Goodness-of-fit tests based on Lorenz curve for progressive censored data from a location-scale distribution

Author and Position

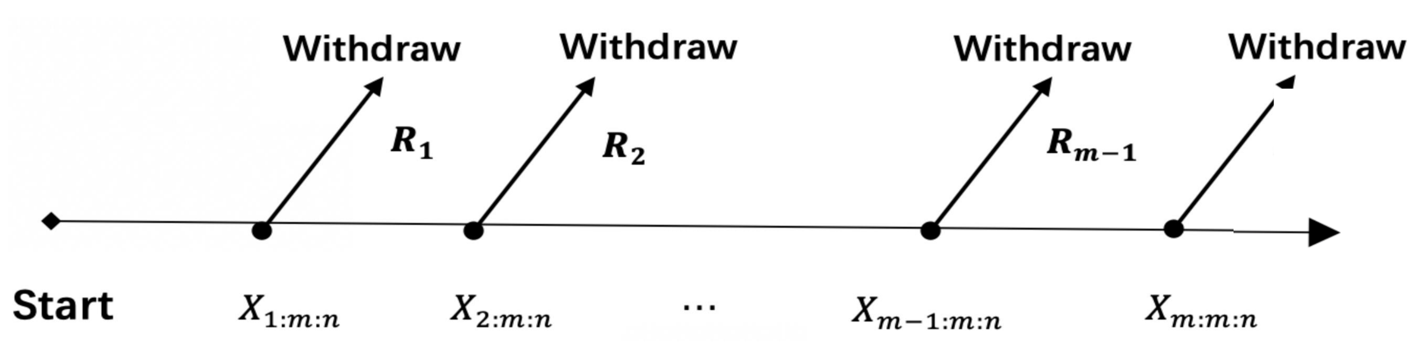
- **Hyein Koo** — Graduate student, Department of Statistics, Daegu University
- **Kyeongjun Lee** — Assistant professor, Division of Mathematics and Big Data Science, Daegu University
- **Namjin Beak** — Undergraduate student, Division of Mathematics and Big Data Science, Daegu University

Text

1. Introduction

❖ The progressively TypeⅡ censored (PC) have become fairly common in a reliability and lifetime-testing experiment as follows. R_1 surviving items are removed from the test at random after 1st observed failure time; furthermore, R_2 surviving items are then removed from the test at random after the 2nd observed failure time. Keep repeating this process until all the remaining $R_m = n - R_1 - \cdots - R_{m-1} - m$ items are removed from the test immediately after the following mth observed failure time. In test, the PC scheme $R = (R_1, R_2, \ldots, R_m)$ is pre-fixed. For this reason the m ordered observed failure times, which we denote by $X_{1:m:n}, X_{2:m:n}, \ldots, X_{m:m:n}$, are referred to as PC.

❖ When the observed failure time data are PC data, the goodness-of-fit tests for perfect data can no longer be used. In this motive, the goodness-of-fit test under PC has received the attention from numerous authors.



2.1 Lorenz curve

❖ In this section explain to Lorenz curve. (LC), which is a standard Lorenz curve scaled up by the mean.

❖ Lorenz curve provides the means to evaluate income or wealth disparity between two distributions. Let $\mathscr{F}$ denotes the CDF of income or wealth distribution. Then the income or wealth is assumed to be non-negative. For a given percentile p, let;

$$\mathscr{F}^{-1}(p) = inf\{y \,|\mathscr{F}(y) \geq p\}, 0 \leq p \leq 1 \quad ,$$

denotes the inverse CDF corresponding to $\mathscr{F}$.

It shall be assumed all through that $\mathscr{F}$ is continuous CDF with finite support.
The Lorenz curves corresponding to the distributions with $\mathscr{F}$ is defined as;

$$\mathrm{L(p)} = \frac{1}{\mu} \int_0^{\mathscr{F}^{-1}} \mathrm{x}\, \mathrm{d}\mathscr{F}(\mathrm{x}) \quad ,$$

where μ denotes the mean of the distribution with $\mathscr{F}$.

With these process the Lorenz curve corresponding to $\mathscr{F}$ is the Lorenz curve scaled up by the mean μ .

