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 fn.]

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 $\delta(X)$ of $g(\Theta)$ is a statistic which is intended to guess $g(\Theta)$

Ex 3.1 (Cont'd) Natural estimator: $\delta_0(X) = \bar{X}$

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:

$L(\Theta, d) \geq 0 \quad \forall \Theta, d$

$L(\Theta, g(\Theta)) = 0 \quad \forall \Theta \quad \text{(no loss from a perfect guess)}$

[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 \( \hat{\delta}(\cdot) \):

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

\( \hat{\delta}(\cdot) \) tells us which parameter value is in effect, NOT "what randomness to integrate over."

**Ex 3.1 (cont'd)**

\[
\hat{\delta}_0(x) = \frac{x}{n}
\]

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

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

**Other choices:**

\[
\hat{\delta}_1(x) = \frac{x + 3}{n}
\]

\[
\hat{\delta}_2(x) = \frac{x + 3}{n + 6}
\]

(stupid) (6 "pseudo-flips," 3 heads)
Comparing Estimators

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

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

a) $R(\theta; \delta^*) \leq R(\theta; \delta)$ \hspace{1cm} \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: $\delta_3(x) = \frac{2}{3}$ best if $\theta = \frac{2}{3}$ ]

Strategies to resolve ambiguity:

1) Summarize risk function by a scalar

   a) Average-case risk: minimize

   $$\int_{\Theta} R(\theta; \delta) d\Lambda(\theta) \hspace{1cm} \text{wrt. some measure } \Lambda$$

   $\Lambda$ 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 \( \delta_1, \delta_2, \delta_3 \).

\( \delta \), is best unbiased estimator.