The Lorenz Curve for Model Assessment in Exponential Order Statistic Models

Jason A. Osborne[a] and Thomas A. Severini[b]

[a]Department of Discrete and Statistical Sciences, Auburn University
120 Mathematics Annex, Auburn, IL 36830
[b]Department of Statistics, Northwestern University
2006 Sheridan Road, Evanston, IL 60208

A goodness-of-fit technique for random samples from the exponential distribution based on the sample Lorenz curve is adapted for use in the exponential order statistic (EOS) model. In the EOS model, only those observations in a random sample from the exponential distribution of unknown size N that are less than some known stopping time T are observable. The model is known as the Jelinski-Moranda model in software reliability, where it is used to estimate the number of bugs in software during development. Distributional results are derived for the distance between the sample Lorenz curve and the population Lorenz curve so that it can be used as a goodness-of-fit test statistic. Simulations show that the test has good power against several alternative distributions. Simulations also indicate that in some cases, model misspecification leads to poor parameter estimation. A plotting procedure provides a means of graphical assessment of fit.

Keywords: Conditional probability integral transform, goodness-of-fit, Kolmogorov-Smirnov distance, Lorenz curve, software reliability

1 Introduction

Consider the problem of estimating the unknown size N of a population based on a search which detects individuals at times $Y_1, \ldots, Y_N$ which are modeled as independent and identically distributed random variables. Suppose that the search ends at a known stopping time $T > 0$ so that Y_j is observable only if $Y_j \leq T$. Raftery (1987) and Miller (1986) refer to these as general order statistic models. The exponential order statistic (EOS) model takes the common probability distribution to be exponential with unknown rate ϕ. It was proposed in Jelinski and Moranda (1972) as a model for the process of fault detections during the debugging stages of software development. It often serves as a point of reference for some of the many other models developed in this area. For an overview of statistical inference for software reliability, see Singpurwalla and Wilson (1999). Agustin and Peña (1999) and Agustin (1999) consider several systems of components each of which fails according to the EOS model. For an application of the EOS model in estimating a population size for market research, see Basu et al. (1995).

Other general order statistics models used in software reliability and other fields include the Pareto order statistic (POS), (see Littlewood (1981)) Weibull order statistic (WOS), and gamma order statistic (GOS) (see Yamada et al. (1983)) models. Many goodness-of-fit techniques have been developed for random samples from these distributions, particularly the exponential distribution, but fewer options are available for general order statistic models, where N is unknown. A test for the exponential distribution based on the sample Lorenz curve was shown in Gail and Gastwirth (1978) to be powerful against a variety of alternatives, robust to measurement error and to be free of the unknown scale parameter. We adapt this procedure based on the Lorenz curve to the truncated exponential distribution and use it to assess the goodness-of-fit of the EOS model. A similar procedure based on the total time on test plot was developed in Klefsjo and Kumar (1992) to assess the fit of a power-law Poisson process model.

Section 2 reviews estimation and existing methods of goodness-of-fit for the EOS model. Section 3 develops a plotting procedure and test statistic based on the sample Lorenz curve. Section 4 presents the results of a simulation study investigating the power of these goodness-of-fit techniques under several alternatives to the EOS model. Section 5 illustrates the methodology using a dataset from software reliability.

2 Existing techniques for goodness-of-fit and estimation

Let the number of detections Y_j occurring before the stopping time T be denoted by R. Let these ordered detection times be denoted by $Y_{(1)} \leq \cdots \leq Y_{(R)} \leq T$ and the times between them by $T_i = Y_{(i)} - Y_{(i-1)}$. Under the EOS model, the interdetection times $T_i,\ i = 1, \ldots, R$ are independent and exponentially distributed with rates $\phi(N - i + 1)$, respectively. Estimated probability transforms for T_i are given by

$$u_i = F(t_i; \hat{N}, \hat{\phi}) = 1 - e^{-\tilde{\phi}(\tilde{N}-i+1)t_i} \quad i = 1, \ldots, r$$

where $(\tilde{N}, \tilde{\phi})$ denotes a parameter estimate based on the observed data. A Kolmogorov-Smirnov (KS) distance can be computed from the u-plot (Musa et al. (1987)): a plot of the empirical distribution function of the estimated probability transforms. Equivalently, this procedure is a test of whether or not the interdetection time residuals defined by

$$e_i = \tilde{\phi}(\tilde{N} - i + 1)t_i \quad i = 1, \ldots, r$$

are a random sample from the standard exponential distribution. Many goodness-of-fit procedures have been developed for exponential random samples (see, e.g. D'Agostino and Stephens (1986).) These can be applied to the interdetection time residuals, but not without some correction for dependence among the residuals caused by estimation of model parameters from data and as in Lilliefors (1969). Lee and Finelli (1989) use conditional probability integral transformations (CPIT) of the interfailure times to develop a powerful test statistic for the EOS model. For similar techniques based on interdetection time residuals applied to the powerlaw Poisson processes, see Gaudoin (1998).