2.2 LC with a location-scale distribution

❖ To test whether the PC data comes from a location-scale distribution, let above the PC data with PC scheme from a location-scale distribution. Also, the PC data have a location-scale distribution with a probability density function (PDF).

$$f(x; \mu, \sigma) = \frac{1}{\sigma}\, g\left(\frac{x-\mu}{\sigma}\right) \quad ,$$

where g(·) is the known function, but each μ and σ is the unknown location scale parameter.

$$H_0 : \mathscr{F} \in f_\theta \;\; for\ some\ \theta \in \Theta = \{(\mu\,,\sigma)|\ -\infty < \mu < \infty\,, \sigma > 0\} \quad ,$$

where $\mathscr{F} = \mathscr{F}(x; \mu, \sigma)$ denote the distribution function.

❖ If $U_{i:m:n} = \mathscr{F}(X_{i:m:n}; \mu, \sigma)$, then $p_{i:m:n} = E(U_{i:m:n})$ denote the expected value of the ith PC order statistics from the standard uniform distribution, which is given by;

$$p_{i:m:n} = 1 - \prod_{j=m-i+1}^{m} \left\{\frac{j+R_{m-j+1}+\cdots+R_m}{j+1+R_{m-j+1}+\cdots+R_m}\right\} \quad .$$

❖ Since a Lorenz curve cannot show the characteristics of the skewed distribution, the above result is multiplied by $(1 - p_{j:m:n})$. Then mLC is obtained as;

$$mLC\left(p_{jm:n}\right) = \frac{\sum_{i=1}^{j} X_{i:m:n} - X_{1:m:n}}{\sum_{i=1}^{m} X_{i:m:n} - X_{1:m:n}} - p_{j:m:n} + 1 \quad .$$

❖ Let $\mathscr{F}$ denote the CDF of location-scale distribution, an nLC^+ and nLC^- are obtained as;

$$nLC^+\left(p_{j:m:n}\right) = \frac{mLC\,(p_{j:m:n})}{mLC_f\,(p_{j:m:n})} \,,$$

$$nLC^-\left(p_{j:m:n}\right) = \frac{mLC_f\,(p_{j:m:n})}{mLC\,(p_{j:m:n})} \,,$$

where

$$mLC_f\left(p_{j:m:n}\right) = \frac{\sum_{i=1}^{j} F^{-1}(p_{i:m:n}) - F^{-1}(p_{1:m:n})}{\sum_{i=1}^{m} F^{-1}(p_{i:m:n}) - F^{-1}(p_{1:m:n})} - p_{j:m:n} + 1 \,.$$

Here, the mLC, nLC^+, and nLC^- are clearly location-scale invariant.

$$G_{m:n}^{+} = \max_{1\le j\le m}\left[\left|1 - nLC^+\left(p_{j:m:n}\right)\right|\right] \,,\quad G_{m:n}^{-} = \max_{1\le j\le m}\left[\left|1 - nLC^-\left(p_{j:m:n}\right)\right|\right] \,,$$

and $G_{m:n}^{+} = G_{m:n}^{+} + G_{m:n}^{-}$.

❖ If the data accurately follows a location-scale distribution we expect the $G_{m:n}^{+}$, $G_{m:n}^{-}$ and $G_{m:n}$ test statistics to be zero.

❖ Suggest new plot methonds using nLC^+ and nLC^-;

$$g^+\left(p_{j:m:n}\right) = \left|1 - nLC^+\left(p_{j:m:n}\right)\right| \,,\quad g^-\left(p_{j:m:n}\right) = \left|1 - nLC^-\left(p_{j:m:n}\right)\right| \quad .$$

3. Comparison of the simulated power values

❖ We generated 10,000 samples for different choices of sample sizes and PC schemes.

First, a normal distribution with the parent distribution with the PDF;

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\left\{\frac{(\mathrm{x}-\mu)^2}{2\sigma^2}\right\} \,,\ -\infty < \mu < \infty\,, \sigma > 0 \quad ,$$

as the parent distribution.

