Application of Convex Optimization in the Empirical Likelihood Ratio Test Statistics

Yan-Yu Chen

Department of Economics, National Taiwan University

b05303134@ntu.edu.tw

June 3, 2021

Overview

Table of Contents

Motivation

Convex optimization techniques play a vital role in statistics for estimation, model fitting, and hypothesis testing. In this project, we apply these methods to

- Compute the empirical likelihood ratio test statistic for a statistical hypothesis;
- Show the connection between convex optimization and statistical inference.

Table of Contents

Nonparametric Likelihood Function

Let $\mathcal{P}$ be a set of information that containing in a prior distribution and p be a distribution in $\mathcal{P}$. Suppose we observe N independent samples $x_1, \ldots, x_N$ from the distribution p. Let k_i denote the number of these samples with value α_i so that $k_1 + \cdots + k_n = N$.

Definition

An empirical likelihood is defined as $L(p) = \prod_{i=1}^n p_i^{k_i}$.

By taking the log of an empirical likelihood, we have a log-likelihood function

$$\ell(p) = \sum_{i=1}^{n} k_i \log p_i,$$

which makes it a concave function of p.

Maximum Likelihood Estimation

Definition

A maximum likelihood estimator $\hat{p}$ is a statistics that satisfies

$$\hat{p} = \arg\max\{L(p)|p \in \mathcal{P}\}.$$

- The set $\mathcal{P}$ should be a set of some known constraints based on the background knowledge about the population.
- Instead of maximizing $L(p)$ directly, it is often easier to find the maximizer $\hat{p}$ that maximizes $\ell(p)$.

Maximum Likelihood Estimation

Example

Since p is a distribution, if we know nothing about the population, then $\mathcal{P} = \{p \in \mathrm{R}_+^n | 1^T p = 1\}$, where $1 \in \mathrm{R}^n$ with all elements being 1.

The Lagrange function is

$$L(p, \nu, \lambda) = \ell(p) - \nu(1^T p - 1) + \lambda^T p.$$

We have the first order condition

$$\begin{aligned}
&\frac{\partial}{\partial p_i} L(p, \nu, \lambda) = \frac{k_i}{p_i} - \nu + \lambda_i = 0; \\
&1^T p - 1 = 0; \\
&p_i > 0, \ \lambda_i \geq 0; \\
&\lambda_i p_i = 0,
\end{aligned}$$

which implies $\lambda_i = 0$ for all i. Then, $p_i = k_i / \nu$. Because $1 = 1^T p = N / \nu$, the optimal solution is $\hat{p}_i = k_i / N$ for $i = 1, \ldots, n$.

Empirical Likelihood Ratio Test (ELRT) Statistic

A hypothesis test contains two conjectures, $H_0 : p \in \mathcal{P}_0$ and $H_a : p \in \mathcal{P} \setminus \mathcal{P}_0$. The goal of hypothesis testing is to determine whether we should reject the hypothesis H_0.

Definition

The likelihood ratio test (LRT) statistic is a useful tool to make such a decision. It is defined as

$$R = \frac{\max\{L(p)|p \in \mathcal{P}_0\}}{\max\{L(p)|p \in \mathcal{P}\}}.$$

When an empirical likelihood $L(p)$ is applied, R is called the empirical likelihood ratio test (ELRT) statistic.

Empirical Likelihood Ratio Test (ELRT) Statistic

It can be shown that $-2\log R \overset{\text{apx}}{\sim} \chi^2(d)$, where $d = \dim\mathcal{P} - \dim\mathcal{P}_0$. Thus, given a significance level α, the LRT statistic provides a decision rule: Reject H_0 if $-2\log R > c_\alpha$, where $c_\alpha = P(\chi^2(d) > \alpha)$; otherwise, we do not reject H_0.

However, the ELRT statistic often has no closed-form solution. In practice, we have to compute the value of the ELRT statistics in order to evaluate the possibility of statistical models. Thus, the computation must be done by optimization techniques.

Table of Contents

Two-Sample Problem

Assume the mean μ exists in the population, denoted by $\mathcal{P}_\mu$. Let $p \in \mathcal{P}_{\mu_1}$ and $q \in \mathcal{P}_{\mu_2}$. We are curious about the relationship between μ_1 and μ_2.

- $X = \{x_1, \ldots, x_{N_1}\}$: N_1 independent samples from the distribution p.
- k_i: the number of the samples X with value α_i so that $k_1 + \cdots + k_{n_1} = N_1$.
- $Y = \{y_1, \ldots, y_{N_2}\}$: N_2 independent samples from the distribution q.
- ℓ_j: the number of the samples Y with value β_j so that $\ell_1 + \cdots + \ell_{n_2} = N_2$.

