PARAMETERS ESTIMATION OF THE WEIBULL DISTRIBUTION ON RANDOMLY CENSORED SAMPLES

M.S. Tikhov, T.S. Borodina

Nizhny Novgorod State University
Nizhny Novgorod, RUSSIA
e-mail: `tikhovm@mail.ru, zhts.260980@mail.ru`

Abstract

We obtained estimation results concerning a randomly censored sample from a two-parameter Weibull distribution. The least square method is used to derive the point estimators of the parameters. Consistency and asymptotic normality of these estimators is established. The best asymptotically normal estimators of the parameters, namely, shape parameter α and scale parameter σ, is offered.

1 Introduction

The Weibull distribution (WD) is the most commonly used distribution in reliability. Hundreds or even thousands of papers have been written on this distribution (see the bibliography in [Rine (2009)]) and the research is ongoing. It is utmost interest to theory-orientated statisticians because of its great number of special features and to practitioners because of its ability to fit to data from various fields, ranging from life data to weather data or observations made in economics and business administration, in hydrology, in biology or in the engineering sciences.

The Weibull distribution with cumulative distribution function

$$F(x) = 1 - \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{\alpha}\right\}, \quad x > \mu, \sigma > 0, \alpha > 0, \tag{1}$$

is a member of the family of extreme value distribution.

These distributions are the limit distribution of the smallest or the greatest value, respectively, in a sample with sample size $n \to \infty$.

Also, for the estimation of Weibull parameters, the least-squares method (LSM) is extensively used in engineering and mathematics problems. We can get a linear relation between the two parameters taking the logarithms of (1) as follows

$$z = \alpha y - \lambda \qquad (2)$$

letting $\mu = 0$, where $z = \ln(-\ln(1 - F(x))$, $\quad y = \ln x$, $\quad \lambda = -\alpha \ln \sigma$.

Thus, in the system of $y0z$ coordinates the equation (2) is the equation of a straight line. Hence estimation of parameters α, λ and σ can be made an a least-square method.

For getting entire progressively censored samples of I and II types such estimations are considered in [1] and have been investigated, basically, using the numerical methods. However the approaches to the creation are not represented and the behaviour of estimations on randomly censored samples is not investigated in the previously published articles, but it matters much for practical applications. In this article we consider a problem of construction of estimations of parameters of Weibull distribution on randomly censored samples. In transformation we use, mainly, of Kaplan-Meier product-limit estimators as well as the method of the least squares. We show, that the received estimations are concistency and asymptotic normal.

2 Parameters estimation

Here we consider randomly censored sample, when instead of $X_i,\ 1 \le i \le n$ variables are observed only (Z_1, δ_1), (Z_2, δ_2),. . . .., (Z_n, δ_n) pairs, $Z_i = \min(X_i\ ,\ T_i)$, $\delta_i = I(X_i \le T_i)$, $i = 1\ ,\ 2\ ,\ ...,$ where $\{T_i\ 1 \le i \le n\}$ is censored variables , $I(A)$ is indicator of A event. Will be creating Weibull distribution of parameters estimators on randomly censored data $\{(Z_i, \delta_i), 1 \le i \le N\}$.

Having in mind practical application of the model randomly censorship, we will admit, that X_i, $1 \le i \le n$ and T_j, $1 \le j \le n$, are the non-negative and independent random variables, besides, each of random T_j variables does not depend on every X_i variables, where X_i has function of distribution $F(x) = 1 - \exp(-(x/\sigma)^\alpha)$. Censoring T_1, T_2,. . . variables are assumed to be independent and identically distributed random variables with *G (x)* function of distribution. In that case function of distribution of Z_i variables is equal

$$H(x) = \mathbf{P}\,(Z_i \le x) = 1 - (1 - F(x))G(x) = 1 - S(x)\bar{G}(x)\,,$$

where $S(x) = 1 - F(x)$, $\bar{G}(x) = \mathbf{P}\,(T_i > x)$.

