Chapter 6

Probability Plotting

William Q. Meeker and Luis A. Escobar
Iowa State University and Louisiana State University

Based on the authors' text *Statistical Methods for Reliability Data*, John Wiley & Sons Inc. 1998.

July 18, 2002
12h 24min

Chapter 6
Probability Plotting
Objectives

- Describe **applications** for probability plots.
- Explain the basic **concepts** of probability plotting.
- Show how to **linearize** a cdf on special plotting scales.
- Explain how to plot a nonparametric estimate $\widehat{F}$ to judge the adequacy of a particular parametric distribution.
- Explain methods of separating **useful** information from **noise** when interpreting a probability plot.
- Use a probability plot to obtain **graphical** estimates of reliability characteristics like failure probabilities and quantiles.

Purposes of Probability Plots

Probability plots are used to:

- Assess the adequacy of a particular distributional model.
- To detect multiple failure modes or mixture of different populations.
- Obtain graphical estimates of model parameters (e.g., by fitting a straight line through the points on a probability plot).
- Displaying the results of a parametric maximum likelihood fit along with the data.
- Obtain, by drawing a smooth curve through the points, a semiparametric estimate of failure probabilities and distributional quantiles.

Probability Plotting Scales: Linearizing a CDF

Main Idea: For a given cdf, $F(t)$, one can **linearize** the $\{ \ t$ versus $F(t) \ \}$ plot by:

- Finding transformations of $F(t)$ and t such that the relationship between the transformed variables is linear.
- The transformed axes can be relabeled in terms of the original probability and time variables.

  The resulting probability axis is generally nonlinear and is called the **probability** scale. The data axis is usually a linear axis or a log axis.

Linearizing the Exponential CDF

CDF: $p = F(t;\theta,\gamma) = 1 - \exp\left[-\frac{(t-\gamma)}{\theta}\right], \quad t \ge \gamma.$

Quantiles : $t_p = \gamma - \theta \log(1-p).$

Conclusion:

The $\{\ t_p$ versus $-\log(1-p)\ \}$ plot is a straight line.

