Probability and Mathematical Statistics in Data Science

Lecture 30: Section 11.1: Bias and Variance

Point Estimation

A point estimate of a parameter θ is a single number that can be regarded as a sensible value for θ.

A **point estimate** is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called **the point estimator** of θ, denoted by $\hat{\theta}$

Example

- An automobile manufacturer has developed a new type of bumper, which is supposed to absorb impacts with less damage than previous bumpers. The manufacturer has used this bumper in a sequence of 25 controlled crashes against a wall, each at 10 mph, using one of its compact car models. Let X = the number of crashes that result in no visible damage to the automobile. The parameter to be estimated is p = the proportion of all such crashes that result in no damage [alternatively, $p = P$(no damage in a single crash)]. If X is observed to be x =15, the most reasonable estimator and estimate are:

$$\text{estimator } \hat{p} = \frac{X}{n} \qquad \text{estimate} = \frac{x}{n} = \frac{15}{25} = .60$$

- If for each parameter of interest there were only one reasonable point estimator, there would not be much to point estimation. In most problems, though, there will be more than one reasonable estimator.

Example

- Consider the accompanying 20 observations on dielectric breakdown voltage for pieces of epoxy resin:

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88

a. Estimator $= \overline{X}$, estimate $= \overline{x} = \Sigma x_i/n = 555.86/20 = 27.793$

b. Estimator $= \widetilde{X}$, estimate $= \widetilde{x} = (27.94 + 27.98)/2 = 27.960$

c. Estimator $= [\min(X_i) + \max(X_i)]/2$ = the average of the two extreme lifetimes, estimate $= [\min(x_i) + \max(x_i)]/2 = (24.46 + 30.88)/2 = 27.670$

d. Estimator $= \overline{X}_{\text{tr}(10)}$, the 10% trimmed mean (discard the smallest and largest 10% of the sample and then average),

$$\begin{aligned}\text{estimate} &= \overline{x}_{\text{tr}(10)} \\ &= \frac{555.86 - 24.46 - 25.61 - 29.50 - 30.88}{16} \\ &= 27.838\end{aligned}$$

Measure of a good Estimator

- Our estimator $\hat{\theta}$ is in fact a function of the sample x_i's, therefore, it is also a random variable. For some samples, $\hat{\theta}$ may yield a value larger than θ, whereas for other samples $\hat{\theta}$ may underestimate θ.

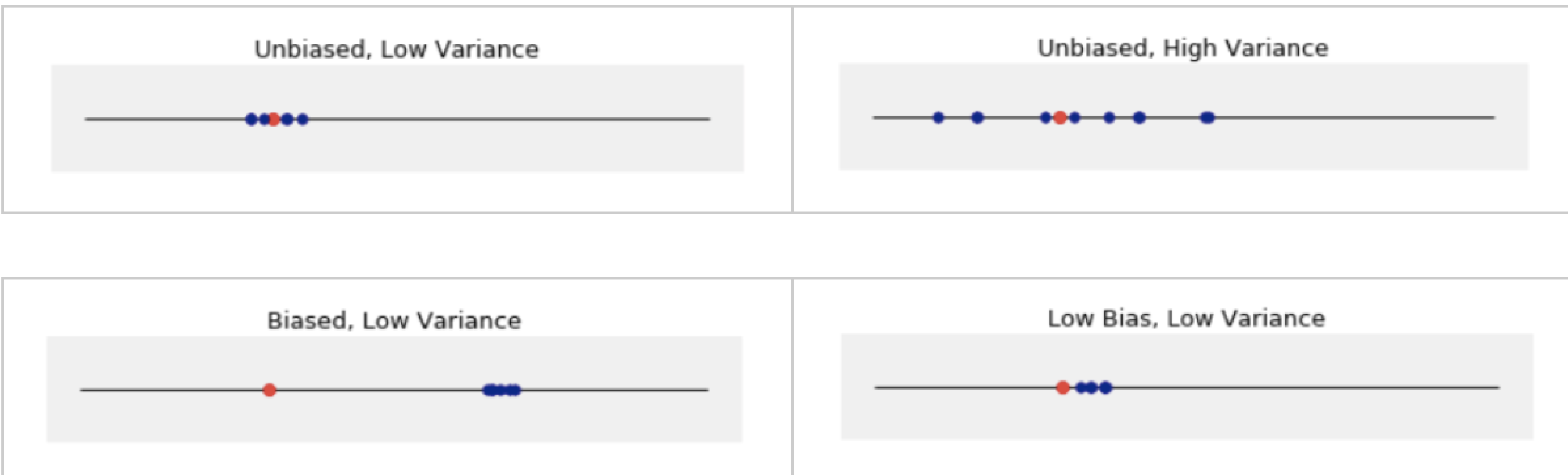
- The quantity $\hat{\theta}$ - θ characterize the error of estimation. A good estimator should result in small estimation errors.