It is well-known that the best estimator of survival function $S(x) = 1 - F(x)$ on $(Z_i, \delta_i), 1 \le i \le n$, is Kaplan-Meier (KM) estimator [Kaplan and Meier (1958)]

$$\hat{S}_n(x) = 1 - \hat{F}_n(x) = \prod_{i=1}^{n}\left(1 - \frac{\delta_{[i:n]}}{n - i + 1}\right)^{I(Z_{i:n} \le x)} = \prod_{u \le x}\left(1 - \frac{d\bar{N}(u)}{\bar{Y}(u)}\right), \qquad (2)$$

where $Z_{1:n} \le ... \le Z_{n:n}$ denote the order statistics pertaining to $Z_1, Z_2, ..., Z_n$ with the corresponding concomitants $\delta_{[1:n]}, ..., \delta_{[n:n]}$, so that $\delta_{[i:n]} = \delta_j$ if $Z_{i:n} = Z_j$, $\bar{N}(t) = \sum_{j=1}^{N} N_j(t)$, $N_j(t) = I\,(X_j \le t\ ,\ \delta_j = 1)$, $\bar{Y}(t) = \sum_{j=1}^{n} Y_j(t)$, $Y_j(t) = I(X_j > t)$.

It is known (see [2]), that for an estimation $\hat{F}_n(x)$ relation takes place:

$$\mathbf{P}\left(\sup_{-\infty < x < +\infty} |\,\hat{F}_n(x) - F(x)\,| \ = O\left(\sqrt{\frac{\ln n}{n}}\right)\right) = 1\,,$$

Also, it is known that these estimators are asymptotic normal with mean $F(x)$ and asymptotic variance $\dfrac{S^2(x)}{n}\int\limits_0^x \dfrac{dF(y)}{S^2(y)\,\bar{G}(y)}$.

2.1 The least squares procedure

In this section, we shall derive the least square estimators (LSEs) of the two parameters α, λ. Given the observed pairs $(Z_i, \delta_i), 1 \le i \le n$, in a censored sample, where $\{X_i\}, 1 \le$

$i \leq n$, from the WD$(\alpha, 0, \sigma)$. Then the least square estimates of the parameters α, λ, denoted $\hat{\alpha}$, $\hat{\lambda}$ respectively, can be obtained by minimizing the following quantity with respect to α, λ:

$$Q = \int_0^\infty (\ln(-\ln(1-\hat{F}_n(x)) - \alpha \ln x + \lambda)^2 \, d\hat{F}_n(x).$$

That is, to get $\hat{\alpha}$, $\hat{\lambda}$, we have to solve the following system of linear equations with respect to α, λ:

$$\begin{cases} \int_0^\infty \ln x \cdot \ln(-\ln(1-\hat{F}_n(x))\, d\hat{F}_n(x) = \alpha \int_0^\infty \ln^2 x \, d\hat{F}_n(x) - \lambda \int_0^\infty \ln x \, d\hat{F}_n(x)\,, \\ \int_0^\infty \ln(-\ln(1-\hat{F}_N(x))\, d\hat{F}_n(x) = \alpha \int_0^\infty \ln x \, d\hat{F}_n(x) - \lambda \end{cases} \quad (5)$$

From this equations we get

$$\hat{\alpha} = \frac{\hat{\nu}_1 \hat{\tau}_1 - \hat{\nu}_2}{\hat{\tau}_1^2 - \hat{\tau}_2}, \quad \hat{\lambda} = \frac{\hat{\nu}_1 \hat{\tau}_2 - \hat{\nu}_2 \hat{\tau}_1}{\hat{\tau}_1^2 - \hat{\tau}_2},$$

where

$$\hat{\nu}_1 = \int_0^\infty \ln(-\ln(1-\hat{F}_n(x))\, d\hat{F}_n(x)\,,$$
$$\hat{\nu}_2 = \int_0^\infty \ln x \cdot \ln(-\ln(1-\hat{F}_n(x))\, d\hat{F}_n(x)\,,$$
$$\hat{\tau}_1 = \int_0^\infty \ln x \, d\hat{F}_n(x)\,, \quad \hat{\tau}_2 = \int_0^\infty \ln^2 x \, d\hat{F}_n(x)\,.$$

3 Consistency of estimators

The following result takes place.