The parameter estimates necessary for the u-plot or other procedures based on the interdetection time residuals can be obtained using maximum likelihood. The likelihood function for the EOS model takes a simple form

$$L(N, \phi) = \frac{N!}{(N-r)!}\phi^r e^{-\phi T(x+N-r)}$$

where $x = \sum_1^r y_{(j)}/T$. The maximum likelihood estimators (MLE) have been shown by many authors (see Joe and Reid (1985) and Littlewood and Verrall (1981)) to have unusual properties in small samples, including point mass at infinity: $\Pr(\hat{N} = \infty) > 0$ and $\Pr(\hat{\phi} = 0) > 0$. This issue also arises in the problem of estimating the index parameter n in binomial random samples where both n and p are unknown (see, Casella (1986).) An alternative method can be based on an integrated likelihood function (see Berger et al. (1999)). The likelihood function under the parameterization which takes δ as the individual detection probability $\delta = 1 - \exp\{-\phi T\}$ is

$$L(N, \delta) = \frac{N!}{(N-r)!}(-\log(1-\delta))^r(1-\delta)^{x+N-r}.$$

The likelihood function for N obtained by integrating $L(N, \delta)$ over a uniform prior for δ can be used for inference about N:

$$L(N) = \int_0^1 L(N, \delta) d\delta = \frac{N!}{(N-r)!}(x+N-r+1)^{-(r+1)}.$$

The integrated likelihood estimator (ILE) $\hat{N}$ which maximizes this function performs as well or better than the MLE and is always finite (Osborne and Severini (2000)).

The value of δ which maximizes $L(N, \delta)$ for N fixed at $\hat{N}$ provides an estimator of δ which performs as well as the MLE but is never degenerate at 0 or 1: $\hat{\delta} = 1 - \exp\{-r/(x + \hat{N} - r)\}$.

3 The sample Lorenz curve for goodness-of-fit

Let $Y_1 \leq Y_2 \leq \cdots Y_r$ denote the ordered observations from a random sample of size r from a non-negative population with continuous probability distribution function $F(y)$, inverse distribution function $G(p)$, mean μ and variance σ^2. The sample Lorenz curve $L_r(\cdot)$ is defined by

$$L_r(p) = \frac{\sum_{j=1}^{[rp]} Y_j}{\sum_{j=1}^{r} Y_j} \quad \text{for } 0 < p < 1$$

where $[np]$ is the largest integer less than or equal to np. The population Lorenz Curve, denoted by $\lambda(p)$ corresponding to $F(y)$ is defined by

$$\lambda(p) = \mu^{-1} \int_0^{G(p)} y dF(y)$$

where $0 < p < 1$. It has been shown independently by both Gail and Gastwirth (1978) and by Goldie (1977) that

$$L_r(p) \xrightarrow{a.s.} \lambda(p)$$

uniformly in p. The former authors show that at any point $p \in (0, 1)$, the numerator and denominator of the Lorenz curve are asymptotically bivariate normal as $r \to \infty$:

$$\sqrt{r} \begin{pmatrix} r^{-1} \sum_{j=1}^{[rp]} Y_j - \mu\lambda(p) \\ \bar{Y} - \mu \end{pmatrix} \xrightarrow{\mathcal{L}} N_2(0, \Sigma(p)),$$

where the components of $\Sigma(p)$ are given by

$$\begin{aligned} \sigma_{11}(p) &= 2 \int_0^{G(p)} \left(\int_0^{G(y)} F(x) dx \right) (1 - F(y)) dy \\ \sigma_{12}(p) &= \sigma_{11}(p) + \int_0^{G(p)} F(x) dx \int_{G(p)}^{\infty} (1 - F(y)) dy \\ \sigma_{22} &= \sigma^2. \end{aligned}$$

It follows that the approximate variance of the sample Lorenz curve at the point p is given by

$$\sigma_L^2(p) = \sigma_{11}(p)/\mu^2 + \sigma_{22}\lambda(p)^2/\mu^2 - 2\sigma_{12}(p)\lambda(p)/\mu^3$$

and a test statistic defined by

$$Z = \frac{\sqrt{r}(L_r(p) - \lambda(p))}{\sigma_L(p)}$$

converges in distribution to a standard normal variate as $r \to \infty$.

