



**Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.**

<http://siba-ese.unisalento.it/index.php/ejasa/index>

e-ISSN: 2070-5948

DOI: 10.1285/i20705948v8n1p122

**Estimation in the Exponentiated Kumaraswamy
Dagum Distribution with Censored Samples**

By Oluyede, Huang

Published: 26 April 2015

This work is copyrighted by Università del Salento, and is licensed under a Creative Commons Attribution - Non commerciale - Non opere derivate 3.0 Italia License.

For more information see:

<http://creativecommons.org/licenses/by-nc-nd/3.0/it/>

Estimation in the Exponentiated Kumaraswamy Dagum Distribution with Censored Samples

Broderick O. Oluyede^{*a} and Shujiao Huang^b

^a*Department of Mathematical Sciences, Georgia Southern University, Statesboro, Georgia, 30458, United States*

^b*Department of Mathematics, University of Houston, Houston, Texas 77204-3008, United States*

Published: 26 April 2015

In a recent note, Huang and Oluyede (2014) proposed a new model called the exponentiated Kumaraswamy Dagum (EKD) distribution with applications to income and lifetime data. In this note, this distribution is shown to be a very competitive model for describing censored observations in lifetime reliability problems. This work shows that in certain cases, the EKD distribution performs better than other parametric model such as the exponentiated Kumaraswamy Weibull distribution and its sub-models, which include some of the commonly used models in survival analysis and reliability analysis, such as the exponentiated Weibull, Weibull and exponential distributions.

Mathematics Subject Classification: 62E15; Secondary 60E05.

keywords: Generalized Distribution; Kumaraswamy Distribution; Dagum Distribution.

1 Introduction

We consider a life-testing experiment where n units is kept under observation until failure. It is possible that these units could be some system, components, or computer

*Corresponding author: boluyede@georgiasouthern.edu

chips in reliability experiments or the units could be patients that are placed under certain drug or clinical conditions. suppose we have to terminate the experiment before all the units have failed, as in the case where individuals in a clinical trial may drop out of the study or the study may be ended or terminated due to economic conditions or lack of funds. In the case of an industrial experiment, units may fail or break accidentally. Data obtained from these settings or under these conditions are referred to as censored data. In this note, we are primarily concerned with the application of a new model called the exponentiated Kumaraswamy Dagum (EKD) distribution to censored data.

We start by looking at the generalized beta distribution of the second kind and the sub-model called Dagum distribution. The generalized beta distribution of the second kind (GB2), McDonald (1984), McDonald and Xu (1995) have application in several areas including modeling of size distribution of personal income. The probability density function (pdf) of the GB2 distribution is given by:

$$f_{GB2}(y; a, b, p, q) = \frac{ay^{ap-1}}{b^{ap}B(p, q)[1 + (\frac{y}{b})^a]^{p+q}}, \quad \text{for } y > 0.$$

Note that $a > 0, p > 0, q > 0$, are the shape parameters and b is the scale parameter and $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ is the beta function. When the parameter $q = 1$, the resulting distribution is called Dagum distribution, Dagum (2008). Kleiber (2008) traced the genesis of Dagum distribution and summarized several statistical properties of this distribution. Domma and Condino (2013) obtained the maximum likelihood estimates of the parameters of Dagum distribution for censored data. Domma and Condino (2013) presented the beta-Dagum distribution. Domma et al. (2011) presented results on the maximum likelihood estimation in Dagum distribution with censored samples.

The main objective here is the application of the EKD model to real lifetime censored data, primarily for describing data which include censored observations in life time reliability problems. In this paper, we present maximum likelihood estimation as well as comparisons with other parametric models in the exponentiated Kumaraswamy Dagum distribution under Type I right censoring and Type II double censoring schemes. This paper is organized as follows. The model is given in section 2. Maximum likelihood estimates of the model parameters under Type I right censoring and Type II double censoring plans are presented in section 3. Applications, case studies and comparisons with the exponentiated Kumaraswamy Weibull distribution are given in section 4, followed by concluding remarks in section 5.

2 The Exponentiated Kumaraswamy-Dagum Distribution

In this section, the EKD distribution is presented for completeness (see Huang and Oluyede (2014) for additional details). Kumaraswamy (1980) introduced a two-parameter distribution on $(0, 1)$. Its cdf is given by

$$G(x) = 1 - (1 - x^\psi)^\phi, \quad x \in (0, 1),$$

for $\psi > 0$ and $\phi > 0$. For an arbitrary cdf $F(x)$ with pdf $f(x) = \frac{dF(x)}{dx}$, the family of Kumaraswamy-G distributions with cdf $G_k(x)$ is given by

$$G_K(x) = 1 - (1 - F^\psi(x))^\phi,$$

for $\psi > 0$ and $\phi > 0$. By letting $F(x) = G_D(x)$, we obtain the Kumaraswamy-Dagum (KD) distribution, with cdf

$$G_{KD}(x) = 1 - (1 - G_D^\psi(x))^\phi.$$

We replace the dependent parameter $\beta\psi$ by α , so that the cdf and pdf of the EKD distribution are given by

$$G_{EKD}(x; \alpha, \lambda, \delta, \phi, \theta) = \{1 - [1 - (1 + \lambda x^{-\delta})^{-\alpha}]^\phi\}^\theta, \quad (1)$$

