STAT 5430
Lecture 08, F, Feb 7

- Homework 1 solution posted

practice on point estimation → - Homework 2 is assigned in Canvas (due by next Monday, Feb 10, by midnight)

Office hours Mine: FM, 12-1 PM & by appointment
TA (Min-Yi): WR 11-12 in Snedecor 2404

STAT 5430: Summary to date
Where we have been
& where we are headed

- Completed
  - Introduction to Statistical Inference
  - Point Estimation
    - MME/MLE as strategies for point estimation
    - Finding MLEs: examples using/without calculus, multivariate case
  - Criteria for Evaluating Point Estimators
    - bias, variance, UMVUE, CRLB, relative efficiency, MSE

- Next: Elements of Decision Theory ← how to choose estimators
  - General concepts: loss, risk, admissibility
  - Minimax Principle
  - Bayes Principle
  - Finding Bayes Estimators

Elements of Decision Theory

Terminology

Definition: A real-valued function $L(t, \theta)$ is called a **loss function** for estimating $\gamma(\theta)$ if

distance function between an estimator T & target $\gamma(\theta)$

1. $L(t, \theta) \geq 0$ for all t and θ
2. $L(t, \theta) = 0$ if $t = \gamma(\theta)$.

i.e., think of $L(t, \theta)$ as a penalty for guessing $\gamma(\theta)$ by a stated value "t"

Definition: For an estimator T of $\gamma(\theta)$, the so-called **risk function** of T is given by

expectation (how data are generated depending on θ)

$$\underbrace{R_T(\theta)}_{\text{risk of estimator } T} \equiv \mathrm{E}_\theta\Big[\underbrace{L(T, \theta)}_{\text{loss using } T \text{ to estimate } \gamma(\theta)}\Big], \quad \theta \in \Theta$$

Example 1. $\quad L(t, \theta) = \big(t - \gamma(\theta)\big)^2 \quad$ squared error loss function

$$R_T(\theta) \;=\; \mathrm{E}_\theta\left[L(T, \theta)\right] \;=\; \mathrm{E}_\theta\left[\big(T - \gamma(\theta)\big)^2\right]$$ ← just MSE

Example 2. $\quad L(t, \theta) = \big|t - \gamma(\theta)\big| \quad$ absolute error loss function

$$R_T(\theta) \;=\; \mathrm{E}_\theta\left[\big|T - \gamma(\theta)\big|\right]$$ mean absolute deviation/error

"0-1 loss"

Example 3. $\quad L(t, \theta) = \begin{cases} 1 & \text{if } |t - \gamma(\theta)| > c \\ 0 & \text{if } |t - \gamma(\theta)| \leq c \end{cases}$ for a given $c > 0$ "threshold"

$$R_T(\theta) = E_\theta[L(T,\theta)]$$
$$= 1 \cdot P_\theta(|T-\gamma(\theta)| > c) + 0 \cdot P_\theta(|T-\gamma(\theta)| \leq c)$$
$$= P_\theta(|T-\gamma(\theta)| > c)$$

