

# Tests based on EDF statistics for randomly censored normal distributions when parameters are unknown

Namhyun Kim<sup>1,a</sup>

<sup>a</sup>Department of Science, Hongik University, Korea

---

## Abstract

Goodness-of-fit techniques are an important topic in statistical analysis. Censored data occur frequently in survival experiments; therefore, many studies are conducted when data are censored. In this paper we mainly consider test statistics based on the empirical distribution function (EDF) to test normal distributions with unknown location and scale parameters when data are randomly censored. The most famous EDF test statistic is the Kolmogorov-Smirnov; in addition, the quadratic statistics such as the Cramér-von Mises and the Anderson-Darling statistic are well known. The Cramér-von Mises statistic is generalized to randomly censored cases by Koziol and Green (*Biometrika*, **63**, 465–474, 1976). In this paper, we generalize the Anderson-Darling statistic to randomly censored data using the Kaplan-Meier estimator as it was done by Koziol and Green. A simulation study is conducted under a particular censorship model proposed by Koziol and Green. Through a simulation study, the generalized Anderson-Darling statistic shows the best power against almost all alternatives considered among the three EDF statistics we take into account.

**Keywords:** Anderson-Darling statistic, Cramér-von Mises statistic, goodness-of-fit tests, Kaplan-Meier estimator, Kolmogorov-Smirnov statistic, normal distribution, random censoring

---

## 1. Introduction

Goodness-of-fit techniques are an important problem in statistical analysis. In survival and reliability experiments, censored data occur frequently. As much as nonparametric methods are widely used, the parametric approach also plays an important role in survival analysis. Hence we need to take into account a testing problem for censored data as well as complete data. Usually the goodness-of-fit tests for complete data are adapted to censored cases according to censoring types. Regarding censoring types, the most common and simplest censoring schemes are type I or type II censoring. In this paper, we deal with random censoring. It is more general censoring type and occurs frequently in medical studies. For several censoring types and general theory of survival analysis, we refer Lee and Wang (2003), and Tableman and Kim (2004).

As for random censoring, Koziol and Green (1976), Koziol (1980), and Nair (1981) modified test statistics based on the empirical distribution function (EDF) or weighted empirical process. In Koziol and Green (1976), the Cramér-von Mises statistic is generalized to randomly censored data using the Kaplan-Meier product limit estimator of the distribution function. Chen (1984) studied a correlation statistic for randomly censored data.

In this paper, we study test statistics for normal distributions with unknown location and scale parameters when data are randomly censored. When lifetimes are assumed to follow a lognormal

---

<sup>1</sup> Department of Science, Hongik University, 94 Wausan-Ro, Mapo-Gu, Seoul 04066, Korea. E-mail: nhkim@hongik.ac.kr

distribution, the logarithm of the lifetimes should follow a normal distribution, and the inference for normal is needed. The lognormal distribution is useful when a hazard rate is increasing at first and then decreasing. We consider the test statistics for normality based on EDF such as the Kolmogorov-Smirnov statistic, the Koziol-Green statistic, and the Anderson-Darling statistic. A more general study about the EDF statistics is given in Stephens (1986), and several topics for testing normality are also studied in Thode (2002).

The Cramer-von Mises statistic and the Anderson-Darling statistic are the quadratic statistics based on EDF. The main difference between two statistics is the weight function in the quadratic form. As it is mentioned, the Koziol-Green statistic is a generalized version of the Cramer-von Mises statistic for randomly censored data. This paper is to generalize the Anderson-Darling statistic to randomly censored data. The Kaplan-Meier estimator is used as in Koziol and Green (1976). The newly defined statistic is applied to test normality with unknown parameters. Kim (2012, 2017) studied the EDF statistic for randomly censored exponential and Weibull distribution, respectively, when some parameters are unknown.

In Section 2, we provide test statistics based on EDF. In Section 3, simulation study and power comparisons are presented. An example is also provided. In Section 4, we mention some concluding remarks.

## 2. Goodness-of-fit test statistics

First, we summarize the goodness-of-fit test statistics based on EDF when data are a complete random sample. Let  $X_1, \dots, X_n$  be a random sample of size  $n$  with a continuous cumulative distribution function  $F$ . We consider the simple hypothesis

$$H_0 : F = F_0 \quad (2.1)$$

with a completely specified distribution function  $F_0$ . The EDF  $F_n(x)$  is defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x).$$

The EDF statistics for goodness-of-fit tests measure the difference between  $F_n(x)$  and  $F_0(x)$ . The most well known EDF statistic is the Kolmogorov-Smirnov statistic

$$D_n = \sup_{x \in \mathbf{R}} |F_n(x) - F_0(x)|. \quad (2.2)$$

A second well known and wide class of statistics are the quadratic statistics. The Cramér-von Mises statistic

$$W_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 dF_0(x) \quad (2.3)$$

and the Anderson-Darling statistic

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x) \quad (2.4)$$

are the most popular quadratic statistics.

We use the probability integral transformation,  $F_0(X_i)$ , to compute the above statistics. Under the null hypothesis in (2.1),  $F_0(X_i)$  follows the uniform distribution between 0 and 1,  $U(0, 1)$ , and we can assume  $F_0$  is the uniform distribution without loss of generality. Consequently comparing the EDF of the  $F_0(X_i)$  with the uniform distribution will take the same values when we calculate the statistics from the EDF of the  $X_i$ ,  $F_n(x)$ , compared with  $F_0$ . It also follows that the distributions of the above statistics do not depend on  $F_0$  and depend only on the sample size  $n$ .