When $F(y) = 1 - e^{-\phi y}$ and $q = 1 - p$, the mean $\lambda(p) = p + q\log(q)$ and variance $\sigma_L^2(p)$ of $L_r(p)$ do not depend on the unknown scale parameter ϕ, which makes this statistic ideal for testing the goodness-of-fit of the exponential distribution in random samples. Simulations and theory in Gail and Gastwirth (1978) indicate that evaluation of the test statistic at $p = 0.5$ has good power against, among others, Weibull, gamma and pareto alternatives as well as robustness to measurement error.

3.1 Moments of the Sample Lorenz Curve for the EOS model

The number of detections R has the binomial distribution with parameters N and δ. Given $R = r$, the conditional probability density for the observed detection times is

$$\begin{aligned} p(y_1, \ldots, y_R | R = r; N, \phi) &= \frac{\frac{N!}{(N-r)!}\phi^r e^{-\phi T(x+N-r)}}{\frac{N!}{(N-r)!r!}(1 - e^{-\phi T})^r e^{-\phi T(N-r)}} \\ &= r! \prod_{j=1}^{r} \frac{\phi e^{-\phi y_j}}{1 - e^{-\phi T}}. \end{aligned}$$

This is the joint density of the order statistics from a random sample from the truncated exponential distribution. Under the δ parameterization, the truncated exponential distribution function is

$$F(y; \phi, T) = \frac{1 - e^{-\phi y}}{1 - e^{-\phi T}} = (1 - (1 - \delta)^{y/T})\delta^{-1} \quad (0 < y < T).$$

Straightforward calculations yield asymptotic moments of the sample Lorenz curve for truncated exponential data and the following approximation (the dependence of λ, μ and σ_{ij} on δ and p is supressed in the notation):

$$\sqrt{r}(L_r(p) - \lambda) \xrightarrow{\mathcal{L}} N(0, \sigma_L^2)$$

with

$$\begin{aligned}
\lambda &= \frac{\delta p + (1-\delta p)(\,\log\,(1-\delta p))}{\delta + (1-\delta)(\,\log\,(1-\delta))} \\
\sigma_L^2 &= \sigma_{11}/\mu^2 + \sigma_{22}\lambda^2/\mu^2 - 2\sigma_{12}\lambda/\mu^3. \\
\mu &= T\{\frac{1}{-\log(1-\delta)} - \frac{1-\delta}{\delta}\} \\
\sigma_{11} &= \{(\delta-1)(\log(1-\delta p))^2 + 2\delta(p-1)(-\log(1-\delta p)) + 2\delta^2 p - \delta^2 p^2\} \\
&\quad \times \frac{T^2}{(\delta\log(1-\delta))^2} \\
\sigma_{22} &= T^2\{\frac{1}{(-\log(1-\delta))^2} - \frac{1-\delta}{\delta^2}\} \\
\sigma_{12} &= \sigma_{11} + \\
&\quad \left(\frac{-\log(1-\delta p)}{\delta} - p\right)\left((1-p) + \frac{1-\delta}{\delta}\log\frac{1-\delta}{1-\delta p}\right)\left(\frac{T}{-\log(1-\delta)}\right)^2.
\end{aligned}$$

This result can be used to formulate a test and plotting procedure for goodness-of-fit for the EOS model based on the sample Lorenz Curve. Any point $p \in (0,1)$ can be chosen and the mean and standard error of $L_r(p)$ can be estimated by substitution of the estimator $\hat{\delta}$: $\hat{\lambda}(p) = \lambda(p, \hat{\delta})$ and $\hat{\sigma}_L^2 = \sigma_L^2(p, \hat{\delta})$. Under the hypothesis that the conditional distribution of the data has been correctly specified as truncated exponential, the test statistic

$$Z = \sqrt{r}(L_r(0.5) - \hat{\lambda}(0.5))\hat{\sigma}_L^{-1}$$

has an approximately standard normal distribution for large r.

