Outline

1) Review
2) Estimation
3) Loss & Risk Function
4) Comparing estimators
Estimation

A statistical model is a family of candidate probability distributions \( \mathcal{P} = \{ P_\theta : \theta \in \Theta \} \)
for some random variable \( X \sim P_\theta \)
\( X \in \mathcal{X} \) called data, observed
\( \Theta \) called parameter, unobserved

[could be infinite dimensional, e.g. density func.]

For now, \( \Theta \) fixed and unknown

Goal of estimation: Observe \( X \sim P_\theta \), guess value of some estimand \( g(\theta) \)

Ex. 3.1  Flip a biased coin \( n \) times
\( \Theta \in [0,1] \) prob. of heads
\( X = \# \) heads after \( n \) flips
\( \sim \) Binom\( (n, \Theta) \)

\[ P_\theta(x) = \Theta^x (1-\Theta)^{n-x} \binom{n}{x} \]
density wrt counting measure on \( x=\{0, \ldots, n\} \)
A statistic is any function $T(X)$ of data $X$ 
NOT of $X$ and $\theta$.

An estimator $\hat{g}(X)$ of $g(\theta)$ is a statistic which is "meant to" guess $g(\theta)$.

Ex 3.1 (Cont’d) Natural estimator: $\hat{\theta}(X) = \bar{X}$

Question: is it a good estimator? Is another better?

Loss & Risk

Loss function $L(\theta, d)$ measures how bad an estimate is.

Ex $L(\theta, d) = (d - g(\theta))^2$ squared-error loss

Typical properties:
$L(\theta, d) \geq 0 \quad \forall \theta, d$
$L(\theta, g(\theta)) = 0 \quad \forall \theta$ [no loss from a perfect guess]

[Loss is random, reflects both whether we choose a good estimator and whether we are lucky]
Risk function is expected loss \( ( \text{risk} ) \) as a function of \( \theta \) for an estimator \( \delta ( \cdot ) \)

\[
R(\theta; \delta(\cdot)) = \mathbb{E}_\theta \left[ L(\theta, \delta(x)) \right]
\]

\( \theta \) tells us which parameter value is in effect, \textbf{NOT} "what randomness to integrate over"

\textbf{Ex 3.1 (cont'd)}

\( \delta_0(x) = \frac{x}{n} \)

\( \mathbb{E}_\theta \left[ \frac{x}{n} \right] = \theta \quad \text{(unbiased)} \)

\( \Rightarrow \text{MSE}(\theta; \delta) = \text{Var}_\theta \left( \frac{x}{n} \right) = \frac{1}{n} \theta (1-\theta) \)

\text{Other choices :}

\( \delta_1(x) = \frac{x+3}{n} \quad \text{(stupid)} \)

\( \delta_2(x) = \frac{x+3}{n+6} \quad \text{(6 "pseudo-flips," 3 heads)} \)

\( \text{MSE} \)

\( \mathbb{R}(\theta; \delta_1) \)

\( \mathbb{R}(\theta; \delta_2) \)

\( \mathbb{R}(\theta; \delta_0) \)
Comparing Estimators

We know $\delta_1(x)$ is bad, $\delta_0$ vs. $\delta_2$ more ambiguous.

An estimator $\delta$ is inadmissible if $\exists \delta^*$ with

a) $R(\theta; \delta^*) \leq R(\theta; \delta) \quad \forall \theta \in \Theta$

b) $R(\theta; \delta^*) < R(\theta; \delta)$ some $\theta \in \Theta$

$\delta_1$ is inadmissible

[We can rule out very bad estimators like $\delta_1$, but it is virtually always impossible to find a single uniformly best estimator. Thought experiment: $\delta_3(x) = \frac{3}{2} \text{ best if } \theta = \frac{2}{3}$]

Strategies to resolve ambiguity:

1) Summarize risk function by a scalar

a) Average-case risk: minimize

$$\int_{\Theta} R(\theta; \delta) d\Lambda(\theta) \quad \text{wrt. some measure } \Lambda$$

$\Lambda$ Bayes estimator, $\Lambda$ called prior

- $\Lambda$ "improper" if not normalizable
- This gives a frequentist motivation for Bayes methods

$\delta_2$ is a Bayes estimator
b) Worst-case risk: minimize
\[ \sup_{\theta \in \Theta} R(\theta; \delta) \]

\[ \Rightarrow \text{Minimax estimator, closely related to Bayes} \]

Why not best-case risk? Again consider \( \delta_3 \)

2) Constrain choice of estimator

a) Only consider unbiased \( \delta \):
\[ \mathbb{E}_\theta [\delta(x)] = g(\theta) \quad \forall \theta \in \Theta \]

Rules out \( \delta_1, \delta_2, \delta_3 \).
\( \delta_1 \) is best unbiased estimator.