PARAMETRIC ESTIMATION AND PROPERTIES OF ESTIMATORS

Recall the goal of Statistics: to learn from random outcomes (data) about the process which generates them.

We often choose to learn about the data generating process by using the data to estimate unknown parameters in a parametric framework, as in the following setup:

PARAMETRIC FRAMEWORK:

Let $X_1, \dots X_n$ be random variables with a joint distribution which depends on the parameters $\theta_1, \dots, \theta_d$, the values of which are unknown, but which lie in the spaces $\Theta_1, \dots, \Theta_d$, respectively, where $\Theta_k \subset \mathbb{R}$ for $k = 1, \dots, d$.

To know the joint distribution of $X_1, \dots X_n$, it is sufficient to know the values of $\theta_1, \dots, \theta_d$, so we choose estimators $\hat{\theta}_1, \dots, \hat{\theta}_d$ of $\theta_1, \dots, \theta_d$ based on $X_1, \dots, X_n$.

E.g. Let $X_1, \dots, X_n$ be a $\overbrace{\text{r.s.}}^{\text{random sample}}$ from the Bernoulli(p) distribution, where p is unknown.

Fits parametric framework with $d = 1$:

$$\theta_1 = p \quad , \quad \Theta_1 = (0, 1)$$

Might choose the estimator $\hat{\theta}_1 = \hat{p} = \bar{X}_n$.

E.g. Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$, with μ, σ^2 unknown.

Fits parametric framework with $d = 2$:

$$\theta_1 = \mu \quad , \quad \Theta_1 = (-\infty, \infty)$$

$$\theta_2 = \sigma^2 \quad , \quad \Theta_2 = (0, \infty)$$

Might choose estimators $\hat{\theta}_1 = \hat{\mu} = \bar{X}_n$

$\hat{\theta}_2 = \hat{\sigma}^2 = S_n^2$.

E.g. Let $X_1, \ldots, X_n$ be independent r.v.s with the Exponential(λ) distribution.

Fits parametric framework with $d = 1$:

$$\theta_1 = \lambda \quad , \quad \Theta_1 = (0, \infty)$$

Might choose estimator $\hat{\theta}_1 = \hat{\lambda} = \bar{X}_n$.

E.g. Let $Y_1, \ldots, Y_n$ be independent r.v.s such that, for some fixed constants $x_1, \ldots, x_n$,

$$Y_i \sim \text{Normal}(\beta_0 + \beta_1 x_i, \sigma^2), \qquad i = 1, \ldots, n,$$

with β_0, β_1, and σ^2 unknown (so $Y_1, \ldots, Y_n$ are <u>not</u> iid).

Fits parametric framework with $d = 3$:

$$\begin{aligned} \theta_1 &= \beta_0 \quad , \quad \Theta_1 = (-\infty, \infty) \\ \theta_2 &= \beta_1 \quad , \quad \Theta_2 = (-\infty, \infty) \\ \theta_3 &= \sigma^2 \quad , \quad \Theta_3 = (0, \infty) \end{aligned}$$

We might choose the estimators

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbb{Y},$$

where $\mathbb{X}$ and $\mathbb{Y}$ are the matrices

$$\mathbb{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbb{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

and $$\hat{\theta}_3 = \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2.$$

NONPARAMETRIC FRAMEWORK:

Let $X_1, \dots, X_n$ have a joint distribution which depends on an infinite number of parameters.

For example, suppose $X_1, \dots, X_n$ is a random sample from a distribution with cdf F_X, for which we do not specify any form. We estimate the function $F_X(x)$ directly from $X_1, \dots, X_n$ for all values of x.

If we regard the value $F_X(x)$ at each x as a parameter, we see that we must estimate an infinite number of parameters (instead of a finite number of parameters, as in the parametric framework).

One might choose $\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$ for all $x \in \mathbb{R}$.

In this course we remain in the parametric framework.

EVALUATING THE QUALITY OF ESTIMATORS

Even within very simple parametric frameworks, it may not be obvious how to estimate the unknown parameters. Moreover, some estimators which seem reasonable may prove to be naive.