and

$$\begin{aligned} g_{EKD}(x; \alpha, \lambda, \delta, \phi, \theta) &= \alpha\lambda\delta\phi\theta x^{-\delta-1}(1 + \lambda x^{-\delta})^{-\alpha-1}[1 - (1 + \lambda x^{-\delta})^{-\alpha}]^{\phi-1} \\ &\times \{1 - [1 - (1 + \lambda x^{-\delta})^{-\alpha}]^\phi\}^{\theta-1}, \end{aligned} \quad (2)$$

for $\alpha, \lambda, \delta, \phi, \theta > 0$, and $x > 0$, respectively. The quantile function of the EKD distribution is in closed form, and is given by

$$G_{EKD}^{-1}(q) = x_q = \lambda^{\frac{1}{\delta}} \left\{ [1 - (1 - q^{\frac{1}{\theta}})^{\frac{1}{\phi}}]^{-\frac{1}{\alpha}} - 1 \right\}^{-\frac{1}{\delta}}. \quad (3)$$

3 Maximum Likelihood Estimation

Different censoring mechanisms lead to different likelihood functions. In the following sections, we construct log-likelihood functions of the EKD distribution to deal with type I right and type II doubly censored observations.

Although the maximum likelihood estimates are not available in closed form, they can be evaluated with the help of numerical techniques. The difficulties in dealing with the EKD distribution due to its complicated mathematical tractability are easily overcome by using iterative methods which do not require high-computational efforts even in the presence of censoring.

3.1 Type I Right Censoring

This is the most common form of incomplete data often encountered in survival analysis. Type I right censoring arises when the study is conducted over a specified time period that can terminate before all the units have failed. Each individual has a fixed censoring time C_i , which would be the time between the date of entry and the end of study, so that the complete failure time of an individual will be known only if it is less than or equal to the censoring time $T_i \leq C_i$; otherwise, only a lower bound of the individual lifetime is available $T_i > C_i$. The data for this design are conveniently indicated by pairs of random

variables (T_i, ϵ_i) , $i = 1, \dots, n$. Consider a sample size n of independent positive random variables T_1, \dots, T_n such that T_i is associated with an indicator variable $\epsilon_i = 0$ if T_i is a censoring time. Let $\Theta = (\alpha, \lambda, \delta, \phi, \theta)^T$, then the likelihood function, $L(\Theta)$, of a type I right censored sample $(t_1, \epsilon_1), \dots, (t_n, \epsilon_n)$ from EKD distribution with pdf $g_{EKD}(\cdot)$ and survival function $S_{EKD}(\cdot)$ can be written as

$$L(\Theta) = \prod_{i=1}^n g_{EKD}(t_i)^{\epsilon_i} S_{EKD}(t_i)^{1-\epsilon_i},$$

where $S_{EKD}(t_i) = 1 - G_{EKD}(t_i)$. The log-likelihood function, $l(\Theta)$, based on data, from Equation (1) and (2) is

$$\begin{aligned} l(\Theta) &= \sum_{i=1}^n \epsilon_i \left\{ \ln \alpha + \ln \lambda + \ln \delta + \ln \phi + \ln \theta - (\delta + 1) \ln t_i \right. \\ &\quad - (\alpha + 1) \ln(1 + \lambda t_i^{-\delta}) + (\phi - 1) \ln[1 - (1 + \lambda t_i^{-\delta})^{-\alpha}] \\ &\quad \left. + (\theta - 1) \ln\{1 - [1 - (1 + \lambda t_i^{-\delta})^{-\alpha}]^\phi\} \right\} \\ &\quad + \sum_{i=1}^n (1 - \epsilon_i) \ln \left\{ 1 - \{1 - [1 - (1 + \lambda t_i^{-\delta})^{-\alpha}]^\phi\}^\theta \right\}. \end{aligned} \tag{4}$$

The MLEs $\hat{\Theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\delta}, \hat{\phi}, \hat{\theta})$ are obtained from the numerical maximization of Equation (4). Let $\Theta = (\alpha, \lambda, \delta, \phi, \theta)^T$ be the parameter vector and $\hat{\Theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\delta}, \hat{\phi}, \hat{\theta})$ be the maximum likelihood estimate of $\Theta = (\alpha, \beta, \theta, \lambda, \delta)$. Under the usual regularity conditions and that the parameters are in the interior of the parameter space, but not on the boundary, (Ferguson, 1996) we have: $\sqrt{n}(\hat{\Theta} - \Theta) \xrightarrow{d} N_5(\underline{0}, I^{-1}(\Theta))$, where $I(\Theta)$ is the expected Fisher information matrix. The asymptotic behavior is still valid if $I(\Theta)$ is replaced by the observed information matrix evaluated at $\hat{\Theta}$, that is $J(\hat{\Theta})$. The multivariate normal distribution $N_5(\underline{0}, J(\hat{\Theta})^{-1})$, where the mean vector $\underline{0} = (0, 0, 0, 0, 0)^T$, can be used to construct confidence intervals and confidence regions for the individual model parameters and for the survival and hazard rate functions.

