MH1820 Week 10

Bias and Standard Error of an Estimator

Bias

Let $\widehat{\theta}$ be an estimator of θ. The **bias** of $\widehat{\theta}$ is defined by

$$\mathrm{Bias}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}] - \theta.$$

Here, the expectation is computed under the population distribution parametrized by θ.

Intepretation: $\mathrm{Bias}(\widehat{\theta})$ is the expected distance of $\widehat{\theta}$ from the true parameter θ.

A good estimator must have bias zero or at least its bias should tend to zero for increasing sample size.

$\widehat{\theta}$ is **unbiased** if $\mathrm{Bias}(\widehat{\theta}) = 0$.

Example 1

- Population distribution: *Bernoulli*(p)
- Estimator: $\widehat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (for p)

Find the bias of $\widehat{p}$.

Solution.

$$\begin{aligned}
\mathrm{Bias}(\widehat{p}) &= \mathbb{E}[\widehat{p}] - p \ = \mathbb{E}[\overline{X}] - p \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - p \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] - p \\
&= \frac{1}{n}\sum_{i=1}^{n} p - p = 0.
\end{aligned}$$

So $\widehat{p} = \overline{X}$ is unbiased. $\square$

Standard Error

Standard error of $\widehat{\theta}$:

$$SE(\widehat{\theta}) = \sqrt{\mathrm{Var}[\widehat{\theta}]}.$$

Here the variance is computed under the population distribution parametrized by θ.

- $SE(\widehat{\theta})$ measures variability of our estimate, i.e. standard deviation of sampling distribution.
- Rule of thumb: For large samples, the true θ will be in the interval
  - $[\widehat{\theta} - SE(\widehat{\theta}), \widehat{\theta} + SE(\widehat{\theta})]$ in around 70% of the cases
  - $[\widehat{\theta} - 2SE(\widehat{\theta}), \widehat{\theta} + 2SE(\widehat{\theta})]$ in around 95% of the cases,

  if $\widehat{\theta}$ is used repeatedly to estimate

Example 2

Let $X_1, \ldots, X_n$ be i.i.d with population distribution $N(0, \sigma^2)$.

Estimator for σ^2: $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

Find $SE(\widehat{\sigma}^2)$.

Solution. Note that

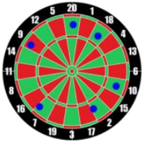
$$\frac{X_i}{\sigma} \sim N(0,1) \Longrightarrow \left(\frac{X_i}{\sigma}\right)^2 \sim \chi^2(1) = Gamma(\frac{1}{2}, 2).$$

Recall that if $X \sim Gamma(\alpha, \theta)$, then $\mathrm{Var}[X] = \alpha\theta^2$.

$$\begin{aligned}
\Longrightarrow \quad & \mathrm{Var}\left[\left(\frac{X_i}{\sigma}\right)^2\right] = \frac{1}{2} \cdot 2^2 = 2 \Longrightarrow \mathrm{Var}[X_i^2] = 2\sigma^4 \\
\Longrightarrow \quad & \mathrm{Var}[\widehat{\sigma}^2] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}[X_i^2] = \frac{1}{n^2} n 2\sigma^4 = \frac{2\sigma^4}{n} \\
\Longrightarrow \quad & SE(\widehat{\sigma}^2) = \sqrt{\frac{2\sigma^4}{n}} = \sigma^2 \sqrt{\frac{2}{n}}.
\end{aligned}$$

□

- Point estimator on average should be close to the true parameter $\Longrightarrow$ bias must be small.
- The values of the estimator should not be spread out too far $\Longrightarrow$ standard error must be small.
- Ideal situation: both *Bias* and *SE* as small as possible.



Maximum Likelihood Estimator

Suppose we observed the following data (sample size 10) drawn from *Bernoulli*(p):

$$x_1 = x_2 = x_3 = 1, \; x_4 = \cdots = x_{10} = 0.$$

- It seems that $p = 0.3$ is quite likely. But we could not decide since the data could have come from $p = 0.5$ or $p = 0.9$, though $p = 0.3$ seems much more plausible.

