2, \ldots, x_n).$$

Estimates

Given a dataset $x_1, x_2, \ldots, x_n$ from a set of random variables $X_1, X_2, \ldots, X_n$,

Estimates

An *estimate* is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$; *i.e.* t is some function of the dataset only:

$$t = h(x_1, x_2, \ldots, x_n).$$

The estimate t can be a single number (e.g. $E[X]$), a vector or a complex distribution function.

Estimates

Given a dataset $x_1, x_2, \ldots, x_n$ from a set of random variables $X_1, X_2, \ldots, X_n$,

Estimates

An *estimate* is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$; *i.e.* t is some function of the dataset only:

$$t = h(x_1, x_2, \ldots, x_n).$$

The estimate t can be a single number (e.g. $E[X]$), a vector or a complex distribution function.

E.g., mean of the Poisson distribution or predicted wood hardness

Estimator

Let $t = h(x_1, x_2, \ldots, x_n)$ be a estimate based on dataset $x_1, x_2, \ldots, x_n$. Then t is a realization of the random variable

$$T = h(x_1, x_2, \ldots, x_n).$$

and random variable T is called an *estimator*.

Estimator

Let $t = h(x_1, x_2, \ldots, x_n)$ be a estimate based on dataset $x_1, x_2, \ldots, x_n$. Then t is a realization of the random variable

$$T = h(x_1, x_2, \ldots, x_n).$$

and random variable T is called an *estimator*.

E.g., two estimators for the same mean of the Poisson distribution or a linear model relating wood hardness and density.

Unbiased Estimators

A estimator T is called an *unbiased estimator* for the parameter θ if

$$E[T] = \theta$$

irrespective of the value of θ. The difference $E[T] - \theta$ is called the *bias* of T; if this difference is non-zero, then T is *biased*.

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Model *variance* is refers to the amount the model changes with different data

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Model *variance* is refers to the amount the model changes with different data

Mode *bias* refers to the error introduced by approximating a solution using a simpler model

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Model *variance* is refers to the amount the model changes with different data

Mode *bias* refers to the error introduced by approximating a solution using a simpler model

Models are trade-offs between *variance* and *bias*.

Biased Estimators: Example

- We built two estimators, S and T, to estimate μ of Poisson

Biased Estimators: Example

- We built two estimators, S and T, to estimate μ of Poisson
- Estimator S is *unbiased*, and $E[S] = p_0$ for all p_0.

Biased Estimators: Example

- We built two estimators, S and T, to estimate μ of Poisson
- Estimator S is *unbiased*, and $E[S] = p_0$ for all p_0.
- Estimator T is *biased*, and over-predicts p_0.

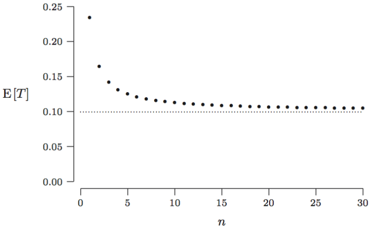

Fig. 19.4. $\mathrm{E}[T]$ as a function of n.

Biased Estimators: Example

- We built two estimators, S and T, to estimate μ of Poisson
- Estimator S is *unbiased*, and $E[S] = p_0$ for all p_0.
- Estimator T is *biased*, and over-predicts p_0.

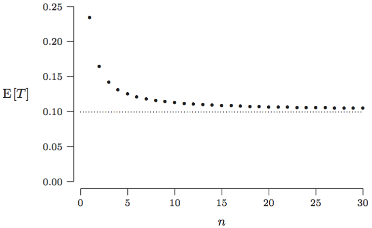

Fig. 19.4. E[T] as a function of n.

The bias occurs because

$$E[T] = E[e^{-\bar{X}_n}] > e^{E[-\bar{X}_n]}$$

due to Jensen's inequality.

Who cares?

- Two common cases:

Who cares?

- Two common cases:
- Unbiased estimate of the mean:

$$E[\bar{X}_n] = \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu$$

Who cares?