Table 1: Alternative Order Statistic Models

Model	Density function	Shape	$\delta = 0.5$ Scale	$\delta = 0.9$ Scale
EOS	$\phi e^{\phi y}$	-	$\phi = 0.69$	$\phi = 2.30$
WOS	$\beta\phi(y\phi)^{\beta-1}e^{(\phi y)^\beta}$	$\beta = 0.5$	$\phi = 0.48$	$\phi = 5.30$
		$\beta = 2.0$	$\phi = 0.83$	$\phi = 1.52$
GOS	$\frac{\phi^\beta y^{\beta-1}}{\Gamma(\beta)}e^{-\phi y}$	$\beta = 0.5$	$\phi = 0.23$	$\phi = 1.35$
		$\beta = 2.0$	$\phi = 1.68$	$\phi = 3.89$
POS	$\beta\phi(1+\phi y)^{-(\beta+1)}$	$\beta = 0.5$	$\phi = 3$	$\phi = 99$
		$\beta = 1.0$	$\phi = 1$	$\phi = 9$

4 Simulation

Two simulation studies are undertaken to investigate the performance of several goodness-of-fit test statistics under the EOS model and the WOS, POS and GOS alternative models. The first study takes fixed values of $N = 50, 100, 200$, a stopping time of $T = 1$ and scale parameters chosen to yield individual detection probabilities of $\delta = \Pr(Y_j \leq 1) = 0.5$ and $\delta = 0.9$. For each alternative, two values for the shape parameter are considered. The choices for the shape of the GOS models are the same as those in Lee and Finelli (1989), while those for the WOS model are similar to those in Gail and Gastwirth (1978). For each of the seven models, as summarized in Table 1, and each of the four choices for (N, δ), 500 datasets are simulated. Four test statistics are computed for each setting of the simulation: the sample Lorenz curve test statistic for truncated exponential data of section 3.1, (LC1), and three statistics based on the interdetection time residuals $\{e_i\}$. These are the KS distance from the u-plot with parameters estimated from data (KS1), the KS distance computed using the fixed values of (N, δ) (KS2) and the sample Lorenz curve for complete exponential data (LC2). Using level $\alpha = 0.05$ critical regions for these test statistics, the power to reject the EOS model was computed and appears in Tables 2-4. The Monte Carlo

Table 2: Power Estimates under WOS Alternatives, fixed (N, δ)

Shape	N	δ	LC1	LC2	KS1	KS2	$\hat{N}$	$\hat{\delta}$
0.5	50	0.5	0.47	0.19	0.04	0.38	26.6	0.94
		0.9	0.98	0.53	0.20	0.92	45.1	1.00
	100	0.5	0.79	0.30	0.09	0.61	52.7	0.95
		0.9	1.00	0.83	0.49	1.00	89.9	1.00
	200	0.5	0.99	0.45	0.14	0.81	105.7	0.95
		0.9	1.00	0.98	0.84	1.00	180.1	1.00
1.0	50	0.5	0.03	0.07	0.00	0.03	44.1	0.61
		0.9	0.04	0.06	0.01	0.04	50.5	0.89
	100	0.5	0.02	0.06	0.00	0.05	95.5	0.56
		0.9	0.05	0.06	0.01	0.04	100.4	0.90
	200	0.5	0.01	0.06	0.00	0.06	196.4	0.54
		0.9	0.03	0.05	0.00	0.04	199.8	0.90
2.0	50	0.5	0.69	0.05	0.01	0.06	116.0	0.23
		0.9	0.76	0.05	0.02	0.03	147.8	0.33
	100	0.5	0.95	0.11	0.03	0.08	387.0	0.13
		0.9	0.98	0.08	0.04	0.08	421.4	0.23
	200	0.5	1.00	0.19	0.10	0.15	1370.1	0.08
		0.9	1.00	0.20	0.07	0.18	1243.5	0.16

standard error of these power estimates from 500 independent simulations is less than 0.01 for a probability near 0.05 and not larger than 0.022 for any probability.

A second study simulates the same seven alternative models and fixes the number of detections at $r = 20$. Values of the scale parameters are again chosen so that individual detection probabilities at the stopping time $T = 1$ are fixed at $\delta = 0.5, 0.9$. For this second study, the KS distance based on the CPIT of the order statistics (LF) derived by Lee and Finelli (1989) is computed and appears in Tables 5-7 along with LC1, LC2 and KS1. Larger values of r present numerical difficulties in obtaining these CPITs.

It is well-known (Ch. 4, D'Agostino and Stephens (1986)) that adjustments to the sampling distribution of the KS distance and similar statistics based on the empirical distribution of probability transforms are needed in "composite" tests where

Table 3: Power Estimates under GOS Alternative, fixed (N, δ)

