

# The Arctangent Survival Distribution

ANDREW G. GLEN

LAWRENCE M. LEEMIS

Department of Mathematics

College of William & Mary

Williamsburg, VA 23187-8795, USA

`{agglen,leemis}@math.wm.edu`

We present a two-parameter survival distribution that has an upside-down bathtub (UBT, or humped-shaped) hazard function. This distribution provides biostatisticians, reliability engineers, and other statisticians with a second two-parameter UBT model whose closed-form survivor function simplifies the analysis of right-censored data sets. We develop the distribution's probabilistic and statistical properties. Maximum likelihood estimators of the parameters are found using numerical methods. Approximate confidence intervals can be determined by using the observed information matrix or the likelihood ratio statistic. We also give examples in which the arctangent distribution is a reasonable alternative to other common lifetime distributions.

---

Andy Glen is a Ph.D. student at the College of William & Mary and is presently on the faculty at the United States Military Academy at West Point, New York.

Larry Leemis is a Professor of Mathematics at the College of William & Mary.

# Introduction

Parametric lifetime distributions have been used in reliability for modeling the time to failure of components and systems. Although emphasis has traditionally been placed on models where the hazard function has a bathtub shape, applications have been found where an upside-down bathtub (UBT), or hump-shaped hazard function is the appropriate model. Kececioglu (1991, page 425) lists transistors, metals subjected to alternating stress levels, insulation degradation, mechanical devices subjected to wear, and bearings as potential UBT applications. Chhikara and Folks (1989, page 156) state that “When early occurrences such as product failures or repairs are dominant in a lifetime distribution, its failure rate is expected to be non-monotonic, first increasing and later decreasing” and cite airborne communication transceivers (page 5, 139–140) as an application. Lee (1992, page 12) further supports the validity of a UBT risk model in describing patients with tuberculosis who have risks that “increase initially, then decrease after treatment.” To further substantiate the usefulness of the UBT model, Donald R. Barr (Professor of Systems Engineering, United States Military Academy) opines that the UBT risk function would apply in modeling the probability of a soldier becoming a casualty as a result of artillery fire. In this example, casualty risk starts out low as the fire is initially inaccurate, increases as the shooter hones in on the target, and then decreases as the remaining soldiers are able to “dig in” for protection. Although reliability engineers generally have an abundance of two-parameter survival distributions to choose from, relatively few have a UBT hazard function. The commonly used UBT distributions are the inverse Gaussian, log normal, and log logistic distributions. Of these, only the log logistic distribution has a closed-form survival function. This distribution is most often used in biostatistical applications, whereas the inverse Gaussian and log normal are typically used in reliability.

The arctangent distribution developed here gives a survival distribution with a UBT hazard function and closed-form survivor function, a useful feature in the analysis of a right-censored data set. Additionally, the survivor function can be inverted in closed-form, which enables synchronization and monotonicity in variate generation. Unlike most survival distributions, the arctangent distribution’s development uses trigonometric functions. We will present the arctangent distribution’s development, probabilistic properties, and statis-

tical inference. Parameter estimation for complete and right-censored data sets is found by maximum likelihood. Finally we will provide three examples where the distribution is a demonstrably reasonable alternative to other survival distributions.

## Development

We noticed that the arctangent function, when negated and shifted vertically, resembles a survivor function. Further, we decided that by shifting the function so that it crossed the vertical axis at 1 and then asymptotically decreased to 0 we then had a flexible survivor function. By adding a phase shift parameter  $\phi$  and a positive scaling parameter  $\alpha$  yields the function

$$g(t) = -\arctan[\alpha(t - \phi)] + \frac{\pi}{2} \quad -\infty < t < \infty,$$

which is a decreasing function with a range of  $(0, \pi)$ . Finally, since  $g(0) = \arctan(\alpha\phi) + \frac{\pi}{2}$ , the appropriate way to adjust this function so that it assumes the value 1 when  $t = 0$  is to divide  $g(t)$  by  $g(0)$ , yielding the survivor function for the random lifetime  $T$

$$S_T(t) = \frac{-\arctan[\alpha(t - \phi)] + \frac{\pi}{2}}{\arctan(\alpha\phi) + \frac{\pi}{2}} \quad t \geq 0.$$