The normal distribution, the alternative distribution is considered t distribution with PDF;

$$f(\mathrm{x}; \gamma) = \frac{\tau\left[\frac{v+1}{2}\right]}{\sqrt{\gamma\pi}\,\tau\left(\frac{\gamma}{2}\right)} \left(1 + \frac{\mathrm{x}^2}{\gamma}\right) \,,\ -\infty < \mathrm{x} < \infty\,, \gamma > \,.$$

Next, a Gumbel distribution with the parent distribution with the PDF;

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma} \exp\left\{\frac{\mathrm{x}-\mu}{\sigma}\right\} \exp\left[-\exp\left[\frac{\mathrm{x}-\mu}{\sigma}\right]\right] \,,\ -\infty < \mathrm{x} < \infty\,,\ -\infty < \mu < \infty\,,\ \sigma > 0 \quad ,$$

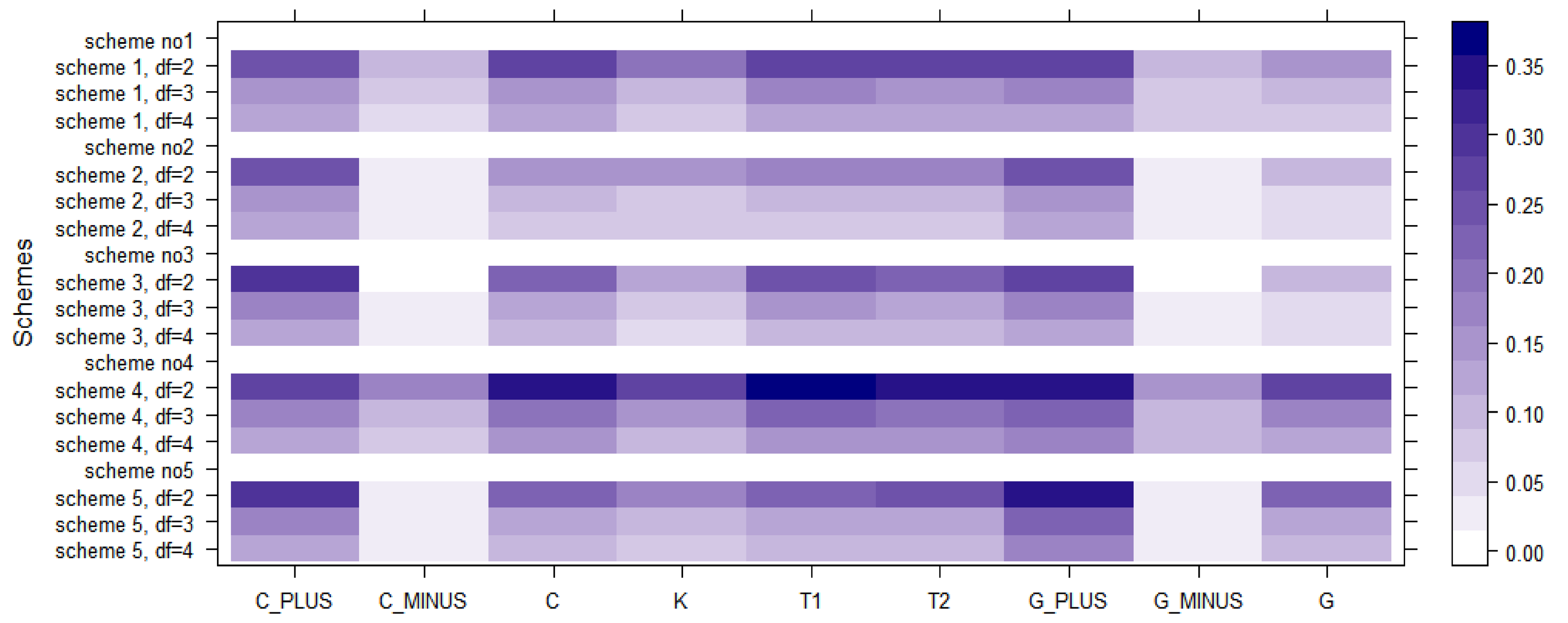
as the parent distribution.