The following properties of estimators give us ways to compare estimators.

Defn: The bias of an estimator $\hat{\theta}$ of a parameter $\theta \in \Theta \subset \mathbb{R}$ is defined as

$$\text{Bias}\, \hat{\theta} = \mathbb{E}\hat{\theta} - \theta .$$

If $\text{Bias}\, \hat{\theta} = 0$, i.e. $\mathbb{E}\hat{\theta} = \theta$, then $\hat{\theta}$ is called an unbiased estimator of θ.

Defn: The standard error (SE) of an estimator $\hat{\theta}$ of a parameter $\theta \in \Theta \subset \mathbb{R}$ is defined as

$$SE\, \hat{\theta} = \sqrt{\text{Var}\, \hat{\theta}},$$

so the standard error of an estimator is merely its standard deviation.

What do we want in an estimator?

* Small or zero bias
* Small variance

The following property of an estimator considers bias and variance together.

Defn: The mean squared error (MSE) of an estimator $\hat{\theta}$ of a parameter θ is defined as

$$\text{MSE}\,\hat{\theta} = \mathbb{E}(\hat{\theta} - \theta)^2,$$

so it is the expected squared distance between $\hat{\theta}$ and θ.

Result: $\text{MSE}\,\hat{\theta} = \text{Var}\,\hat{\theta} + (\text{Bias}\,\hat{\theta})^2.$

Proof:

$$\begin{aligned}
\text{MSE}\,\hat{\theta} &= \mathbb{E}(\hat{\theta} - \theta)^2 \\
&= \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta} + \mathbb{E}\hat{\theta} - \theta)^2 \\
&= \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})^2 + 2\,\underbrace{\mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})}_{=0}(\mathbb{E}\hat{\theta} - \theta) + \mathbb{E}(\underbrace{\mathbb{E}\hat{\theta} - \theta}_{\text{Bias }\hat{\theta}})^2 \\
&= \text{Var}\,\hat{\theta} + (\text{Bias}\,\hat{\theta})^2.
\end{aligned}$$

We usually want estimators with a small MSE: we may even prefer a biased estimator with small variance to an unbiased estimator with large variance.

The MSE thus offers a criterion for choosing between estimators.

E.g. Let $X_1, \ldots, X_n$ be independent r.v.s with the same distribution as $X \sim \text{Uniform}(0, \theta)$.

Consider two possible estimators of θ:

$$\hat{\theta} = X_{(n)}$$

$$\tilde{\theta} = 2\bar{X}_n$$

(i) Find Bias $\hat{\theta}$ and Bias $\tilde{\theta}$:

Bias $\hat{\theta}$: To find $\mathbb{E}\hat{\theta} = \mathbb{E}X_{(n)}$, we need the pdf of $X_{(n)}$.

$$f_{X_{(n)}}(x) = n\left[F_X(x)\right]^{n-1} f_X(x)$$

(from what we know of order statistics)

$$= n \underbrace{\left[\frac{x}{\theta}\right]}_{\text{cdf of Uniform}(0,\theta) \text{ for } 0<x<\theta.}^{n-1} \frac{1}{\theta} \mathbb{1}(0<x<\theta),$$

So

$$\mathbb{E}X_{(n)} = \int_0^\theta x \; n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta}\, dx$$

$$= \frac{n}{\theta^n} \frac{x^{n+1}}{n+1}\Bigg|_0^\theta$$

$$= \frac{n}{n+1}\theta .$$

This gives

So $\hat{\theta}$ is not unbiased.

$$\text{Bias } \hat{\theta} = \frac{n}{n+1}\theta - \theta = \theta\left[\frac{n}{n+1} - \frac{n+1}{n+1}\right] = \theta\left(\frac{1}{n+1}\right).$$

Bias $\tilde{\theta}$: We have $\mathbb{E}\tilde{\theta} = \mathbb{E}\,2\bar{X}_n$

$$= 2\,\mathbb{E}X$$

$$= 2\int_0^\theta x\,\frac{1}{\theta}\,dx$$

$$= 2\,\frac{x^2}{2\theta}\Bigg|_0^\theta$$

$$= \theta,$$

$\tilde{\theta}$ is unbiased.