Since the arctangent is an odd function, the form of the survivor function that will be used here is

$$S_T(t) = \frac{\arctan[\alpha(\phi - t)] + \frac{\pi}{2}}{\arctan(\alpha\phi) + \frac{\pi}{2}} \quad t \geq 0, \quad (1)$$

where  $\alpha > 0$  and  $-\infty < \phi < \infty$ . This survivor function satisfies the three existence conditions:  $S(0) = 1$ ,  $\lim_{t \rightarrow \infty} S(t) = 0$ , and  $S(t)$  is non-increasing. Furthermore, the distribution's probability density function and hazard function are

$$f_T(t) = -S'_T(t) = \frac{\alpha}{\left[\arctan(\alpha\phi) + \frac{\pi}{2}\right] \left[1 + \alpha^2(t - \phi)^2\right]} \quad t \geq 0, \quad (2)$$

$$h_T(t) = \frac{f_T(t)}{S_T(t)} = \frac{\alpha}{\left[\arctan[\alpha(\phi - t)] + \frac{\pi}{2}\right] \left[1 + \alpha^2(t - \phi)^2\right]} \quad t \geq 0. \quad (3)$$

This arctangent distribution is equivalent to a Cauchy distribution truncated on the left at  $t = 0$ . Thus the parameter  $\alpha$  is similar to a scale parameter and the parameter  $\phi$  is similar to a phase shift or location parameter. Figure 1 shows three different pdfs of the arctangent

distribution. Notice that as  $\phi$  becomes negative, the distribution changes from a bell-shaped distribution to a distribution with a mode of 0. Also notice that the parameter  $\alpha$  controls the “peakedness” of the distribution: the dispersion of the distribution is a decreasing function of  $\alpha$ .

We initially chose the name “Arctangent Distribution,” because of the similarities that the arctangent function has with a generic survivor function. The relationship to the Cauchy distribution was noticed later. The cdf of the Cauchy distribution is also a shifted and scaled arctangent function.

## Probabilistic Properties

The arctangent distribution has several useful probabilistic properties that make it a viable distribution for lifetime data analysis. Specifically, it enjoys closed-form survivor and hazard functions, unlike most distributions in the UBT class. The closed-form survivor function simplifies parameter estimation for censored data sets and allows for variate generation via inversion. The probabilistic properties include:

- The distribution is in the UBT class when  $\alpha\phi > c$  and is in the decreasing failure rate (DFR) class when  $\alpha\phi \leq c$  where

$$1 + 2c \arctan(c) + c\pi = 0,$$

found by simplifying  $h'_T(0) = 0$ . Using numerical methods,  $c \approx -0.42898$ .

- The mode of the distribution is  $t_{\text{mode}} = \phi$ ,  $\forall \phi > 0$ . For the case where the probability density function is monotonically decreasing,  $t_{\text{mode}} = 0$ ,  $\forall \phi \leq 0$ .
- The  $p^{\text{th}}$  fractile of the distribution is

$$t_p = \phi + \frac{1}{\alpha} \tan \left[ \frac{\pi}{2} - (1-p) \left( \arctan(\alpha\phi) + \frac{\pi}{2} \right) \right], \quad (4)$$

which yields a median of

$$t_{0.5} = \phi + \frac{1}{\alpha} \tan \left[ \frac{\pi}{4} - \frac{1}{2} \arctan(\alpha\phi) \right].$$

This closed-form expression for the fractiles will be useful in determining initial estimates for the numerical methods required to determine the maximum likelihood estimates.

- Variates can be generated via inversion by

$$T \leftarrow \phi + \frac{1}{\alpha} \tan \left[ \frac{\pi}{2} - (1 - U) \left( \arctan(\alpha\phi) + \frac{\pi}{2} \right) \right],$$

where  $U$  is uniformly distributed between 0 and 1.

