Parametric Bootstrap Procedures for Censored Data.

Russell Zaretzki

rzaretzk[at]utk.edu

University of Tennessee, Knoxville, USA

Introduction

- Bootstrap methods for regular parametric models have been heavily studied over the previous 25 years.
- Coverage probabilities of bootstrap based confidence intervals have been investigated by numerous researchers.
- Jeng, Lahiri and Meeker (2005), investigated the use of the bootstrap in parametric models when the data is censored by time.

Outline.

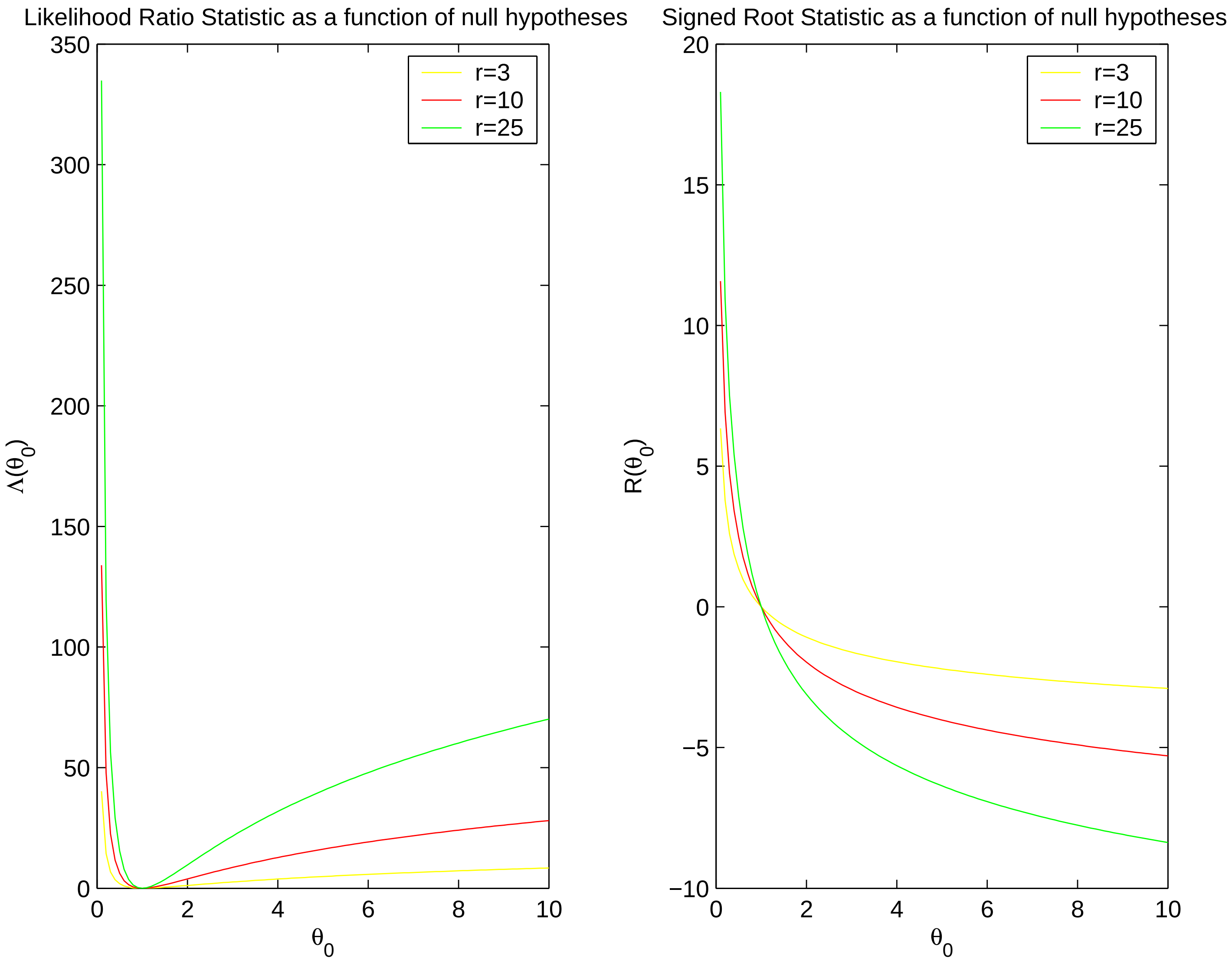
- Models and Statistics of Interest.
- Bootstrap Theory in Uncensored Case.
- Results of Jeng, Lahiri and Meeker(1999,2005)
- The Null Hypothesis Bootstrap.
- Results for Exponential Distribution.
- Results for the Weibull/Extreme Value Distribution.

