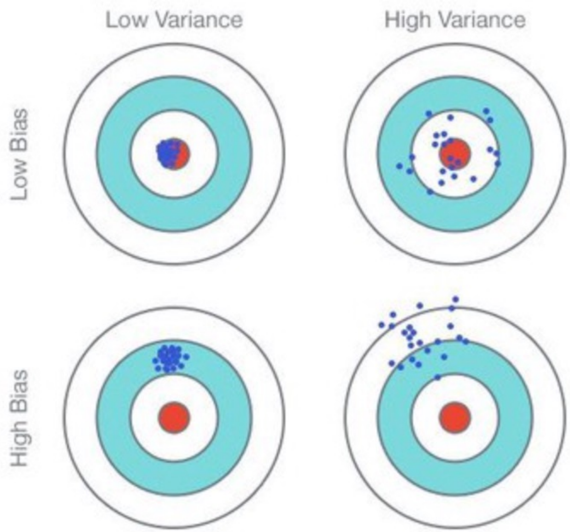
Stat 88: Probability & Mathematical Statistics in Data Science



https://medium.com/@mp32445/understanding-bias-variance-tradeoff-ca59a22e2a83

Fig. 1: Graphical Illustration of bias-variance trade-off , Source: Scott Fortmann-Roe., Understanding Bias-Variance Trade-off

Lecture 36 : 4/21/2021

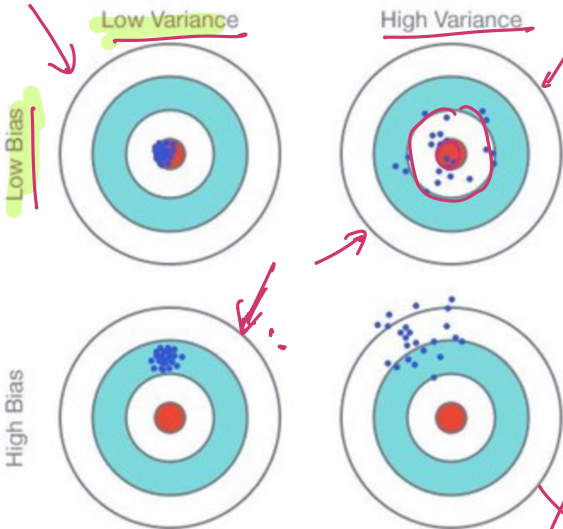
Chapter 11

Bias, Variance, and Least Squares

Understanding Bias and Variance

low variance : precise

accurate



T: estimator (rv)

θ: parameter (target, constant)

Say T is *unbiased* if $E(T) = \theta$

Bias $= E(T) - \theta$

$$\boxed{MSE_\theta(T) = E[(T - \theta)^2]}$$

$E[(T - \theta)^2]$

Bias, Variance, and Mean Squared Error

- Bias: $\boldsymbol{B_\theta(T) = E_\theta(T) - \theta}$ (note that $B_\theta(T)$ is a constant)
- Bias is difference between expected value of the estimator and the target.
- Suppose B_θ is positive, what does this mean?

$B(T) > 0$, $E(T) > \theta$ overestimate on avg.

$B(T) < 0$ $E(T) < \theta$ underestimate on avg.

- Deviation (from the mean): $\boldsymbol{D_\theta(T) = T - E_\theta(T)}$ (note that $D_\theta(T)$ is a r.v.)

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

- Error: $\boldsymbol{T - \theta} =$ $\boxed{T - E(T)} + \boxed{E(T) - \theta} = D(T) + B(T)$
- Mean Squared Error: $\boldsymbol{MSE_\theta(T) = E[(T - \theta)^2]}$

$MSE(T) = E[(T-\theta)^2] = E((D(T) + B(T))^2)$

$E(T) = E_\theta(T)$

- What is the expected value of $D_\theta(T)$? What about $\left(D_\theta(T)\right)^2$?