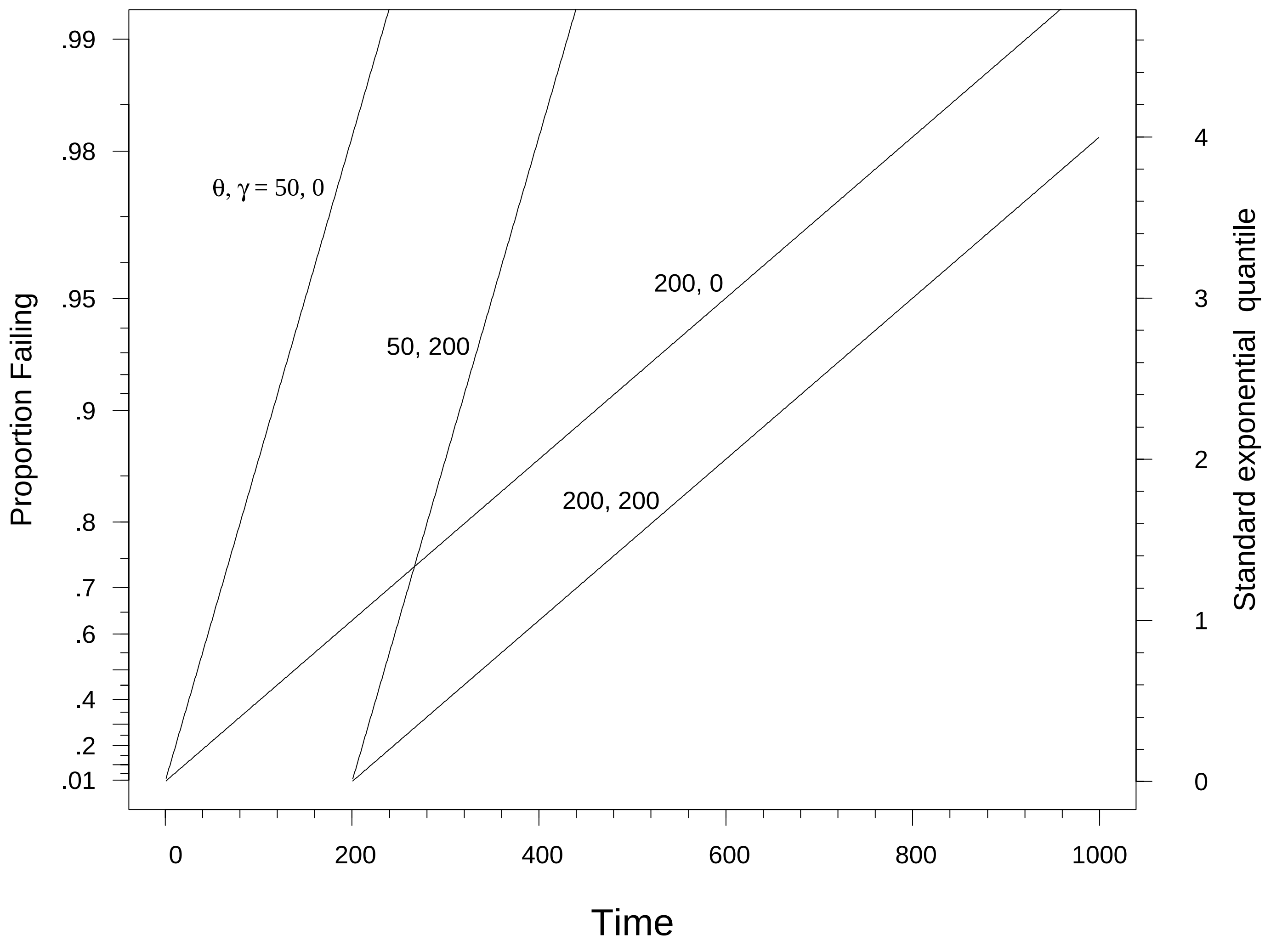
We plot t_p on the horizontal axis and p on the vertical axis. γ is the **intercept** on the time axis and $1/\theta$ is equal to the slope of the cdf line.

Note:

Changing θ changes the slope of the line and changing γ changes the position of the line.

Plot with Exponential Distribution Probability Scales Showing Exponential cdfs as Straight Lines for Combinations of Parameters $\theta = 50, 200$ and $\gamma = 0, 200$

$$t_p = \gamma - \theta \log(1 - p)$$



Linearizing the Normal CDF

CDF: $p = F(y; \mu, \sigma) = \Phi_{\text{nor}}\left(\frac{y-\mu}{\sigma}\right), \quad -\infty < y < \infty.$

Quantiles : $y_p = \mu + \sigma \Phi_{\text{nor}}^{-1}(p).$

$\Phi_{\text{nor}}^{-1}(p)$ is the p quantile of the standard normal distribution.