Theorem 1. *Assume that* $S(x) = \exp(-(x/\sigma)^\alpha)$ *and* $G(x)$ *are continuous functions. Then* $\hat{\alpha} \xrightarrow[n\to\infty]{p} \alpha\,, \quad \hat{\lambda} \xrightarrow[n\to\infty]{p} \lambda\,.$

Proof. Taking into account the following notation

$$\hat{\nu}_1 = \int_0^\infty \ln(-\ln(1-\hat{F}_n(x)))\, d\hat{F}_n(x) = \int_0^\infty \ln(-\ln(1-F(x)))\, d\hat{F}_n(x) +$$
$$+ \int_0^\infty \ln(-\ln(1-\hat{F}_n(x)) - \ln(-\ln(1-F(x)))\, d\hat{F}_n(x)\,,$$

and having $n \to \infty$,

$$\sup_{x \leq x_0} |\ln(-\ln(1-\hat{F}_n(x) - \ln(-\ln(1-F(x))| \leq C_1 \sup_{x \leq x_0} |\ln(1-\hat{F}_n(x)) - \ln(1-F(x))| \leq$$
$$\leq C_2 \sup_{x \leq x_0} |\hat{F}_n(x) - F(x)|\,, \quad (3)$$

Using results of work [4], , we receive: $\hat{\nu}_1 - \nu_1 \xrightarrow[n\to\infty]{p} 0$, where

$$\nu_1 = \int_0^\infty \ln(-\ln(1-F(x))\, dF(x) = \int_0^\infty (\alpha \ln x + \lambda)\, dF(x)\,.$$

Similarly,

$$\hat{\nu}_2 = \int_0^\infty \ln x \cdot \ln(-\ln(1-\hat{F}_n(x)))\, d\hat{F}_n(x) \xrightarrow[n\to\infty]{p} \nu_2,$$

where

$$\nu_2 = \int_0^\infty \ln x \ln(-\ln(1-F(x)))\, dF(x) = \int_0^\infty \ln x (\alpha \ln x + \lambda)\, dF(x)\,,$$
$$\hat{\tau}_1 = \int_0^\infty \ln x \, d\hat{F}_n(x) \xrightarrow[n\to\infty]{p} \tau_1 = \int_0^\infty \ln x \, dF(x)\,,$$
$$\hat{\tau}_2 = \int_0^\infty \ln^2 x \, d\hat{F}_n(x) \xrightarrow[n\to\infty]{p} \tau_2 = \int_0^\infty \ln^2 x \, dF(x)\,.$$

From these relations follows, that $\hat{\alpha} \xrightarrow[n\to\infty]{p} \dfrac{\nu_1 \tau_1 - \nu_2}{\tau_1^2 - \tau_1} = \alpha$, and $\hat{\lambda} \xrightarrow[n\to\infty]{p} \dfrac{\nu_1 \tau_2 - \nu_2 \tau_1}{\tau_1^2 - \tau_1} = \lambda\,.$

4 Asymptotic normal estimators.

As $H(x)$ is a function of distribution of $Z_i = \min(X_i\,,\ T_i)$ variables, we will admit that $\bar{H} = 1 - H$ and define $\tau_H = \inf\{x \geq 0:\ H(x) = 1\}$.

Theorem 2. *Under conditions of the theorem* 1 $\sqrt{n}(\hat{\alpha} - \alpha)$ *and* $\sqrt{n}\,(\hat{\lambda} - \lambda)$ *be asymptotic normal* (*as* $n \to \infty$).

Proof. For a convergence establishment normalized differences estimations $\sqrt{n}(\hat{\alpha} - \alpha)$ and $\sqrt{n}\,(\hat{\lambda} - \lambda)$ to normal random variables, we will use advantage of the results of works [4] in which it is proved, that

$$\sqrt{n}\,\left(\int \varphi\, d\hat{F}_n - \int_0^{\tau_H} \varphi\, dF\right) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, \sigma_1^2),$$

$$\sigma_1^2 = \int\limits_0^{\tau_H} \frac{\varphi^2(x)}{\bar{G}(x)}\, dF(x) - \left(\int\limits_0^{\tau_H} \varphi\,(x)\, dF(x)\right)^2 - \int\limits_0^{\tau_H} \frac{S(x)}{1 - H(x)} \left\{\int\limits_x^{\tau_H} \varphi\,)y)\, dF(y)\right\} dG(x)\,,$$

where φ: $R \to R$ measuring function provides $\int \varphi^2\, dF < \infty$.