3.2 Type II Double Censoring

Type II double censoring is used to indicate that, in an ordered sample of size n , a known number of observations is missing at both ends, while in type I censoring, the number of censored observations is a random variable and the time of study is fixed. The data consist of the remaining ordered observations t_{r+1}, \dots, t_m when the r smallest observations and the $n - m$ largest observations are out of a sample of size n from the EKD distribution. The likelihood function, $L(\Theta)$, of the type II doubly censored sample $t_{(r+1)}, \dots, t_{(m)}$ from EKD distribution with pdf $g_{EKD}(\cdot)$, cdf $G_{EKD}(\cdot)$ and survival function $S_{EKD}(\cdot)$ is given by

$$L(\Theta) = \frac{n!}{r!(n-m)!} \{G_{EKD}(t_{(r+1)})\}^r \{S_{EKD}(t_{(m)})\}^{n-m} \prod_{i=r+1}^m g_{EKD}(t_{(i)}).$$

The log-likelihood function, $l(\Theta)$, for a type II doubly censored sample $t_{(r+1)}, \dots, t_{(m)}$ from EKD distribution is given by

$$\begin{aligned}
 l(\Theta) &= \ln \left(\frac{n!}{r!(n-m)!} \right) + r\theta \ln \{1 - [1 - (1 + \lambda t_{(r+1)}^{-\delta})^{-\alpha}]^{\phi}\} \\
 &+ (n-m) \ln \left(1 - \{1 - [1 - (1 + \lambda t_{(m)}^{-\delta})^{-\alpha}]^{\phi}\}^{\theta} \right) \\
 &+ \sum_{i=r+1}^m \left(\ln \alpha + \ln \lambda + \ln \delta + \ln \phi + \ln \theta - (\delta + 1) \ln t_{(i)} \right. \\
 &\quad \left. - (\alpha + 1) \ln(1 + \lambda t_{(i)}^{-\delta}) + (\phi - 1) \ln[1 - (1 + \lambda t_{(i)}^{-\delta})^{-\alpha}] \right. \\
 &\quad \left. + (\theta - 1) \ln \{1 - [1 - (1 + \lambda t_{(i)}^{-\delta})^{-\alpha}]^{\phi}\} \right). \tag{5}
 \end{aligned}$$

As in the type I right censoring scheme, the MLEs $\hat{\Theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\delta}, \hat{\phi}, \hat{\theta})$ are only obtained by numerical methods.

4 Application

In this section, we give some applications to real life data. Maximum likelihood estimates of the model parameters under type I right and type II double censored data are obtained and comparisons with the exponentiated Kumaraswamy Weibull distribution and its sub-models, which are widely used in reliability and survival data analysis are presented. We compared EKD and its sub-models, as well as exponentiated Kumaraswamy Weibull (EKW) distribution, with the aid of the statistics: $-2 \text{ Log-likelihood statistic}$, Akaike Information Criterion, $AIC = 2p - 2 \ln(L)$, Bayesian Information Criterion, $BIC = p \ln(n) - 2 \ln(L)$, and Consistent Akaike Information Criterion, $AICC = AIC + 2 \frac{p(p+1)}{n-p-1}$, where $L = L(\hat{\Theta})$ is the value of the likelihood function evaluated at the parameter estimates, n is the number of observations, and p is the number of estimated parameters in the model. The exponentiated Kumaraswamy Weibull (EKW) pdf given by

$$\begin{aligned}
 f_{EKW}(x) &= \theta abc \lambda^c x^{c-1} e^{-(\lambda x)^c} \left[1 - e^{-(\lambda x)^c} \right]^{a-1} \left\{ 1 - \left[1 - e^{-(\lambda x)^c} \right]^a \right\}^{b-1} \\
 &\times \left[1 - \left\{ 1 - \left[1 - e^{-(\lambda x)^c} \right]^a \right\}^b \right]^{\theta-1},
 \end{aligned}$$

for $\theta \geq 0, \lambda, a, b, c > 0$, was also compared to the EKD distribution via the statistics states above. Probability plots (Chambers (1983)) are also presented. For the probability plot, we plotted $G_{EKD}(x_{(j)}; \hat{\alpha}, \hat{\lambda}, \hat{\delta}, \hat{\phi}, \hat{\theta})$ against $\frac{j - 0.375}{n + 0.25}$, $j = 1, 2, \dots, n$, where $x_{(j)}$ are the ordered values of the observed data. We also computed a measure of closeness of each plot to the diagonal line. This measure of closeness of the plot to the diagonal

Table 1: Estimation of models for remission times data

Distribution	Estimates					Statistics				
	α	λ	δ	ϕ	θ	-2 Log Likelihood	AIC	AICC	BIC	SS
EKD	4.0674 (0.6287)	52.0264 (32.0067)	1.7826 (0.2623)	1.425 (0.6799)	0.1792 (0.02053)	836.7	846.7	847.2	861.3	0.0410
KD	0.1862 (0.05143)	0.5639 (0.2804)	7.8559 (0.1391)	0.0577 (0.005263)	1 -	944.6	952.6	952.9	964.3	2.8410
D	2.5445 (0.8068)	2.5344 (1.3466)	1.1709 (0.1174)	1 -	1 -	857.1	863.1	863.2	871.8	0.1870
EKW	a 1.1633 (0.2468)	b 0.05785 (0.01698)	c 2.4018 (0.1602)	λ 0.05674 (0.005661)	θ 0.1918 (0.04479)	927.2	937.2	937.7	951.8	4.7914

line given by the sum of squares

$$SS = \sum_{j=1}^n \left[G(x_{(j)}) - \left(\frac{j - 0.375}{n + 0.25} \right) \right]^2,$$

where the data censor=0 have been deleted in type I right censored data.

