I. Estimation

A. Aim
1. Want to guess some number $\theta$, called a parameter.
   a. Fraction of population supporting a candidate
   b. Population Average effect of some cholesterol–lowering medication.
   c. Mass of an electron
2. Want it based on some data.
   
   F: 9.1, 9.3–9.4

B. Preliminaries: What makes a good estimator
1. Quantify what happens if you make a wrong decision
   a. Suppose that you pay a penalty $L(a, \theta)$ if your guess is $a$ when the truth is $\theta$
      i. Penalty is called Loss function.
      ii. Most typically, $L(a, \theta) = (a - \theta)^2$
   b. Typically choose $a$ to make worst case for $L$ as $\theta$ changes as small as possible
      i. That is, for each $a$ choose worst $\theta$, called $\theta(a)$
      ii. Minimize $L(a, \theta(a))$
Lecture 1

iii. Called Minimax rule

c. Example:

i. $\theta$ is a proportion between 0 and 1

- $\theta(a) = \begin{cases} 
  1 & \text{if } a \leq \frac{1}{2} \\
  0 & \text{if } a \geq \frac{1}{2}
\end{cases}$

- $L(a, \theta(a)) = \begin{cases} 
  (1 - a)^2 & \text{if } a \leq \frac{1}{2} \\
  a^2 & \text{if } a \geq \frac{1}{2}
\end{cases}$

- Minimized at $a = \frac{1}{2}$

ii. $\theta$ any real number

- $\theta(a) = \pm\infty$

- $L(a, \theta(a)) = \infty \forall a$

- Minimized at $a = \frac{1}{2}$

- Useless

2. Typically, want rule that depends on data, $\delta(X)$

a. Example

i. If $X \sim \text{Bin}(n, \theta)$, $\delta(X)$ might be $X/n$

ii. If $X$ is a sample from a population with mean $\theta$, then $\delta(X)$ might be $\bar{X}$

iii. If $X$ is a sample from a population with median $\theta$, then $\delta(X)$ might be sample median
Lecture 1

b. Want to consider rule before you have any data

c. Since we don’t have data, consider average of loss function

\[ R(\delta, \theta) = \mathbb{E}[L(\delta(X), \theta)] : \text{called risk function.} \]

i. Review expectation

d. Choose \( \delta \) to minimize the maximum of this.

e. With squared-error loss, \( R(\delta, \theta) = \mathbb{E}[(\delta(X) - \theta)^2] \)

i. Ex, \( X \sim \text{Bin}(n, \pi) \)

ii. \( \delta(X) = (X + a)/(n + b) \)

iii. \( R(\delta, \pi) \) is

\[
= \mathbb{E}\left[(\delta(X) - \pi)^2\right] \\
= \text{Var}[\delta(X)] + (\mathbb{E}[\delta(X)] - \pi)^2 \\
= n\pi(1 - \pi)/(n + b)^2 + ((n\pi + a)/(n + b) - \pi)^2 \\
= n\pi(1 - \pi) + (n\pi + a - (n + b)\pi)^2/(n + b)^2 \\
= n\pi(1 - \pi) + (a - b\pi)^2/(n + b)^2 \\
= a^2 + (n - 2ab)\pi + (b^2 - n)\pi^2/(n + b)^2
\]

iv. Risk maximized when

- \( |b| < \sqrt{n} \) and \( n - 2ab - 2(n - b^2)\pi = 0 \), or
\[ \pi = \frac{(n - 2ab)}{[2(n - b^2)]} \] (as long as this is in \([0, 1]\)).

- \(|b| \geq \sqrt{n}\) and \(\pi = 0\) or \(\pi = 1\)

v. Maximized risk is

- \(|b| < \sqrt{n}\) then

\[
\frac{n \left( (2a - b)^2 + n - b^2 \right)}{4 \left( n - b^2 \right) \left( b + n \right)^2},
\]

\(\triangleright\) minimized with \(a = b/2\) to give \(n/[4(b + n)^2]\),

\(\triangleright\) minimized at \(b = \sqrt{n}\).

- \(|b| \geq \sqrt{n}\) then \(\frac{\max(a^2, (b-a)^2)}{(b+n)^2}\)

\(\triangleright\) Minimized re \(a\) when \(a = b/2\) to get \(\frac{b^2}{4(b+n)^2}\),

\(\triangleright\) minimized at \(b = \sqrt{n}\).

vi. Let \(\mu = \mathbb{E}[\delta(X)]\)

vii. \(R(\delta, \theta)\) is
Lecture 2

\[ \begin{align*}
&= E \left[ (\delta(X) - \mu + \mu - \theta)^2 \right] \\
&= E \left[ (\delta(X) - \mu)^2 + 2(\delta(X) - \mu)(\mu - \theta) + (\mu - \theta)^2 \right] \\
&= E \left[ (\delta(X) - \mu)^2 \right] + E \left[ 2(\delta(X) - \mu)(\mu - \theta) \right] \\
&\quad + E \left[ (\mu - \theta)^2 \right] \\
&= \text{Var} [\delta(X)] + 2(\mu - \theta)E [\delta(X) - \mu] + (\mu - \theta)^2 \\
&= \text{Var} [\delta(X)] + (\mu - \theta)^2
\end{align*} \]

F: 10.1–10.2

viii. We can often make the second part 0: If \( E [\delta(X)] = \theta \), then \( \delta(X) \) is called unbiased.

• and \( E [\delta(X)] - \theta \) is called the bias.