Statistical Estimation

Tunde Ajayi
Emmanuel Owusu Ahenkan
Nando Tezoh Franky kévin

Lecturer: Comfort Mintah

AMMI-Ghana

December 12, 2019

OUTLINE

Introduction

Goal

Leveraging statistical and optimisation principles to design the optimal strategy to adopt in the light of real world problems with numerous alternative solutions.

Random Variable and Probability Density Function

- A variable X is a random variable (r.v)if :

$$X : \Omega \to \mathbb{R},$$

  where: Ω is the set of possible outcomes and $\mathbb{R}$ is a measurable space.
- Example,Toss of coin, the set of possibles outcome is $\Omega = \{T, H\}$ and $X(\Omega) = \{1, 0\}$.
- Probability density function of a r.v denoted by f_X is defined as:

$$f_X(x) = \frac{d}{dx} F_X(x), \quad \text{with} \quad F_X(x) = \mathbb{P}(X \leq x).$$

MLE

- Let us consider $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, where x is the set of parameters and y are observed values.
- MLE: tool use to find the parameters that maximize the probability. It is given by :

$$\max_x p_x(y)$$

.

- To finding parameters that maximize the likelihood can be transformed to an optimization problem defined by the following:

$$\begin{array}{ll} \text{Maximize} & \ell(x) = \log p_x(y) \\ \text{subject to} & x \in \mathcal{C}. \end{array}$$

Contd

- For a convex optimization problem, $p_x(y)$ should be a convex function, which means $\log p_x(y)$ should be concave for each value of y. Moreover the set $\mathcal{C}$ can be described by set of affine equality and convex inequalities constraints.

Example

- A linear measurement problem is defined as:

$$y_i = a_i^T x + v_i, \quad i = 1 \cdots m.$$

- The corresponding optimization problem is:

$$\text{Maximize} \quad \ell(x) = \sum_{i=1}^{m} \log p(y_i - a_i^T x)$$

- If $v_i \sim \mathcal{N}(0, \sigma^2)$, then the optimization problem can be written as:

$$\text{Maximize} \quad l(x) = -\sum_{i=1}^{m} \left(a_i^T x - y_i\right)^2$$

Hypothesis Testing

Given observation of r.v $X \in \{1, \cdots, n\}$, it can be generated either by p or q distributions.

- Hypothesis 1: X was generated by distribution $p = (p_1, \cdots, p_n)$.
- hypothesis 2: X was generated by distribution $q = (q_1, \cdots, q_n)$.

Probability Matrix

- It is a nonnegative matrix $P \in \mathbb{R}^{n\times 2}$ with $\mathbb{1}^T P = 1$ with elements defined by equation (9).

$$p_{kj} = \text{prob}(X = k|\theta = j).$$

- p_{k1} gives the probability distribution associated to the hypothesis 1 meanwhile p_{k2} is related to hypothesis 2.

Randomized Detector

- It is a nonnegative matrix $T \in \mathbb{R}^{2\times n}$ with $1^T T = 1$, where its elements are defined by equation (1)

$$t_{ik} = \text{prob}(\hat{\theta} = i | X = k), i = 1, 2 \quad \text{and} \quad t = 1, \cdots, n \qquad (1)$$

- t_{1k} represents the probability of detect $\hat{\theta} = 1$ (Hypothesis 1) when we observe $X = k$, while t_{2k} is related to the second hypothesis.
- Deterministic detector: It is observed when t_{1k} and t_{2k} are either 0 or 1 for $k = 1, \cdots, n$. An example is the maximum likelihood detector.

AIMS African Institute for Mathematical Sciences GHANA

Properties

- Product between the Detection and probability matrices with elements given by equation :

$$D_{ij} = \mathrm{prob}(\hat{\theta} = i | \theta = j) \quad i, j = 1, 2. \tag{2}$$

- Equation (2) denotes the probability of guessing $\hat{\theta} = i$ when $\theta = j$. For $i = j$, it represents the probability of correctly detect $\theta = i$.
- The Detection probability matrix is defined as:

$$D = \begin{bmatrix} Tp & Tq \end{bmatrix} = \begin{bmatrix} 1 - P_{fp} & P_{fn} \\ P_{fp} & 1 - P_{fn} \end{bmatrix}$$

.

Contd

- Matrix D contains on its non diagonal element type I and II errors denotes respectively P_{fp} and P_{fn}.
  - P_{fp} is The false positive.
  - P_{fn} is The false negative.
- If matrix D is a diagonal matrix $(I_{2\times2})$, then we have a perfect detector. That means we will always correctly guess $\hat{\theta} = \theta$.

Formulation of Problem

- Multicriterion formulation of detector design is given as follows:

$$\begin{aligned} \text{Minimize}\left(w.r.t \quad \mathbb{R}_+^2\right) \quad & (P_{fp}, P_{fn}) = ((Tp)_2, (Tq)_1) \\ \text{subject to} \quad & t_{1k} + t_{2k} = 1, k = 1, \cdots, n \\ & t_{ik} \geq 0, i = 1, 2; k = 1, \cdots, n \end{aligned}$$

- Multicriterion Problem has double objective function which represents the errors that we would like to minimize.

Problem Reformulation

- To reformulate the above problem, we use scalarization that enables us to convert a multi-objective problem to a single one.
- Then the new optimization problem becomes

$$\begin{array}{ll} \text{Minimize} & ((Tp)_2 + \lambda(Tq)_1) \\ \text{subject to} & t_{1k} + t_{2k} = 1, k = 1, \cdots, n \\ & t_{ik} \geq 0, i = 1, 2; k = 1, \cdots, n \end{array}$$

- λ is positive constant.
- It corresponds to a Linear programming Problem.