$E(D(T)) = 0$

$(D(T) = T - E(T)$

$E(D(T)) = E(T - E(T))$

$= E(T) - E(T) = 0$

$E(D(T))^2 = Var(T)$

Mean Squared Error & the Bias-Variance Decomposition

- $MSE_\theta(T) = E[(T-\theta)^2] = \mathbb{E}[(D(T) + B(T))^2]$

$$\mathbb{E}(B(T)) = B(T)$$

$$= \mathbb{E}[(D(T))^2 + 2D(T)\cdot \underbrace{B(T)}_{} + (B(T))^2]$$

$$= \underbrace{\mathbb{E}(D(T))^2}_{} + 2B(T)\cdot \cancelto{0}{\mathbb{E}(D(T))} + (B(T))^2$$

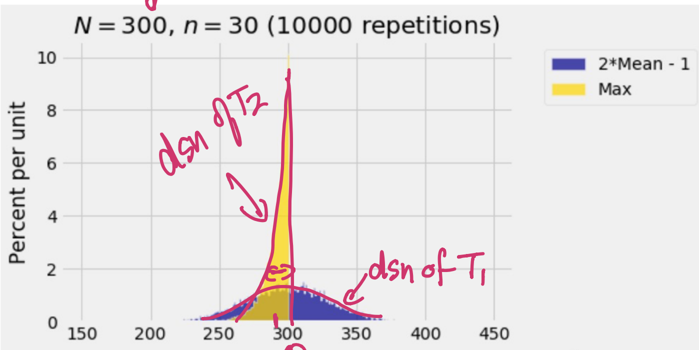
$$= \text{Variance of } T + (\text{Bias}(T))^2$$

MSE = variance + bias^2

Small mean squared error implies low variance AND low bias

German Tank Problem: $T_1, T_2, \& T_3$

For example, say $N=300$



$\theta = N$

$Var(T_1) > Var(T_2)$

$E(T_1) = N$, $E(T_2) \le N$

\# of Tanks manufactured by Germans $= N \leftarrow$ unknown

\#'s observed $X_1, \ldots X_n$ drawn at random from $\{1, 2, \ldots N\}$

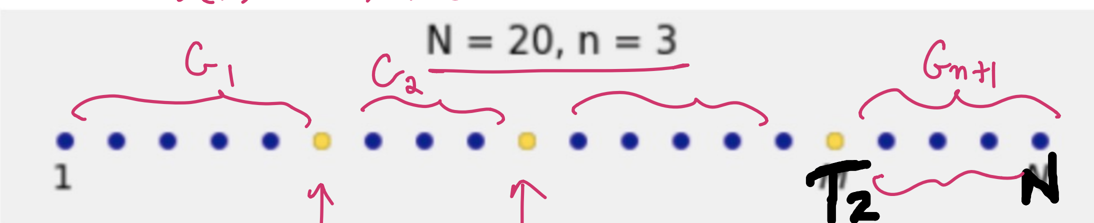
Assume that $X_1, \ldots X_n$ is a SRS from $\{1, 2, \ldots N\}$

By symmetry, X_i's have the same dsn $E(X_i) = \frac{N+1}{2} \Rightarrow E(\bar{X}) = \frac{N+1}{2}$

Let $T_1 = 2\bar{X} - 1$, then $E(T_1) = N$

$T_2 = \max\{X_1, \ldots X_n\}$

$T_2 \le 300 \Rightarrow E(T_2) \le 300$



$\text{Bias}(T_1) = 0$

$\text{Bias}(T_2) = E(T_2) - N$

sampled values

$T_2 = \max\{X_1, X_2, X_3\}$

Gold dots are sampled values

Blue dots are not sampled.

