Parameter Estimation

Chapters 13-15

Stat 346 - Short-term Actuarial Math

Methods for parameter estimation

Methods for estimating parameters in a parametric model:

- method of moments
- matching of quantiles (or percentiles)
- maximum likelihood estimation
  - full/complete, individual data
  - complete, grouped data
  - truncated or censored data
- Bayes estimation

We are only going to discuss MLEs.

The method of maximum likelihood

The maximum likelihood estimate of parameter vector θ is obtained by maximizing the likelihood function. The likelihood contribution of an observation is the probability of observing the data.

In many cases, it is more straightforward to maximize the logarithm of the likelihood function.

The likelihood function will vary depending on whether it is completely observed, or if not, then possibly truncated and/or censored.

Complete, individual data

In the case where observations in a data set were observed without no truncation and no censoring and each observation X_i, for $i = 1, \ldots, n$, is recorded, the likelihood function is

$$L(\theta) = \prod_{j=1}^{n} f_{X_j}(x_j|\theta).$$

The corresponding log-likelihood function is

$$\ell(\theta) = \log[L(\theta)] = \sum_{j=1}^{n} \log f_{X_j}(x_j|\theta).$$

This type of data is called **complete, individual data**.

MLE for Discrete Case: Poisson Distribution

Poisson distribution is used for modeling count data. Given a sample $x_1, x_2, \ldots, x_n$ from a Poisson distribution with parameter λ, the likelihood function is:

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

The log-likelihood function is:

$$\ell(\lambda) = \sum_{i=1}^{n}(-\lambda + x_i \log(\lambda) - \log(x_i!))$$

Maximizing $\ell(\lambda)$ w.r.t. λ gives the MLE of λ as the sample mean:

$$\hat{\lambda} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Illustrations

Assume we have n samples completely observed. Derive expressions for the maximum likelihood estimates for the following distributions:

- **Exponential:** $f_X(x) = \lambda e^{-\lambda x}$, for $x > 0$.
- **Uniform:** $f_X(x) = \frac{1}{\theta}$, for $0 < x < \theta$.

Closed-Form MLE Solutions for Key Distributions

1. **Uniform Distribution** $(0, \theta)$**:**
   - MLE: $\hat{\theta} = x_{\max}$, the maximum observed value.
2. **Normal Distribution** (μ, σ^2)**:**
   - MLEs: $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.
3. **Binomial Distribution** (m, q)**:**
   - MLE: $\hat{q} = \frac{\bar{x}}{n}$, the sample proportion.
4. **Exponential Distribution** (λ)**:**
   - MLE: $\hat{\theta} = \bar{x}$, the sample mean.
5. **Poisson Distribution** (λ)**:**
   - MLE: $\hat{\lambda} = \bar{x}$, the sample mean.

Constructing Likelihood Functions for Heterogeneous Observations

When observations come from different distributions or parameters, the overall likelihood is the product of individual likelihoods.

Example: Suppose we have two sets of observations, $X_1, ..., X_m$ from a Normal distribution with parameters μ and σ^2, and $Y_1, ..., Y_k$ from a Binomial distribution with parameters m and q. The likelihood functions are:

- For X: $L_X(\mu, \sigma^2) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$
- For Y: $L_Y(m, q) = \prod_{j=1}^{k} \binom{m}{y_j} q^{y_j} (1-q)^{m-y_j}$

The overall likelihood function is $L(\mu, \sigma^2, n, p) = L_X(\mu, \sigma^2) \times L_Y(n, p)$. This approach allows for the simultaneous estimation of parameters from diverse data sources.

SOA Example #26

You are given:

- Low-hazard risks have an exponential claim size distribution with mean θ.
- Medium-hazard risks have an exponential claim size distribution with mean 2θ.
- High-hazard risks have an exponential claim size distribution with mean 3θ.
- No claims from low-hazard risks are observed.
- Three claims from medium-hazard risks are observed, of sizes 1, 2 and 3.
- One claim from a high-hazard risk is observed, of size 15.