- A commonly used measure of accuracy is the mean square error.

$$\mathrm{MSE} = \mathrm{E}(\hat{\theta} - \theta)^2$$

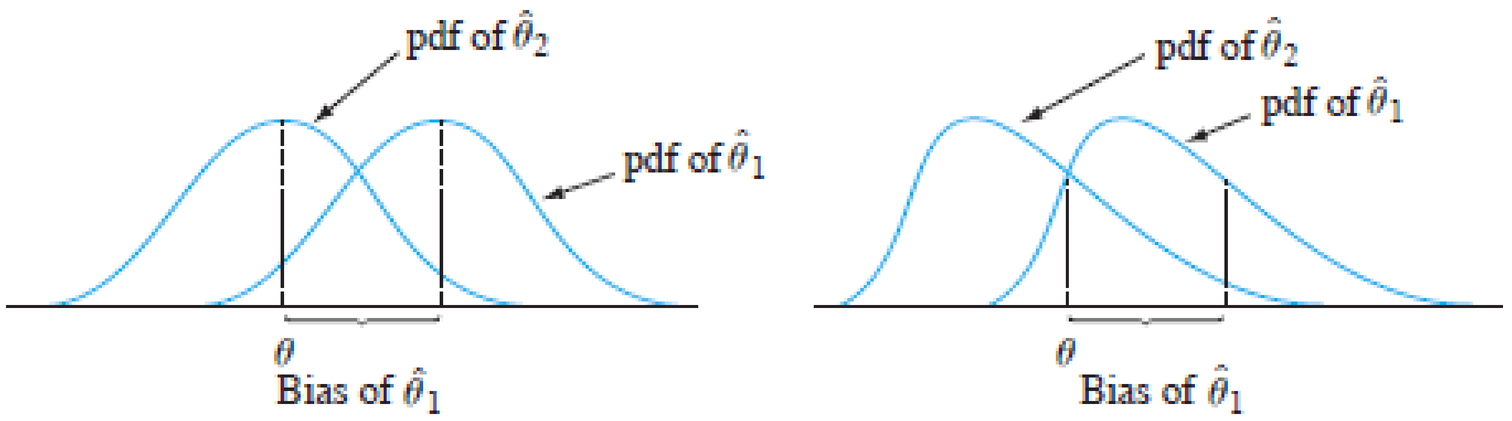
- However, since MSE will generally depend on the value of θ, finding an estimator with smallest MSE is typically NOT possible.

Understanding Bias and Variance



Unbiased Estimators

A point estimator $\hat{\theta}$ is said to be an **unbiased estimator** of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ. If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the **bias** of $\hat{\theta}$.



Unbiased Estimators

- It may seem as though it is necessary to know the value of θ (in which case estimation is unnecessary) to see whether $\hat{\theta}$ is unbiased.

- This is not usually the case, though, because unbiasedness is a general property of the estimator's sampling distribution.

- For example, the sample mean is an unbiased estimate of the population mean.

Example

- In previous example, the sample proportion X/n was used as an estimator of p, where X, the number of sample successes, had a binomial distribution with parameters n and p. Thus

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n}(np) = p$$

When X is a binomial rv with parameters n and p, the sample proportion $\hat{p} = X/n$ is an unbiased estimator of p.

MSE: Bias-Variance Decomposition

The *bias-variance decomposition* says

$$\text{mean squared error} = \text{variance} + \text{bias}^2$$

Principle of Unbiased Estimation

When choosing among several different estimators of θ, select one that is unbiased.

Principle of Minimum Variance Unbiased Estimation

Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ is called the **minimum variance unbiased estimator (MVUE)** of θ.

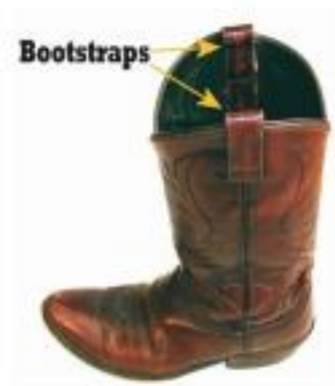
Bootstrapping

- Quantitative research: testing and estimation require information about the population distribution.

- Bootstrap: Estimates the sampling distribution by using the information based on a number of resamples from the single sample you obtained.

Bootstrap

- Wikipedia: Boots may have a tab, loop or handle at the top known as a bootstrap, allowing one to use fingers or a tool to provide greater force in pulling the boots on.