Basic Statistics of Interests.

The likelihood ratio statistic, $\Lambda(\theta)$, and its signed root counterpart, $R(\theta)$, are two of the most basic tools for statistical inference. When interest focuses on a scaler parameter of interest θ_0 in the presence of a vector nuisance parameter ψ, we write

$$\begin{aligned}
\Lambda(\theta_0) &= 2\left\{l(\hat{\theta},\hat{\psi}) - l(\theta_0,\hat{\psi})\right\} \\
R(\theta_0) &= sign(\hat{\theta}-\theta_0)\sqrt{\Lambda(\theta_0)} \\
W(\theta_0) &= (\hat{\theta}-\theta_0)/i(\hat{\theta},\hat{\psi})^{-1/2}
\end{aligned}$$

Graphics for Λ and R



Basic Asymptotic Theory.

In the case of no censoring:

- When using LR statistics with the chi-squared approximation to create 2-sided confidence intervals coverage errors decrease with rate $O(n^{-1})$.
- When using signed root statistic or the Wald statistic with a normal approximation, coverage errors decrease with rate $O(n^{-1/2})$.

Parametric Bootstrap CI.

- Fit models using maximum likelihood.
- Simulate data sets using model at fitted MLE.
- Compute statistic on each simulated data set. Use $\theta = \hat{\theta}$ as the true value.
- Use distribution of simulated statistics to find CI endpoints.

Bootstrap Asymptotic Theory.

In the case of no censoring:

- When using LR statistics with the bootstrap approximation to create 2-sided confidence intervals coverage errors decrease with rate $O(n^{-2})$.
- When using signed root statistic or the Wald statistic with a normal approximation, coverage errors decrease with rate $O(n^{-1})$.

Earlier Work.

- Jeng, Lahiri, and Meeker(2005). Extreme Value Failure Times. Inference for scale parameters and percentiles of the distribution.
- Conditioned on the situation where the number of failures $r \geq 2$.
- One sided confidence bounds. The parametric bootstrap signed root statistic is adequate when the number of failures is 15 or greater.
- For two sided intervals, the LLR is adequate when expected failures is greater than 15. If Bartlett correction is used then the LLR statistic is adequate when expected number of failures is as small as 7.

The Exponential Distribution.

Two common models for lifetime data.

1. Exponential.

$$f(t) = \theta^{-1} \exp(-t/\theta), \qquad t \geq 0$$

2. Log-likelihood.

$$l(\theta) = -r \log \theta - \frac{1}{\theta} \sum_{i=1}^{n} t_i$$

r is the number of non-censored observations.

Exp. Cont.

1. Maximum Likelihood of Estimators.

$$\hat{\theta} = \sum_{i=1}^{n} t_i / r, \quad r > 0$$

2. Likelihood Ratio Statistic,

$$\Lambda(\theta_0) = 2r\{(\hat{\theta}/\theta_0) - 1 - \log(\hat{\theta}/\theta_0)\}$$