- The conditional survivor function is given by

$$\begin{aligned} S_{T|T>a}(t) &= \frac{S_T(t)}{S_T(a)} = \frac{\arctan[\alpha(\phi - t)] + \frac{\pi}{2}}{\arctan[\alpha(\phi - a)] + \frac{\pi}{2}} = \frac{\arctan[\alpha(\phi - a) - (t - a)] + \frac{\pi}{2}}{\arctan[\alpha(\phi - a)] + \frac{\pi}{2}} \\ &= \frac{\arctan[\alpha(\gamma - y)] + \frac{\pi}{2}}{\arctan[\alpha\gamma] + \frac{\pi}{2}}, \end{aligned}$$

for  $y \geq 0$ , which is again an arctangent distribution with the same  $\alpha$  as the unconditional distribution and where  $\gamma = \phi - a$  and  $y = t - a$ .

- The limiting distribution as  $\alpha \rightarrow \infty$  is a degenerate distribution at  $\phi$ .
- The arctangent distribution is related to the Cauchy distribution. Specifically, a Cauchy distribution truncated on the left at zero will yield the arctangent distribution. Not surprisingly, the mean and higher order moments of the distribution are undefined. This poses a challenge when discussing mean time to failure (MTTF) of components. Since the mean is undefined, central tendencies of the distribution could be discussed in terms of its mode and median. The disadvantage of this limitation is that practitioners are generally more comfortable with the mean as the primary measure of central tendency.
- The arctangent distribution has a heavy right tail which makes it useful for evaluating items that fail with less risk once it has survived a certain time threshold. Certain biostatistical data sets indicate such heavy right tails in cancer data. The arctangent distribution is capable of modeling lifetime distributions with a heavier tail than the log normal or log logistic. We will show this in an example where the distribution models the survival time of rats given a cancer accelerating drug, where there is a heavy right tail. Competing risks models are also useful in modeling heavy right tails.

# Statistical Inference

We now present some straightforward statistical inference of the distribution for complete and right-censored data sets. We then compare the results with other distributions, including other commonly used UBT distributions. The arctangent distribution requires numerical methods to determine the maximum likelihood estimators of its parameters, which is typical of most two-parameter lifetime models. First, we present statistical inference procedures for uncensored data using a reliability example. Second, we illustrate inferences with censored data using a biostatistical data set. Finally, we will present a biostatistical example that illustrates the heavy tail of the distribution.

For the uncensored case, let  $t_1, t_2, \dots, t_n$  be the failure times. The likelihood function is

$$L(\alpha, \phi) = \prod_{i=1}^n f(t_i, \alpha, \phi) = \prod_{i=1}^n \frac{\alpha}{\left[ \arctan(\alpha\phi) + \frac{\pi}{2} \right] \left[ 1 + \alpha^2(t_i - \phi)^2 \right]}.$$

The first partial derivatives of  $\log L(\alpha, \phi)$  with respect to the two parameters yield

$$\frac{\partial \log L(\alpha, \phi)}{\partial \alpha} = \frac{-n\phi}{\left[ 1 + (\alpha\phi)^2 \right] \left[ \frac{\pi}{2} + \arctan(\alpha\phi) \right]} + \frac{n}{\alpha} + \sum_{i=1}^n \frac{-2\alpha(t_i - \phi)^2}{\left[ 1 + \alpha^2(t_i - \phi)^2 \right]} \quad (5)$$

and

$$\frac{\partial \log L(\alpha, \phi)}{\partial \phi} = \frac{-n\alpha}{\left[ 1 + (\alpha\phi)^2 \right] \left[ \frac{\pi}{2} + \arctan(\alpha\phi) \right]} + \sum_{i=1}^n \frac{2\alpha^2(t_i - \phi)}{\left[ 1 + \alpha^2(t_i - \phi)^2 \right]}. \quad (6)$$

Equating (5) and (6) to zero does not yield closed-form solutions for the maximum likelihood estimators  $\hat{\alpha}$  and  $\hat{\phi}$ . For the numerical methods to be effective in finding  $\hat{\alpha}$  and  $\hat{\phi}$  it is necessary to have appropriate initial estimates of the parameters. Since the mean and higher order moments are undefined, one must rely on a “method of fractiles”, as opposed to the method of moments, to find initial estimates  $\hat{\alpha}_0$  and  $\hat{\phi}_0$ . This entails an initial system of two equations based on the  $p^{\text{th}}$  fractile of the the distribution, where the fractiles are chosen based on the test data.

