2020-02-03 # Parametric Statistics

Core mission: From a population $\{\mathcal{D}_\omega\}_{\omega \in \Omega}$ find a distribution $\mathcal{D}_\omega$ that best fits some set $x_1, \dots, x_N$ of observations of some random variable X.

Once we have $\mathcal{D}_\omega$, we can do many things:

- prediction (what are likely new observations?)
- inference (what does ω <u>mean</u>? what does $\mathcal{D}_\omega$ look like?)

⋱

---

population – the source of the r.v. X. $\{\mathcal{D}_\omega\}$
parameter – value(s) determining a single dist. for pop.
sample – one set of actual observations $x_1, \dots, x_N$
statistic – some function of a sample w/ no use of non-sample values

<u>Example</u> We roll a D6 and note the face up number.
X: outcome of the dice roll (values in .. ?)
useful model?
pop? param? sample? statistic?

<u>Example</u> 100 coin flips, count heads.

same Qs

What do we mean by best fitting ϑ here?
We want to not be wrong.
For some notion of error, we want the error in guessing $\hat{\vartheta}$ for a "true" value ϑ to be low.

<u>Dfn</u> An <u>estimator</u> $\hat{\vartheta}$ of a parameter ϑ is some statistic chosen with the ambition for $\hat{\vartheta}$ to be close to ϑ in value with high probability

A <u>loss function</u> is some function $\mathcal{L}(\hat{\vartheta}, \vartheta)$ that measures how wrong the estimate was.

<u>Example</u>
- $\mathcal{L}(\hat{\vartheta}, \vartheta) = |\hat{\vartheta} - \vartheta|$ absolute error loss
- $\mathcal{L}(\hat{\vartheta}, \vartheta) = |\hat{\vartheta} - \vartheta|^2$ squared error loss

Since statistics are themselves random variables, $\mathcal{L}(\hat{\vartheta}, \vartheta)$ is also a random variable. We might get a high loss even for a good estimator if we are unlucky with our sample.

The <u>risk</u> associated to a loss function $\mathcal{L}$ is the expected value $\mathbb{E}\mathcal{L}(\hat{\vartheta}, \vartheta)$.

The risk can be estimated on a sample as the mean loss value.
In this course we will deal almost exclusively with <u>mean squared error</u> MSE.

Given a sample $X_1, \dots, X_N$, the mean squared error is

$$MSE(X) = \mathbb{E}(\hat{\vartheta} - \vartheta)^2$$

For an estimator, there are two fundamentally important descriptors we will talk about here:

- Variance $\mathbb{V}\hat{\vartheta}$ $= \mathbb{E}(\hat{\vartheta} - \mathbb{E}\hat{\vartheta})^2$ is the variance of $\hat{\vartheta}$ as a r.v.

Dfn - Bias $B(\hat{\vartheta}) = \mathbb{E}\hat{\vartheta} - \vartheta$ is how away, on average, the estimator is.

Consider MSE:

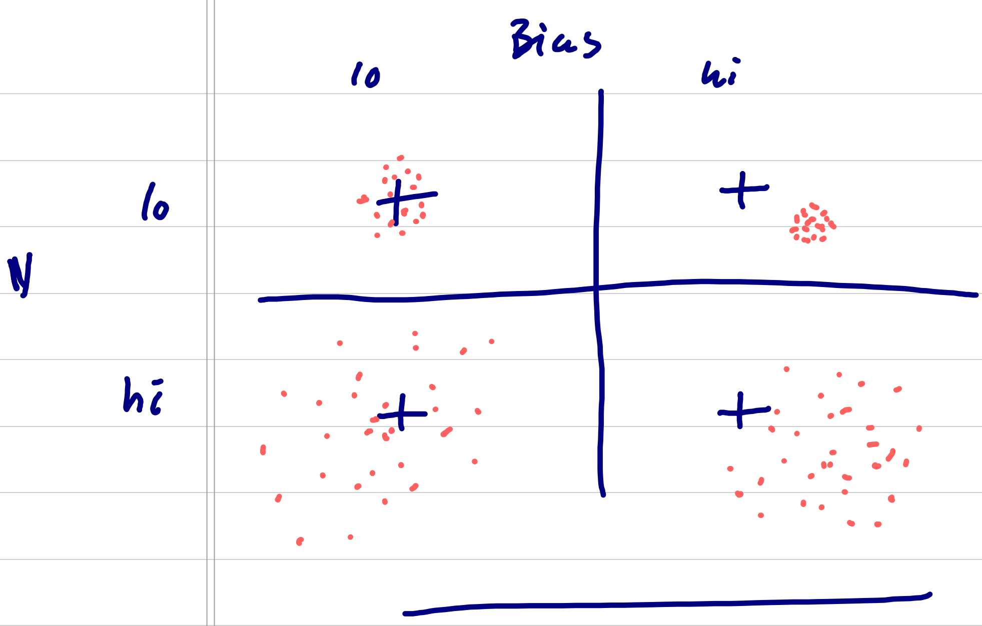
$$\mathbb{E}(\hat{\vartheta} - \vartheta)^2 = \mathbb{E}\left((\hat{\vartheta} - \mathbb{E}\hat{\vartheta}) + (\mathbb{E}\hat{\vartheta} - \vartheta)\right)^2 =$$