Calculate the maximum likelihood estimate of θ. [2]

Calculating MLE for θ

Likelihood Function: The likelihood $L(\theta)$ is the product of individual likelihoods:

$$\begin{aligned} L(\theta) &= \left(\frac{1}{2\theta}e^{-\frac{1}{2\theta}} \cdot \frac{1}{2\theta}e^{-\frac{2}{2\theta}} \cdot \frac{1}{2\theta}e^{-\frac{3}{2\theta}}\right) \cdot \left(\frac{1}{3\theta}e^{-\frac{15}{3\theta}}\right) \\ &= \frac{e^{-\frac{8}{\theta}}}{24\theta^4} \end{aligned}$$

Log-Likelihood Function:

$$\ell(\theta) = \log(L(\theta)) = -\log(24) - 4\log(\theta) - \frac{8}{\theta}$$

Finding MLE of θ: Differentiate $\ell(\theta)$ with respect to θ and set to zero:

$$\frac{d\ell(\theta)}{d\theta} = -\frac{4}{\theta} + \frac{8}{\theta^2} = 0$$

Complete, grouped data

Starting with a set of numbers $c_0 < c_1 < \cdots < c_k$ where from the sample, we have n_j observed values in the interval $(c_{j-1}, c_j]$.

For such data, the likelihood function is

$$L(\theta) = \prod_{j=1}^{k} \left[F(c_j|\theta) - F(c_{j-1}|\theta) \right]^{n_j}.$$

Example

Suppose you are given the following observations for a loss random variable X:

Interval	Number of Observations
(0, 4]	6
(4, 8]	10
(8, ∞)	3
Total	19

Determine the log-likelihood function of the sample if X has a Pareto with parameters α and θ. If it is possible to maximize this log-likelihood and solve explicitly, determine the MLE of the parameters.

Likelihood Function for Grouped Data

Grouped Data Likelihood:

$$L(\alpha,\theta) = [F(4|\alpha,\theta)]^6 \times [F(8|\alpha,\theta) - F(4|\alpha,\theta)]^{10} \times [1 - F(8|\alpha,\theta)]^3$$

$$L(\alpha,\theta) = \left[1 - \left(\frac{\theta}{\theta+4}\right)^{\alpha}\right]^6 \times \left[\left(1 - \left(\frac{\theta}{\theta+8}\right)^{\alpha}\right) - \left(1 - \left(\frac{\theta}{\theta+4}\right)^{\alpha}\right)\right]^{10} \times \left[1 - \left(1 - \left(\frac{\theta}{\theta+8}\right)^{\alpha}\right)\right]^3$$

$$= \left[1 - \left(\frac{\theta}{\theta+4}\right)^{\alpha}\right]^6 \times \left[\left(\frac{\theta}{\theta+4}\right)^{\alpha} - \left(\frac{\theta}{\theta+8}\right)^{\alpha}\right]^{10} \times \left[\left(\frac{\theta}{\theta+8}\right)^{\alpha}\right]^3$$

SOA Example #44

You are given:

- Losses follow an exponential distribution with mean θ.
- A random sample of 20 losses is distributed as follows:

Loss Range	Frequency
[0,1000]	7
(1000, 2000]	6
$(2000, \infty)$	7

Which of the following could be the MLE?

(a) 1701.04

(b) 1876.43

(c) 1996.90

Truncated vs. Censored Data: Definitions and Examples

Truncated Data occurs when observations falling outside a certain range are not included in the analysis at all. For example, a study on adult heights excluding individuals below 5 feet.
Censored Data involves observations that are only partially known. For instance, in survival analysis, we may know an individual has survived beyond a certain time but not the exact time of death.
Real World Scenarios:

- An employment study only including individuals with incomes above a certain threshold (Truncated).
- A clinical trial where patients are followed up for a fixed period, and some are still alive at the end of the study (Right Censored).

Identifying Truncated vs. Censored Data

Consider the following scenarios. Are they examples of truncated or censored data?

1. A survey on household income where only incomes above $50,000 are reported.
2. A study on the lifespan of appliances where the study ends after 10 years, and some appliances are still operational.
3. An insurance claim dataset that only includes claims above a certain deductible.

Discussion:

- Scenario 1 and 3 involve truncation since data below a threshold is not included.
- Scenario 2 involves censoring because we have partial information for some observations (i.e., they are still operational at the study's end).

Likelihood Contribution for Truncated Data

Truncated Data occurs when observations outside a certain range are not included in the analysis.