To illustrate the applicability of the arctangent distribution, we will use Lieblein and Zelen’s (1956) data set of  $n = 23$  ball bearing failure times (each measurement in  $10^6$

revolutions):

17.88	28.92	33.00	41.52	42.12	45.60	48.48
51.84	51.96	54.12	55.56	67.80	68.64	68.64
68.88	84.12	93.12	98.64	105.12	105.84	127.92
128.04	173.40					

Although it's an old data set, we start with this example because Crowder et al. (1991, page 63) conjectured that the UBT shaped distributions may fit the ball bearing data better than the IFR distributions, based on the values of the log likelihood function at the maximum likelihood estimators.

Using the “method of fractiles” to find initial parameter estimates for the parameters, we note from the empirical survivor function for the data (Figure 2) that time 42.12 corresponds to the  $\frac{5}{23} \cdot 100 = 22.7^{\text{th}}$  percentile of the distribution and that time 105.12 corresponds to the  $\frac{19}{23} \cdot 100 = 82.6^{\text{th}}$  percentile of the distribution. Thus using (4), initial estimates for the MLEs are found by solving

$$42.12 = \phi + \frac{1}{\alpha} \tan \left[ \frac{\pi}{2} - \left( 1 - \frac{5}{23} \right) \left( \arctan(\alpha\phi) + \frac{\pi}{2} \right) \right],$$

$$105.12 = \phi + \frac{1}{\alpha} \tan \left[ \frac{\pi}{2} - \left( 1 - \frac{19}{23} \right) \left( \arctan(\alpha\phi) + \frac{\pi}{2} \right) \right]$$

for  $\alpha$  and  $\phi$ . This system yields our initial estimates of the parameters as follows:  $\hat{\alpha}_0 = 0.04102$  and  $\hat{\phi}_0 = 57.96$ . Using these values as initial estimates, we may now solve equations (5) and (6) numerically yielding  $\hat{\alpha} = 0.04238$  and  $\hat{\phi} = 58.08$ . Taking second partial derivatives of the log likelihood function and evaluating at the MLEs yields the  $2 \times 2$  observed information matrix (see Cox and Oakes, 1984)

$$I = \begin{pmatrix} 5989 & 1.305 \\ 1.305 & 0.021 \end{pmatrix}.$$

Inverting the matrix and taking the square roots of the diagonal elements gives asymptotically valid 95% approximate confidence intervals for  $\alpha$  and  $\phi$

$$0.01727 < \alpha < 0.06753$$

$$44.46 < \phi < 71.70.$$

Figure 2 gives a graphical comparison of the arctangent fit versus the Weibull fit of the empirical data. We can see the fits are virtually identical in the early stages, then the arctangent fits more closely than does the Weibull in the center. At the right tail, the Weibull fits closer, due to the arctangent distribution's propensity for a heavy tail. Finally, we can compare the arctangent distribution's model adequacy with that of other popular two-parameter lifetime distributions. The Kolmogorov–Smirnov (K–S) goodness-of-fit statistic for the arctangent distribution is  $D_n = 0.093$ . Table 1 gives  $D_n$  values for some popular distributions fitted to the ball bearing data evaluated at the maximum likelihood estimators of their parameters. (Chhikara and Folks (1989, page 74) fit the inverse Gaussian distribution to this data set.) The K-S statistic is one measure to gauge quality of fit. Thus, as we see in Table 1, the lower K-S values of the three UBT distributions, the inverse Gaussian, arctangent, and log normal, add further credibility to Crowder's conjecture that UBT models seem to fit this data set better than those with increasing failure rates (i.e., the Weibull and gamma distributions). Using a K–S statistic with estimated parameters is problematic. Therefore, we use this statistic not for a formal test, but to further the conjecture that UBT distributions fit this data better.