Minimax Detector Design

- We would like to minimize the Worst case.
- The Minimax detector problem can be expressed as an optimization problem given by:

$$\begin{array}{ll} \text{Minimize} & \max\left(P_{fp}, P_{fn}\right) = \max((Tp)_2, (Tq)_1) \\ \text{subject to} & t_{1k} + t_{2k} = 1, k = 1, \cdots, n \\ & t_{ik} \geq 0, i = 1, 2; k = 1, \cdots, n \end{array}$$

We consider the problem of estimating a vector $x \in \mathbb{R}$ from measurements or experiments

$$y_i = a_i^T x + w_i, \qquad i = 1, 2, ..., m, \quad x \in \mathbb{R}$$

- m is the number of experiments.
- Measurement errors w_i are IID $\mathcal{N}(0, 1)$
- Maximum likelihood (least squares) estimate is given as

$$\hat{x} = \left(\sum_{i=1}^{m} a_i^T a_i \right)^{-1} \sum_{i=1}^{m} a_i y_i$$

.

- error $e = \hat{x} - x$ has zero mean and covariance matrix

$$E = \mathbf{E}ee^T = \left(\sum_{i=1}^{m} a_i^T a_i\right)^{-1}.$$

- The goal of experiment design is to choose the vectors a_i , from among the possible choices, so that the error covariance E is small.
- m_j represents the number of experiments for which a_i is chosen to have the value v_j.

Experiment Design

- We can express the error covariance matrix as,

$$E = \left(\sum_{i=1}^{m} a_i^T a_i\right)^{-1} = \left(\sum_{j=1}^{p} m_j v_j v_j^T\right)^{-1}.$$

- This leads to the optimization problem,

$$\begin{aligned} &\text{Minimize} \quad (\text{w.r.t } S_+^n+)\ E = \left(\sum_{j=1}^{p} m_j v_j v_j^T\right)^{-1} \\ &\text{Subjected to} \quad m_i \geq 0, \quad m_1 + \ldots + m_p = m \\ &\qquad\qquad\qquad\qquad\qquad m_i \in \mathbf{Z}. \end{aligned}$$

Relaxed Experiment Design

- Experiment design can be sometimes a hard combinatorial problem to solve when m is comparable to n.
- Assume $m >> p$ and $\lambda = \frac{m_i}{m}$, then the optimization problem is given as,

$$\text{Minimize} \quad (\text{w.r.t } S_+^n)\ E = \frac{1}{m}\left(\sum_{i=1}^{p} \lambda_i v_i v_i^T\right)^{-1}$$

$$\text{Subjected to} \quad \lambda \geq 0, \quad \mathbf{1}\lambda = 1.$$

Scalarization

There are different kinds of scalarization methods, the common ones are;

- A-optimal Design
  In A-optimal experiment design, we minimize the trace of the covariance.
- E-optimal Design
  In E-optimal design, we minimize the norm of the error covariance matrix.
- D-optimal Design
  The most widely used scalarization is called D-optimal design, in which we minimize the determinant of the error covariance matrix E.

D-optimal Design

- The optimization problem is given as,

$$\begin{array}{ll} \text{Minimize} & \log\det\left(\sum_{i=1}^{p}\lambda_i v_i v_i^T\right)^{-1} \\ \text{Subjected to} & \lambda \geq 0, \quad \mathbf{1}\lambda = 1. \end{array}$$

- We reformulated the primal problem with the new variable X.

- The primal problem is then given as;

$$\begin{aligned} \text{Minimize} \quad & \log\det(X)^{-1} \\ \text{Subjected to } X = \lambda_i v_i v_i^T \quad & \lambda \geq 0, \quad \mathbf{1}\lambda = 1. \end{aligned}$$

- We formulate the dual of the primal problem above.
  - The Lagrangian is

$$L(X, \lambda, Z, z, \nu) = \log\det X^{-1} + tr\left(Z\left(X - \sum_{i=1}^{p} \lambda_i v_i v_i^T\right)\right) - z^T\nu + \nu(\mathbf{1}\lambda - 1).$$

AIMS African Institute for Mathematical Sciences GHANA

- The conjugate function is:

$$g(X,\lambda,Z,z,\nu) = \inf \left(\log\det X^{-1} + tr\left(Z \left(X - \sum_{i=1}^{p} \lambda_i v_i v_i^T \right)\right) \right.$$
$$\left. - z^T \nu + \nu(\mathbf{1}\lambda - 1) \right).$$

-

$$\begin{aligned} \nabla_X L &= \frac{-|X^{-1}|X^{-1}}{|X^{-1}|} + Z = 0, \\ &\Longrightarrow X^{-1} = Z. \end{aligned}$$

- $$\begin{aligned} \nabla_{\lambda_i} &= -\nu_i^T Z \nu_i + \nu - z = 0, \\ &\Longrightarrow -\nu_i^T Z \nu_i + \nu = z. \end{aligned}$$

- $$g(X, \lambda, Z, z, \nu) = n + \log\det\, Z - \nu.$$

- The dual problem is given by;

$$\begin{aligned} \text{Minimize} \quad & n + \log\det Z - \nu, \\ \text{Subjected to} \quad & \nu_i^T Z \nu_i \leq \nu. \end{aligned}$$


THANKS
FOR YOUR ATTENTION!!
ANY QUESTIONS?
NO?
GREAT! BYE!