Total length $= N$ (total # of dots)

3 gold dots, 17 blue dots, avg length of gap $= \frac{17}{4} \leftarrow$

In general, expected gap length $= \frac{N-n}{n+1}$

Comparing $MSE(T_2)$ & $MSE(T_3)$

$= \frac{\text{total \# of blue dots}}{\text{total \# of "gaps"}}$

$N = T_2 + G_{n+1}$ $\qquad E(G_{n+1}) = \frac{N-n}{n+1}$

$B(T_2) = E(T_2) - N = -E(G_{n+1})$

$N = E(T_2) + E(G_{n+1}) \Rightarrow B(T_2) = E(T_2) - N = -E(G_{n+1})$

$\text{Bias}(T_2) = -\frac{(N-n)}{n+1}$

Further let's try to figure out an expression for # of dots up to T_2

n total gold dots

n "gaps" of blue dots of expected length $\frac{N-n}{n+1}$

$$E(T_2) = n\left(\frac{N-n}{n+1}\right) + n$$

$$\mathbb{E}(T_2) = \frac{nN - n^2 + n(n+1)}{n+1} = \frac{nN - \cancel{n^2} + \cancel{n^2} + n}{n+1}$$

$$\mathbb{E}(T_2) = \frac{n(N+1)}{n+1}$$

$$N = \frac{n+1}{n}\mathbb{E}(T_2) - 1$$

Define T_3 (based on T_2) by

$$T_3 = \left(\frac{n+1}{n}\right) T_2 - 1 \qquad \left(\text{so that } \mathbb{E}(T_3) = \left(\frac{n+1}{n}\right)\mathbb{E}(T_2) - 1 = N\right)$$

$$\mathbb{E}(T_3) = N$$

$$SD(T_3) = \left(\frac{n+1}{n}\right) SD(T_2)$$

so if n is very large, then

$$SD(T_3) \approx SD(T_2) \qquad \left(\frac{n+1}{n} \approx 1\right)$$

$Var(T_3) \approx Var(T_2)$ for reasonably large n

$Bias(T_3) = 0$, $Bias(T_2) = \dfrac{-(N-n)}{n+1}$

$$MSE(T_3) = Var(T_3) + bias^2 = Var(T_3) + 0$$

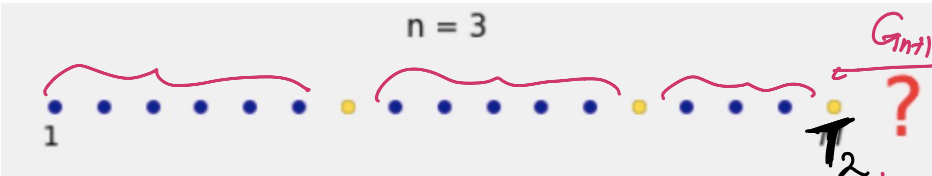
$$MSE(T_2) = Var(T_2) + bias^2 = Var(T_2) + \frac{(N-n)^2}{(n+1)^2}$$

T_3 is unbiased & has smaller MSE,
so more accurate.

T_3 is called the "Augmented Maximum"

$$T_3 = \left(\frac{n+1}{n}\right) T_2 - 1 \qquad \left(\frac{n+1}{n} > 1\right)$$

The Augmented Maximum



G_{n+1}

T_2 Max. observed value

N

We see only T_2
we have to guess
how many dots
come after T_2 = max of
observed sample.

Estimate the next gap length
G_{n+1} by the avg. gap length up to T_2

Total # of blue dots $= \dfrac{T_2 - n}{n}$

# of gold dots (sample size)

# of gaps up to T_2

We can estimate N as

$$N = T_2 + \text{avg. gap length} = T_2 + \frac{T_2 - n}{n} = \overbrace{T_2\left(\frac{n+1}{n}\right) - 1}^{T_3}$$

$\boxed{17}$ $\boxed{5}$ $\boxed{2}$ $\boxed{11}$

1 $\boxed{2}$ 3 4 $\boxed{5}$ 6 7 8 9 10 $\boxed{11}$ --- $\boxed{17}$