3. Signed Root.

$$R(\theta) = sign(\hat{\theta} - \theta_0)\sqrt{\Lambda(\theta_0)}$$

4. No nuisance parameters in this model.

Important Notes

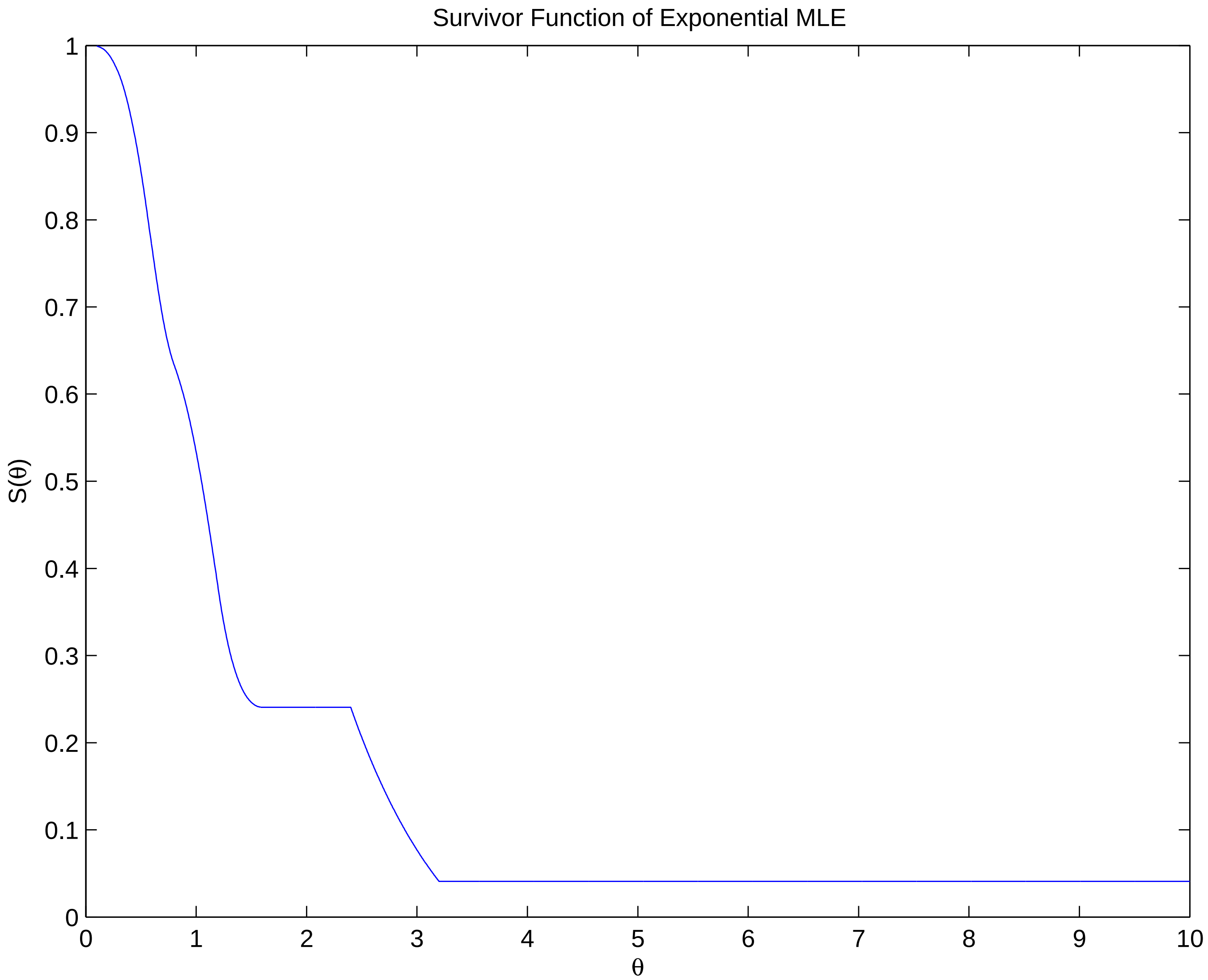
1. No nuisance parameters in this model.
2. For fixed θ_0, the signed root is maximized when r=0.
3. Signed Root is a one-to-one function of the MLE.