Shape	N	δ	LC1	LC2	KS1	KS2	Sample mean $\hat{N}$	$\hat{\delta}$
0.5	50	0.5	0.37	0.09	0.03	0.23	27.8	0.90
		0.9	0.72	0.22	0.05	0.50	45.5	0.98
	100	0.5	0.64	0.20	0.06	0.38	55.1	0.91
		0.9	0.96	0.41	0.12	0.75	91.3	0.98
	200	0.5	0.91	0.33	0.06	0.56	110.1	0.91
		0.9	1.00	0.64	0.25	0.96	183.5	0.98
2.0	50	0.5	0.39	0.02	0.00	0.04	89.4	0.30
		0.9	0.36	0.04	0.01	0.07	73.2	0.64
	100	0.5	0.70	0.06	0.02	0.06	269.6	0.20
		0.9	0.73	0.07	0.01	0.08	149.1	0.62
	200	0.5	0.94	0.07	0.03	0.07	874.8	0.12
		0.9	0.96	0.12	0.04	0.12	302.0	0.61

Table 4: Power Estimates under POS Alternative, fixed (N, δ)

Shape	N	δ	LC1	LC2	KS1	KS2	$\hat{N}$	$\hat{\delta}$
0.5	50	0.5	0.05	0.05	0.00	0.14	28.9	0.87
		0.9	1.00	0.69	0.47	1.00	45.1	1.00
	100	0.5	0.08	0.06	0.01	0.19	57.9	0.88
		0.9	1.00	0.95	0.84	1.00	89.8	1.00
	200	0.5	0.09	0.04	0.00	0.30	114.6	0.88
		0.9	1.00	1.00	0.99	1.00	179.8	1.00
1.0	50	0.5	0.02	0.06	0.00	0.06	33.8	0.76
		0.9	0.48	0.07	0.02	0.86	45.1	1.00
	100	0.5	0.03	0.06	0.00	0.06	68.5	0.75
		0.9	0.77	0.08	0.01	0.99	90.0	1.00
	200	0.5	0.03	0.03	0.01	0.10	134.8	0.75
		0.9	0.98	0.14	0.04	1.00	180.1	1.00

parameters are estimated from data. Comparing KS1 and KS2 in the part of Table 2 corresponding to a shape parameter of 1.0 shows that this is also the case in the EOS model. The size of the KS2 test is close to the nominal 0.05 level, but substantially lower for KS1. While KS1 almost never rejects a correctly specified EOS model, inspection of the rest of Tables 2-4 shows that the u-plot suffers from a lack of power to detect most of the alternatives considered here. LC1 has good power against the GOS and WOS alternatives. The power of these procedures generally increases with N and δ. LC1 has good power against the POS alternative when δ is large, but does poorly (as do LC2 and KS1) when $\delta = 0.5$. This may be due to the fact that the Pareto densities considered here have shape similar to the exponential density but with heavier tails, so that differences are only detected when δ is large. The average of the ILEs $(\hat{N}, \hat{\delta})$ from the simulations are provided to gauge the impact of model misspecification on parameter estimation. For example under misspecification of the EOS model for Pareto data, the estimate of δ is badly positively biased, which would yield overly optimistic predictions about the proportion of the population which is undetected at the stopping time T. For Weibull data with shape parameter 0.5, estimates are too optimistic for $\delta = 0.5$, but are relatively acceptable for $\delta = 0.9$. For the Weibull data with shape parameter 2.0, estimation is too pessimistic, drastically overestimating N and underestimating δ. The gamma data exhibit the same configuration with regard to estimation for the small and large values of the shape parameter, although the absolute value of the bias is not as bad. The effect of EOS model misspecification for Pareto data does not appear to be as bad, at least not for the two values of the shape parameter considered here.

Tables 5-7 indicate that both LC1 and LF perform better than the u-plot. LF does better than LC1 for the small choice of the shape parameter considered in both the GOS and WOS model and vice-versa for the larger shape parameter. Under the

Table 5: Power Estimates under WOS Alternative, $r = 20$

r	shape	δ	LC1	LC2	KS1	KS-LF	$\hat{N}$	$\hat{\delta}$
20	0.5	0.5	0.32	0.12	0.03	0.56	21.2	0.94
20	0.5	0.9	0.66	0.21	0.06	0.73	20.0	0.99
20	1.0	0.5	0.04	0.08	0.01	0.05	33.5	0.63
20	1.0	0.9	0.05	0.08	0.00	0.04	22.8	0.88
20	2.0	0.5	0.73	0.03	0.01	0.09	78.0	0.26
20	2.0	0.9	0.50	0.04	0.01	0.14	48.2	0.43

Table 6: Power Estimates under GOS Alternative, $r = 20$

r	shape	δ	LC1	LC2	KS1	KS-LF	$\hat{N}$	$\hat{\delta}$
20	0.5	0.5	0.22	0.07	0.03	0.46	22.5	0.90
20	0.5	0.9	0.34	0.08	0.02	0.52	20.4	0.97
20	2.0	0.5	0.42	0.03	0.01	0.09	61.6	0.34
20	2.0	0.9	0.28	0.06	0.00	0.14	31.2	0.66