- **Left Truncated:** Observations are only included if they exceed a certain value, L.

$$L(\theta) = \frac{f(x|\theta)}{1 - F(L|\theta)}$$

  where $f(x|\theta)$ is the PDF and $F(L|\theta)$ is the CDF at the truncation point L.

- **Right Truncated:** Observations are only included if they are below a certain value, U.

$$L(\theta) = \frac{f(x|\theta)}{F(U|\theta)}$$

  where $F(U|\theta)$ is the CDF at the truncation point U.

Truncation affects the range of observable data and adjusts the likelihood function accordingly.

Likelihood Contribution for Censored Data

Censored Data occurs when the observation is only partially known.

- **Left Censored:** The exact value is unknown, but it is known to be less than a certain value, L.

$$L(\theta) = F(L|\theta)$$

- **Right Censored:** The exact value is unknown, but it is known to be greater than a certain value, U.

$$L(\theta) = 1 - F(U|\theta)$$

Censoring indicates the presence of partial information about observations, impacting the likelihood function's formulation.

Likelihood Contribution for Left Truncated and Right Censored Data

When data are **Left Truncated and Right Censored**, observations are only included if they exceed a lower bound, L, and are known only to be above a certain value, U, if they exceed an upper bound.

$$L(\theta) = \begin{cases} \frac{f(x|\theta)}{1-F(L|\theta)}, & \text{if } x \text{ is observed exactly} \\ \frac{1-F(U|\theta)}{1-F(L|\theta)}, & \text{if } x \text{ is right censored} \end{cases}$$

This case combines the adjustments for both truncation and censoring, modifying the likelihood function to accommodate the reduced and partially known observation ranges.

Consider an exponential distribution with parameter θ, left truncation at 2, observed data points at 3 and 5, and two data points right-censored at 6.

Exponential PDF and CDF:

$$f(x|\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \quad F(x|\theta) = 1 - e^{-\frac{x}{\theta}}$$

- For observed data (3 and 5):

$$L_{\text{obs}}(\theta) = \left(\frac{f(3|\theta)}{1 - F(2|\theta)}\right) \times \left(\frac{f(5|\theta)}{1 - F(2|\theta)}\right)$$

- For right-censored data (¿6):

$$L_{\text{cens}}(\theta) = \left(\frac{1 - F(6|\theta)}{1 - F(2|\theta)}\right)^2$$

Combined Likelihood:

$$L(\theta) = L_{\text{obs}}(\theta) \times L_{\text{cens}}(\theta)$$

Sometimes the math works out nicely

$$L(\theta) = \left(\frac{\frac{1}{\theta}e^{-\frac{3}{\theta}}}{e^{-\frac{2}{\theta}}}\right) \times \left(\frac{\frac{1}{\theta}e^{-\frac{5}{\theta}}}{e^{-\frac{2}{\theta}}}\right) \times \left(\frac{e^{-\frac{6}{\theta}}}{e^{-\frac{2}{\theta}}}\right)^2$$

$$L(\theta) = \left(\frac{1}{\theta^2}e^{-\frac{3+5-4}{\theta}}\right) \times \left(e^{-\frac{12-4}{\theta}}\right)$$

$$L(\theta) = \frac{1}{\theta^2}e^{-\frac{16}{\theta}}$$

Distributions with a CDF of the form $1 - \ldots$ are good candidates for these types of problems.

SOA Example #4

You are given:

- Losses follow a single-parameter Pareto distribution with density function:
$$f(x) = \frac{\alpha}{x^{\alpha+1}} \qquad x > 1 \qquad 0 < \alpha < \infty$$
- A random sample of size five produced three losses with values 3, 6 and 14, and two losses exceeding 25.

Calculate the maximum likelihood estimate of α. [0.2507]

Illustrative example

An insurance company records the claim amounts from a portfolio of policies with a current deductible of 100 and policy limit of 1100. Losses less than 100 are not reported to the company and losses above the limit are recorded as 1000. The recorded claim amounts as

$$120,\ 180,\ 200,\ 270,\ 300,\ 1000,\ 1000.$$

Assume ground-up losses follow a Pareto with parameters α and $\theta = 400$.

Use the maximum likelihood estimate of α to estimate the Loss Elimination Ratio for a policy with twice the current deductible.