Now let us consider censored data cases. Let  $Y_1^0, \dots, Y_n^0$  be lifetimes with a continuous distribution function  $F_Y$ , and  $C_1, \dots, C_n$  be random censoring times drawn independently of the  $Y_1^0, \dots, Y_n^0$  from a distribution function  $F_C$ . We assume the  $Y_i^0$  is censored on the right by  $C_i$ . Hence we observe  $n$  iid random pairs  $(Y_i, \delta_i)$ ,  $i = 1, \dots, n$  with

$$Y_i = \min(Y_i^0, C_i) \quad \text{and} \quad \delta_i = \begin{cases} 1, & \text{if } Y_i^0 \leq C_i, \\ 0, & \text{if } Y_i^0 > C_i. \end{cases}$$

We write  $(Y_{(i)}, \delta_{(i)})$  when the data  $Y_1, \dots, Y_n$  are ordered, where  $Y_{(1)} \leq \dots \leq Y_{(n)}$  are the ordered observations, and  $\delta_{(i)}$  is the indicator corresponding to  $Y_{(i)}$ . We want to test the null hypothesis

$$H_0 : F_Y = F_Y^0. \quad (2.5)$$

Since the censored data do not have the full knowledge of the EDF, we use the product limit estimator  $\hat{F}_n$

$$1 - \hat{F}_n(t) = \begin{cases} 1, & t < Y_{(1)}, \\ \prod_{Y_{(j)} \leq t} \left( \frac{n-j}{n-j+1} \right)^{\delta_{(j)}}, & t \leq Y_{(n)}, \\ 0, & t > Y_{(n)} \end{cases}$$

to estimate  $F_Y$ . The estimator is studied in Kaplan and Meier (1958), Efron (1967), Meier (1975), and Breslow and Crowley (1974), and we usually call it Kaplan-Meier estimator. It is usually written as

$$1 - \hat{p}_i = \prod_{j \leq i} \left( \frac{n-j}{n-j+1} \right)^{\delta_{(j)}}, \quad i = 1, \dots, n, \quad (2.6)$$

for simplicity. By Michael and Schucany (1986), the  $\hat{p}_i$  could be modified by

$$\hat{p}_{i,c} = 1 - \frac{n-c+1}{n-2c+1} \prod_{j \leq i} \left( \frac{n-j-c+1}{n-j-c+2} \right)^{\delta_{(j)}}, \quad 0 \leq c \leq 1, \quad i = 1, \dots, n,$$

and it reduces to  $(i-c)/(n-2c+1)$  for a complete sample. The popular value of  $c$  is  $c = 0$  or  $c = 0.5$ .

As in the complete sample, we still need to use the probability integral transformation to compute the statistics. Hence we define the product limit estimator  $\hat{G}_n$  for

$$Z_{(i)} = F_Y^0(Y_{(i)}) \quad (2.7)$$

as

$$\hat{G}_n(z) = \begin{cases} 0, & z < Z_{(1)}, \\ 1 - \prod_{Z_{(j)} \leq z} \left( \frac{n-j}{n-j+1} \right)^{\delta_{(j)}}, & z \leq Z_{(n)}, \\ 1, & z > Z_{(n)}. \end{cases} \quad (2.8)$$

Using (2.8) and (2.6), the Kolmogorov-Smirnov statistic based on the EDF in (2.2) could be generalized to randomly censored data by

$$DC_n = \sup_{0 < z < 1} |\hat{G}_n(z) - z| = \max(DC_n^+, DC_n^-), \quad (2.9)$$

where  $Z_{(n+1)} = 1$ ,  $\hat{p}_0 = 0$ ,  $\hat{p}_{n+1} = 1$ , and

$$DC_n^+ = \max_{1 \leq j \leq n+1, \delta_{(j)}=1} \{\hat{p}_j - Z_{(j)}\}, \quad DC_n^- = \max_{1 \leq j \leq n+1, \delta_{(j)}=1} \{\hat{Z}_{(j)} - \hat{p}_{j-1}\}$$

with  $\hat{p}_i$  in (2.6). Koziol (1980) proposed a similar statistic to  $DC_n$  based on the weighted empirical process.

Koziol and Green (1976) generalized the Cramér-von Mises statistic in (2.3) to

$$\psi_n^2 = n \int_0^1 (\hat{G}_n(z) - z)^2 dz \quad (2.10)$$

for randomly censored data. It measures the discrepancy between  $\hat{G}_n$  in (2.8) and  $U(0, 1)$ . By Koziol and Green (1976), the statistic in (2.10) can be computed as

$$\begin{aligned} \psi_n^2 &= n \sum_{j=1}^n (\hat{G}_n(Z_{(j)}))^2 (Z_{(j+1)} - Z_{(j)}) - n \sum_{j=1}^n \hat{G}_n(Z_{(j)}) (Z_{(j+1)}^2 - Z_{(j)}^2) + \frac{1}{3}n \\ &= n \sum_{j=1}^n \hat{G}_n(Z_{(j)}) (Z_{(j+1)} - Z_{(j)}) \left\{ \hat{G}_n(Z_{(j)}) - (Z_{(j+1)} + Z_{(j)}) \right\} + \frac{1}{3}n \end{aligned} \quad (2.11)$$

with  $Z_{(0)} = 1$ ,  $Z_{(n+1)} = 1$ .