Elements of Decision Theory

Terminology, cont'd

More Definitions

1. An estimator T_1 is **at least as good as** T_2 if

$$R_{T_1}(\theta) \leq R_{T_2}(\theta) \text{ for all } \theta \in \Theta$$

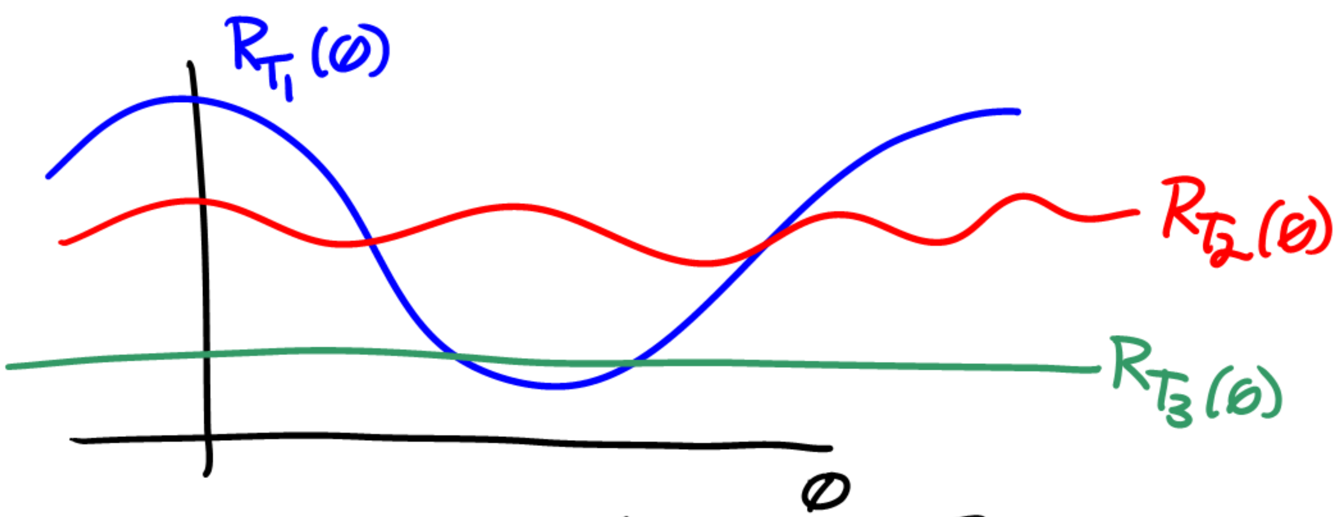
want small risk

2. An estimator T_1 is called **better** than T_2 if
   - (a) $R_{T_1}(\theta) \leq R_{T_2}(\theta)$ for all $\theta \in \Theta$ *and* ← least as good
   - (b) if $R_{T_1}(\theta_0) < R_{T_2}(\theta_0)$ for some $\theta_0 \in \Theta$

3. An estimator T is called **admissible** if there does not exist an estimator that is better than T. Also, T is **inadmissible** if T is not admissible.

means that there exists a "better" estimator

Little sketch here



Here T_3 is better estimator than T_2
(T_2 is inadmissible)

Elements of Decision Theory

Ordering Estimators by Risk

Example: Suppose X_1, X_2 iid Poisson(θ). $n=2$

Consider estimators $T_1 = \bar{X}_2$ and $T_2 = 10$ for $\gamma(\theta) = \theta$ with $L(t,\theta) = (t-\theta)^2$.

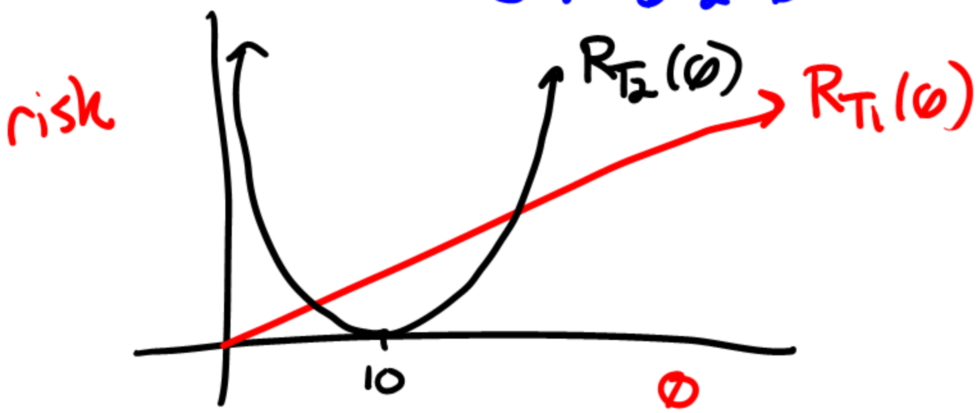
$$R_{T_2}(\theta) \;=\; \mathrm{E}_\theta\left[\left(T_2-\theta\right)^2\right] \;=\; (10-\theta)^2$$

"ignorant estimator"

$$R_{T_1}(\theta) \;=\; \mathrm{E}_\theta\left[\left(T_1-\theta\right)^2\right] \;=\; \mathrm{MSE}_\theta(T_1) \;=\; \mathrm{Var}_\theta(\bar{X}_2) = \frac{\mathrm{Var}_\theta(X_1)}{2} = \frac{\theta}{2}$$

n

$E_\theta T_1 = E_\theta \bar{X}_2 = \theta$



Remark: If T_1 is inadmissible, then we can find an estimator T that is better than T_1. *Hence, we need only consider the set of admissible estimators.*

Remark: In general, a "best" estimator does NOT exist. Instead, one may

1. restrict the class of estimators (e.g., consider only UEs) and look for the best estimator within the smaller class (e.g., UMVUE).

