ECON220B Discussion Section 9

M-Estimation

Lapo Bini

Roadmap

1. Introduction
2. General Consistency Theorem
3. Uniform Law of Large Numbers
4. Exercise: Poisson Survival Model

Motivation

- We want to find a solution to the following statistical problem:

$$\theta = \arg\max_{c \in \Theta} M(c)$$

  where $M(\cdot)$ is the population criterion function. To form an estimator, one usually solves a sample analogue, and study its properties.

- Sample analogue $\hat{\theta} = \arg\max_{c \in \Theta} M_n(c)$ has no closed-form solution.

- Consistency result: $M_n(c) \xrightarrow{p} M(c)$ is not sufficient for $\hat{\theta} \xrightarrow{p} \theta$. Moreover, here convergence involves two directions: $M_n(\hat{\theta}) \xrightarrow{p} M(\theta)$

General Consistency Theorem

Consider the setting:

$$\theta = \arg\max_{c \in \Theta} M(c) \quad \text{and} \quad \hat{\theta} = \arg\max_{c \in \Theta} M_n(c)$$

with Θ parameter space, if the following condition holds:

1. Identification: $\forall \delta > 0, \exists \varepsilon > 0$ s.t. if $\|c - \theta\| > \delta$ then $M(c) < M(\theta) - \varepsilon$
2. Uniform Convergence: $M_n(c) \xrightarrow{u} M(c)$, i.e. $sup_{c \in \Theta} \|M_n(c) - M(c)\| \xrightarrow{p} 0$

then we have the following result:

$$\hat{\theta} \xrightarrow{p} \theta$$

General Consistency Theorem

Let's analyze the two conditions:

1. If you run away from your true population parameter, the value of the criterion function is strictly lower than its value at the true parameter.

2. Fix tolerance level ε, there exists a cutoff $\bar{n}$ such that for every point c in the domain, given $n > \bar{n}$ the criterion function converges to the true one.

Theorem: Uniform Law of Large Numbers

- Let $m : \mathcal{X} \text{ x } \Theta \to \mathbb{R}$, with Θ parameter space, consider the setting:

$$\theta = \arg\max_{c \in \Theta} E[m(x_i; c)] \qquad \text{and} \qquad \hat{\theta} = \arg\max_{c \in \Theta} \frac{1}{n} \sum_{i=1}^{n} m(x_i; c)$$

- if the following condition holds:

  1. Parameter space Θ is compact
  2. Function $m(x_i; c)$ is continuous in the second argument
  3. $E[sup_{c \in \Theta} \| m(x_i; c) \|] < \infty$ (envelope condition)

- then we have the following results:
  (i) $M_n(c) \xrightarrow{u} M(c)$
  (ii) $M(c) = E[m(x_i; c)]$ is continuous.

Estimating a Right-Censored Poisson Survival Model

- Data: the data contain 228 subjects with advanced lung cancer from the North Central Cancer Treatment Group. It includes the following variables: age, sex in years, survival time in days, censoring status.

- Goal: we want to estimate the survival probabilities of lung cancer patients by sex and age, and quantify the uncertainty around our estimates.

- Strategy: we will estimate the model by MLE using M-estimation.

Statistical Framework: Survival Analysis

- The random variable T_i represents the time until death occurs. It follows an exponential distribution with constant rate λ_i.
- The rate λ_i is known as hazard rate is a measure of the instantaneous risk of an event (death in our case) occurring at any istant t.
- We are interested in the vector of parameters β: $\lambda_i = exp\{x_i^T \beta\}$. Hazard rate is constant over time, but individual-specific.
- The survival function $S(t; i) = \mathbb{P}(T_i > t)$ tells us the probability that the survival time exceeds some time t.

Statistical Framework: Right-Censoring

- Data are rigth-censored: instead of the sequence of survival times $\{T_1, \ldots, T_n\}$ we observe $\{(U_1, y_1), \ldots, (U_n, y_n)\}$.
- We have the auxiliary variable $U_i = T_i$ if $y_i = 1$ (death occurred before the end of the study), and $U_i < T_i$ if $y_i = 0$ otherwise.
- We now derive the likelihood function for individual i based on the information that we have:

Q1 - Distribution Censoring Status

Assume you have sequence $\{Y_i\}_{i=1}^{n}$ of iid random variables $N_i \sim \text{Poisson}(\alpha_i)$:

$$f_{Y_i}(y_i; \alpha_i) = \frac{\alpha_i^{y_i} e^{-\alpha_i}}{y_i!}$$

Derive the log likelihood function for the sample $(Y_1, \ldots, Y_n)$ and substitute $\alpha_i = e^{x_i^T \beta} t_i$. Remember, your sample is iid.