In this paper, let us think about a generalization of the Anderson-Darling statistic to randomly censored data. In this case, the statistic in (2.4) becomes

$$AC_n^2 = n \int_0^1 \frac{(\hat{G}_n(z) - z)^2}{z(1-z)} dz,$$

for  $\hat{G}_n$  in (2.8). Since

$$\int_0^1 \frac{1}{z(1-z)} dz, \quad \int_0^1 \frac{1}{1-z} dz, \quad \int_0^1 \frac{z}{1-z} dz$$

are not integrable on  $[0, 1]$ , we define  $AC_n^2$  as the limit

$$AC_n^2 = \lim_{\epsilon, \delta \rightarrow 0} n \int_{\epsilon}^{1-\delta} \frac{(\hat{G}_n(z) - z)^2}{z(1-z)} dz,$$

for fixed  $0 < \epsilon < Z_{(1)}$ ,  $0 < \delta < 1 - Z_{(n)}$  as in a complete sample. For computing formula in a complete sample, see Stephens (1986) or Bagdonavičius *et al.* (2001). Now we think about the computational

form of the statistic  $AC_n^2$ .

$$\begin{aligned}
 AC_n^2 &= \lim_{\epsilon, \delta \rightarrow 0} n \left( \int_{\epsilon}^{Z_{(1)}} \frac{z}{1-z} dz + \sum_{j=1}^{n-1} \int_{Z_{(j)}}^{Z_{(j+1)}} \frac{(\hat{G}_n(z) - z)^2}{z(1-z)} dz + \int_{Z_{(n)}}^{1-\delta} \frac{1-z}{z} dz \right) \\
 &= \lim_{\epsilon, \delta \rightarrow 0} n \left( \int_{\epsilon}^{Z_{(1)}} \frac{z}{1-z} dz + \int_{Z_{(n)}}^{1-\delta} \frac{1-z}{z} dz \right. \\
 &\quad + \sum_{j=1}^{n-1} \left( \hat{G}_n(Z_{(j)}) \right)^2 \left( -\ln(1 - Z_{(j+1)}) + \ln Z_{(j+1)} + \ln(1 - Z_{(j)}) - \ln Z_{(j)} \right) \\
 &\quad \left. - 2 \sum_{j=1}^{n-1} \hat{G}_n(Z_{(j)}) \left( -\ln(1 - Z_{(j+1)}) + \ln(1 - Z_{(j)}) \right) + \int_{Z_{(1)}}^{Z_{(n)}} \frac{z}{1-z} dz \right).
 \end{aligned}$$

Since

$$\begin{aligned}
 \int_{\epsilon}^{Z_{(1)}} \frac{z}{1-z} dz &= -Z_{(1)} - \ln(1 - Z_{(1)}) + \epsilon + \ln(1 - \epsilon), \\
 \int_{Z_{(n)}}^{1-\delta} \frac{1-z}{z} dz &= \ln(1 - \delta) - (1 - \delta) - \ln Z_{(n)} + Z_{(n)}, \\
 \int_{Z_{(1)}}^{Z_{(n)}} \frac{z}{1-z} dz &= -Z_{(n)} - \ln(1 - Z_{(n)}) + Z_{(1)} + \ln(1 - Z_{(1)}),
 \end{aligned}$$

$AC_n^2$  becomes

$$\begin{aligned}
 AC_n^2 &= n \sum_{j=1}^{n-1} \left( \hat{G}_n(Z_{(j)}) \right)^2 \left( -\ln(1 - Z_{(j+1)}) + \ln Z_{(j+1)} + \ln(1 - Z_{(j)}) - \ln Z_{(j)} \right) \\
 &\quad - 2n \sum_{j=1}^{n-1} \hat{G}_n(Z_{(j)}) \left( -\ln(1 - Z_{(j+1)}) + \ln(1 - Z_{(j)}) \right) - n \ln(1 - Z_{(n)}) - n \ln Z_{(n)} - n. \quad (2.12)
 \end{aligned}$$

Let us consider the case  $F_Y^0 = \Phi$  in (2.5), and it includes some unknown parameters. The null hypothesis is

$$H_0 : F_Y(y) = \Phi\left(\frac{y - \mu}{\sigma}\right), \quad \text{for some } \mu, \sigma > 0.$$

When we assume the distribution of  $Y_i^0$  follows a normal,  $Y_i^0$  itself cannot be a lifetime. It means  $Y_i^0$  is the logarithm of the lifetime. For the composite null hypothesis, we need to estimate the unknown parameters  $\mu$  and  $\sigma$ . We usually use the maximum likelihood estimators (MLEs) of  $\mu$  and  $\sigma$ . Kim (2018) studied the estimation of the parameters for normal distributions under random censoring. In that paper, explicit forms of the approximate MLEs are provided by expanding nonlinear parts of the likelihood equations in Taylor series around some suitable points, which are closely related to the Kaplan-Meier estimators in (2.6). If we estimate  $\mu$  and  $\sigma$  by  $\hat{\mu}$  and  $\hat{\sigma}$ , we can take  $Z_{(i)}$  in (2.7) as  $\hat{Z}_{(i)}$  with the estimated parameters,

$$\hat{Z}_{(i)} = \Phi\left(\frac{Y_{(i)} - \hat{\mu}}{\hat{\sigma}}\right), \quad (2.13)$$