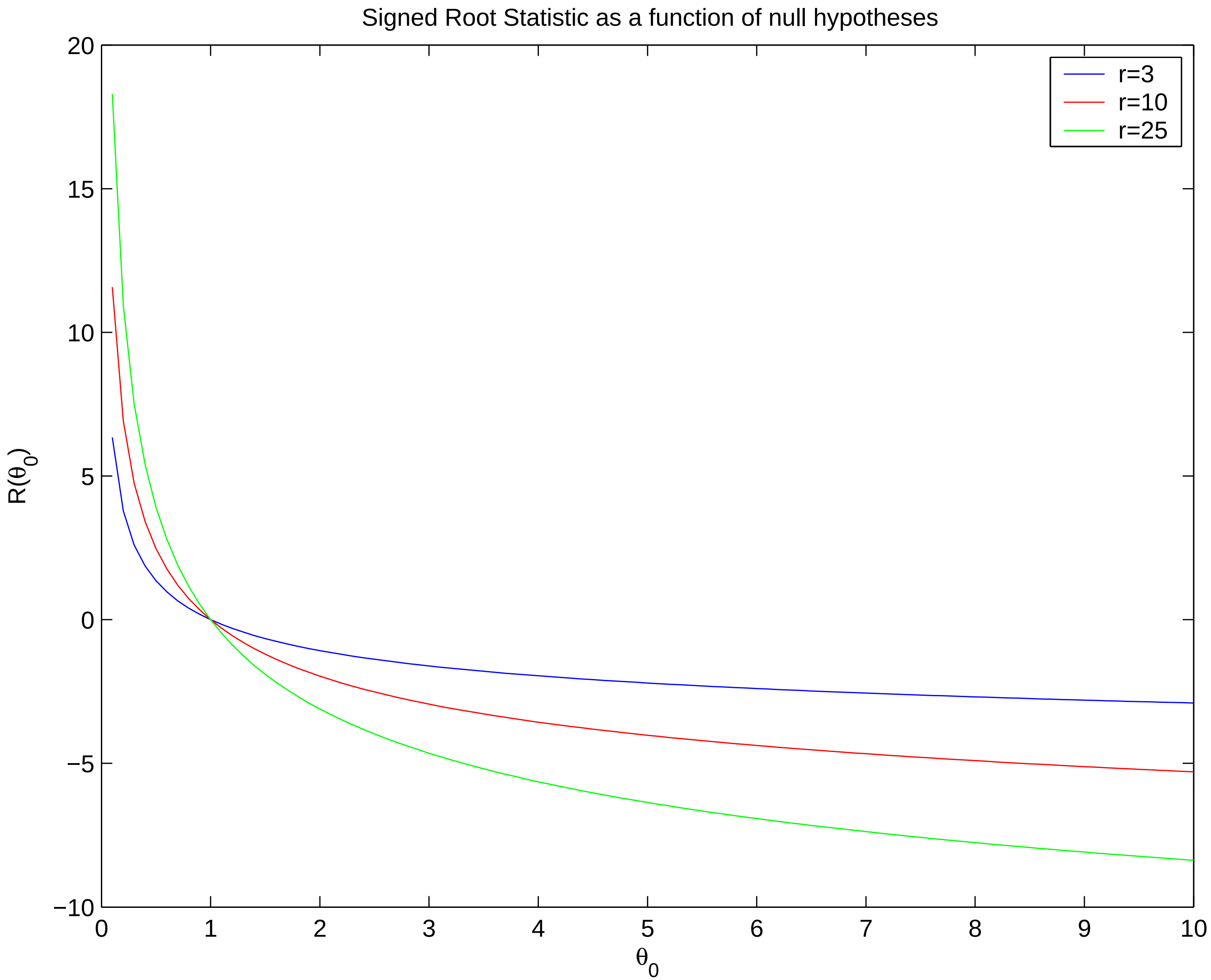
Computing CI's for R.

Given confidence level α, compute upper and lower confidence bounds for the parameter θ we need to solve two equations:

- $$\theta_L = \{\theta : P(R(\theta) > r_{data}(\theta)) = \alpha\}$$
- $$\theta_U = \{\theta : P(R(\theta) > r_{data}(\theta)) = 1 - \alpha\}$$
- Note: A difficulty may arise if a point mass exists in the distribution.

Probability Mass.