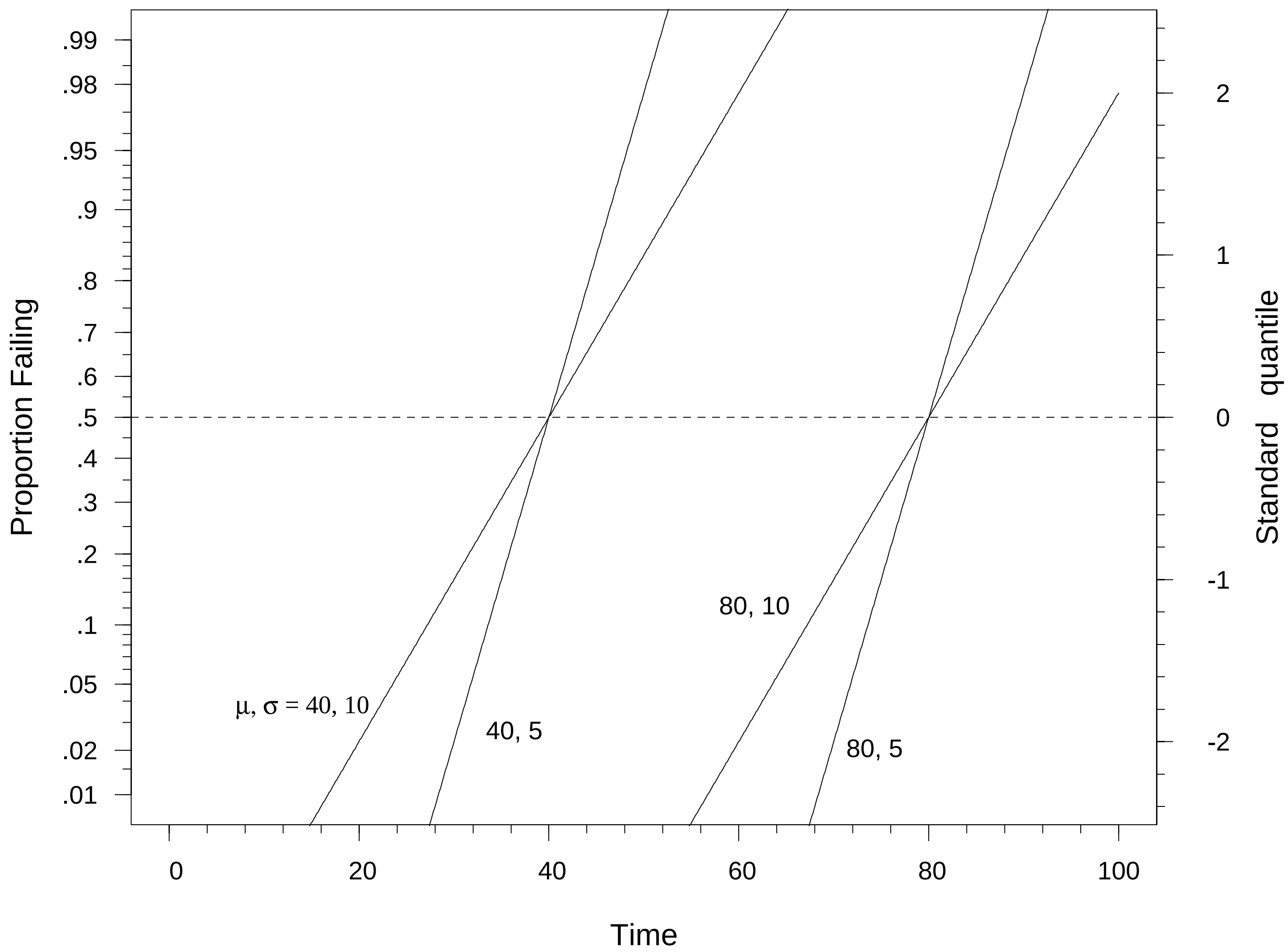
Conclusion:
$\{\ y_p$ versus $\Phi_{\text{nor}}^{-1}(p)\ \}$ will plot as a straight line.

μ is the point at the time axis where the cdf intersects the $\Phi^{-1}(p) = 0$ line (i.e., $p = .5$). The slope of the cdf line on the graph is $1/\sigma$.

Note:
Any normal cdf plots as a straight line with positive slope. Also, any straight line with positive slope corresponds to a normal cdf.