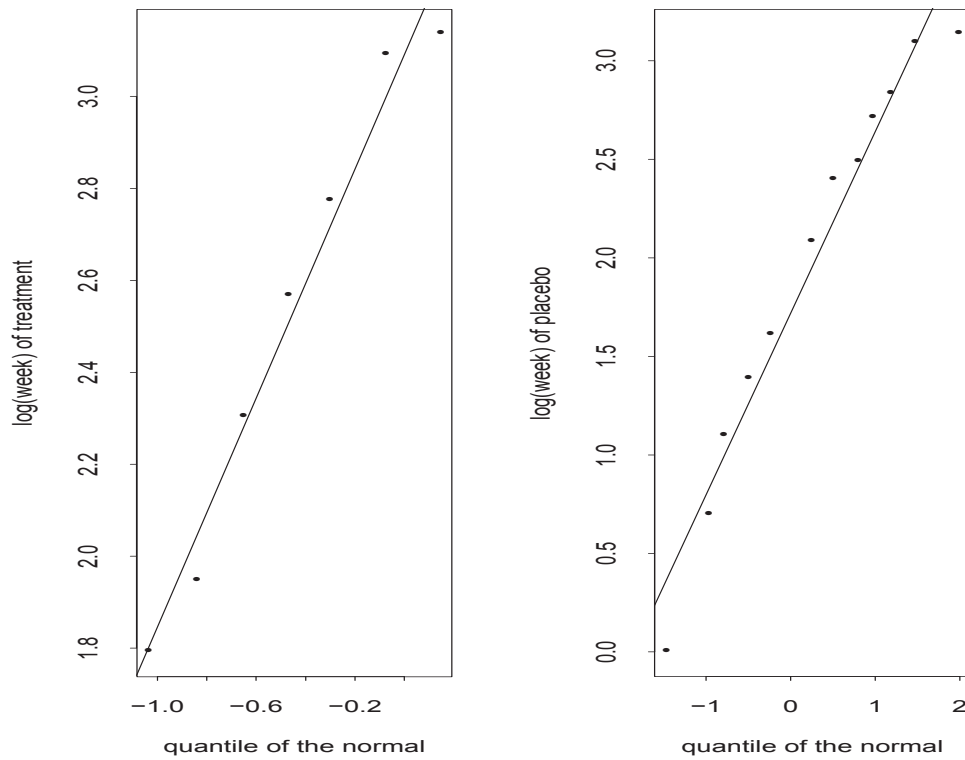


Figure 1: Q-Q plots of the  $\log(\text{remission times})$  for a normal distribution. The left plot is for the treatment group, and the right plot is for the placebo.

and consider  $\widehat{DC}_n$ ,  $\hat{\psi}_n^2$ ,  $\widehat{AC}_n^2$ , which have exactly the same form with the statistics  $DC_n$ ,  $\psi_n^2$ ,  $AC_n^2$  in (2.9), (2.11), and (2.12) replacing  $Z_{(i)}$  with  $\hat{Z}_{(i)}$  in (2.13).

### 3. Simulation and examples

#### 3.1. Examples

We consider the following set of remission times for two groups of acute leukemia patients. In this clinical trial, one group of 21 patients received a treatment called 6-mercaptopurine (6-MP), and the other group of 21 patients received a placebo. Each patient was randomized to receive a treatment or a placebo; the study ended after one year. The data set originally comes from Freireich *et al.* (1963) in a clinical trial. It is also discussed in Lee and Wang (2003) and Kleinbaum and Klein (2005). Here are the remission times, in weeks.

Treatment	6,	6,	6,	7,	10,	13,	16,	22,	23,					
	6+,	9+,	10+,	11+,	17+,	19+,	20+,	25+,	32+,	32+,	34+,	35+		
Placebo	1,	1,	2,	2,	3,	4,	4,	5,	5,					
	8,	8,	8,	8,	11,	11,	12,	12,	15,	17,	22,	23		

By looking at Figure 1, the straight lines fit quite closely to both groups. It indicates a normal distribution appears to give fairly good fit; however, the way of judging may be subjective. First, let

Table 1: Upper tail percentage points of the test statistic  $\sqrt{n}\widehat{DC}_n$  with  $r$  the ratio of censored data