We now consider statistical inference for censored data. The arctangent distribution's closed-form survivor function yields a closed-form likelihood function, thus simplifying the analysis of right-censored data. The only other UBT distribution with this property is the log logistic distribution. The statistical methods are similar to those of the uncensored case; however, the numerical methods are a bit more tedious. We will use Gehan's (1965) test data of remission times from the drug 6-MP when used on  $n = 21$  leukemia patients of which there were  $r = 9$  observed remissions and 12 individuals who were randomly right censored. Letting an asterisk denote a right-censored observation, the remission times in weeks are:

6	6	6	6*	7	9*	10
10*	11*	13	16	17*	19*	20*
22	23	25*	32*	32*	34*	35*

To fit this data to the arctangent distribution, let  $t_1, t_2, \dots, t_n$  be the remission times and  $c_1, c_2, \dots, c_n$  be the associated censoring times. Our maximum likelihood estimation

is now be based on the likelihood function

$$\begin{aligned}
 L(\alpha, \phi) &= \prod_{i \in U} f(t_i, \alpha, \phi) \prod_{i \in C} S(c_i, \alpha, \phi) \\
 &= \prod_{i \in U} \frac{\alpha}{\left[ \arctan(\alpha\phi) + \frac{\pi}{2} \right] \left[ 1 + \alpha^2(t_i - \phi)^2 \right]} \prod_{i \in C} \frac{\arctan(\alpha(\phi - c_i)) + \frac{\pi}{2}}{\arctan(\alpha\phi) + \frac{\pi}{2}},
 \end{aligned}$$

where  $U$  and  $C$  are the sets of indices of uncensored and censored observations, respectively.

The log likelihood function is

$$\begin{aligned}
 \log L(\alpha, \phi) &= r \log \alpha - r \log \left[ \arctan(\alpha\phi) + \frac{\pi}{2} \right] - \sum_{i \in U} \log[1 + \alpha^2(t_i - \phi)^2] \\
 &\quad + \sum_{i \in C} \log[2 \arctan[\alpha(\phi - c_i)] + \pi] - (n - r) \log[2 \arctan(\alpha\phi) + \pi].
 \end{aligned}$$

A “method of fractiles” initial estimate for the parameters yields:  $\hat{\alpha}_0 = 0.0562$  and  $\hat{\phi}_0 = 9.58$ . Now we take the two partial derivatives of  $\log L$  with respect to  $\alpha$  and  $\phi$ , set them equal to zero, and compute  $\hat{\alpha} = 0.0455$  and  $\hat{\phi} = 11.2$ . This example illustrates the methodology used to fit the arctangent distribution to censored data sets.

A third example illustrates the usefulness of the distribution’s heavy right tail. Cox and Snell (1981, page 169) present data on life span of rats who have been given a cancer accelerator. The following complete data set gives number of days the rats survived:

37 38 42 43 43 43 43 43 48 49  
51 51 55 57 59 62 66 69 86 177

The arctangent MLE estimates for this data set yield  $\hat{\alpha} = 0.127$  and  $\hat{\phi} = 48.0$  which is plotted against the empirical survivor function in Figure 3. Note how the heavy right tail of the arctangent distribution models the heavy right tail of the rat lifetimes. The rat with the survival time of 177 is driving the fit. Thus, for data sets with heavy tails such as this one, the practitioner may make use of the arctangent distribution’s propensity for a heavy tail.

## Conclusion

The arctangent distribution is a new two-parameter lifetime distribution in the UBT class with closed-form survivor functions. It gives reliability engineers, biostatisticians, and others another tool in the complex task of statistical modeling. Although a UBT model has a

smaller number of applications than does the IFR or bathtub-shaped models, there are enough references in the literature to indicate a need for more distributions in this class. This paper is presented to give the UBT model a second distribution that enjoys a closed-form survivor function and has been demonstrated to adequately describe well-known data sets.