$$= \mathbb{E}(\hat{\vartheta} - \mathbb{E}\hat{\vartheta})^2 + \mathbb{E}(\mathbb{E}\hat{\vartheta} - \vartheta)^2 + \underbrace{\mathbb{E}}_{\text{linearity of } \mathbb{E}}\left[2(\hat{\vartheta} - \mathbb{E}\hat{\vartheta})\underbrace{(\mathbb{E}\hat{\vartheta} - \vartheta)}_{\text{constant}}\right]$$

$$= \mathbb{V}\hat{\vartheta} + B(\hat{\vartheta})^2 + 2\cancel{(\mathbb{E}\hat{\vartheta} - \mathbb{E}\hat{\vartheta})}(\mathbb{E}\hat{\vartheta} - \vartheta)$$

This is the variance/bias decomposition - or var/bias tradeoff.

Dfn If $\mathbb{E}(\hat{\vartheta}) = \vartheta$ then $\hat{\vartheta}$ is <u>unbiased</u>.



Sidebar to var ex.

$\mathbb{E}(X^2) = \mathbb{V}X + (\mathbb{E}X)^2.$

$$\mathbb{V}X = \mathbb{E}(X-\mathbb{E}X)^2 = \mathbb{E}\left(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2\right) = \mathbb{E}(X^2) - 2\mathbb{E}X\cdot\mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

$\bar X$ is an estimator of $\mathbb{E}X$

Unbiased?

$$\mathbb{E}\bar X = \mathbb{E}\frac{X_1+\dots+X_N}{N} = \frac{\mathbb{E}X_1+\dots+\mathbb{E}X_N}{N} \overset{\text{iid}}{\underset{\text{sample}}{=}} \frac{\mathbb{E}X+\dots+\mathbb{E}X}{N} = \mathbb{E}X. \checkmark$$

Variance is $\mathbb{E}(X-\mathbb{E}X)^2$. So should be estimated as a mean, right?

$$\mathbb{E}\sum(X_i-\bar X)^2 = \mathbb{E}\left[\sum X_i^2 - 2\bar X \overbrace{\sum X_i}^{=n\bar X} + n\bar X^2\right] =$$

$$= \mathbb{E}\left[\sum X_i^2 - n\bar X^2\right] \overset{\text{iid}}{=} n\mathbb{E}[X^2] - n\mathbb{E}[\bar X^2] =$$

$$= n\left(\mathbb{V}X + (\mathbb{E}X)^2\right) - n\left(\underbrace{\mathbb{V}\bar X}_{=\frac{\mathbb{V}X}{n}} + \underbrace{(\mathbb{E}\bar X)^2}_{=(\mathbb{E}X)^2}\right) = n\mathbb{V}X + \cancel{n(\mathbb{E}X)^2} - \cancel{n}\frac{\mathbb{V}X}{\cancel{n}} - \cancel{n(\mathbb{E}X)^2}$$

$$= (n-1)\mathbb{V}X.$$

So taking $\frac{1}{n}\sum(X_i-\bar X)^2$ gives the wrong answer! Biased!

Biased estimators might still have lower MSE sometimes.

Example Toss a coin 100x. Call y the # of heads

$\hat{p}_1 = y/100$

$\hat{p}_2 = (y+3)/106$ ie add 3x H and 3xT virtually

Model?

Bias? $B(\hat{p}_1) = \mathbb{E}\hat{p}_1 - p = \frac{\mathbb{E}y}{100} - p = \frac{100p}{100} - p = 0$ unbiased

$B(\hat{p}_2) = \frac{\mathbb{E}y + 3}{106} - p = \frac{100p+3}{106} - p = \frac{3-6p}{106}$ biased

Variance? $\mathbb{V}\hat{p}_1 = p(1-p)/100$

$\mathbb{V}\hat{p}_2 = \mathbb{V}\left(\frac{y+3}{106}\right) = \mathbb{V}\left(\frac{y}{106} + \frac{3}{106}\right) = \mathbb{V}\left[\frac{y}{106}\right] = \frac{1}{106^2}\mathbb{V}[y] = \frac{100(p(1-p))}{106^2}$

MSE? $MSE(\hat{p}_1) = p(1-p)/100$

$MSE(\hat{p}_2) = \mathbb{V} + B^2 = \frac{1}{106^2}\left(100p(1-p) + (3-6p)^2\right) = \frac{1}{106^2}\left(100p(1-p) + 9 - 36p + 36p^2\right)$

$= \frac{1}{106^2}\left(64p(1-p) + 9\right)$

MSEs intersect at $\frac{1}{2} \pm \frac{\sqrt{103/403}}{2}$. $\hat{p}_2$ lower MSE near $\frac{1}{2}$.