Table 7: Power Estimates under POS Alternative, $r = 20$

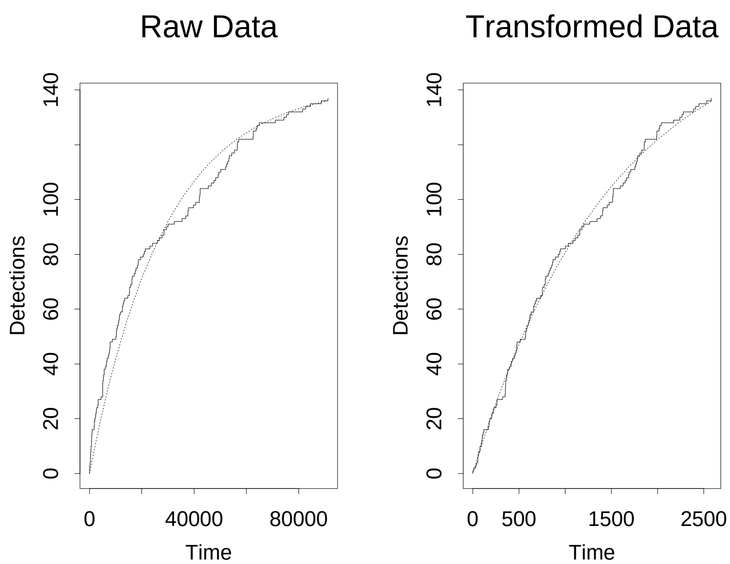
r	shape	δ	LC1	LC2	KS1	KS-LF	$\hat{N}$	$\hat{\delta}$
20	0.5	0.5	0.03	0.06	0.00	0.04	23.1	0.87
20	0.5	0.9	0.77	0.29	0.19	0.42	20.0	1.00
20	1.0	0.5	0.04	0.07	0.00	0.03	27.1	0.76
20	1.0	0.9	0.14	0.04	0.01	0.11	20.0	0.99

Table 8: Software Reliability Dataset

3	1726	5089	8738	14708	14708	21012	37915	52489	71043
33	1846	5097	10089	15251	15251	21308	39715	52875	74364
146	1872	5324	10237	15261	15261	23063	40580	53321	75409
227	1986	5389	10258	15277	15277	24127	42015	53443	76057
342	2311	5565	10491	15806	15806	25910	42045	54433	81542
351	2366	5623	10625	16185	16185	26770	42188	55381	82702
353	2608	6080	10982	16229	16229	27753	42296	56463	84566
444	2676	6380	11175	16358	16358	28460	42296	56485	88682
556	3098	6477	11411	17168	17168	28493	45406	56560	-2526
571	3278	6740	11442	17458	17458	29361	46653	57042	
709	3288	7192	11811	17758	17758	30085	47596	62551	
759	4434	7447	12559	18287	18287	32408	48296	62651	
836	5034	7644	12559	18568	18568	35338	49171	62661	
860	5049	7837	12791	18728	18728	36799	49416	63732	
968	5085	7843	13121	19556	19556	37642	50145	64103	
1056	5089	7922	13486	20567	20567	37654	52042	64893	

EOS model, both have levels close to the nominal level 0.05, while KS1 does not. Again, none are especially powerful against the POS alternative. The failure of LC1 for these cases indicates a limitation of the procedure. There are datasets in the literature (see Musa et al. (1987)) where the sample Lorenz curve does not indicate any lack-of-fit, but the u-plot does. LC1 is not particularly sensitive to discreteness in the data, for example, but the u-plot can be. If detections are made several at a time, so that many interdetection times T_i are zero, the u-plot takes a big jump at the first nonzero value of u_i, often leading to a significant KS distance, while LC1 can fail to detect this discreteness. Short of any gold standard, use of more than a single goodness-of-fit test in data analysis seems warranted.