Bootstrap Method

- Use the information of a number of resamples from the sample to estimate the sampling distribution

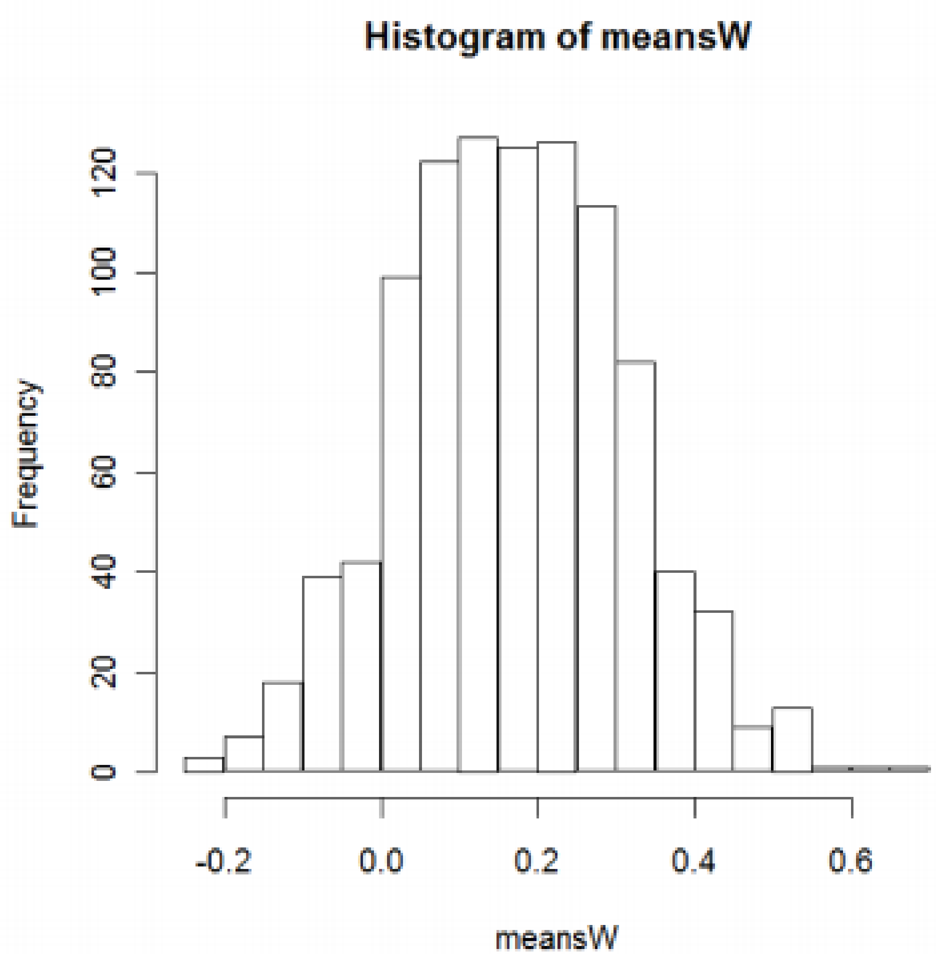
Procedure: Given a sample of size n:

- treat the sample of measurements as if it were the population
- Draw B samples of size n with replacement from your sample (the bootstrap samples)
- For each bootstrap sample, compute the statistic of interest (for example, the sample mean)
- Estimate the sampling distribution by the bootstrap sampling distribution of sample means

Example: Sample from Standard Normal Distribution – Mean=0, SD=1

Case number	Original data	B_sample 1	B_sample 2	B_sample 3
1	1.22	-0.49	1.30	1.40
2	-0.49	-0.93	-1.20	-0.07
3	0.61	0.72	-0.26	-0.55
4	0.72	1.22	-0.08	0.13
5	0.58	-0.49	0.40	1.27
6	-0.08	-0.49	-0.15	-0.08
7	0.07	1.22	-0.55	-0.08
8	-0.26	-0.17	1.22	0.61
9	-0.07	0.40	0.11	-0.15
.			.	.
.			.	.
25	-0.17	1.27	0.52	0.15
26	-0.41	-0.52	1.40	0.72
27	1.40	-1.20	1.30	1.22
28	1.27	-0.15	-0.55	0.07
29	0.26	0.26	0.07	0.40
30	-0.55	-0.17	1.40	0.13
Mean	0.17	0.20	1.11	0.32
Sd	0.81	0.73	0.83	0.85

Bootstrap Result



Example

- Population: standard normal distribution: mean=0 std. dev = 1
- Sample with sample size = 30. 1000 bootstrap samples.
- Results sample (**traditional analysis**)
- sample mean: 0.1698
- 95% Confidence interval: (-0.12,0.46)
- Bootstrap results:
- sample mean: 0.1689;
- 95% Confidence interval : (-0.10, 0.45)

Example

Consider a random sample $X_1, \ldots, X_n$ from the pdf

$$f(x; \theta) = .5(1 + \theta x) \qquad -1 \leq x \leq 1$$

where $-1 \leq \theta \leq 1$ (this distribution arises in particle physics). Show that $\hat{\theta} = 3\overline{X}$ is an unbiased estimator of θ. [*Hint*: First determine $\mu = E(X) = E(\overline{X})$.]