Plot with Normal Distribution Probability Scales Showing Normal cdfs as Straight Lines for Combinations of Parameters $\mu = 40, 80$ and $\sigma = 5, 10$

$$y_p = \mu + \sigma \Phi^{-1}_{\text{nor}}(p)$$



Linearizing the Lognormal CDF

CDF: $$p = F(t;\mu,\sigma) = \Phi_{\text{nor}}\left[\frac{\log(t)-\mu}{\sigma}\right], \quad t > 0.$$

Quantiles : $$t_p = \exp\left[\mu + \sigma\Phi_{\text{nor}}^{-1}(p)\right].$$

Then $\log(t_p) = \mu + \Phi_{\text{nor}}^{-1}(p)\sigma$

Conclusion:

$\{\ \log(t_p)$ versus $\Phi_{\text{nor}}^{-1}(p)\ \}$ will plot as a straight line.

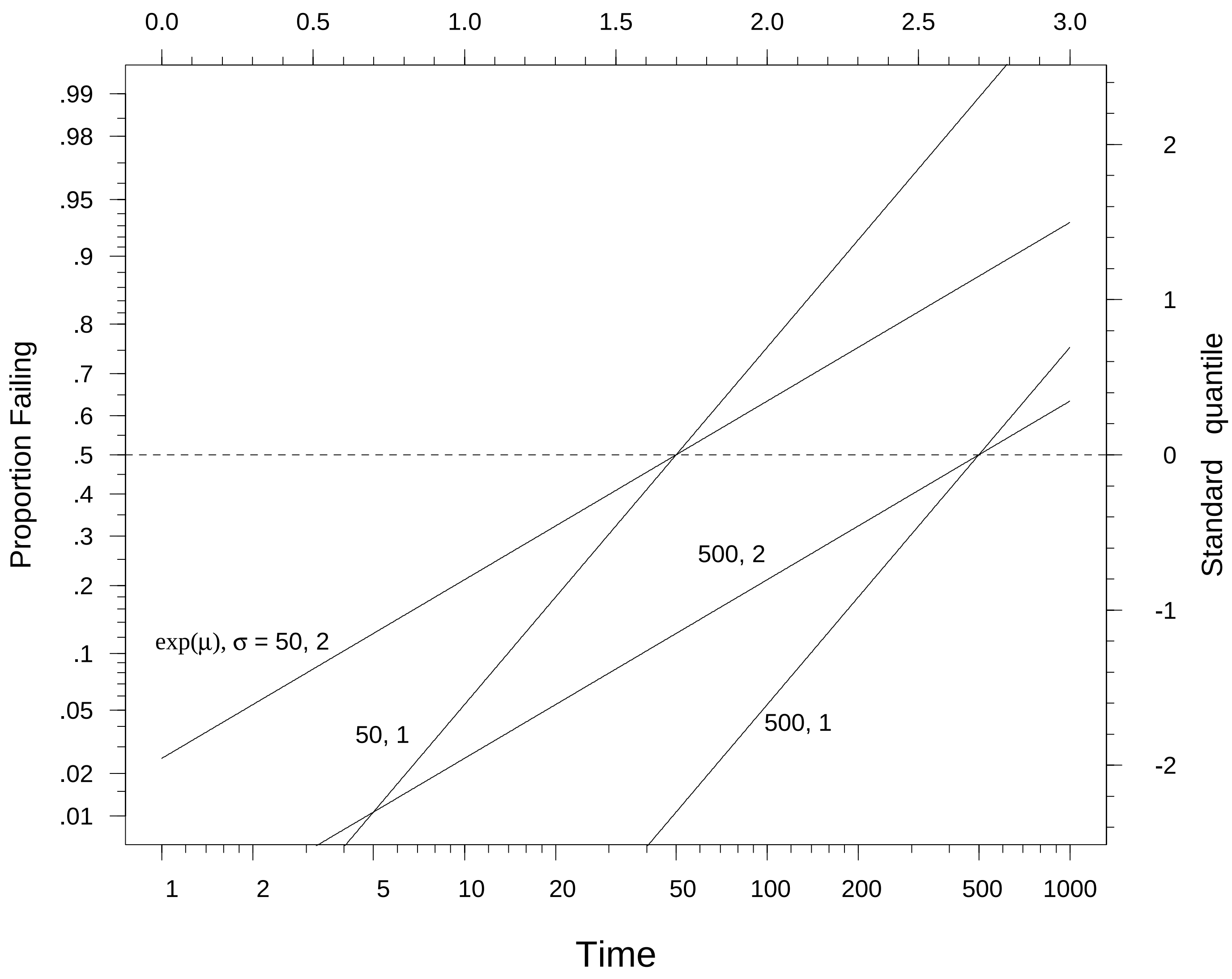
$\exp(\mu)$ can be read from the time axis at the point where the cdf intersects the $\Phi_{\text{nor}}^{-1}(p) = 0$ line. The slope of the cdf line on the graph is $1/\sigma$ (but in the computations use base e logarithms for the times rather than the base 10 logarithms used for the figures).