4.1 Case I: Remission times of cancer patients

For the first example, we consider the data set on remission times (in months) for 137 cancer patients, Lee and Wang (2003). Estimates of the parameters for type I right censoring models, AIC, AICC, BIC, and SS are given in Table 1. In the plot comparing the survival functions of the EKD, KD, D and EKW distributions with the Kaplan-Meier curve, we see that the EKD distribution is preferred, while the other models tend to over or underestimate, mostly overestimate the empirical curve.

Plots of the fitted densities and the histogram, hazard functions, survival functions, and observed probability vs predicted probability for the remission times data are given in Figures 1, 2, 3 and 4, respectively.

The LR test statistic of the hypothesis $H_0 : EKD$ against $H_a : KD$ and $H_0 : EKD$ against $H_a : D$ are 107.9 (p-value < 0.0001) and 20.4 (p-value < 0.0001). Also, EKD distribution gives the smallest SS value. We can conclude that EKD distribution provides a very good fit for remission times data.

4.2 Case II: 2004 New Car and Truck Data

The second example consists of price of 428 new vehicles for the 2004 year (Kiplinger's Personal Finance, Dec 2003). After sorting the data, we note the 4 largest numbers are far away from others. So we drop the last 4 number. As a result, 424 numbers are study, where $n = 428, r = 0, m = 424$. Estimates of the parameters, $-2 \log(L)$, AIC, AICC, BIC, and SS are given in Table 2.

Table 2: Estimation of models for price of car ($\times 10^4$)

Distribution	Estimates					Statistics				
	α	λ	δ	ϕ	θ	-2 Log Likelihood	AIC	AICC	BIC	SS
EKD	0.006567 (0.000265)	0.3440 (0.03825)	13.7987 (0.1624)	0.1609 (0.00789)	16.6667 (2.5266)	1489.9	1499.9	1500.0	1520.1	0.2816
KD	9.4739 (1.2372)	0.0027 (0.000734)	13.5703 (0.1214)	0.0537 (0.00266)	1 -	1979.1	1987.1	1987.2	2003.3	14.8571
D	2.453 (0.7916)	6.3435 (3.6931)	2.9237 (0.2089)	1 -	1 -	1472.4	1478.4	1478.4	1490.5	0.0478
EKW	a 6.0407 (0.1765)	b 0.0273 (0.00367)	c 4.1272 (0.07324)	λ 0.2520 (0.003682)	θ 0.0705 (0.003311)	1582.5	1592.5	1592.7	1612.8	3.5570

Plots of the fitted density and the histogram, and observed probability versus predicted probability are given in Figures 5 and 6. The sub-model, Dagum distribution seem to be the “superior” fit for this data, based on the plots and the statistics given in Table 2.

The LR test statistic of the hypothesis $H_0 : \text{EKD}$ against $H_a : \text{KD}$ is 489.2 (p-value < 0.0001). Dagum gives the smallest AIC, AICC, BIC and SS values.

5 Concluding Remarks

We have proposed and presented results on a new class of distributions called the Exponentiated Kumaraswamy Dagum (EKD) distribution. We employed this family of EKD distributions to fit censored data. We showed that this class of distributions is competitive class of models as far as censored observations in lifetime and reliability analysis are concerned. Estimation of the parameters of the models under Type I right and Type II double censoring plans and applications are also given.

References

- Chambers, J. M. (1983). *Graphical methods for data analysis*.
- Dagum, C. (2008). *A new model of personal income distribution: specification and estimation*. Springer.
- Domma, F. and Condino, F. (2013). The beta-dagum distribution: definition and properties. *Communications in Statistics-Theory and Methods*, 42(22):4070–4090.
- Domma, F., Giordano, S., and Zenga, M. (2011). Maximum likelihood estimation in dagum distribution with censored samples. *Journal of Applied Statistics*, 38(12):2971–2985.
- Huang, S. and Oluyede, B. O. (2014). Exponentiated kumaraswamy-dagum distribution with applications to income and lifetime data. *Journal of Statistical Distributions and Applications*, 1(1):8.
- Kleiber, C. (2008). *A guide to the Dagum distributions*. Springer.

- Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1):79–88.
- Lee, E. T. and Wang, J. (2003). *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons.
- McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica: Journal of the Econometric Society*, pages 647–663.
- McDonald, J. B. and Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66(1):133–152.