$n$	$r$	$\alpha$						
		0.01	0.025	0.05	0.10	0.15	0.25	0.5
20	0.6	3.23	2.90	2.62	2.25	2.02	1.70	1.27
	0.5	2.54	2.24	1.95	1.67	1.50	1.28	0.98
	0.4	1.88	1.63	1.43	1.24	1.13	1.00	0.81
	0.2	1.13	1.04	0.96	0.88	0.83	0.76	0.64
30	0.6	3.44	3.06	2.73	2.37	2.13	1.79	1.35
	0.5	2.54	2.24	1.97	1.68	1.52	1.31	1.02
	0.4	1.83	1.62	1.43	1.24	1.15	1.02	0.83
	0.2	1.13	1.03	0.96	0.88	0.83	0.76	0.65
40	0.6	3.52	3.13	2.82	2.44	2.19	1.88	1.41
	0.5	2.59	2.24	1.97	1.71	1.54	1.34	1.05
	0.4	1.82	1.60	1.42	1.25	1.14	1.02	0.84
	0.2	1.11	1.03	0.95	0.88	0.82	0.75	0.65
50	0.6	3.58	3.24	2.89	2.51	2.27	1.95	1.47
	0.5	2.56	2.22	1.97	1.70	1.56	1.35	1.07
	0.4	1.84	1.60	1.42	1.24	1.14	1.02	0.84
	0.2	1.11	1.03	0.95	0.87	0.83	0.76	0.65
100	0.6	3.86	3.46	3.09	2.71	2.45	2.13	1.63
	0.5	2.59	2.28	2.01	1.78	1.62	1.42	1.13
	0.4	1.75	1.51	1.37	1.22	1.14	1.03	0.86
	0.2	1.11	1.02	0.96	0.88	0.83	0.77	0.66

us examine the treatment group. To test that a lognormal distribution fits the data, we need to estimate unknown parameters. The MLEs of the parameters are

$$\hat{\mu} = 3.2030, \quad \hat{\sigma} = 0.9787. \quad (3.1)$$

They are computed by the S-plus function `survReg`. When we compute the statistics  $\sqrt{n}\widehat{DC}_n$ ,  $\hat{\psi}_n^2$ , and  $\widehat{AC}_n^2$  in this paper, we have

$$\begin{aligned} \sqrt{n}\widehat{DC}_n &= 2.0410 \text{ (} p\text{-value} \approx 0.15), \\ \hat{\psi}_n^2 &= 0.6303 \text{ (} p\text{-value} \approx 0.15), \\ \widehat{AC}_n^2 &= 2.0394 \text{ (} p\text{-value} \approx 0.25). \end{aligned}$$

The  $p$ -values are found approximately by looking at Tables 1–3 when  $n = 20$ ,  $r = 0.6$ . By the results, we conclude that the lognormal distribution could fit the data. Using the parameter estimation in (3.1), the probability that the remission time  $T$  is longer than 10 weeks can be estimated as

$$P(T > 10) = P(\log T > 2.3) = 1 - \Phi\left(\frac{2.3 - \hat{\mu}}{\hat{\sigma}}\right) = 0.8219.$$

The value is almost the same when we use the Weibull distribution (Lee and Wang, 2003, Chapter 3).

As for the placebo group, the data are complete. Hence we can compute either  $\widehat{DC}_n$ ,  $\hat{\psi}_n^2$ ,  $\widehat{AC}_n^2$  or  $D_n$ ,  $W_n^2$ ,  $A_n^2$  in (2.2), (2.3), (2.4) by using the computational forms and  $\Phi((\log t_{(i)} - \hat{\mu})/\hat{\sigma})$ , where  $t_{(i)}$ 's are the order statistics of the remission time. The computational forms for a complete sample can be

Table 2: Upper tail percentage points of the test statistic  $\hat{\psi}_n^2$  with  $r$  the ratio of censored data

$n$	$r$	$\alpha$						
		0.01	0.025	0.05	0.10	0.15	0.25	0.5
20	0.6	2.52	1.85	1.37	0.90	0.68	0.46	0.25
	0.5	1.25	0.88	0.61	0.42	0.34	0.25	0.15
	0.4	0.56	0.39	0.31	0.23	0.19	0.15	0.10
	0.2	0.22	0.18	0.15	0.13	0.11	0.09	0.06
	0.0	0.17	0.14	0.12	0.10	0.09	0.07	0.05
30	0.6	2.51	1.78	1.31	0.88	0.68	0.46	0.25
	0.5	1.05	0.75	0.55	0.39	0.31	0.24	0.15
	0.4	0.45	0.35	0.28	0.22	0.19	0.15	0.10
	0.2	0.22	0.18	0.15	0.12	0.11	0.09	0.06
	0.0	0.17	0.14	0.12	0.10	0.09	0.07	0.05
40	0.6	2.32	1.68	1.24	0.85	0.66	0.47	0.26
	0.5	0.99	0.67	0.51	0.37	0.31	0.23	0.15
	0.4	0.43	0.33	0.27	0.21	0.18	0.14	0.10
	0.2	0.21	0.17	0.15	0.12	0.11	0.09	0.06
	0.0	0.17	0.14	0.12	0.10	0.09	0.07	0.05
50	0.6	2.19	1.66	1.22	0.84	0.67	0.48	0.27
	0.5	0.88	0.62	0.47	0.35	0.29	0.23	0.15
	0.4	0.41	0.31	0.26	0.20	0.17	0.14	0.10
	0.2	0.21	0.17	0.15	0.12	0.11	0.09	0.06
	0.0	0.17	0.14	0.12	0.10	0.09	0.07	0.05
100	0.6	2.00	1.48	1.09	0.79	0.64	0.48	0.29
	0.5	0.70	0.53	0.41	0.33	0.28	0.22	0.15
	0.4	0.33	0.27	0.23	0.18	0.16	0.13	0.09
	0.2	0.20	0.17	0.15	0.12	0.10	0.09	0.06
	0.0	0.17	0.15	0.13	0.10	0.09	0.07	0.05

found in Stephens (1986), for example. The value of each statistic is as follows.

