Boundary-Free Estimators for Mean Residual Life Function by Logarithmic Transformation

Rizky Reza Fauzi, Graduate School of Mathematics Kyushu University
Maesono Yoshihiko, Faculty of Mathematics Kyushu University

In this talk we discuss the kernel-type estimators of mean residual life function $m_X(t) = E(X - t|X > t)$. New estimators that can eliminate the boundary bias effect are proposed. Let $X_1, X_2, ..., X_n$ be independently and identically distributed nonnegative random variables with an absolutely continuous survival function S_X and a density f_X. If we define $\mathbb{S}_X(t) = \int_t^\infty S_X(x)\mathrm{d}x$, then the mean residual life function can be written as $m_X(t) = \mathbb{S}_X(t)[S_X(t)]^{-1}$.

The naive kernel estimator for $m_X(t)$ has been discussed by Guillamón *et al.* (1998) and they defined it as $\widehat{m}_X(t) = \widehat{\mathbb{S}}_X(t)[\widehat{S}_X(t)]^{-1}$, where

$$\widehat{S}_X(t) = \frac{1}{n}\sum_{i=1}^{n} V\left(\frac{t - X_i}{h}\right), \quad \widehat{\mathbb{S}}_X(t) = \frac{h}{n}\sum_{i=1}^{n} \mathbb{V}\left(\frac{t - X_i}{h}\right), \tag{1}$$

with $V(x) = \int_x^\infty K(y)\mathrm{d}y$ and $\mathbb{V}(x) = \int_x^\infty V(y)\mathrm{d}y$. The function K is called as kernel and $h > 0$ is called as bandwidth. Though under some weak regularity conditions the naive kernel estimator performs well with the bias and the variance are in the order of $O(h^2)$ and $O(n^{-1})$, respectively, but the boundary bias problem arises because the analysis of mean residual life function is mostly used for nonnegative data. Especially when $f_X(0) > 0$ such as in exponential distribution, the boundary effect is so severe.

Our purpose is to eliminate the boundary bias problem using two new kernel-type estimators. The proposed estimators are, with $j = 1, 2$, $\widetilde{m}_{X,j}(t) = \widetilde{\mathbb{S}}_{X,j}(t)[\widetilde{S}_{X,j}(t)]^{-1}$, where $\widetilde{S}_{X,j}(t) = \frac{1}{n}\sum_{i=1}^n V_{j,h}(\log x, \log X_i)$ and $\widetilde{\mathbb{S}}_{X,j}(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{V}_{j,h}(\log x, \log X_i)$, with

$$V_{1,h}(x, y) = V\left(\frac{x - y}{h}\right), \quad \mathbb{V}_{1,h}(x, y) = e^{x+y}\int_{-\infty}^{y} e^{-z} V\left(\frac{x - z}{h}\right)\mathrm{d}z, \tag{2}$$

and

$$V_{2,h}(x, y) = \frac{1}{h}\int_{-\infty}^{y} K\left(\frac{x - z}{h}\right)\mathrm{d}z, \quad \mathbb{V}_{2,h}(x, y) = \int_{-\infty}^{y} e^{z} V\left(\frac{x - z}{h}\right)\mathrm{d}z. \tag{3}$$

As we can see, our new estimators are based on transformation of data using logarithmic function. Hence, even though the rate of convergence of the variances do not change, the biases are stay in the order of $O(h^2)$ near 0 because no weight is put on the negative real line. As a result, there is no boundary effect.

References

[1] Guillamón, A., Navarro, J., and Ruiz, J. M. (1998) Nonparametric estimator for mean residual life and vitality function. *Statistical Papers* Vol. 39, 263-276.

Keywords: boundary bias problem, data transformation, kernel estimator, logarithmic function, mean residual life function, nonparametric estimation, survival function.