We further assume that p and q are mutually independent. Thus, the empirical likelihood function can be expressed as $L(p, q) = L(p)L(q)$.

A Hypothesis in the Means of Populations

Let $M_1(t)$ and $M_2(t)$ be affine functions from R to R. We develop a hypothesis test:

$$H_0 : M_1(\mu_1) \leq M_2(\mu_2) \text{ v.s. } H_a : M_1(\mu_1) > M_2(\mu_2).$$

Let $\alpha = (\alpha_1, \ldots, \alpha_{n_1})$ and $\beta = (\beta_1, \ldots, \beta_{n_2})$. In this case, the set $\mathcal{P}_0$ will be

$$\{(p, q) \in \mathrm{R}_+^{n_1} \times \mathrm{R}_+^{n_2} | 1^T p = 1, 1^T q = 1, M_1(\hat{\mu}_1) \leq M_1(\hat{\mu}_2)\},$$

where $\hat{\mu}_1 = \alpha^T p$ and $\hat{\mu}_2 = \beta^T q$ are sample means for each population, and the set $\mathcal{P}$ is simply $\{(p, q) \in \mathrm{R}_+^{n_1} \times \mathrm{R}_+^{n_2} | 1^T p = 1, 1^T q = 1\}$. To make such a hypothesis testing, we have to compute the ELRT statistic

$$R = \frac{\max\{L(p, q) | (p, q) \in \mathcal{P}_0\}}{\max\{L(p, q) | (p, q) \in \mathcal{P}\}}.$$

Fortunately, both nominator and denominator of R are convex problems!

Equivalence Form of Convex Problems

Numerator:

$$\begin{aligned}
\min_{p\in\mathbb{R}_+^{n_1},\ q\in\mathbb{R}_+^{n_2}} \quad & -\sum_{i=1}^{n_1} k_i \log p_i - \sum_{j=1}^{n_2} \ell_j \log q_j \\
\text{subject to} \quad & \sum_{i=1}^{n_1} p_i = 1 \\
& \sum_{j=1}^{n_2} q_j = 1 \\
& M_1\left(\sum_{i=1}^{n_1} \alpha_i p_i\right) - M_2\left(\sum_{j=1}^{n_2} \beta_j q_j\right) \le 0. \qquad (1)
\end{aligned}$$

Denominator: Analogous as the numerator except for the inequality constraint (1).

Computing the ELRT Statistic Using Convex Optimization Techniques

By the independent assumption,

$$
\begin{aligned}
R &= \frac{\max\{L(p,q) | (p,q) \in \mathcal{P}_0\}}{\max\{L(p,q) | (p,q) \in \mathcal{P}\}} \\
&= \frac{\max\{L(p,q) | (p,q) \in \mathcal{P}_0\}}{\max\{L(p) | p \in \mathcal{P}_{\mu_1}\} \ \max\{L(q) | q \in \mathcal{P}_{\mu_2}\}} \\
&= \frac{\max\{L(p,q) | (p,q) \in \mathcal{P}_0\}}{L(\hat{p})L(\hat{q})},
\end{aligned}
$$

where $\hat{p}_i = k_i/N_1$ for $i = 1, \ldots, n_1$ and $\hat{q}_j = \ell_j/N_2$ for $j = 1, \ldots, n_2$. The solution follows from the previous example. The rest of computation can only be solved numerically. One practical approach is the infeasible start Newton Method[1].

[1] Stephen Boyd and Lieven Vandenberghe (2004), *Convex Optimization*, Cambridge University Press, p.531-556

Table of Contents

Normal Populations with Different Means

Let X and Y be draw from the normal populations $N(\mu_1, 1)$ with $\mu_1 = 1$ and $N_1 = 100$, and $N(\mu_2, 1)$ with $\mu_2 = 3$ and $N_2 = 120$, respectively. For simplicity, we take $M_1(t) = ct$ and $M_2(t) = t$, where c is a constant. Now, we have

$$R = \frac{\max\{L(p,q)|(p,q) \in \mathrm{R}_+^{n_1} \times \mathrm{R}_+^{n_2}, 1^T p = 1, 1^T q = 1, c\hat{\mu}_1 \leq \hat{\mu}_2\}}{\prod_{i=1}^{n_1}(1/k_i)^{k_i} \prod_{j=1}^{n_2}(1/\ell_j)^{\ell_j}}.$$

We take $c = 1, 2, \ldots, 7$ to see its impact toward the ELRT statistic.