Kolmogorov-Smirnov = 0.1797 ( $p$ -value  $\approx 0.0751$ ),

Cramer-von Mises = 0.0617 ( $0.25 < p$ -value  $< 0.5$ ),

Anderson-Darling = 0.5064 ( $p$ -value  $\approx 0.15$ ).

The  $p$ -value of the Kolmogorov-Smirnov statistic is computed by the S-plus function `ks.gof`. As for the other statistics, Table 2, Table 3 are used for  $n = 20$ ,  $r = 0$ . The Kolmogorov-Smirnov statistic gives a small  $p$ -value; however, the  $p$ -values of the other statistics support that the data fit the lognormal distribution.

### 3.2. Simulation results

A simulation study is conducted to give the null distributions of the test statistics  $\widehat{DC}_n$ ,  $\hat{\psi}_n^2$ , and  $\widehat{AC}_n^2$  described in Section 2. Here we estimate the unknown parameters  $\mu$  and  $\sigma$  by the MLEs. The S-plus function `survReg` is used to compute them. The power of the statistics is also compared through a simulation. The upper percentage points of the test statistics are given in Table 1 to Table 3 for sample sizes  $n = 20, 30, 40, 50, 100$ , censored ratio  $r = 0.2, 0.4, 0.5, 0.6$ , and the significance level  $\alpha = 0.01, 0.025, 0.05, 0.10, 0.15, 0.25, 0.50$ .  $N = 10,000$  runs have been done to have null distributions.

We use the random censorship model proposed in Koziol and Green (1976) to control the censored ratio. It is

$$1 - F_C = (1 - F_Y)^\beta, \quad \text{for some } \beta > 0,$$

where  $F_C$  is the distribution function of the censoring time  $C_i$ , and  $\beta$  is called a censoring parameter.



Table 3: Upper tail percentage points of the test statistic  $\widehat{AC}_n^2$  with  $r$  the ratio of censored data

$n$	$r$	$\alpha$						
		0.01	0.025	0.05	0.10	0.15	0.25	0.5
20	0.6	7.05	5.09	3.96	3.11	2.65	2.07	1.32
	0.5	3.79	2.97	2.42	1.94	1.67	1.33	0.90
	0.4	2.47	1.94	1.63	1.34	1.16	0.95	0.67
	0.2	1.40	1.14	0.97	0.82	0.73	0.61	0.44
	0.0	0.99	0.83	0.71	0.60	0.53	0.45	0.32
30	0.6	6.75	5.42	4.34	3.33	2.84	2.27	1.48
	0.5	3.90	3.15	2.55	2.01	1.75	1.41	0.95
	0.4	2.54	2.04	1.68	1.37	1.20	0.99	0.71
	0.2	1.41	1.16	0.99	0.81	0.72	0.61	0.44
40	0.6	6.82	5.43	4.30	3.48	3.01	2.40	1.60
	0.5	4.03	3.09	2.61	2.11	1.83	1.50	1.02
	0.4	2.68	2.10	1.73	1.41	1.23	1.00	0.71
	0.2	1.36	1.13	0.97	0.81	0.71	0.60	0.43
50	0.6	7.01	5.46	4.48	3.66	3.19	2.58	1.71
	0.5	3.87	3.18	2.66	2.12	1.85	1.50	1.04
	0.4	2.68	2.09	1.75	1.40	1.23	1.02	0.73
	0.2	1.38	1.15	0.97	0.81	0.72	0.61	0.44
100	0.6	7.78	6.22	5.20	4.32	3.76	3.05	2.09
	0.5	4.64	3.61	2.95	2.36	2.07	1.70	1.18
	0.4	2.89	2.18	1.76	1.46	1.27	1.05	0.76
	0.2	1.34	1.12	0.97	0.82	0.72	0.61	0.45
	0.0	1.00	0.86	0.74	0.62	0.55	0.46	0.33

We have

$$P(Y_i^0 > C_i) = \int_{-\infty}^{\infty} (1 - F_Y(y)) dF_C(y) = \int_0^1 \beta(1-x)^\beta dx = \frac{\beta}{\beta+1}$$

under this model. It is the expected proportion of the censored observations. In Table 1 to Table 3,  $r$  is the expected ratio of the censored data, and it is equal to  $\beta/(\beta+1)$ . Csörgő and Horváth (1981), Chen *et al.* (1982), and Kim (2011, 2012) mentioned the motivation or characterization of this model.

As we explained in Section 1, the EDF statistics we introduced are distribution free under the simple null hypothesis, since we can use the probability integral transformation. However they depend on the distribution tested when unknown parameters are estimated (Stephens, 1986). If we investigate the null distribution of  $\sqrt{n}\widehat{DC}_n$  and  $\hat{\psi}_n^2$  in Table 2 and Table 3, the values are very similar to the upper tail percentage points for randomly censored Weibull distributions in Kim (2017), especially when the censoring ratio  $r$  is small. Apparently the null distributions of the statistics do not change significantly by the tested distribution and the unknown parameter estimation.

In Table 2 and Table 3,  $r = 0$  for  $n = 20$ ,  $n = 100$  means no censoring, i.e., complete data. The numbers are from Stephens (1986, Table 4.10) for comparison. He presented the upper tail percentage points for type II censoring and complete data.