Figure 1: Cumulative Errors Over Time

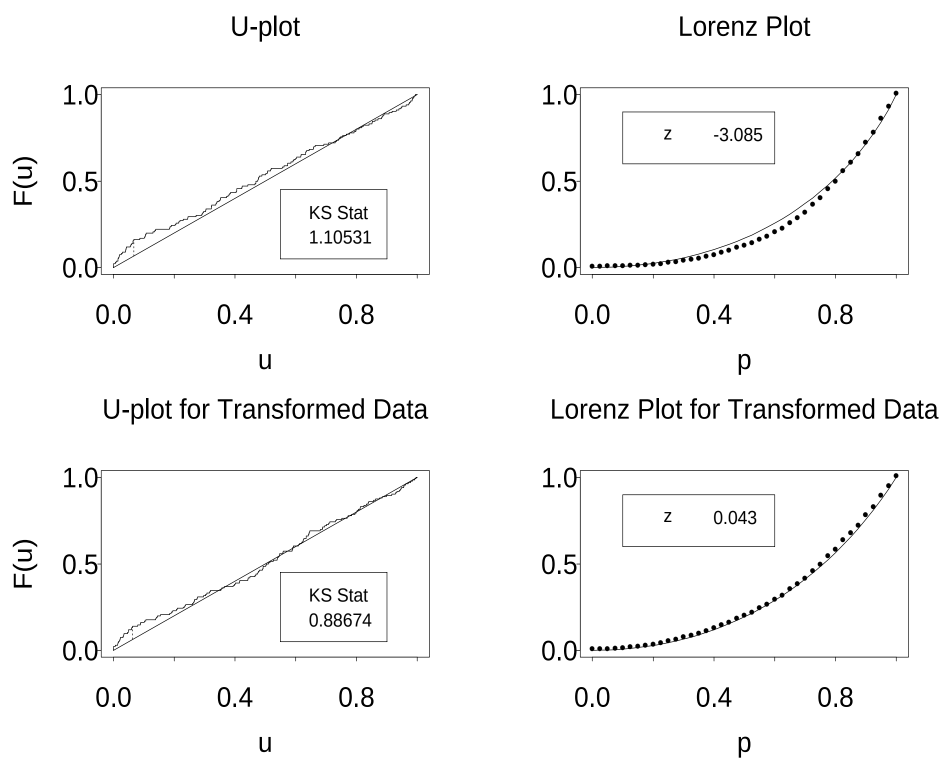


5 Example

The failure times in wall-clock seconds from a systems test and operational phases of a real time command and control system are given in Table 8. The negative value indicates failure free operation of the program for 2526 seconds. The EOS model was considered for these data in Musa et al. (1987), Raftery (1987) and Meinhold and Singpurwalla (1983) The sufficient statistic for estimation of (N, δ) is $(x, r) = (36.9, 136)$. The ILE is $(\hat{N}, \hat{\delta}) = (142, 0.96)$. The left plot in Figure 1 shows the cumulative occurrence of failures over time. The estimated mean failures over time t is given by $\hat{N}(1 - \exp\{-\hat{\phi}t\})$ and is overlaid on the plot.

To assess the fit of the model, the u-plot is shown in the top left quadrant of Figure 2. The normalized KS distance between the empirical distribution function of the estimated probability transforms of the interfailure times and the line of unit slope is $KS = 1.11$ This is not significant at level $\alpha = 0.10$. Alternatively, the sample

Figure 2: U-plot and Lorenz curve for model assessment



Lorenz curve shown in the upper right quadrant of Figure 2 and the associated test statistic can be used to assess the fit of the EOS model for these data. There appears to be some disparity between the estimated curve, $\hat{\lambda}(p)$, shown by the solid line, and the observed curve, $L_{136}(p)$, shown by the dotted line. The normalized difference between these curves at the point $p = 0.5$ takes the value $z = -3.085$ which is highly significant (p-value < 0.002), suggesting a poor fit. This lack of fit appears to have been detected by the test statistic based on the sample Lorenz curve, but "missed" by the KS statistic.

In Raftery (1988), it is suggested that among the exponential, Weibull, and Pareto order statistic models for these data, the Weibull model provides the best fit. Under the (β, ϕ) parameterization of the WOS model in Table 1, the power transformed data $Y_1^\beta, \ldots, Y_N^\beta$ follow the EOS model. Under the WOS model, the conditional log-likelihood given $R = r$:

$$l(\beta, \phi | R = r) = \sum_{j=1}^{r} \log(f(y_j; \beta, \phi)) - r \log(F(T; \beta, \phi))$$

is maximized at $(\hat{\beta}, \hat{\phi}) = (0.69, 0.00002)$. If the WOS model for the raw data is appropriate, then the EOS model may provide an approximate fit for the power transformed data obtained via $w_j = y_j^{\hat{\beta}}$ for $j = 1, \ldots, r$. To assess the fit of the newer model, the u-plot and the sample Lorenz curve for w_j are shown in the lower left and right quadrants, respectively, in Figure 2. Both are in agreement with what is expected under the EOS model for the transformed data. The associated test statistics are calculated to be $Z = 0.043$ and $KS = 0.887$, neither of which provides evidence of any lack of fit. Cumulative failures and their expected values over time for the raw and transformed data are plotted on the right in Figure 1. Inspection of the plots for these processes indicates a nicer fit for the transformed data. The sufficient statistic for the EOS model parameter for the transformed data is $(x_w, r) = (49.5, 136)$, corresponding to the ILE $(\hat{N}, \hat{\delta}) = (165, 0.82)$. This estimate is considerably less optimistic about the number of undetected faults than the one obtained by fitting the EOS model to the raw data.

