Statistics

Setup

Suppose $Y_1, Y_2, \ldots, Y_n$ are random variables with CDF F_θ, for some $\theta \in \Theta$, where Θ denotes the set of legitimate parameter values, called *parameter space*.

We would like to estimate the unknown parameter θ from the data $y_1, y_2, \ldots, y_n$.

e.g. $Y_1, Y_2, \ldots, Y_n \sim N(\mu, \sigma^2)$

parameters: $\theta = \{\mu, \sigma^2\}$

parameter space: $\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

Definition 1.1

A function $T(Y_1, Y_2, \ldots, Y_n)$ is called a *statistic*.

A statistic T that takes values in Θ is called an *estimator* for θ.

e.g. $T = \sum_{i=1}^{n} Y_i$ is a statistic.

$T = \min(Y_1, Y_2, \ldots, Y_n)$

$T = \max(Y_1, Y_2, \ldots, Y_n)$ are also statistics.

e.g. $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ is an estimator for μ.

Y_1 is also an estimator for μ.

Bias

Definition 1.2

Let T be an estimator for θ, the *bias* of T is defined by

$$b_T(\theta) = E[T] - \theta.$$

If $b_T(\theta) = 0$ for all θ, then T is said to be an *unbiased estimator* for θ.

Let T be an estimator for θ, the *bias* of T is defined by
$$b_T(\theta) = E[T] - \theta.$$
If $b_T(\theta) = 0$ for all θ, then T is said to be an *unbiased estimator* for θ.

Example 1.1

Suppose$Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed (IID) $N(\mu, \sigma^2)$ random variables and let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

be an estimator for μ. Calculate $b_{\bar{Y}}(\mu)$.

Example 1.1 Solutions

$$b_{\bar{Y}}(\mu) = E(\bar{Y}) - \mu$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] - \mu$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[Y_i] - \mu$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu - \mu$$

$$= \frac{1}{n}(n\mu) - \mu$$

$$= \mu - \mu$$

$$= 0$$

So $\bar{Y}$ is an unbiased estimator of μ.

Example 1.2

Suppose $Y_1, Y_2, \ldots, Y_n$ are IID Bernoulli random variables with probability of success θ and let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

be an estimator for θ. Calculate $b_{\bar{Y}}(\theta)$.

Example 1.2 Solutions

$$b_{\bar{Y}}(\theta) = E(\bar{Y}) - \theta$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(Y_i) - \theta$$

$$= \frac{1}{n}\sum_{i=1}^{n} \theta - \theta$$

$$= \frac{1}{n}(n\theta) - \theta$$

$$= \theta - \theta$$

$$= 0$$

So $\bar{Y}$ is an unbiased estimator of θ.

MSE

Definition 1.3

Let T be an estimator for θ, the *mean squared error* (MSE) of T is defined by

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

The smaller the MSE, the better the estimator.

Example 1.3

Suppose$Y_1, Y_2, \dots, Y_n$ are IID $N(\mu, \sigma^2)$ random variables and let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

be an estimator for μ.

Calculate $MSE_{\bar{Y}}(\mu)$.

Which estimator has a lower MSE, Y_1 or $\bar{Y}$?

Example 1.3 Solutions

$$
\begin{aligned}
MSE_{\bar{Y}}(\mu) &= E[(\bar{Y}-\mu)^2] \\
&= E[(\bar{Y}-E(\bar{Y}))^2] \\
&= var(\bar{Y}) \\
&= var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} var(Y_i) \quad (\text{by independence of } Y_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{1}{n^2}(n\sigma^2) \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

$$
\begin{aligned}
MSE_{Y_1}(\mu) &= E[(Y_1-\mu)^2] \\
&= E[(Y_1-E(Y_1))^2] \\
&= var(Y_1) \\
&= \sigma^2
\end{aligned}
$$

$\therefore$ $\bar{Y}$ has a lower MSE than Y_1 if $n > 1$.

Example 1.4

Suppose $Y_1, Y_2, \dots, Y_n$ are IID Bernoulli random variables with probability of success θ and let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

be an estimator for θ.

Prove that $MSE_{\bar{Y}}(\theta) = \frac{\theta(1-\theta)}{n}$.

