Lecture 2  (8/29/2023)

Outline

1) Statistical models
2) Estimation
3) Decision theory
Statistical Models

Probability vs statistics

Probability: Distribution $P$ fully specified
What can we say about $X \sim P$?
Deductive

Statistics: Observe data $X$ from unknown dist. $P$
What can we conclude about $P$?
Inductive

Statistical model
Family $P$ of candidate probability distributions for data $X$

Assume $X \sim P$ for some $P \in \mathcal{P}$
$X$ yields evidence about which $P$ (hopefully)
**Parametric vs. Nonparametric**

**Parametric model** dists indexed by parameter $\Theta \in \Theta$

$$\mathcal{P} = \{ P_{\Theta} : \Theta \in \Theta \}$$

Typically $\Theta \subseteq \mathbb{R}^d$, $d$ called **model dimension**

**Example** $X \sim \text{Binom}(n, \Theta)$ for $\Theta \in [0,1]$ $n$ "known," $\Theta$ "unknown" (by analyst)

$$\mathcal{P} = \{ \text{Binom}(n, \Theta) : \Theta \in [0,1] \}$$

**Nonparametric model** no natural way to index $\mathcal{P}$

Still usually makes assumptions, e.g.

- independence
- shape constraints (e.g. unimodal density)

**Example** $X_1, \ldots, X_n \overset{iid}{\sim} P$ $P$ any distr. on $\mathbb{R}$

$$\mathcal{P} = \{ P^n : P \text{ is a distr. on } \mathbb{R} \}$$ (for $X=(X_1, \ldots, X_n)$)

We can use "parametric notation" $\mathcal{P} = \{ P_{\Theta} : \Theta \in \Theta \}$ wlog (could take $\Theta = P$, $\Theta = \mathcal{P}$).
Bayesian vs. "Frequentist" Inference

Assume \( X \sim P_\theta \) \( \theta \) unknown

Bayesian assumption: \( \theta \) random with known dist.

Inference = calculating dist. \( (\theta | x) \) (posterior)

Considered a strong assumption
(will consider interp., pros & cons later)

Alternate perspective: treat \( \theta \) as fixed, unknown

Methods designed without knowledge of \( \theta \)

Study frequency properties as \( \theta \) varies
**Estimation**

**Setup**

\[ P = \{ P_\theta : \Theta \in \Theta \} \] (wlog)

**Estimand**

\[ g(\theta) \] (something we want to know)

Observe \( X \), calculate estimate \( \hat{\theta}(X) \)

\( \hat{\theta}(\cdot) \) called estimator.

We want to evaluate & compare estimators

**Example**

Flip a biased coin \( n \) times

\( \Theta \in [0,1] \) probability of heads

\( X = \# \) heads

\( \sim \) \( \text{Binom}(n, \Theta) \)

**Goal**: estimate \( \Theta \)

Natural estimator is \( \hat{\theta}_0(X) = \frac{X}{n} \)

How good is it?
Loss and Risk

**Loss function** $L(\theta, d)$

Disutility of guessing $g(\theta) = d$

Typically non-negative, with $L(\theta, d) = 0$ iff $d = g(\theta)$

[Different for every realization]

**Squared error loss:** $L(\theta, d) = (d - g(\theta))^2$

**Risk function:** expected loss of an estimator

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

* tells us which parameter value is in effect, NOT “what randomness to integrate over”

Risk for squared error loss is mean squared error (MSE)

$$\text{MSE}(\theta; \delta(.)) = \mathbb{E}_\theta \left[ (\delta(x) - g(\theta))^2 \right]$$
What is \( \text{MSE}(\theta; \hat{\theta}_0) \)?

\[
E_{\theta} \left[ \frac{X}{n} \right] = \theta \quad \text{(unbiased)}
\]

\[
\Rightarrow \text{MSE}(\theta; \hat{\theta}_0) = E_{\theta} \left[ \left( \frac{X}{n} - \theta \right)^2 \right]
\]

\[
= \text{Var}_{\theta} \left( \frac{X}{n} \right)
\]

\[
= \frac{1}{n} \theta (1 - \theta)
\]

Other possibilities (based on adding "pseudo-flips")

\[
\hat{\theta}_1(X) = \frac{X+1}{n+2} \quad \hat{\theta}_2(X) = \frac{X+2}{n+4} \quad \hat{\theta}_3(X) = \frac{X+1}{n}
\]

Mean squared error for binomial estimators (n=16)
Comparing estimators

We want to choose \( \delta \) to minimize \( R \)
... but this is generally not possible.

An estimator \( \delta \) is \underline{inadmissible} if \( \exists \delta^* \) with

a) \( R(\theta; \delta^*) \leq R(\theta, \delta) \) for all \( \theta \)

b) \( R(\theta, \delta^*) < R(\theta, \delta) \) for some \( \theta \)

We say \( \delta^* \) \underline{strictly dominates} \( \delta \)

\( \delta \) \underline{is inadmissible because} \( \delta_0 \) \underline{dominates} it

(Is there any \underline{uniformly} best estimator
for the binomial example?)
Resolving ambiguity

Main strategies to resolve ambiguity:

1) Summarize risk function by a scalar:

a) **Average-case risk**

\[ \text{Minimize } \int_{\Theta} R(\theta; \delta) \, d\pi(\theta) \]

for some measure \( \pi \), called **prior**

If \( \pi \) is probability measure,

\[ \text{same as } \mathbb{E}_{\theta \sim \pi} [R(\theta; \delta)] \]

\( \rightarrow \) **Bayes estimator**

**Binomial:** \( \delta_1 \) is Bayes w.r.t. \( \pi = \lambda \) on \([0,1]\)

\( \delta_2 \) is also Bayes w.r.t. \( \pi = \text{Beta}(2,2) \)

b) **Worst-case risk**

\[ \text{Minimize } \sup_{\theta} R(\theta; \delta) \]

\( \rightarrow \) **Minimax estimator**

Closely related to **Bayes**

**Binomial:** \( \delta_2 \) is minimax (for \( n=16 \))
2) Restrict choices of estimator
   a) Restrict to unbiased estimators:
      \[ E_{\theta}[\delta(x)] = g(\theta) \text{ for all } \theta \]

      Binomial: \( \delta_0 \) is best unbiased estimator