- Is there a way to pick the "most probable" p?

- Problem: There is no "most probable" value of p (since p is not a random variable!)

Main idea of maximum likelihood:
Reverse the approach: Find a value of p that makes the observations most likely, i.e. maximizing the probability of observing the data!

Set-up for maximum likelihood:

- Let $X_1, \ldots, X_n$ be i.d. with PMF or PDF $f(x|\theta)$, depending on an unknown parameter θ.
- Observations $x_1, \ldots, x_n$ are given.
- The idea of the maximum likelihood method is to choose the value for θ as estimator which maximizes the following **maximum likelihood function**:

$$L(X_1 = x_1, \ldots, X_n = x_n|\theta) \quad = \quad \prod_{i=1}^{n} f(x_i|\theta)$$

Interpretation. The maximum likelihood function $L(x_1, \ldots, x_n|\theta)$ is the probability of observing the data assuming θ is the real value. E.g. in the discrete case, we have

$$\begin{aligned}
\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n|\theta) &= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i|\theta) \quad \text{(since } X_i \text{ are independent)} \\
&= \prod_{i=1}^{n} f(x_i|\theta) \quad \text{(since } X_i\text{'s are identical)} \\
&= L(x_1, \ldots, x_n|\theta).
\end{aligned}$$

Finding Maximum Likelihood Estimator (MLE):

The maximum likelihood estimator i.e. the value of θ that maximizes $L(x_1, \ldots, x_n|\theta)$ can be found by solving

- $\dfrac{d}{dp}(L) = 0$ or
- $\dfrac{d}{dp}(\ln L) = 0$

Both solution methods are valid, but sometimes the second method often is faster. There are likelihood functions for which the maximizer cannot be found in this way, but such cases will not occur in this course.

Example 3

- $X_1, \ldots, X_{10}$ i.i.d. $\sim$ *Bernoulli*(p), $0 < p < 1$.
- Observations: $x_1 = x_2 = x_3 = 1$, $x_4 = \cdots = x_{10} = 0$

Find p that maximizes the likelihood function.

Solution. Recall that the PMF for *Bernoulli*(p) is $f(x|p) = p^x(1-p)^{1-x}$, $x = 0, 1$. The likelihood function is

$$L = L(x_1, \ldots, x_{10}|p) = \prod_{i=1}^{10} f(x_i|p) = f(1|p)^3 f(0|p)^7 = p^3(1-p)^7.$$

To find the maximum of L, we set the derivative (with respect to p) to zero:

$$\begin{aligned} \frac{dL}{dp} &= 0 \\ 3p^2(1-p)^7 - 7p^3(1-p)^6 &= 0 \\ p^2(1-p)^6(3(1-p)-7p) &= 0 \end{aligned}$$

This implies that $p = 0$, or $p = 1$, or $p = \frac{3}{10}$.

Thus, $p = \frac{3}{10}$ is the maximizer, i.e. the maximum likelihood estimator for p is $\frac{3}{10}$. $\square$

Example 4

- $X_1, \ldots, X_n$ i.i.d $\sim Geom(p)$, $0 < p < 1$.
- PMF: $f(x) = (1-p)^{x-1}p$, $x = 1, 2, \ldots$.
- Given the observation $x_1 = 2$ $(n = 1)$, what is the *MLE* for p?

Solution. We have $f(x|p) = (1-p)^{x-1}p$. Since $n = 1$, the maximum likelihood function is

$$L(x|p) = f(x|p) = (1-p)^{x-1}p \implies \ln L(x|p) = (x-1)\ln(1-p) + \ln(p).$$

Solving $\frac{d}{dp}\ln L = 0$, we have

$$\frac{1-x}{1-p} + \frac{1}{p} = 0 \implies p = \frac{1}{x}.$$

Substituting $x = 2$, the MLE for p is $p = \frac{1}{2}$. $\square$

Example 5

- $X_1, \ldots, X_n$ i.i.d $\sim Exp(\theta)$, $\theta > 0$.
- PDF: $f(x|\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x > 0$ and $f(x|\theta) = 0$ otherwise.
- Find MLE for θ based on the observations 1, 2, 5, 1, 1 ($n = 5$).