8.8 $Y_1, Y_2, Y_3 \sim \mathrm{Exp}(\vartheta)$

$$f(y) = \begin{cases} e^{-y/\vartheta}/\vartheta & y>0 \\ 0 & y\le 0 \end{cases}$$

mean ϑ
var ϑ^2

$\hat\vartheta_1 = Y_1 \quad \hat\vartheta_2 = \frac{Y_1+Y_2}{2} \quad \hat\vartheta_3 = \frac{Y_1+2Y_2}{3} \quad \hat\vartheta_4 = \min(Y_1,Y_2,Y_3) \quad \hat\vartheta_5 = \bar Y$

a) biases?

b) among unbiased, which has lowest variance?

$\mathbb{E}\hat\vartheta_1 = \vartheta \quad \mathbb{E}\hat\vartheta_2 = \frac{\vartheta+\vartheta}{2} = \vartheta \quad \mathbb{E}\hat\vartheta_3 = \frac{\vartheta+2\vartheta}{3} = \vartheta \quad \mathbb{E}\hat\vartheta_5 = \frac{\vartheta+\vartheta+\vartheta}{3} = \vartheta$

§6.7 tells us: $f_{(1)}(y) = n f(y)[1-F(y)]^{n-1} = 3\cdot\frac{1}{\vartheta}\cdot e^{-y/\vartheta}\cdot e^{-2y/\vartheta} = \mathrm{Exp}(\vartheta/3)$

So $\mathbb{E}\hat\vartheta_4 = \vartheta/3 \neq \vartheta$.

$\mathbb{V}\hat\vartheta_1 = \vartheta^2 \quad \mathbb{V}\hat\vartheta_2 = \frac{1}{4}\vartheta^2 + \frac{1}{4}\vartheta^2 = \frac{1}{2}\vartheta^2 \quad \mathbb{V}\hat\vartheta_3 = \frac{1}{9}\vartheta^2 + \frac{4}{9}\vartheta^2 = \frac{5}{9}\vartheta^2$

$\mathbb{V}\hat\vartheta_5 = \frac{1}{3}\vartheta^2$

8.9 Notice this is the exponential distr. again, with scale $=\vartheta+1$

So $\mathbb{E}Y=\vartheta+1$, so $\bar{Y}-1$ might work

8.10 Recall if $Y\sim \text{Poisson}(\lambda)$ then $\mathbb{E}Y=\lambda$, $\mathbb{V}Y=\lambda$

Notice also:

$B(\bar{Y})=?$ $\mathbb{V}(\bar{Y})=?$

---

a) We can use $\bar{Y}$ as an unbiased estimator.

b) Recall $\mathbb{V}Y=\mathbb{E}[Y^2]-\mathbb{E}[Y]^2$ so $\mathbb{E}[Y^2]=\mathbb{V}Y+\mathbb{E}[Y]^2$.

$\mathbb{E}(3Y+Y^2)==3\mathbb{E}[Y]+\mathbb{V}Y+\mathbb{E}[Y]^2=3\cdot\lambda+\lambda+\lambda^2=4\lambda+\lambda^2.$

c) Maybe $4\bar{Y}+\bar{Y}^2$ works?

8.12 a) $\mathbb{E}\bar{Y}=\frac{n\mathbb{E}Y}{n}=\vartheta+\frac{1}{2}$ $B(\bar{Y})=\frac{1}{2}$.

b) $\bar{Y}-\frac{1}{2}$ would work

c) Recall if $X\sim\text{Uniform}(a,b)$ then $\mathbb{V}X=\frac{(b-a)^2}{12}$.

$MSE=\mathbb{V}\bar{Y}+B(\bar{Y})^2=\frac{n\mathbb{V}Y}{n^2}+1/4=$

$$=\frac{(\vartheta+1-\vartheta)^2}{12n}+1/4=\frac{1+3n}{12n}$$

8.13 This estimator is $n\hat{p}\hat{q}$

$E[n\hat{p} - n\hat{p}^2] = E\left[Y - \frac{Y^2}{n}\right] = EY - \frac{1}{n}EY^2 =$

$= np - \frac{1}{n}\left(VY + (EY)^2\right) = np - \frac{1}{n}\left(npq + n^2p^2\right) = np - pq - np^2 =$

$= npq - pq = (n-1)pq = \frac{n-1}{n} \cdot npq$

b) Use $(n-1)\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)$.