Next, we examine the power of the test statistics. Table 4 and Table 5 provide the power of the statistics at the significance level  $\alpha = 0.10$  for sample sizes  $n = 50, 100$ .  $N = 5,000$  samples are generated for each alternative. We take into account the following alternatives.

- exponential distribution with pdf  $f(t) = e^{-t}$ ,  $t > 0$ .
- gamma distribution,  $\text{Gamma}(\alpha)$ ,  $\alpha = 0.5, 2$ , with pdf  $f(t; \alpha) = t^{\alpha-1} e^{-t} / \Gamma(\alpha)$ ,  $t > 0$ .

Table 4: Power comparison of  $\widehat{DC}_n$ ,  $\hat{\psi}_n^2$ , and  $\widehat{AC}_n^2$  for  $\alpha = 0.10$  and  $n = 50$ 

Distribution	Censoring ratio ( $r$ )	$\widehat{DC}_n$	$\hat{\psi}_n^2$	$\widehat{AC}_n^2$
log-normal	0.6	0.10	0.10	0.10
	0.5	0.11	0.11	0.10
	0.4	0.11	0.10	0.11
	0.2	0.11	0.10	0.10
exponential	0.6	0.15	0.16	0.37
	0.5	0.25	0.30	0.44
	0.4	0.31	0.38	0.47
	0.2	0.47	0.55	0.59
Gamma(0.5)	0.6	0.18	0.21	0.51
	0.5	0.35	0.43	0.64
	0.4	0.50	0.60	0.71
	0.2	0.71	0.78	0.82
Gamma(2)	0.6	0.12	0.13	0.24
	0.5	0.17	0.19	0.28
	0.4	0.20	0.25	0.30
	0.2	0.29	0.34	0.36
Weibull(0.5)	0.6	0.16	0.17	0.37
	0.5	0.24	0.29	0.42
	0.4	0.32	0.40	0.49
	0.2	0.45	0.54	0.59
Weibull(2)	0.6	0.16	0.17	0.36
	0.5	0.23	0.29	0.43
	0.4	0.32	0.39	0.49
	0.2	0.47	0.55	0.59
log-logistic	0.6	0.11	0.11	0.16
	0.5	0.13	0.14	0.16
	0.4	0.13	0.15	0.17
	0.2	0.16	0.19	0.21
log-DE	0.6	0.14	0.16	0.32
	0.5	0.19	0.28	0.34
	0.4	0.23	0.37	0.37
	0.2	0.44	0.52	0.51
half-logistic	0.6	0.17	0.20	0.47
	0.5	0.30	0.38	0.55
	0.4	0.44	0.54	0.64
	0.2	0.59	0.69	0.72

- Weibull distribution, Weibull( $\alpha$ ),  $\alpha = 0.5, 2$  with pdf  $f(t; \alpha) = \alpha t^{\alpha-1} e^{-t^\alpha}$ ,  $t > 0$ .
- log-logistic distribution with pdf  $f(t) = 1/(1+t)^2$ ,  $t > 0$ .
- log double exponential, log-DE, DE with pdf  $f(t) = e^{-|t|}/2$ .
- half-logistic distribution with  $f(t) = 2e^{-t}/(1+e^{-t})^2$ ,  $t > 0$ .

Note that the  $Y_{(i)}$  in (2.13) are the logarithm of the lifetime. Therefore, we should take the logarithm of the simulated data for each alternative before we compute the test statistics  $\widehat{DC}_n$ ,  $\hat{\psi}_n^2$ , and  $\widehat{AC}_n^2$ . The distribution of the first row of Table 4 and Table 5 is written as the lognormal for comparison. We see the following from the power results. First of all,  $\widehat{AC}_n^2$  shows the best power,  $\hat{\psi}_n^2$ , the second, and  $\widehat{DC}_n$  the lowest for each alternative considered except the log-DE. Even in that case,  $\widehat{AC}_n^2$  has comparable power to  $\hat{\psi}_n^2$ , and  $\widehat{DC}_n$  still has the lowest. According to D'Agostino (1986),

Table 5: Power comparison of  $\widehat{DC}_n$ ,  $\hat{\psi}_n^2$ , and  $\widehat{AC}_n^2$  for  $\alpha = 0.10$  and  $n = 100$ 

Distribution	Censoring ratio ( $r$ )	$\widehat{DC}_n$	$\hat{\psi}_n^2$	$\widehat{AC}_n^2$
log-normal	0.6	0.10	0.10	0.10
	0.5	0.10	0.09	0.10
	0.4	0.11	0.10	0.11
	0.2	0.10	0.10	0.09
exponential	0.6	0.23	0.31	0.57
	0.5	0.42	0.55	0.68
	0.4	0.62	0.71	0.78
	0.2	0.74	0.82	0.87
Gamma(0.5)	0.6	0.36	0.49	0.81
	0.5	0.65	0.79	0.89
	0.4	0.87	0.92	0.95
	0.2	0.95	0.97	0.99
Gamma(2)	0.6	0.16	0.20	0.37
	0.5	0.24	0.33	0.43
	0.4	0.38	0.46	0.51
	0.2	0.45	0.55	0.61
Weibull(0.5)	0.6	0.24	0.32	0.59
	0.5	0.42	0.56	0.69
	0.4	0.61	0.70	0.76
	0.2	0.73	0.81	0.87
Weibull(2)	0.6	0.23	0.31	0.60
	0.5	0.42	0.55	0.69
	0.4	0.61	0.71	0.76
	0.2	0.73	0.82	0.87
log-logistic	0.6	0.12	0.14	0.19
	0.5	0.12	0.16	0.18
	0.4	0.15	0.20	0.20
	0.2	0.20	0.24	0.27
log-DE	0.6	0.18	0.31	0.45
	0.5	0.25	0.50	0.48
	0.4	0.41	0.65	0.58
	0.2	0.69	0.77	0.76
half-logistic	0.6	0.31	0.43	0.73
	0.5	0.56	0.72	0.84
	0.4	0.78	0.86	0.90
	0.2	0.86	0.92	0.95