Solution. The maximum likelihood function is

$$L(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = \frac{1}{\theta^n} e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}.$$

From the obeservations, we have $n = 5$ and $\sum_{i=1}^{5} x_i = 1 + 2 + 5 + 1 + 1 = 10$, and so

$$L = \theta^{-5} e^{-10/\theta} \implies \ln L = -5 \ln \theta - 10/\theta.$$

Differentiating $\ln L$ with respect to θ and set it to 0, we get

$$\begin{aligned}
\frac{d}{d\theta}(\ln L) &= 0 \\
-\frac{5}{\theta} + \frac{10}{\theta^2} &= 0 \\
\theta &= 2.
\end{aligned}$$

The MLE for θ is 2. $\square$

Interval Estimator

An estimator $\widehat{\theta}$ only provides a single "best guess" for θ ("point estimator"), based on a random sample.

Bias and standard error measure average precision of $\widehat{\theta}$. Both types of information, the best guess and average precision, can be combined into a **confidence interval** ("**interval estimation**").

In almost all applications of parameter estimation, confidence intervals are used (point estimation is not enough).

More precisely, we want to find values $\widehat{\theta}_L$ and $\widehat{\theta}_U$ such that

$$\mathbb{P}(\widehat{\theta}_L < \theta < \widehat{\theta}_U) = 1 - \alpha, \quad \text{where } 0 < \alpha < 1.$$

- The interval $[\widehat{\theta}_L, \ \widehat{\theta}_U]$ computed from the selected sample is called a **confidence interval** for θ.
- The fraction $1 - \alpha$ is called the **confidence level**. Some common values of α are 0.01, 0.05, 0.025.
- The endpoints $\widehat{\theta}_L$, $\widehat{\theta}_U$ are called the **lower and upper confidence limits**.

Strategy to Construct Confidence Intervals

- $X_1, \ldots, X_n$ i.i.d with some distribution depending on an unknown parameter θ.
- Goal: Find $100(1-\alpha)\%$ confidence interval for θ.
- Idea: Find a statistic Q involving $X_1, \ldots, X_n$ and θ such that the distribution of Q is known.
- Find a, b such that $\mathbb{P}(a < Q < b) = 1 - \alpha$.
- Transform $a < Q < b$ to an equivalent condition $\theta_L < \theta < \theta_U$. Then $[\theta_L, \theta_U]$ is the required confidence interval.

Summary: $100(1-\alpha)\%$ Confidence Interval or μ of $N(\mu,\sigma^2)$:

Case	Statistic Q	Dist. of Q	CI
Known σ^2	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$N(0,1)$	$\left[\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
Unknown σ^2 ($n \geq 30$)	$\frac{\overline{X}-\mu}{S/\sqrt{n}}$	$\approx N(0,1)$	$\left[\overline{X}-z_{\alpha/2}\frac{S}{\sqrt{n}},\ \overline{X}+z_{\alpha/2}\frac{S}{\sqrt{n}}\right]$
Unknown σ^2 ($n < 30$)	$\frac{\overline{X}-\mu}{S/\sqrt{n}}$	$t-distribution$ $degree\ freedom\ n-1$	$\left[\overline{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}},\ \overline{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right]$

Theorem 6 (Confidence Interval for μ of normal distribution with known σ^2)

Let $X_1, \ldots, X_n$ i.i.d $\sim N(\mu, \sigma^2)$. The $100(1-\alpha)\%$ confidence interval for μ is given by

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Recall z_α is the upper $100(1-\alpha)\%$ point of the standard normal, i.e. $\mathbb{P}(\phi > z_\alpha) = \alpha$.