Note:

Any given lognormal cdf plots as a straight line with positive slope. Also, any straight line with positive slope corresponds to a lognormal distribution.

Plot with Lognormal Distribution Probability Scales Showing Lognormal cdfs as Straight Lines for Combinations of $\exp(\mu) = 50, 500$ and $\sigma = 1, 2$

$$\log(t_p) = \mu + \Phi_{\text{nor}}^{-1}(p)\sigma$$



Linearizing the Weibull CDF

CDF: $\quad p = F(t;\mu,\sigma) = \Phi_{\text{sev}}\left[\frac{\log(t)-\mu}{\sigma}\right], \quad t > 0.$

Quantiles : $\quad t_p = \exp\left[\mu + \sigma\Phi_{\text{sev}}^{-1}(p)\right] = \eta[-\log(1-p)]^{1/\beta},$

where $\Phi_{\text{sev}}^{-1}(p) = \log[-\log(1-p)]$, $\eta = \exp(\mu)$, $\beta = 1/\sigma$.

This leads to

$$\log(t_p) = \mu + \log[-\log(1-p)]\sigma = \log(\eta) + \log[-\log(1-p)]\frac{1}{\beta}$$

Conclusion:

$\{$ $\log(t_p)$ versus $\log[-\log(1-p)]$ $\}$ will plot as a straight line.