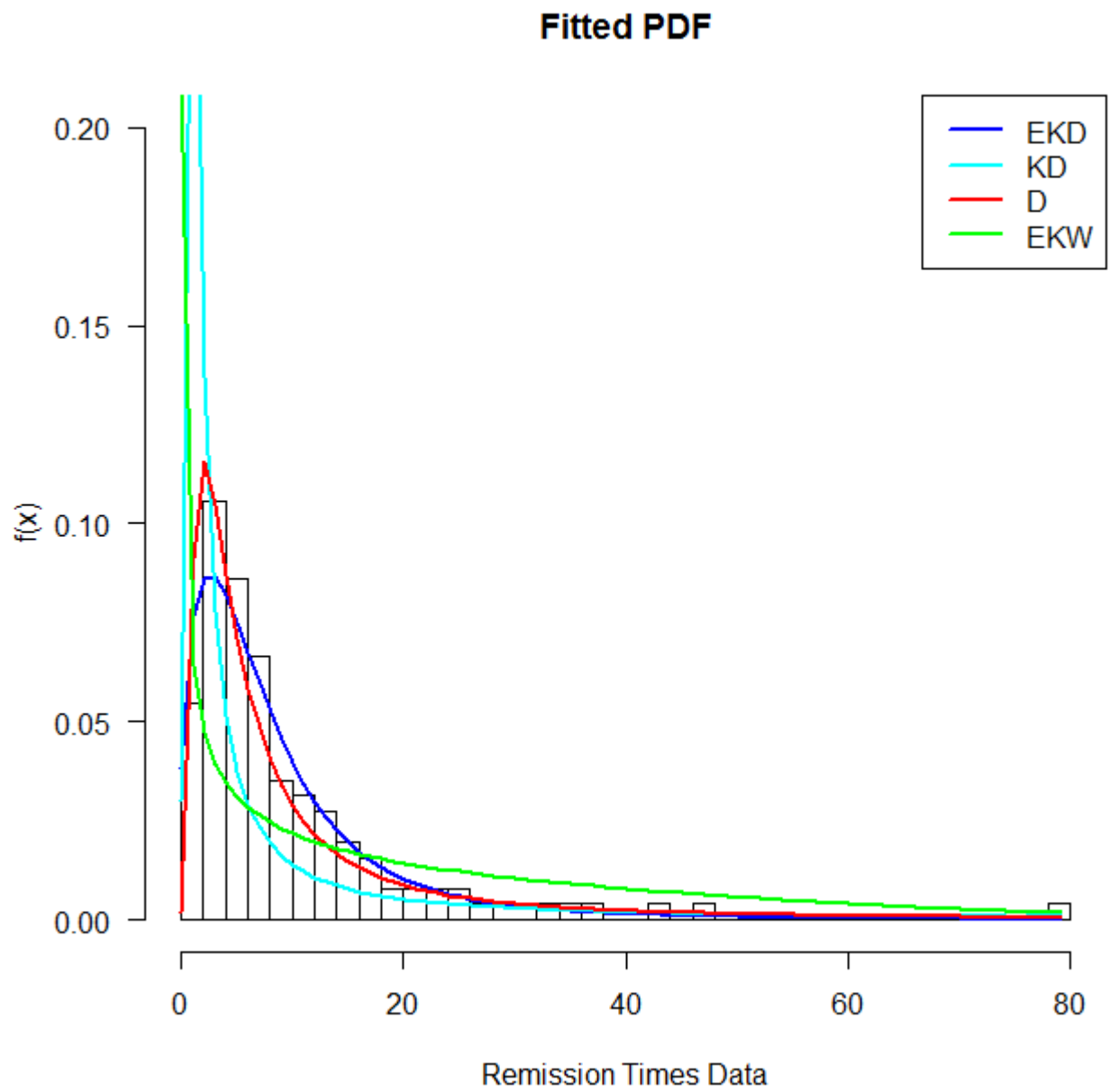


Figure 1: Fitted Densities for Remission Data

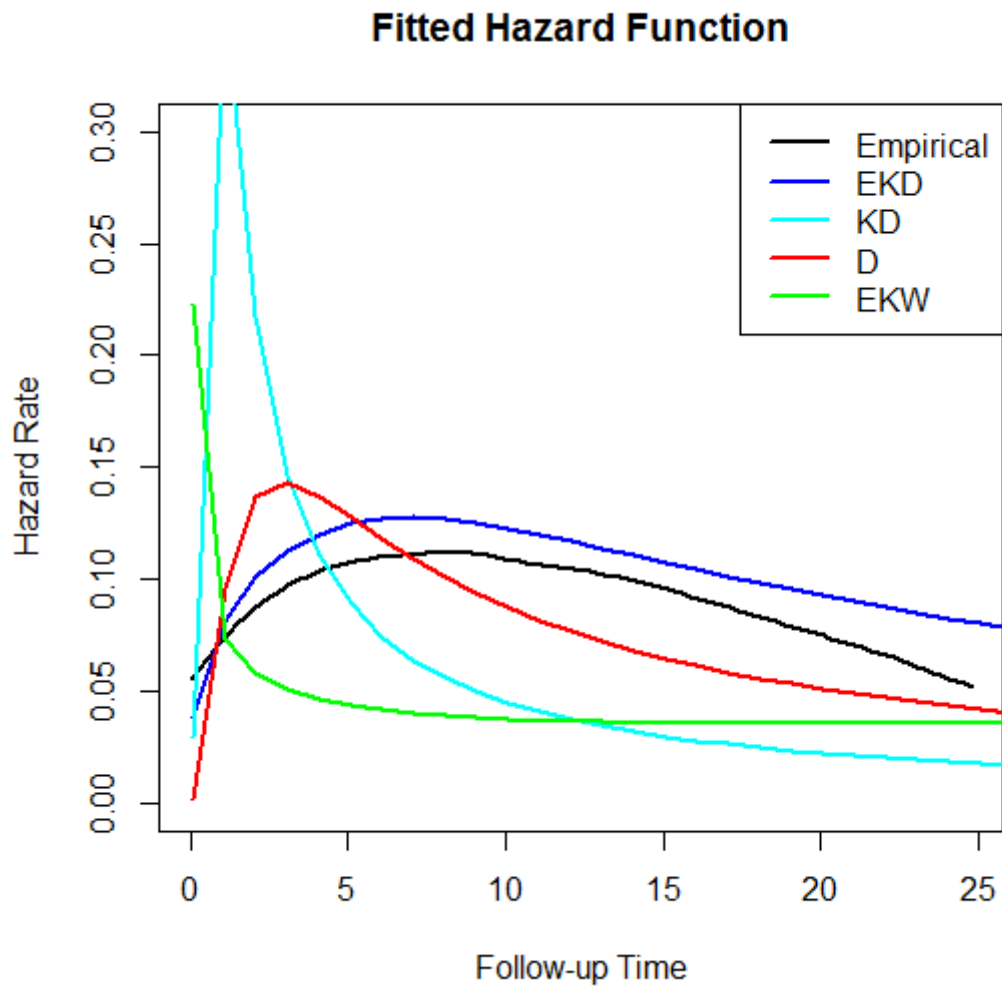