## Acknowledgements

The authors acknowledge the contributions of Dr. Donald Barr and two anonymous referees to this paper.

## References

- CHHIKARA, R. S. and FOLKS, L. S. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications*. Marcel Dekker, Inc., New York, NY.
- COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data*. Chapman and Hall, New York, NY.
- COX, D. R. and SNELL, E. J. (1981). *Applied Statistics*. Chapman and Hall, New York, NY.
- CROWDER, M. J., KIMBER, A. C., SMITH, R. L. and SWEETING, T. J. (1991). *Statistical Analysis of Reliability Data*. Chapman and Hall, New York, NY.
- GEHAN, E. A. (1965). "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples." *Biometrika* 52, pp. 203–223.
- KECECIOGLU, D. (1991). *Reliability Engineering Handbook, Volume I*. Prentice–Hall, Inc., Englewood Cliffs, NJ.
- LEE, E. T. (1992). *Statistical Methods for Survival Data Analysis*, Second Edition. John Wiley and Sons, New York, NY.
- LIEBLEIN, J. and ZELEN, M. (1956). "Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings." *Journal of Research of the National Bureau of Standards* 57, pp. 273–316.

Key Words: *Censored Data, Hump-Shaped Hazard Functions, Lifetime Distributions, Maximum Likelihood, Survival Analysis, Trigonometric Functions.*

TABLE 1. Kolmogorov–Smirnov Goodness-of-fit Statistics for the Ball Bearing Data.

Distribution	$D_{23}$
Exponential	0.301
Weibull	0.152
Gamma	0.123
Inverse Gaussian	0.099
Arctangent	0.093
Log normal	0.090

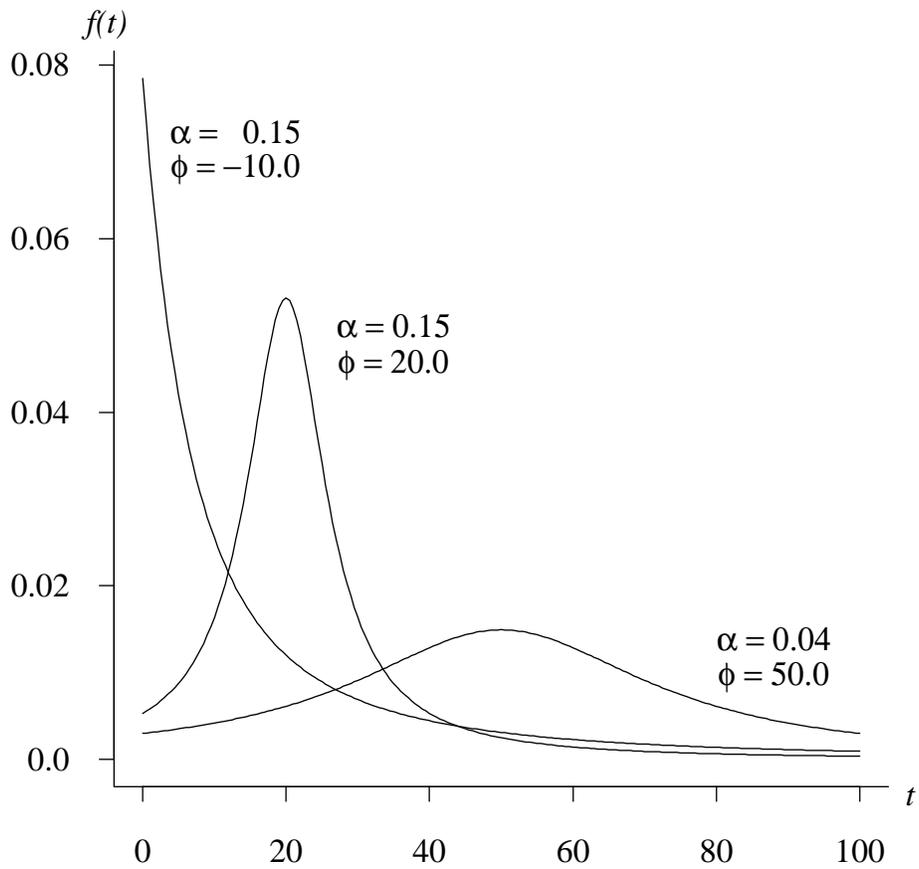


Figure 1: Examples of the arctangent probability density function.