So $\text{Bias } \tilde{\theta} = \theta - \theta = 0.$

(ii) Find $\operatorname{Var}\hat{\theta}$ and $\operatorname{Var}\tilde{\theta}$:

$\operatorname{Var}\hat{\theta}$: We have $\operatorname{Var}\hat{\theta} = \mathbb{E}\hat{\theta}^2 - (\mathbb{E}\hat{\theta})^2$, where

$$\mathbb{E}\hat{\theta}^2 = \mathbb{E}X_{(n)}^2$$

$$= \int_0^\theta x^2 n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta}\, dx$$

$$= \frac{n}{\theta^n} \frac{x^{n+2}}{n+2}\Big|_0^\theta$$

$$= \left(\frac{n}{n+2}\right)\theta^2,$$

so

$$\operatorname{Var}\hat{\theta} = \left(\frac{n}{n+2}\right)\theta^2 - \left[\left(\frac{n}{n+1}\right)\theta\right]^2$$

$$= \theta^2\left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right]$$

$$= \theta^2\left[\frac{(n+1)^2 n - n^2(n+2)}{(n+2)(n+1)^2}\right]$$

$$= \theta^2\left[\frac{(n^2+2n+1)n - n^3 - 2n^2}{(n+2)(n+1)^2}\right]$$

$$= \theta^2\left[\frac{n}{(n+2)(n+1)^2}\right].$$

$\operatorname{Var}\tilde{\theta}$: We have

$$\operatorname{Var}\tilde{\theta} = \operatorname{Var}[2\bar{X}_n] = 4\operatorname{Var}\bar{X}_n = \frac{4}{n}\operatorname{Var}X,$$

where

$$\operatorname{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

$$= \int_0^\theta x^2 \frac{1}{\theta}\, dx - \left(\frac{\theta}{2}\right)^2$$

$$= \left.\frac{x^3}{3\theta}\right|_0^{\theta} - \frac{\theta^2}{4}$$

$$= \frac{\theta^2}{12},$$

larger than $\operatorname{Var}\hat\theta$ for all $n \geq 1$

so $$\operatorname{Var}\tilde\theta = \frac{4}{n}\,\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

(iii) Find $\text{MSE}\,\hat\theta$ and $\text{MSE}\,\tilde\theta$:

$$\text{MSE}\,\hat\theta = \operatorname{Var}\hat\theta + (\text{Bias}\,\hat\theta)^2$$

$$= \theta^2\left[\frac{n}{(n+2)(n+1)^2}\right] + \left[\theta\left(\frac{1}{n+1}\right)\right]^2$$

$$= \theta^2\left[\frac{2n+2}{(n+2)(n+1)^2}\right]$$

$$= \theta^2\left[\frac{2}{(n+2)(n+1)}\right]$$

$$\text{MSE}\,\tilde\theta = \operatorname{Var}\tilde\theta + (\text{Bias}\,\tilde\theta)^2$$

$$= \frac{\theta^2}{3n} + (0)^2$$

$$= \frac{\theta^2}{3n}$$

Conclusion? Consider $\text{MSE}\,\hat\theta$ and $\text{MSE}\,\tilde\theta$ as function of n:

n	$\text{MSE}\,\hat\theta$	$\text{MSE}\,\tilde\theta$
1	$\theta^2/3$	$\theta^2/3$
2	$\theta^2/6$	$\theta^2/6$
3	$\theta^2/10$	$\theta^2/9$
$\vdots$	$\vdots$	$\vdots$

$\hat\theta = X_{(n)}$ has smaller MSE for $n \geq 3$

Also: With $\hat\theta = X_{(n)}$, we never "overshoot" θ, since $P_{X_{(n)}}(X_{(n)} < \theta) = 1$.

(iv) We can also modify $\hat{\theta} = X_{(n)}$ to correct the bias:

Define $\hat{\theta}^{\text{unbiased}} = \left(\frac{n+1}{n}\right) \hat{\theta}$.