Figure 2: Fitted Hazard Rate Functions for Remission Data

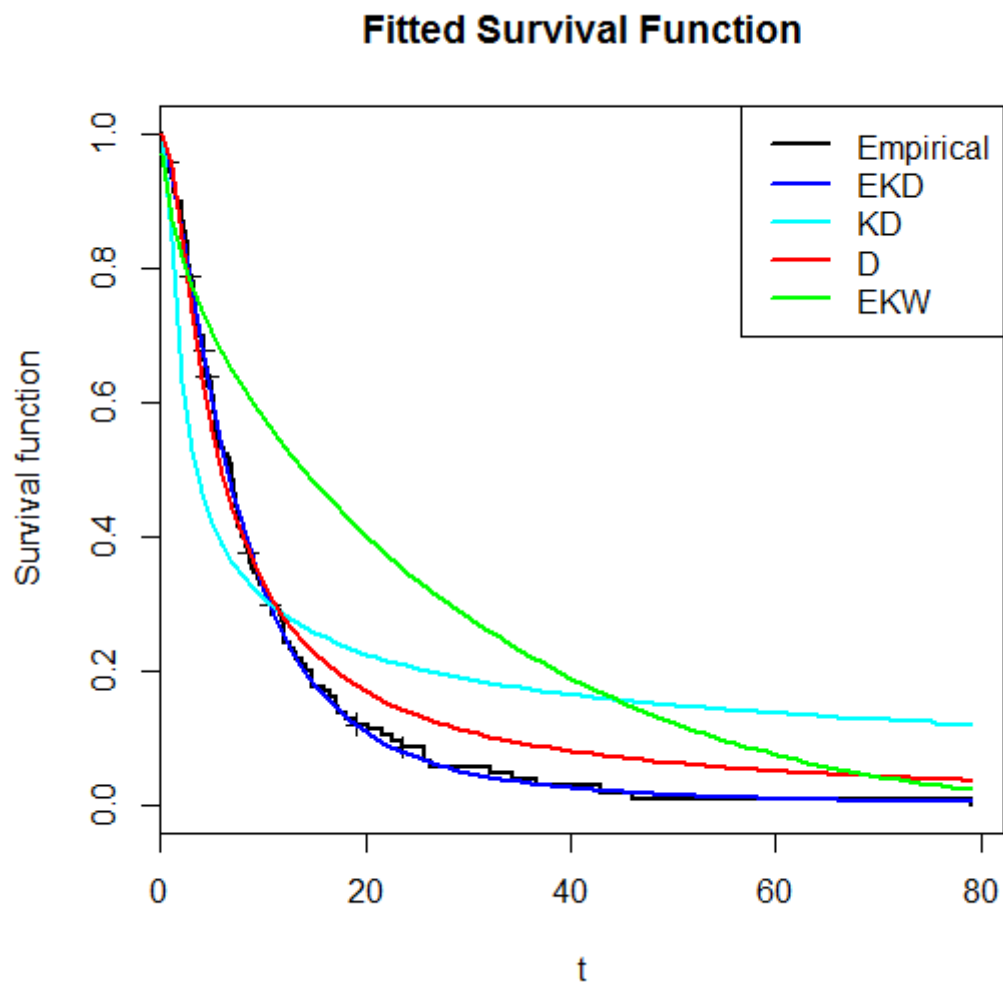


Figure 3: Fitted Survival Functions for Remission Data