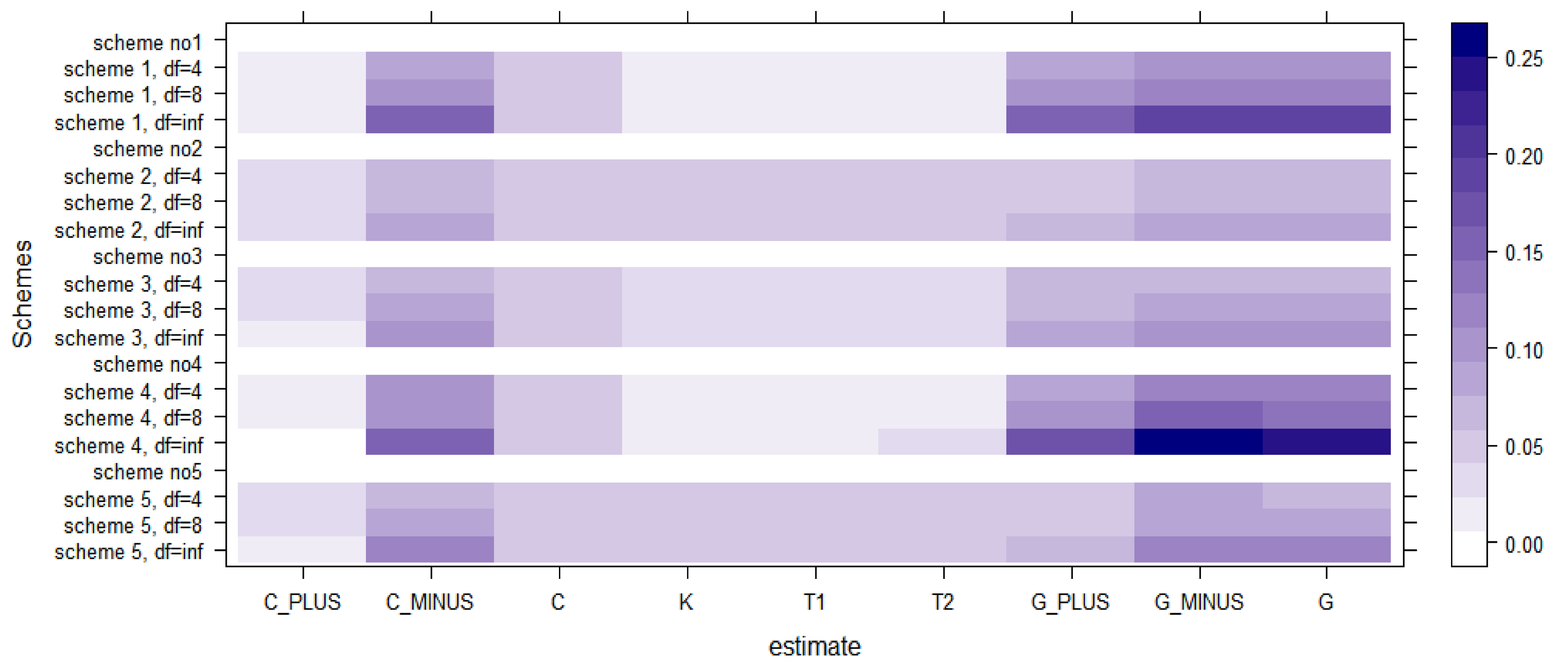
For testing the normal distribution, the alternative distribution is considered log-gamma distribution with PDF;

$$f(\mathrm{x}; \mathrm{k}) = \frac{\mathrm{k}^{\mathrm{k}-\frac{1}{2}}}{\tau(\mathrm{k})} \exp\left[\sqrt{\mathrm{kx} - \mathrm{k}\exp\left\{\frac{\mathrm{x}}{\sqrt{\mathrm{k}}}\right\}}\right] \,,\ -\infty < \mathrm{x} < \infty\,, \mathrm{k} > 0 \quad .$$

4. Simulation Study

❖ The mean squared error (MSE) of the estimators are simulated by Monte Carlo method based on 10,000 runs sample 5 type scheme and 3 different df. We compare the estimators in the sense of the estimation for different censored schemes.





Insulating fluid log data and progressive Type Ⅱ censoring scheme

i	1	2	3	4	5	6	7	8
$X_{i:m:n}$	−1.661	−0.249	−0.041	0.270	1.022	1.579	2.872	1.99
R_i	0	0	3	0	3	0	0	5

Test statistics and the corresponding p-values for the insulating fluid log data

Criterion	$C_{m:n}^{+}$	$C_{m:n}^{-}$	$C_{m:n}$	$K_{m:n}$	$T_{m:n}^{(1)}$	$T_{m:n}^{(2)}$	$G_{m:n}^{+}$	$G_{m:n}^{-}$	$G_{m:n}$
Test statistic	0.080	0.068	0.080	0.148	0.001	0.042	0.076	0.071	0.147
$p-value$	0.563	0.450	0.648	0.564	0.535	0.516	0.794	0.813	0.804