- Two common cases:
- Unbiased estimate of the mean:

$$E[\bar{X}_n] = \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu$$

- Unbiased estimate of the variance:

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

Who cares?

- Two common cases:
- Unbiased estimate of the mean:

$$E[\bar{X}_n] = \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu$$

- Unbiased estimate of the variance:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

- Why do we divide by $n-1$ rather than n?

Unbiased estimate for μ

Suppose $X_1 + X_2 + \ldots + X_n$ is a random sample from a distribution with mean μ and variance σ^2.

$$\begin{aligned} E[\bar{X}_n] &= \frac{X_1 + X_2 + \ldots + X_n}{n} \\ &= \frac{1}{n}\left(E[X_1] + E[X_2] + \ldots + E[X_n]\right) \\ &= \frac{1}{n}\left(\mu + \mu + \ldots + \mu\right) = \mu \end{aligned}$$

Unbiased estimate for μ

Suppose $X_1 + X_2 + \ldots + X_n$ is a random sample from a distribution with mean μ and variance σ^2.

$$\begin{aligned} E[\bar{X}_n] &= \frac{X_1 + X_2 + \ldots + X_n}{n} \\ &= \frac{1}{n}\left(E[X_1] + E[X_2] + \ldots + E[X_n]\right) \\ &= \frac{1}{n}\left(\mu + \mu + \ldots + \mu\right) = \mu \end{aligned}$$

This holds for all μ and thus $\bar{X}_n$ is an unbiased estimator for μ.

Variance of Sample *vs* Estimated Variance

N.B.

$$Var[\bar{X}_n] \neq S_n^2$$

$Var[\bar{X}_n]$ is variance of a sample of X_i with mean μ and variance σ^2.

$$\begin{aligned} Var[\bar{X}_n] &= Var[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)] \\ &= \left(\frac{1}{n}\right)^2 Var[X_1 + X_2 + \ldots + X_n] \\ &= \frac{1}{n^2}(\sigma + \ldots + \sigma) = \frac{\sigma^2}{n} \end{aligned}$$

Variance of Sample *vs* Estimated Variance

N.B.

$$Var[\bar{X}_n] \neq S_n^2$$

$Var[\bar{X}_n]$ is variance of a sample of X_i with mean μ and variance σ^2.

$$\begin{aligned}
Var[\bar{X}_n] &= Var[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)] \\
&= \left(\frac{1}{n}\right)^2 Var[X_1 + X_2 + \ldots + X_n] \\
&= \frac{1}{n^2}(\sigma + \ldots + \sigma) = \frac{\sigma^2}{n}
\end{aligned}$$

S_n^2 is an estimator for the σ^2 of X_i.

Estimating S_n^2

We stated:

$$E[S_n^2] = \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \bar{X}_n)^2]$$

Estimating S_n^2

We stated:

$$E[S_n^2] = \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \bar{X}_n)^2]$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} Var[X_i - \bar{X}_n] \quad \text{Magic \#1}$$

Estimating S_n^2

We stated:

$$
\begin{aligned}
E[S_n^2] = \sigma^2 &= \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \bar{X}_n)^2] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} Var[X_i - \bar{X}_n] \qquad \text{Magic \#1} \\
&= \frac{1}{n-1} * n * \left(\frac{n-1}{n}\sigma^2\right) \qquad \text{Magic \#2} \\
&= \sigma^2
\end{aligned}
$$

Magic #1

We need to show that

$$Var[X_i - \bar{X}_n] = E[(X_i - \bar{X}_n)^2]$$

Magic #1

We need to show that

$$Var[X_i - \bar{X}_n] = E[(X_i - \bar{X}_n)^2]$$

Side trip!

Earlier def'n of variance sez:

$$Var[Y] = E[Y^2] - (E[Y])^2$$

Magic #1

We need to show that

$$Var[X_i - \bar{X}_n] = E[(X_i - \bar{X}_n)^2]$$

Side trip!