Given $\alpha = 0.05$, we have $c_\alpha = 3.8415$. Once the ELRT statistic is known, our decision rule is to reject $H_0 : c\mu_1 \leq \mu_2$ if $-2\log R > 3.8415$. For the rest of computation, we use CVX in MATLAB to complete the computation.

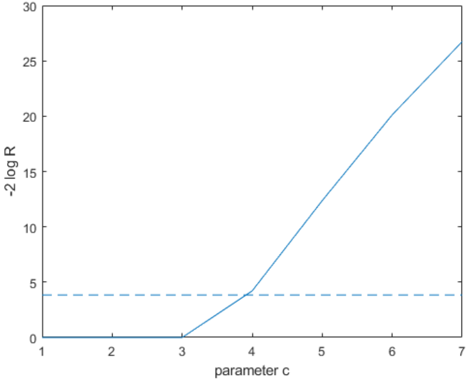
MATLAB Code

```
N1 = 100;
N2 = 120;

x = normrnd(1,1,1,N1);
[x_u, ia_x, ic_x] = unique(x);
k = accumarray(ic_x, 1);
out_x = [x_u' k];
n1 = size(x_u);
n1 = n1(2);

y = normrnd(3,1,1,N2);
[y_u, ia_y, ic_y] = unique(y);
ell = accumarray(ic_y, 1);
out_y=[y_u' ell];
n2 = size(y_u);
n2 = n2(2);
```

MATLAB Code

```
logELRT = zeros(1,7);
for c = 1:7
    cvx_begin quiet
        variables p(n1) q(n2);
        maximize det_rootn(diag([p' q']));
        ones(1,n1) * p == 1;
        ones(1,n2) * q == 1;
        c * (x_u * p) - 1 * (y_u * q) <= 0;
        p >= 0;
        q >= 0;
    cvx_end
    logELRT(c)= -2 * (log(p)' * k + log(q)' * ell -
k' * log(k/N1) - ell' * log(ell/N2));
end

alpha = 0.05;
c_alpha = chi2inv(1-alpha, 1);
```

A Connection between Optimization and Statistics



- The inequality constraint $c\mu_1 \leq \mu_2$ is binding when $c = 4, \ldots, 7$.
- Given $\alpha = 0.05$, we reject the null hypothesis $H_0 : c\mu_1 \leq \mu_2$ for $c = 4, \ldots, 7$.

Table of Contents

Infeasible Start Newton Method

Infeasible Start Newton Method is a generalization of Newton's method that works with initial points, and iterates, that are not feasible.

When trying to tackle the optimization

$$\max\{L(p,q) | (p,q) \in \mathrm{R}_+^{n_1} \times \mathrm{R}_+^{n_2}, 1^T p = 1, 1^T q = 1, c\hat{\mu}_1 \leq \hat{\mu}_2\}$$

directly, we have to select a feasible starting point. Instead of doing feasible analysis, we can use infeasible start Newton method to solve the problem.

Infeasible Start Newton Method

Our problem can be modified into the form:

$$\min_{p\in\mathbb{R}_+^{n_1},\ q\in\mathbb{R}_+^{n_2}} -\sum_{i=1}^{n_1} k_i \log p_i - \sum_{j=1}^{n_2} \ell_j \log q_j$$
$$\text{subject to } Ax = b,$$

where $x = (p, q)$, $b = (1, 1, 0)^T$, and A is a matrix in $\mathrm{R}^{3\times(n_1+n_2)}$ that represents equality constraints and the inequality constraint (1) when it is binding. Based on the previous observation, we have an alternative way to compute the ELRT statistic.

1. Check whether $\hat{p}_i = k_i/N_1$ and $\hat{q}_j = \ell_j/N_2$ is feasible.
2. If it is feasible, then we are done. Otherwise, apply the infeasible start Newton method with the infeasible starting point $x^{(0)} = (\hat{p}, \hat{q})$.

Infeasible Start Newton Method

Let $r(x, \nu)$ be the residual function defined as
$r(x, \nu) = (\nabla f(x) + A^T \nu, Ax - b)$.

Algorithm 1 infeasible start Newton method

Input: starting point $x \in \mathbf{dom} f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$

Output: prime optimizer x^*, dual optimizer ν^*

while $Ax \neq b$ or $\|r(x, \nu)\| > \epsilon$ **do**
    Compute primal and dual Newton steps Δx_{nt}, $\Delta \nu_{nt}$
    Backtracking line search on $\|r\|$. $t := 1$
    while $\|r(x + t\Delta x_{nt}, t\Delta \nu_{nt})\| > (1 - \alpha t)\|r(x, \nu)\|$ **do**
        $t := \beta t$
    end while
    Update $x := x + t\Delta x_{nt}$, $\nu := \nu + t\Delta \nu_{nt}$
end while