Constrained Optimization in Quality Control Testing

Department of Statistics and Data Science

November 6, 2023

1 Introduction

Consider a quality control scenario involving a machine designed to test the leakiness of vials. The machine detects the presence of a certain substance, and if the substance's concentration exceeds a threshold T, it indicates a leak, and the entire batch of vials is considered defective. Let X represent the random variable for the concentration of the substance. The batch fails the test if the probability $\Pr(X > T)$ exceeds a critical level ϵ.

2 Problem Statement

The machine's accuracy is limited, and it cannot detect leaks for concentrations below a censoring point $C < T$. The Federal Drug Administration (FDA) mandates that for a distribution of X to be considered unbiased, it must satisfy the following conditions:

1. $\Pr(X > T) = \epsilon$
2. $\mathbb{E}[X] < C$

3 Distributions and Constraints

We will explore two distributions for X: the Exponential and the Generalized Pareto distribution with $\mu = 0$. The Cumulative Distribution Function (CDF) and expected value $E[X]$ for both distributions are given as:

3.1 Exponential Distribution

For the Exponential distribution with parameter λ:

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

$$\mathbb{E}[X] = \frac{1}{\lambda} \tag{2}$$

3.2 Generalized Pareto Distribution

For the Generalized Pareto distribution with scale σ and shape ξ:

$$F(x;\sigma,\xi) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}}, & x \geq 0, \xi \neq 0 \\ 1 - e^{-\frac{x}{\sigma}}, & x \geq 0, \xi = 0 \end{cases} \tag{3}$$

$$\mathbb{E}[X] = \begin{cases} \frac{\sigma}{1-\xi}, & \xi < 1 \\ \text{undefined}, & \xi \geq 1 \end{cases} \tag{4}$$

4 Given Scenario

For a given scenario with $C = 1$, $T = 10$, and $\epsilon = 0.001$, we have observed 10 vials, out of which 5 were censored at 1 and the rest had values of $(1.1, 1.1, 1.1, 1.2, 1.2)$.

5 Likelihood with Partially Observed Data

The likelihood function for the partially observed data, considering the Generalized Pareto distribution, is given by:

$$L(\sigma,\xi) = \prod_{i=1}^{5} f(x_i;\sigma,\xi) \prod_{j=1}^{5} F(c;\sigma,\xi) \tag{5}$$

where $f(x;\sigma,\xi)$ is the probability density function, and $F(c;\sigma,\xi)$ is the CDF evaluated at the censoring point C.

6 Tasks

- Can you use the exponential to solve this problem? If so, do it.
- Derive the Maximum Likelihood Estimators (MLEs) for the parameters of the Generalized Pareto distribution under the constraint that $\Pr(X > T) = \epsilon$ and $\mathbb{E}[X] < C$.
- Obtain the MLEs for the parameters using only the second constraint $\mathbb{E}[X] < C$.

7 Assignment

Using the provided data and constraints, compute the MLEs for the parameters of the Generalized Pareto distribution. Consider both the constrained optimization problem and the problem with only the expectation constraint. Discuss the implications of each constraint on the estimation process and the resulting parameter estimates.