Justification:

- Select statistic $Q = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ the standardized sample mean. We know that

$$Q \sim N(0,1).$$

- Choose $a = -z_{\alpha/2}$ and $b = z_{\alpha/2}$. Then

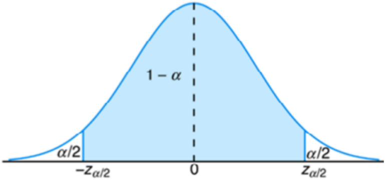
$$\mathbb{P}\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha.$$

- Rearranging the terms, the interval $-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}$ is equivalent to

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Some commonly used z_α:

α	z_α
0.05	1.645
0.025	1.96
0.005	2.575



Example 7

The average zinc concentration recovered from a sample of measurements taken from 36 different locations in a river is found to be 2.6 g/ml.

Find the 95% **confidence interval** for the mean concentration in the river based on the data above. Assume the population of zinc concentration is normally distributed with standard deviation of 0.3 g/ml.

Let $X \sim N(\mu, \sigma^2 = 0.3^2)$ be the random variable for the population zinc concentration in the river.

We have $1 - \alpha = 0.95 \implies \alpha = 0.05$. Thus, the 95% confidence interval for μ is

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$2.6 - z_{0.025}\frac{0.3}{\sqrt{36}} < \mu < 2.6 + z_{0.025}\frac{0.3}{\sqrt{36}}$$

$$2.5 < \mu < 2.7.$$

$\square$

Remarks.

- Ideally, we prefer short interval of high level of confidence.
- Confidence interval computed from a given set of observations either contain the true value or it does not.
- The level of confidence is about proportions of samples of a given size that may be expected to contain the true value.
- That is, for a 95% level of confidence, if many samples of a given size are collected and the confidence intervals are computed, about 95% of these intervals would contain the true value.

Theorem 8 (Confidence Interval for μ of normal distribution with unknown σ^2, $n \geq 30$)

Let $X_1, \ldots, X_n$ i.i.d $\sim N(\mu, \sigma^2)$, $n \geq 30$, and σ^2 is unknown. The $100(1-\alpha)\%$ confidence interval for μ is given by

$$\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}},$$

where S^2 is the sample variance.

When population variance σ^2 is unknown, sample variance is sufficiently good estimator of σ^2 for $n \geq 30$.

For small sample size $n < 30$, the value of sample variance S^2 fluctuate considerably from sample to sample.
To deal with inference on μ, we consider the following statistic

$$Q = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

The distribution of Q is a t-**distribution with degree of freedom** $n - 1$, under the assumption that the population is (approximately) normal.

The upper $100(1-\alpha)\%$ point of a t-distribution with degree r, denoted by $t_\alpha(r)$, is defined similatly. That is

$$\mathbb{P}(Q > t_\alpha(r)) = \alpha.$$

The values of $t_\alpha(r)$ can be found in the Table (NTULearn $->$ Content $->$ TABLES.pdf).

Using a similar argument, we have the following:

Theorem 9 (Confidence Interval for μ of normal distribution with unknown σ^2, $n < 30$)

Let $X_1, \ldots, X_n$ i.i.d $\sim N(\mu, \sigma^2)$, $n < 30$, and σ^2 is unknown. The $100(1-\alpha)\%$ confidence interval for μ is given by

$$\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}},$$

where S^2 is the sample variance.

Example 10

The contents of seven similar containers of sulphuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9, 6 liters. Find a 95% confidence interval for the mean contents μ of population, assuming a normal distribution.

It is a small sample with $n = 7$, with $\overline{x} = 10$, $s = 0.283$. Also, $\alpha = 0.05$ and $t_{0.025}(6) = 2.447$. A 95% confidence interval for μ is

$$\overline{x} - t_{\alpha}(n-1)\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha}(n-1)\frac{s}{\sqrt{n}}$$

$$10 - 2.447\frac{0.283}{\sqrt{7}} < \mu < 10 - 2.447\frac{0.283}{\sqrt{7}}$$

$$9.74 < \mu < 10.26.$$

$\square$