Using this result we can show, for example, that $\sqrt{n}\,(\hat{\nu}_1 - \nu_1)$ and $\sqrt{n}\,(\hat{\tau}_1 - \tau_1)$ be asymptotic normal $\mathcal{N}\,(0, \sigma_2^2)$ and $\mathcal{N}\,(0, \sigma_3^2)$, if as $\varphi\,(x)$ function we take $\varphi_2(x) = = \ln(-\ln(1 - F(x)))$ and $\varphi_3(x) = \ln x$ respectively.

From decomposition $\hat{\nu}_1\hat{\tau}_1 - \nu_1\tau_1 = \hat{\nu}_1\,(\hat{\tau}_1 - \tau_1) + \tau_1(\hat{\nu}_1 - \nu_1)$ we come to conclusion, that sequence $\sqrt{n}\,(\hat{\nu}_1\hat{\tau}_1 - \nu_1\tau_1)$ at $n \to \infty$ will be asymptotic normal $\mathcal{N}(0, \sigma_4^2)$, where

$$\sigma_4^2 = \sigma_2^2 + \sigma_3^2 + 2\sigma_{23}, \qquad \sigma_{23} = \lim_{N\to\infty} \sqrt{n}\,\mathbf{E}\,((\hat{\nu}_1 - \nu_1)(\hat{\tau}_1 - \tau_1)).$$

Similarly, $\sqrt{n}\,(\hat{\nu}_1\hat{\tau}_2 - \nu_1\tau_2)$, $\sqrt{n}\,(\hat{\nu}_2\hat{\tau}_1 - \nu_2\tau_1)$ and $\sqrt{n}\,(\hat{\tau}_1^2 - \tau_1^2)$ will be asymptotic normal $\mathcal{N}\,(0, \sigma_5^2)$, $\mathcal{N}\,(0, \sigma_6^2)$ and $\mathcal{N}\,(0, \sigma_7^2)$. Further, let $\sqrt{n}\,(\hat{T}_{1n} - T_1)$ and $\sqrt{n}\,(\hat{T}_{2n} - T_2)$ be asymptotic normal $\mathcal{N}\,(0, \sigma_8^2)$, $\mathcal{N}\,(0, \sigma_9^2)$. We have :

$$\frac{\hat{T}_{2n}}{\hat{T}_{1n}} - \frac{T_2}{T_1} = \frac{\hat{T}_{2n} - T_2}{T_1} - \frac{T_2\,(\hat{T}_{1n} - T_1)}{T_1^2} + O\left(\frac{(\hat{T}_{2n} - T_2)(\hat{T}_{1n} - T_1)}{T_1^2}\right) + O\left(\frac{T_2\,(\hat{T}_{1n} - T_1)^2}{T_1^3}\right),$$

from which follows asymptotic normality of sequences $\sqrt{n}\,(\hat{\alpha} - \alpha)$ and $\sqrt{n}\,(\hat{\lambda} - \lambda)$.

Let's construct the best asymptotically normal estimator (*b.a.n.e.*).

Put $\tilde{\theta}_n^{(1)} = (\hat{\alpha}, \hat{\sigma})$ and we will define $\mathbf{B}_n(\theta)$ and $\mathbf{a}_n(\theta)$ as in Zacks (1971, 5.5.5).

Then as $n \to \infty$,

$$\tilde{\theta}_n^{(2)} = \tilde{\theta}_n^{(1)} - \frac{1}{\sqrt{n}}\,\mathbf{B}_n^{-1}(\tilde{\theta}_n^{(1)})\,\mathbf{a}_n(\tilde{\theta}_n^{(1)}) \quad \text{is} \quad b.a.n.e.$$

References

[1] Rinne H. (2009). *The Weibull Distribution A Handbook*. CRC Press, N.Y. 764 p.

[2] Foldes A., Rejto L. (1981) Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist.* Vol. **9**, pp. 122-129.

[3] Kaplan E.L., Meier P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* Vol. **53**, pp. 457-481.

[4] Stute W. (1995). The statistical analysis of Kaplan-Meier integrals. *Lecture Notes.* Vol. **27**, pp. 231-254.

[5] Zacks S. (1971). *The Theory of Statistical Inference.* John Willey and Sons, N.Y.