Earlier def'n of variance sez:

$$Var[Y] = E[Y^2] - (E[Y])^2$$

Let $Y = (X_i - \bar{X}_n)$.

Magic #1

We need to show that

$$Var[X_i - \bar{X}_n] = E[(X_i - \bar{X}_n)^2]$$

Side trip!

Earlier def'n of variance sez:

$$Var[Y] = E[Y^2]$$

Let $Y = (X_i - \bar{X}_n)$.

$$E[Y] = E[(X_i - \bar{X}_n] = E[X_i] - E[\bar{X}_n] = \mu - \mu = 0$$

Magic #1

We need to show that

$$Var[X_i - \bar{X}_n] = E[(X_i - \bar{X}_n)^2]$$

Side trip!

Earlier def'n of variance sez:

$$Var[Y] = E[Y^2]$$

Let $Y = (X_i - \bar{X}_n)$.

$$E[Y] = E[(X_i - \bar{X}_n] = E[X_i] - E[\bar{X}_n] = \mu - \mu = 0$$

Thus,

$$Var[Y] = Var[X_i - \bar{X}_n] = E[(X_i - \bar{X}_n)^2]$$

Magic #2

We want to show $Var[X_i - \bar{X}_n] = \frac{n-1}{n}\sigma^2$.

Magic #2

We want to show $Var[X_i - \bar{X}_n] = \frac{n-1}{n}\sigma^2$.

Side trip!

$$X_i - \bar{X}_n = \frac{n}{n}X_i - \left(\frac{X_i}{n} + \frac{1}{n}\sum_{j \neq i} X_j\right)$$

$$= \frac{n-1}{n}X_i - \frac{1}{n}\sum_{j \neq i} X_j$$

Magic #2

We want to show $Var[X_i - \bar{X}_n] = \frac{n-1}{n}\sigma^2$.

$$Var[X_i - \bar{X}_n] = Var\left[\frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j\right]$$

Magic #2

We want to show $Var[X_i - \bar{X}_n] = \frac{n-1}{n}\sigma^2$.

$$Var[X_i - \bar{X}_n] = Var\left[\frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j\right]$$

$$= \left(\frac{n-1}{n}\right)^2 Var[X_i] + \left(\frac{1}{n}\right)^2 Var[\sum_{j\neq i} X_j]$$

Magic #2

We want to show $Var[X_i - \bar{X}_n] = \frac{n-1}{n}\sigma^2$.

$$
\begin{aligned}
Var[X_i - \bar{X}_n] &= Var\left[\frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j\right] \\
&= \left(\frac{n-1}{n}\right)^2 Var[X_i] + \left(\frac{1}{n}\right)^2 Var[\sum_{j\neq i} X_j] \\
&= \left(\left(\frac{n-1}{n}\right)^2 + \frac{n-1}{n^2}\right)\sigma^2 \\
&= \frac{n^2 - 2n + 1 + n - 1}{n^2}\sigma^2 = \frac{n(n-1)}{n * n}\sigma^2 \quad = \frac{n-1}{n}\sigma^2
\end{aligned}
$$

Back to Estimating S_n^2

Back to where we were..

$$
\begin{aligned}
E[S_n^2] = \sigma^2 &= \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \bar{X}_n)^2] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} Var[X_i - \bar{X}_n] \qquad \text{Magic \#1} \\
&= \frac{1}{n-1} * n * \left(\frac{n-1}{n}\sigma^2\right) \qquad \text{Magic \#2} \\
&= \sigma^2
\end{aligned}
$$

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Model *variance* refers to the amount the model changes with different data

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Model *variance* refers to the amount the model changes with different data

Mode *bias* refers to the error introduced by approximating a solution using a simpler model

Bias and Machine Learning

Machine learning (e.g. a regression model) is trying learn a rule from data

It's *estimating* the parameters from the data

Model *variance* refers to the amount the model changes with different data

Mode *bias* refers to the error introduced by approximating a solution using a simpler model

Models are trade-offs between *variance* and *bias*.