the Anderson-Darling statistic is still the most powerful EDF test for normal when data are complete. Second, the difference in power between  $\widehat{AC}_n^2$  and  $\hat{\psi}_n^2$  becomes large for the big values of  $r$ . Third, every statistic has relatively low power for the log-logistic alternative. In this alternative, the statistics have just a negligible increase of the power even if the sample size becomes bigger or the censoring ratio is smaller.

#### 4. Concluding remarks

In this paper, we have studied goodness-of-fit test statistics for normal distributions with an unknown location and scale parameter when data are randomly censored. In this case the distributional assumption for lifetime itself is lognormal distributions. We take into account test statistics based on EDF statistics such as the Kolmogorov-Smirnov statistic, the Koziol-Green statistic, and the Anderson-Darling statistic. We have generalized the Anderson-Darling statistic to randomly censored data, and found a computational form. We have used the Kaplan-Meier product limit as it was done in Koziol

and Green (1976).

Based on the simulation studies, the generalized Anderson-Darling statistic has shown the best power among EDF statistics that we have considered under almost all alternatives. The power results are consistent with complete sample cases in normal distributions as it is mentioned in D'Agostino (1986).

## Acknowledgements

This work was supported by a 2019 Hongik University Research Fund.

## References

- Bagdonavičius V, Kruopis J, and Nikulin MS (2011). *Non-Parametric Tests for Complete Data*, John Wiley & Sons, New Jersey.
- Breslow N and Crowley J (1974). A large sample study of the life table and product limit estimates under random censorships, *The Annals of Statistics*, **2**, 437–453.
- Chen C (1984). A correlation goodness-of-fit test for randomly censored data, *Biometrika*, **71**, 315–322.
- Chen YY, Hollander M, and Langberg NA (1982). Small-sample results for the Kaplan Meier estimator, *Journal of the American Statistical Association*, **77**, 141–144.
- Csörgő S and Horváth L (1981). On the Koziol-Green Model for random censorship, *Biometrika*, **68**, 391–401.
- D'Agostino RB (1986). Test for the normal distribution. In D'Agostino RB and Stephens MA (Eds), *Goodness-of-Fit Techniques* (Chapter 9), Marcel Dekker, New York.
- Efron B (1967). The two sample problem with censored data. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 831–853.
- Freireich EJ, Gehan EA, Frei E, *et al.* (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: a model for evaluation of other potential useful therapy, *Blood*, **21**, 699–716.
- Kaplan EL and Meier P (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457–481.
- Kim N (2011). Testing log normality for randomly censored data, *The Korean Journal of Applied Statistics*, **24**, 883–891.
- Kim N (2012). Testing exponentiality based on EDF statistics for randomly censored data when the scale parameter is unknown, *The Korean Journal of Applied Statistics*, **25**, 311–319.
- Kim N (2017). Goodness-of-fit tests for randomly censored Weibull distributions with estimated parameters, *Communications for Statistical Applications and Methods*, **24**, 519–531.
- Kim N (2018). On the maximum likelihood estimation for a normal distribution under random censoring, *Communications for Statistical Applications and Methods*, **25**, 647–658.
- Kleinbaum DG and Klein M (2005). *Survival Analysis: A Self-Learning Test*, Springer, New York.
- Koziol JA (1980). Goodness-of-fit tests for randomly censored data, *Biometrika*, **67**, 693–696.
- Koziol JA and Green SB (1976). A Cramér-von Mises statistic for randomly censored data, *Biometrika*, **63**, 465–474.
- Lee ET and Wang JW (2003). *Statistical Methods for Survival Data Analysis*, John Wiley & Sons, New Jersey.
- Meier P (1975). Estimation of a distribution function from incomplete observations. In J. Gani (Ed), *Perspectives in Probability and Statistics*, Academic Press, London.

- Michael JR and Schucany WR (1986). Analysis of data from censored samples. In D'Agostino RB and Stephens MA (Eds), *Goodness of Fit Techniques* (Chapter 11), Marcel Dekker, New York.
- Nair VN (1981). Plots and tests for goodness of fit with randomly censored data, *Biometrika*, **68**, 99–103.
- Stephens MA (1986). Tests based on EDF statistics. In D'Agostino RB and Stephens MA (Eds), *Goodness-of-Fit Techniques* (Chapter 4), Marcel Dekker, New York.
- Tableman M and Kim JS (2004). *Survival Analysis using S: Analysis of Time-to-Event Data*, Chapman & Hall CRC, Florida.
- Thode HC (2002). *Testing for Normality*, Marcel Dekker, New York.

*Received February 3, 2019; Revised July 23, 2019; Accepted August 26, 2019*