Then $\mathbb{E}\,\hat{\theta}^{\text{unbiased}} = \left(\frac{n+1}{n}\right) \mathbb{E}\,\hat{\theta} = \left(\frac{n+1}{n}\right)\left(\frac{n}{n+1}\right)\theta = \theta$.

and
$$\begin{aligned} \operatorname{Var} \hat{\theta}^{\text{unbiased}} &= \left(\frac{n+1}{n}\right)^2 \operatorname{Var}\hat{\theta} \\ &= \left(\frac{n+1}{n}\right)^2 \theta^2 \frac{n}{(n+2)(n+1)^2} \\ &= \frac{\theta^2}{n(n+2)}. \end{aligned}$$

Thus also $\text{MSE}\ \hat{\theta}^{\text{unbiased}} = \frac{\theta^2}{n(n+2)}$.

E.g. Let $X_1, \ldots, X_n$ be independent Bernoulli(p) r.v.s, with p unknown. Set $Y = X_1 + \ldots + X_n$ and consider two estimators of p:

$$\hat{p} = Y/n$$

$$\tilde{p} = \frac{Y+2}{n+4}$$

(add two successes and two failures to the r.s.)

(i) Find Bias $\hat{p}$ and Bias $\tilde{p}$:

Bias $\hat{p}$: We have $\mathbb{E}\,\hat{p} = \mathbb{E}\, Y/n = \frac{np}{n} = p$, so Bias $\hat{p} = 0$

Bias $\tilde{p}$: We have $\mathbb{E}\,\tilde{p} = \frac{\mathbb{E}Y + 2}{n+4} = \frac{np+2}{n+4}$, so

$$\text{Bias}\ \tilde{p} = \frac{np+2}{n+4} - p = \frac{2-4p}{n+4}$$

$\tilde{p}$ is biased

(ii) Find Var $\hat{p}$ and Var $\tilde{p}$:

Var $\hat{p}$: We have $\text{Var}\,\hat{p} = \text{Var}\left[Y/n\right] = \left(\frac{1}{n}\right)^2 n p(1-p) = \frac{p(1-p)}{n}$.

Var $\tilde{p}$: We have
$$\text{Var}\,\tilde{p} = \text{Var}\left[\frac{Y+2}{n+4}\right]$$
$$= \text{Var}\left[\frac{Y}{n+4}\right]$$
$$= \left(\frac{1}{n+4}\right)^2 n p(1-p)$$
$$= \left(\frac{n}{n+4}\right)^2 \frac{p(1-p)}{n}.$$

(iii) Find MSE $\hat{p}$ and MSE $\tilde{p}$:

$$\text{MSE}\,\hat{p} = \text{Var}\,\hat{p} + (\underbrace{\text{Bias}\,\hat{p}}_{=0})^2$$
$$= \frac{p(1-p)}{n}$$

$$\text{MSE}\,\tilde{p} = \text{Var}\,\tilde{p} + (\text{Bias}\,\tilde{p})^2$$
$$= \left(\frac{n}{n+4}\right)^2 \frac{p(1-p)}{n} + \left(\frac{2-4p}{n+4}\right)^2$$

Conclusion? It turns out that neither estimator has an MSE which is lower than that of the other for all $p \in (0,1)$; For some values of p, MSE $\tilde{p}$ is less than MSE $\hat{p}$, while for other values of p, the opposite is true.

Which estimator has a smaller MSE depends on the true value of p (See R supplement for some exploration of this)!