Example 1.4 Solutions

$Y_i \sim Ber(\theta)$

$var(Y_i) = \theta(1-\theta)$

$$MSE_{\bar{Y}}(\theta) = E[(\bar{Y}-\theta)^2]$$

$$= var(\bar{Y})$$

$$= var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} var(Y_i) \quad (\text{by independence of } Y_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \theta(1-\theta)$$

$$= \frac{1}{n^2} n\theta(1-\theta)$$

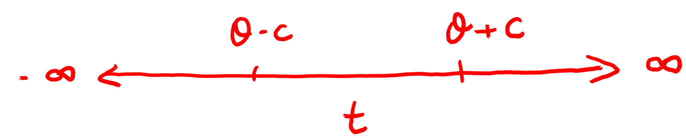
$$= \frac{\theta(1-\theta)}{n}$$

Theorem 1

For any $k > 0$,

$$P\left(|T-\theta| \geq k\sqrt{MSE}\right) \leq \frac{1}{k^2}.$$

Proof of Theorem 1



For any $c > 0$,

$$
\begin{aligned}
MSE_T(\theta) &= E\left[(T-\theta)^2\right] \\
&= \int_{-\infty}^{\infty} (t-\theta)^2 f(t)\,dt \\
&= \int_{-\infty}^{\theta-c} (t-\theta)^2 f(t)\,dt + \int_{\theta-c}^{\theta+c} (t-\theta)^2 f(t)\,dt + \int_{\theta+c}^{\infty} (t-\theta)^2 f(t)\,dt \\
&\geq \int_{-\infty}^{\theta-c} \boxed{(t-\theta)^2} f(t)\,dt + \int_{\theta+c}^{\infty} \boxed{(t-\theta)^2} f(t)\,dt
\end{aligned}
$$

in this domain, $t \leq \theta - c$
$t - \theta \leq -c$
$(t-\theta)^2 \geq c^2$

in this domain, $t \geq \theta + c$
$t - \theta \geq c$
$(t-\theta)^2 \geq c^2$

$$
\begin{aligned}
&\geq c^2 \int_{-\infty}^{\theta-c} f(t)\,dt + c^2 \int_{\theta+c}^{\infty} f(t)\,dt \\
&= c^2 \left[\int_{-\infty}^{\theta-c} f(t)\,dt + \int_{\theta+c}^{\infty} f(t)\,dt \right] \\
&= c^2 \left[P(T \leq \theta - c) + P(T \geq \theta + c) \right]
\end{aligned}
$$

Proof of Theorem 1 (cont.)

$$c^2 \left[P(T \le \theta - c) + P(T \ge \theta + c) \right]$$

$$= c^2 \left[P(T - \theta \le -c) + P(T - \theta \ge c) \right]$$

$$= c^2 \left[P(-(T - \theta) \ge c) + P(T - \theta \ge c) \right]$$

$$= c^2 P(|T - \theta| \ge c)$$

$$MSE_T(\theta) \ge c^2 P(|T - \theta| \ge c)$$

$$\frac{MSE_T(\theta)}{c^2} \ge P(|T - \theta| \ge c)$$

$$\frac{\cancel{MSE_T(\theta)}}{k^2 \cancel{MSE_T(\theta)}} \ge P\left(|T - \theta| \ge k\sqrt{MSE_T(\theta)}\right) \qquad \left(\text{take } c = k\sqrt{MSE_T(\theta)}\right)$$

$$\frac{1}{k^2} \ge P\left(|T - \theta| \ge k\sqrt{MSE_T(\theta)}\right)$$

Theorem 2

$$MSE_T(\theta) = \mathrm{var}(T) + b_T(\theta)^2.$$

If T is unbiased, then $MSE_T(\theta) = \mathrm{var}(T)$.

Proof of Theorem 2

variance formula:

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

$$E(X^2) = \mathrm{var}(X) + E(X)^2$$

$$\begin{aligned}
MSE_T(\theta) &= E[\underbrace{(T-\theta)}_{X}{}^2] \\
&= \mathrm{var}(T-\theta) + E(T-\theta)^2 \\
&= \mathrm{var}(T) + [E(T)-\theta]^2 \\
&= \mathrm{var}(T) + b_T(\theta)^2
\end{aligned}$$

Proof of Theorem 2

Alternative proof

$$
\begin{aligned}
& MSE_T(\theta) \\
&= E[(T-\theta)^2] \\
&= E[(\underbrace{\boxed{T-E(T)}}_{a} + \underbrace{\boxed{E(T)-\theta}}_{b})^2] \\
&= E\left[\overbrace{\left(T-E(T)\right)^2}^{a^2}\right] + \overbrace{2\,E\left[\left(T-E(T)\right)(E(T)-\theta)\right]}^{2ab} + \overbrace{E[E(T)-\theta]^2}^{b^2} \\
&= \text{var}(T) + 2\underline{E[T-E(T)]}[E(T)-\theta] + E[b_T(\theta)]^2
\end{aligned}
$$

$$E[T-E(T)] = E[T] - E\big[E[T]\big] = E[T] - E[T] = 0$$

$$= \text{var}(T) + b_T(\theta)^2$$