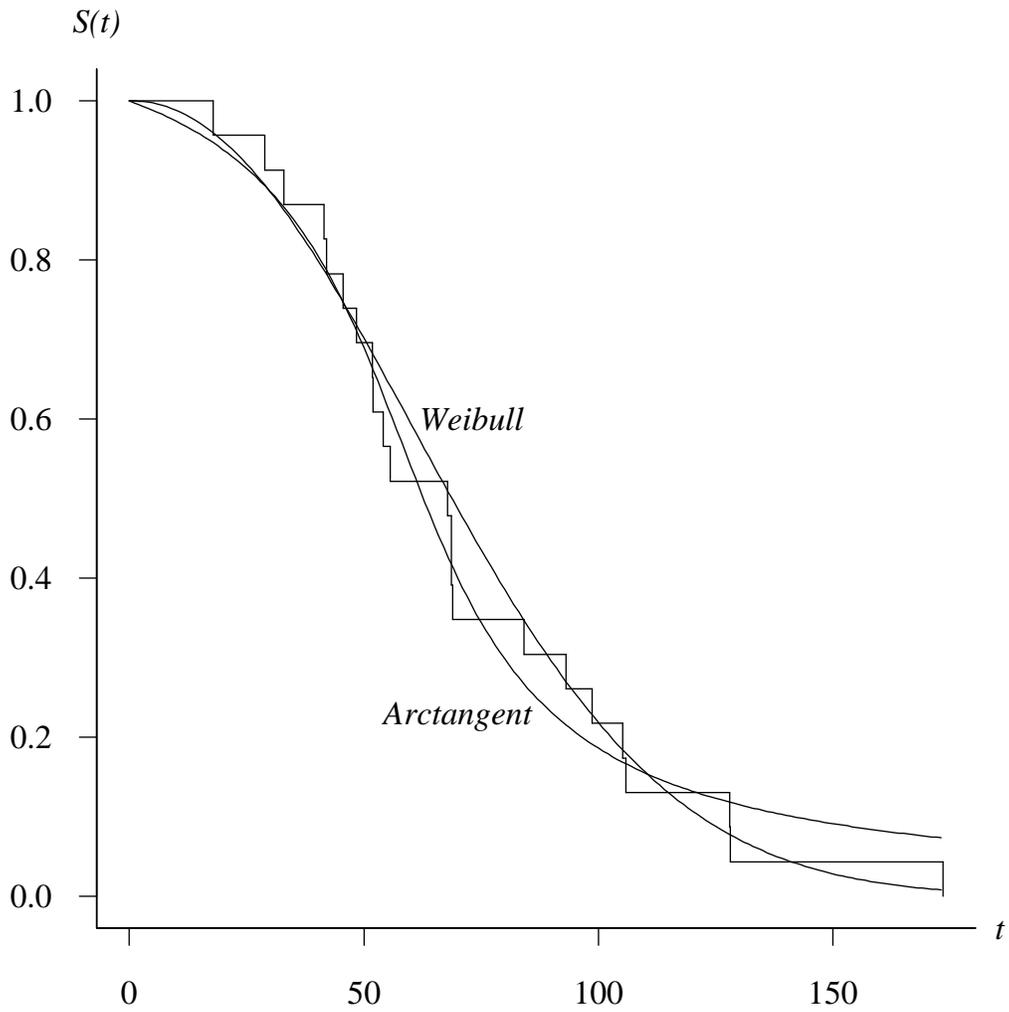


Figure 2: Empirical, fitted arctangent, and fitted Weibull survivor functions for the ball bearing lifetimes.