References

Agustin, M. A. (1999), "Asymptotic and Finite-sample Properties of a Reliability Estimator for a Competing Risks System," *Journal of Statistical Computation and Simulation*, 64, 221–234.

Agustin, M. A. and Peña, E. A. (1999), "A Dynamic Competing Risks Model," *Probability in the Engineering and Informational Sciences*, 13, 333–358.

Basu, A. K., A., B., and Batra, R. (1995), "Modeling the Response Pattern to Direct Marketing Campaigns," *Journal of Marketing Research*, 32, 204–212.

Berger, J. O., Liseo, B., and Wolpert, R. L. (1999), "Integrated Likelihood Methods for Eliminating Nuisance Parameters," *Statistical Science*, 14, 1–22.

Casella, G. (1986), "Stabilizing Binomial n Estimators," *Journal of the American Statistical Association*, 81, 172–175.

D'Agostino, R. and Stephens, M. (1986), *Goodness-of-fit Techniques*, New York: Marcel Dekker, Inc.

Gail, M. H. and Gastwirth, G. L. (1978), "A Scale-free Goodness-of-fit Test for the Exponential Distribution Based on the Lorenz Curve," *Journal of the American Statistical Association*, 73, 787–793.

Gaudoin, O. (1998), "CPIT Goodness-of-fit Tests for the Power-law Process," *Communications in Statistics-Theory and Methods*, 27, 165–180.

Goldie, C. (1977), "Convergence Theorems for Empirical Lorenz Curves and Their Inverses," *Advances in Applied Probability*, 9, 393–401.

Jelinski, Z. and Moranda, P. B. (1972), "Software Reliability Research." in *Statistical Computer Performance Evaluation*, ed. Freiberger, W., New York: Academic, pp. 465–497.

Joe, H. and Reid, N. (1985), "Estimating the Number of Faults in a System," *Journal of the American Statistical Association*, 80, 222–226.

Klefsjo, B. and Kumar, U. (1992), "Goodness-of-Fit Tests for the Power-Law Process Based on the TTT-Plot," *IEEE Transactions on Reliability*, 41, 593–598.

Lee, L. and Finelli, G. (1989), "A transformation for testing the fit of a Exponential Order Statistic Model," *Stochastic Processes and their Applications*, 33, 299–307.

Lilliefors, H. (1969), "On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown," *Journal of the American Statistical Association*, 64, 387–389.

Littlewood, B. (1981), "Stochastic reliability growth: a model for fault-removal in computer programs and hardware designs," *IEEE Transactions on Reliability*, 30, 313–320.

Littlewood, B. and Verrall, J. L. (1981), "Likelihood Function of a Debugging Model for Computer Software Reliability," *IEEE Transactions on Reliability*, 30, 145–148.

Meinhold, R. and Singpurwalla, N. D. (1983), "Bayesian Analysis of a Commonly Used Model for Describing Software Failures," *The Statistician*, 32, 168–173.

Miller, D. R. (1986), "Exponential Order Statistic Models of Software Reliability Growth," *IEEE Transactions on Software Engineering*, 12, 12–24.

Musa, J. D., Iannino, A., and Okumoto, K. (1987), *Software Reliability: Measurement, Prediction, Application*, New York: McGraw Hill.

Osborne, J. A. and Severini, T. A. (2000), "Inference for Exponential Order Statistic Models Based on an Integrated Likelihood Function," *Journal of the American Statistical Association*, 95, 1220–1128.

Raftery, A. E. (1987), "Inference and Prediction for a General Order Statistic Model with Unknown Population Size," *Journal of the American Statistical Association*, 82, 1163–1168.

— (1988), "Analysis of a Simple Debugging Model," *Journal of the Royal Statistical Society Ser. B*, 37, 12–22.

Singpurwalla, N. D. and Wilson, S. P. (1999), *Statistical Methods in Software Engineering: Reliability and Risk*, New York: Springer-Verlag.

Yamada, S., Ohba, M., and Osaki, S. (1983), "S-Shaped Reliability Growth Modeling for Software Error-Detection," *IEEE Transactions on Reliability*, 32, 475–478.