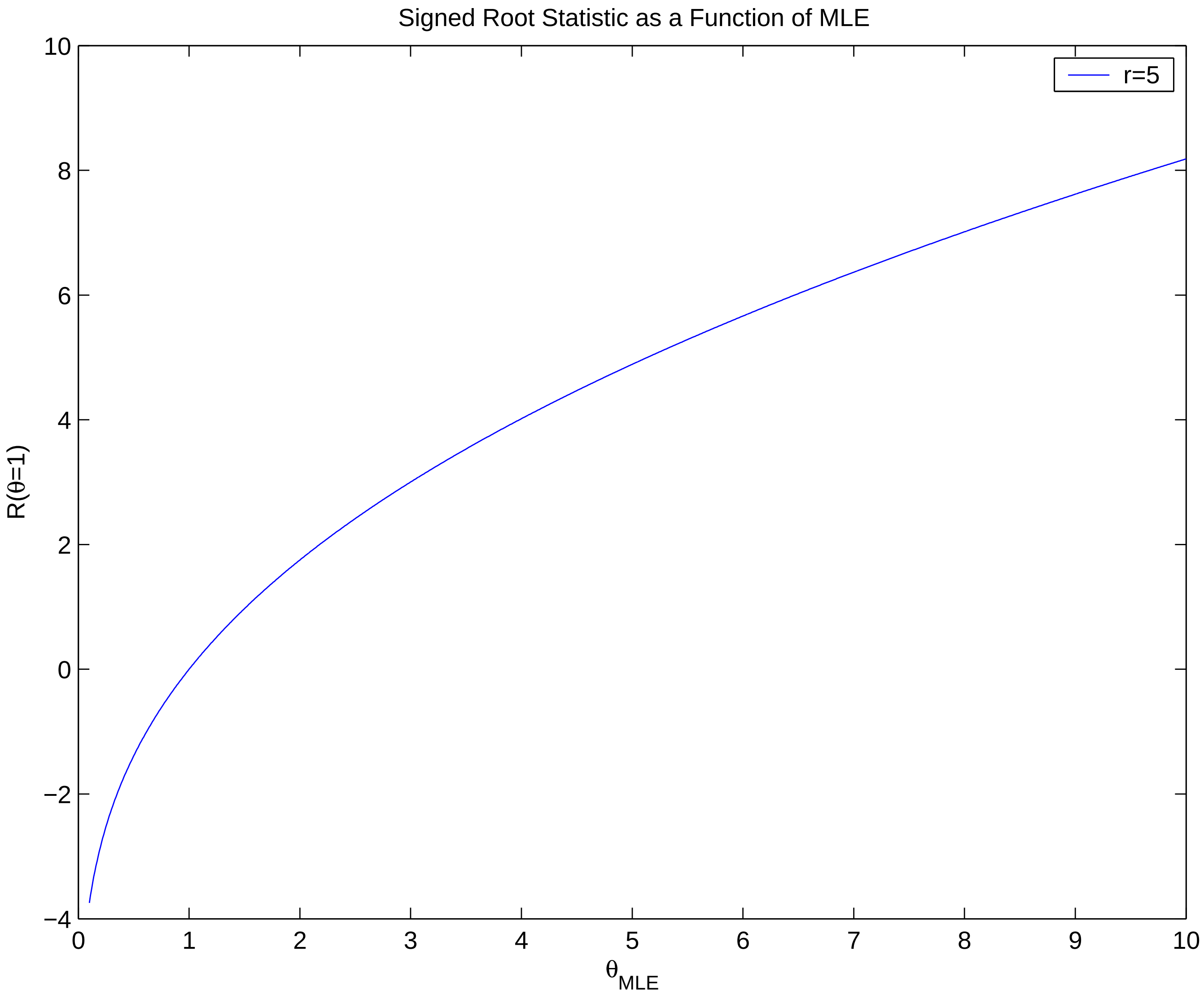


Signed Root Statistic as a function of null hypotheses
r=3
r=10
r=25
R(θ0)
θ0
20
15
10
5
0
−5
−10
0
1
2
3
4
5
6
7
8
9
10

Null Parametric Bootstrap.

- Fit models using maximum likelihood.
- Simulate data sets using a grid of values of θ as the true parameter.
- Compute the statistic on each simulated data set.
- Use distribution of simulated statistics to find CI endpoints.

Signed Root Function



Simulation Results: N=4, Right Tail Probabilities

	Exact			Null SR		
Prob Cens.	.8	.5	.368	.8	.5	.368
.01	0.1362	0.1924	0.2313	0.1363	0.1908	0.228
.025	**0.1654**	**0.2233**	**0.2676**	**0.1653**	**0.2222**	**0.2638**
.1	0.2490	0.3014	0.3584	0.2488	0.3004	0.3543
.90	3.9398	1.4075	1.5706	3.9317	1.4052	1.5844
.95	7.8843	1.8885	2.0799	7.8949	1.8905	2.1016
.99	39.3345	3.5167	3.7808	39.1998	3.55	3.7877

	ML SR		
Prob Cens.	.8	.5	.368
.01	0	0.2003	0.2178
.025	**0**	**0.2134**	**0.2387**
.1	0	0.3040	0.3445
.90	3.3193	1.4778	1.7484
.95	7.4784	2.0059	2.3465
.99	30.5026	3.6635	4.3886

The Weibull Distribution

- A popular distribution for fitting lifetime data ,

$$f(t) = \lambda\beta(\lambda t)^{\beta-1}\exp\Big[-(\lambda t)^{\beta}\Big], \quad t > 0$$

- The parameter β is known as the shape parameter while λ is called the scale parameter.
- The shape parameter controls the overall shape of the distribution while λ just changes the scale of the horizontal axis t.

More about the Weibull

- If $T \sim W(\lambda, \beta)$ then $Y = \log T$ has the extreme value distribution.

$$f(x) = \frac{1}{b} \exp\left[\frac{x-\mu}{b} - \exp\left(\frac{x-\mu}{b}\right)\right];$$

$-\infty < x < \infty$ and $b > 0$

- This distribution is useful because it is a location-scale family.