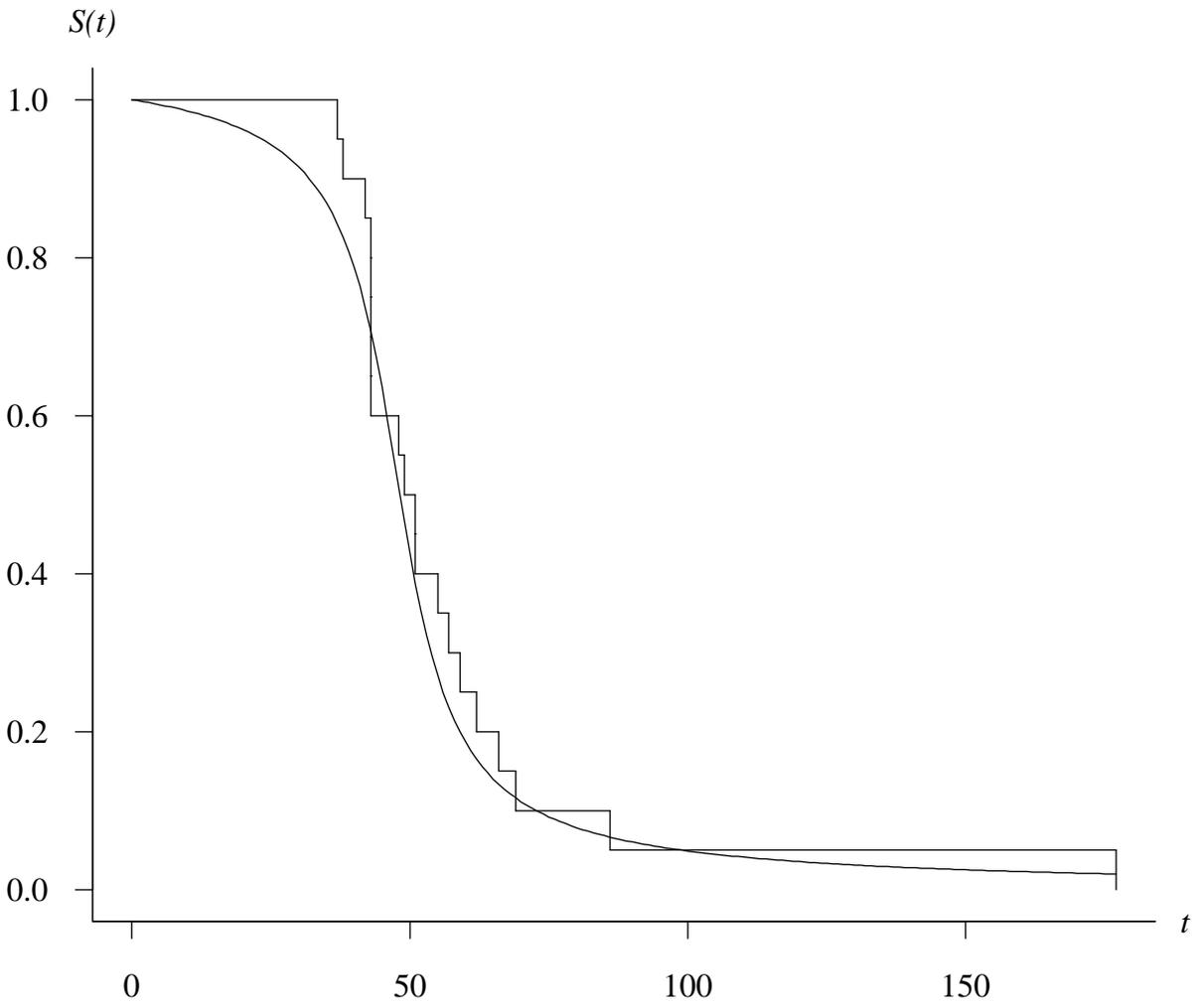


Figure 3: The arctangent distribution fit to the rat cancer data.