   "best estimator among UEs"

2. or define another optimality criterion for ordering the risk function, such as

   (a) Bayes principle,

   (b) or the minimax principle,

   reduce the risk function $R_T(\theta), \theta \in \Theta$ (curve) to a single number & compare estimators

   and select the best estimator under the new criterion (e.g., Bayes estimators, minimax estimators).

Elements of Decision Theory

Minimax Principle (Game theory)

Player I (Nature/Adversary) | Player II (Statistician)

choose parameter $\downarrow$ θ $\longrightarrow$ data $X_1, \ldots, X_n$ $\longrightarrow$ T $\uparrow$ choose estimator

Rationalization of Minimax Principle:

- If statistician chooses estimator T_1, then nature will pick θ_1 so that $R_{T_1}(\theta_1) = \max_{\theta \in \Theta} R_{T_1}(\theta)$. θ_1 is worst scenario for T_1
- If statistician chooses estimator T_2, then nature will pick θ_2 so that $R_{T_2}(\theta_2) = \max_{\theta \in \Theta} R_{T_2}(\theta)$.
- and so on... so that, for the statistician, the strategy becomes choosing a **minimax estimator**.

Definition: An estimator $\underline{T}$ is called **mi<u>nim</u>ax** if

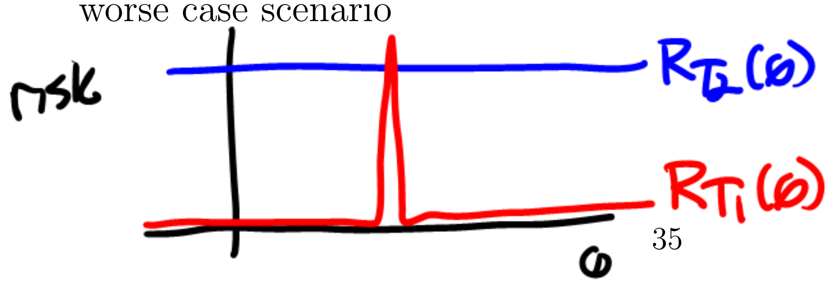
$$\max_{\theta \in \Theta} R_T(\theta) = \min_{T_1} \max_{\theta \in \Theta} R_{T_1}(\theta)$$

$\uparrow$ all estimators

smallest possible "worst case" risk

Notes:

1. If the maximum is not attained on Θ, replace "max" with "sup"
2. Minimax is a <u>conservative</u> optimality criterion since it guards against the worse case scenario



Which estimator is preferred by minimax? pick T_2 because $\max_\theta R_{T_2}(\theta) < \max_\theta R_{T_1}(\theta)$

Elements of Decision Theory

Bayes Principle: Terminology

Definitions:

1. Let $\pi(\theta)$ be a pdf/pmf on Θ. Then, $\pi(\theta)$ is called a **prior**. ↙ distribution on parameter space Θ

2. Then, the **Bayes risk** of an estimator T with respect to $\pi(\theta)$ and loss function $L(t,\theta)$ is

expectation of risk $R_T(\theta)$ w.r.t. prior $\pi(\theta)$

$$\mathrm{BR}_T = \begin{cases} \int_\Theta R_T(\theta)\pi(\theta)d\theta & \text{if } \pi(\cdot) \text{ is continuous} \\ \sum_{\theta\in\Theta} R_T(\theta)\pi(\theta) & \text{if } \pi(\cdot) \text{ is discrete} \end{cases}$$

(average risk)

3. An estimator T_0 is called the **Bayes estimator** (with respect to the prior $\pi(\theta)$) if

$$\mathrm{BR}_{T_0} = \min_T \mathrm{BR}_T$$