(iv) Find values of p such that $\text{MSE}\,\tilde{p} \le \text{MSE}\,\hat{p}$:

$$\text{MSE}\,\tilde{p} \le \text{MSE}\,\hat{p}$$

$$\Leftrightarrow \quad \left(\frac{n}{n+4}\right)^2 \frac{p(1-p)}{n} + \left(\frac{2-4p}{n+4}\right)^2 \le \frac{p(1-p)}{n}$$

$$\Leftrightarrow \quad \frac{4-16p+16p^2}{(n+4)^2} \le \frac{p(1-p)}{n}\left[1-\left(\frac{n}{n+4}\right)^2\right]$$

$$\Leftrightarrow \quad \frac{4-16p(1-p)}{(n+4)^2} \le \frac{p(1-p)}{n}\left[\frac{(n+4)^2-n^2}{(n+4)^2}\right]$$

$$\Leftrightarrow \quad 4-16p(1-p) \le p(1-p)\left[\frac{n^2+8n+16-n^2}{n}\right]$$

$$\Leftrightarrow \quad 1-4p(1-p) \le p(1-p)\left(\frac{2n+4}{n}\right)$$

$$\Leftrightarrow \quad 1-p(1-p)\left(\frac{4n+2n+4}{n}\right) \le 0$$

$$\Leftrightarrow \quad 1-2\left(\frac{3n+2}{n}\right)p + 2\left(\frac{3n+2}{n}\right)p^2 \le 0$$

Equality occurs at

$$p^* = \frac{2\left(\frac{3n+2}{n}\right) \pm \sqrt{\left[2\left(\frac{3n+2}{n}\right)\right]^2 - 4(1)2\left(\frac{3n+2}{n}\right)}}{2\cdot 2\left(\frac{3n+2}{n}\right)}$$

$$= \frac{1}{2} \pm \frac{1}{2}\sqrt{1-2\left(\frac{n}{3n+2}\right)}$$

So $\text{MSE}\,\tilde{p} < \text{MSE}\,\hat{p}$ for

$$p \in \left(\frac{1}{2}-\frac{1}{2}\sqrt{1-2\left(\frac{n}{3n+2}\right)},\ \frac{1}{2}+\frac{1}{2}\sqrt{1-2\left(\frac{n}{3n+2}\right)}\right).$$

A WORD OF CAUTION ABOUT ESTIMATING FUNCTIONS OF PARAMETERS

Suppose we want to estimate a function $\tau(\theta)$ of θ, $\tau: \mathbb{R} \to \mathbb{R}$. If $\hat{\theta}$ is an unbiased estimator of θ, it is not generally true that $\tau(\hat{\theta})$ is unbiased for θ.

E.g. Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exponential}(\lambda)$.

Then $\hat{\lambda} = \frac{1}{n}(X_1, \ldots, X_n)$ is an unbiased estimator of λ.

However, $1/\hat{\lambda}$ is a <u>biased</u> estimator of $1/\lambda$:

We have $\mathbb{E}\, 1/\hat{\lambda} = \mathbb{E}\, n\,(X_1 + \cdots + X_n)^{-1} = n\,\mathbb{E}\, Y^{-1}$, where

$$Y = X_1 + \ldots + X_n \sim \text{Gamma}(n, \lambda),$$

since $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = (1 - \lambda t)^{-n}$.

So

$$\mathbb{E}\, Y^{-1} = \int_0^\infty \frac{1}{y} \frac{1}{\Gamma(n)\lambda^n} y^{n-1} e^{-y/\lambda}\, dy$$

$$= \frac{\Gamma(n-1)\lambda^{n-1}}{\Gamma(n)\lambda^n} \underbrace{\int_0^\infty \frac{1}{\Gamma(n-1)\,\lambda^{n-1}} y^{(n-1)-1} e^{-y/\lambda}\, dy}_{=1 \text{, integral over pdf of Gamma}(n-1,\lambda)}$$

$$= \frac{1}{(n-1)\lambda}.$$

Therefore $\mathbb{E}[1/\hat{\lambda}] = \left(\frac{n}{n-1}\right)\frac{1}{\lambda} \neq \frac{1}{\lambda}$.

However, we see that $\mathbb{E}\left(\frac{n-1}{n}\right)\frac{1}{\hat{\lambda}} = \frac{1}{\lambda}$, so that

$$\left(\frac{n-1}{n}\right)\frac{1}{\hat{\lambda}} = (n-1)(X_1 + \cdots + X_n)^{-1}$$

is an unbiased estimator of $1/\lambda$.