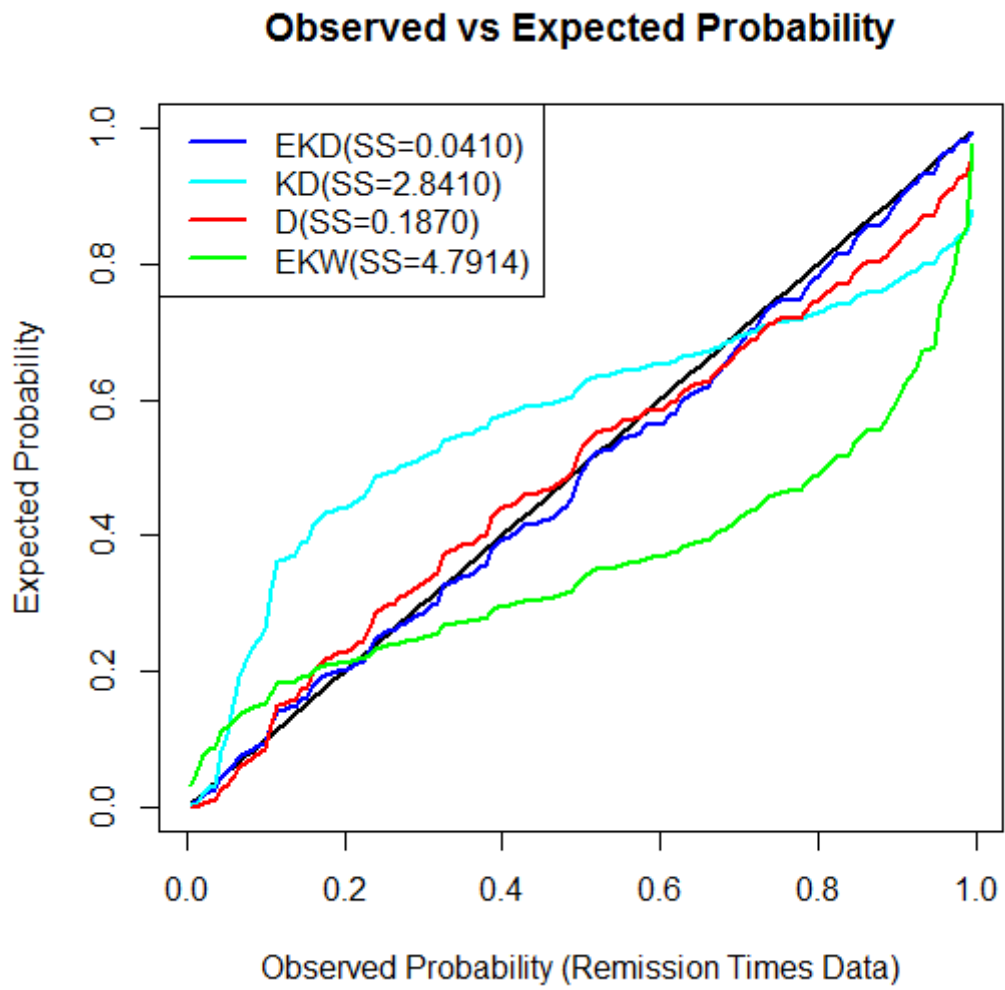


Figure 4: Probability Plots for Remission Data

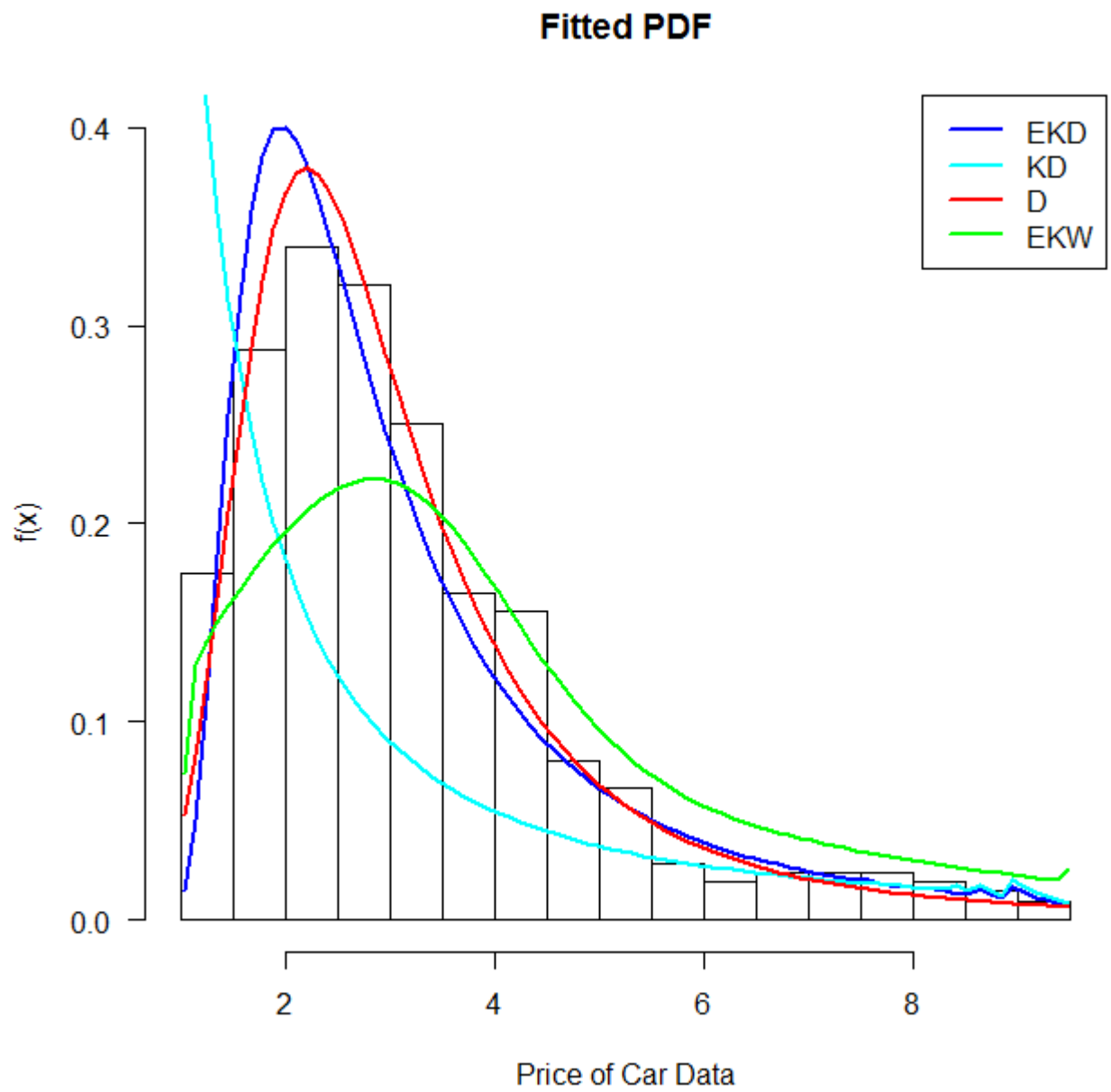