Q2 - Score Equations

Derive the empirical score equations, defined as $\frac{1}{n}\sum \dot{m}(\mathcal{O}_i, \beta) = 0$, using the log-likelihood function for the poisson random sample. Remember:
$m(\mathcal{O}_i, \beta) = log\, f_{Y_i}(y_i; x_i, t_i, \beta)$.

Q3 - Score Equations for Censoring Likelihood

Derive the empirical score equations for the log-likelihood of the censoring dataset. What do you observe?

Q4 - Consistency Result

Provide sufficient condition and show that the uniform law of large numbers holds for the criterion function $M_n(c) = \frac{1}{n} \sum m(\cdot\,; c) \xrightarrow{u} M(c) = E[m(\cdot\,; c)]$.
Then, conclude that our estimator $\hat{\beta} \xrightarrow{p} \beta$.

Interpreting the results

Suppose our estimates are $\hat{\beta} = (-6.84, 0.01, -0.48)$, let's interpret the last estimated coefficient given our log linear regression:

$$log(\lambda_i) = \beta_0 + \text{AGE}_i\beta_1 + \text{FEMALE}_i\beta_2$$

Q5 - Derive Empirical Hessian

Derive the empirical hessian matrix defined as $\hat{H}(\beta) = \frac{1}{n} \sum \ddot{\mathrm{m}}(\mathcal{O}_i, \beta)$

Q6 - Asymptotic Distribution

Given the following asymptotic linear representation:

$$\sqrt{n}\left(\hat{\beta} - \beta\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[-H(\beta)\right]^{-1} \dot{m}(\mathcal{O}_i, \beta) + o_p(1)$$

Find the asymptotic distribution of our estimator and the asymptotic variance V in terms of the standard sandwich.

Q7 - Asymptotic Variance

Claim: $V = -H(\beta)^{-1}$. True or False? Justify.

Survival Probability

- Finally, we can compute the survival probability given the individual characteristic of individual i. Remember:

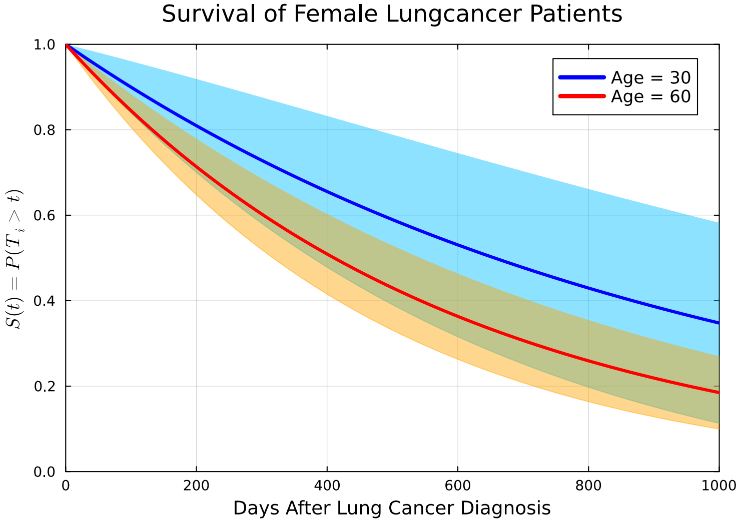
$$g(\beta; t, i) = S(t; i, \beta) = P(T_i > t) = e^{-\lambda_i t}$$

- We can easily use delta method to derive the asymptotic distribution:

$$\sqrt{n}\left(g(\hat{\beta}) - g(\beta)\right) \sim \mathcal{N}(0, \nabla g(\beta)' V \nabla g(\beta))$$

- Let's define $g(\beta; t, i) = exp\{-exp\{x_i^T \beta\} t\}$ where x_i^T and t are fixed, not input of the function. Therefore, the gradient is:

Survival Probability Female Patients



Survival 60 Years Old Patients