Figure 5: Fitted Densities for Price of Car Data

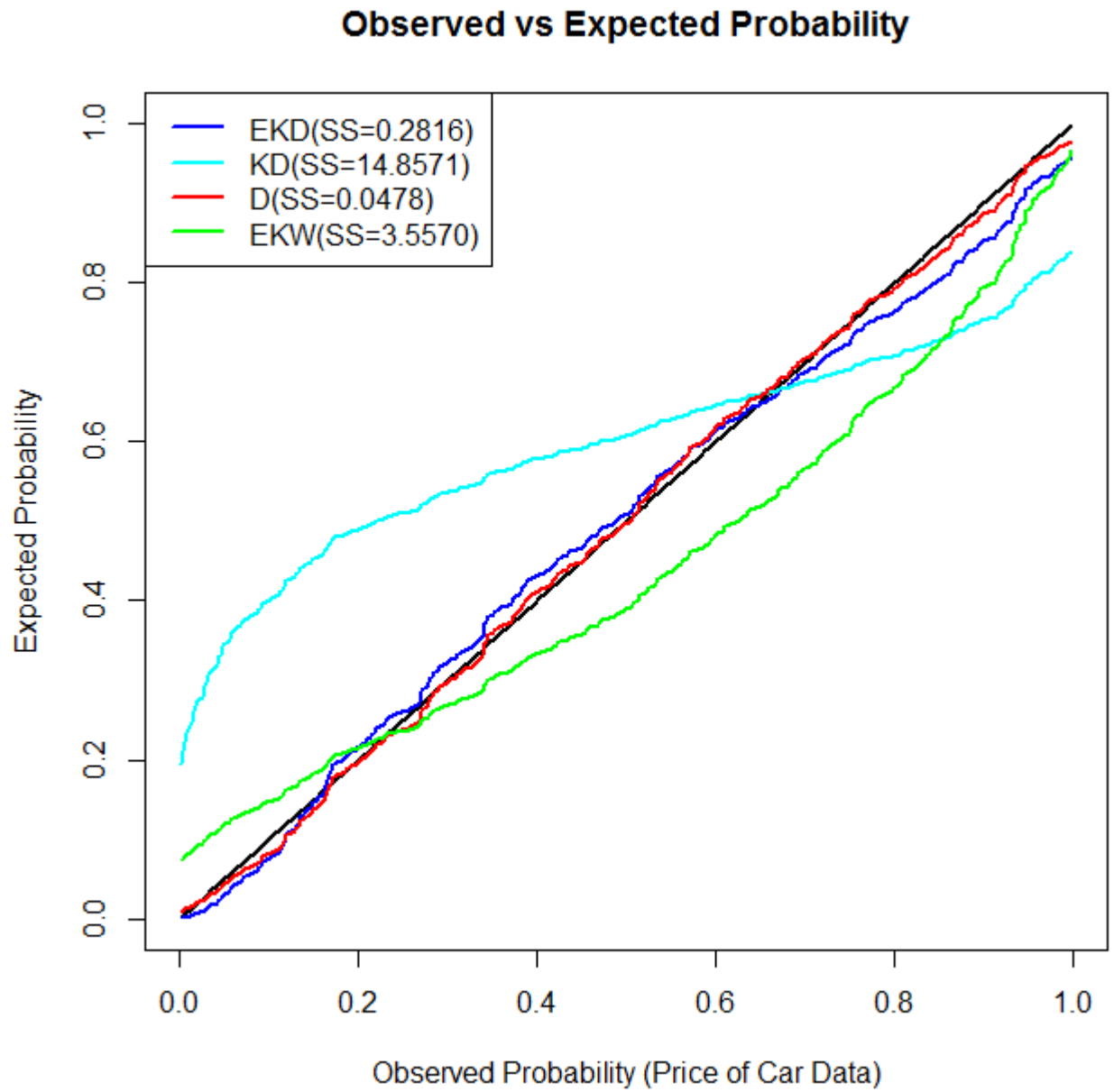


Figure 6: Probability Plots for Price of Car Data