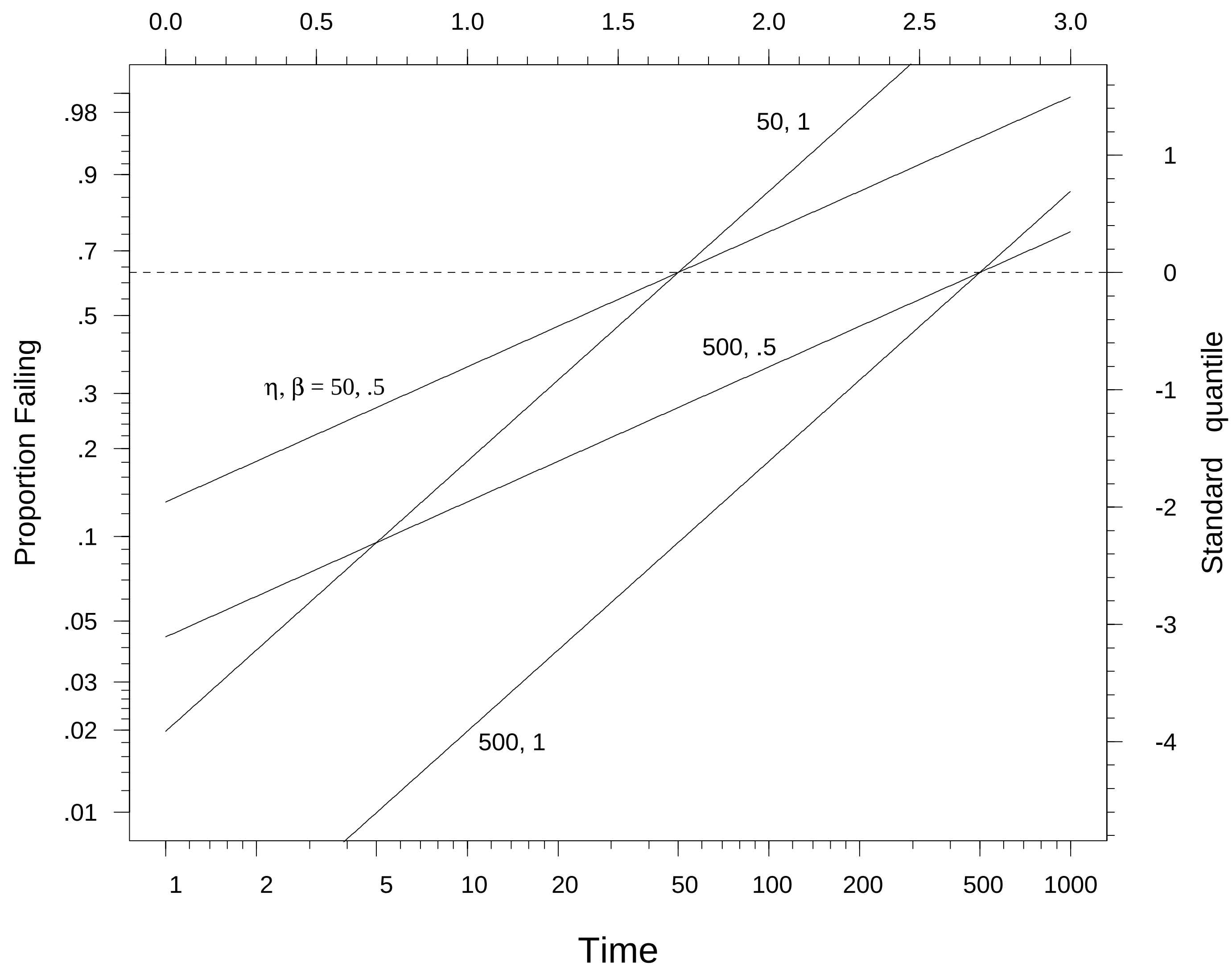
Linearizing the Weibull CDF-Continued

Comments:

- $\eta = \exp(\mu)$ can be read from the time axis at the point where the cdf intersects the $\log[-\log(1-p)] = 0$ line, which corresponds to $p \approx 0.632$.
- The slope of the cdf line on the graph is $\beta = 1/\sigma$ (but in the computations use base e logarithms for the times rather than the base 10 logarithms used for the figures).
- Any Weibull cdf plots as a straight line with positive slope. And any straight line with positive slope corresponds to a Weibull cdf.
- Exponential cdfs plot as straight lines with slopes equal to 1.

Plot with Weibull Distribution Probability Scales Showing Weibull cdfs as Straight Lines for Combinations of $\eta = 50, 500$ and $\beta = .5, 1$

$$\log(t_p) = \log(\eta) + \log[-\log(1-p)]\frac{1}{\beta}$$



Choosing Plotting Positions to Plot the Nonparametric Estimate of F

- The **discontinuity** and **randomness** of $\widehat{F}(t)$ make it difficult to choose a definition for pairs of points $(t, \widehat{F})$ to plot.
- With times reported as **exact**, it is has been traditional to plot $\{\ t_i$ versus $\widehat{F}(t_i)\ \}$ at the observed failure times.

General Idea: Plot an estimate of F at some specified set of points in time and define **plotting** positions consisting of a corresponding estimate of F at these points in time.

Criteria for Choosing Plotting Positions

Criteria for choosing plotting positions should depend on the **application** or **purpose** for constructing the probability plot.

Some applications that suggest criteria:

- Checking distributional assumptions.
- Estimation of parameters.
- Display of maximum likelihood results with data.

Plotting Positions: Continuous Inspection Data and Multiple Censoring

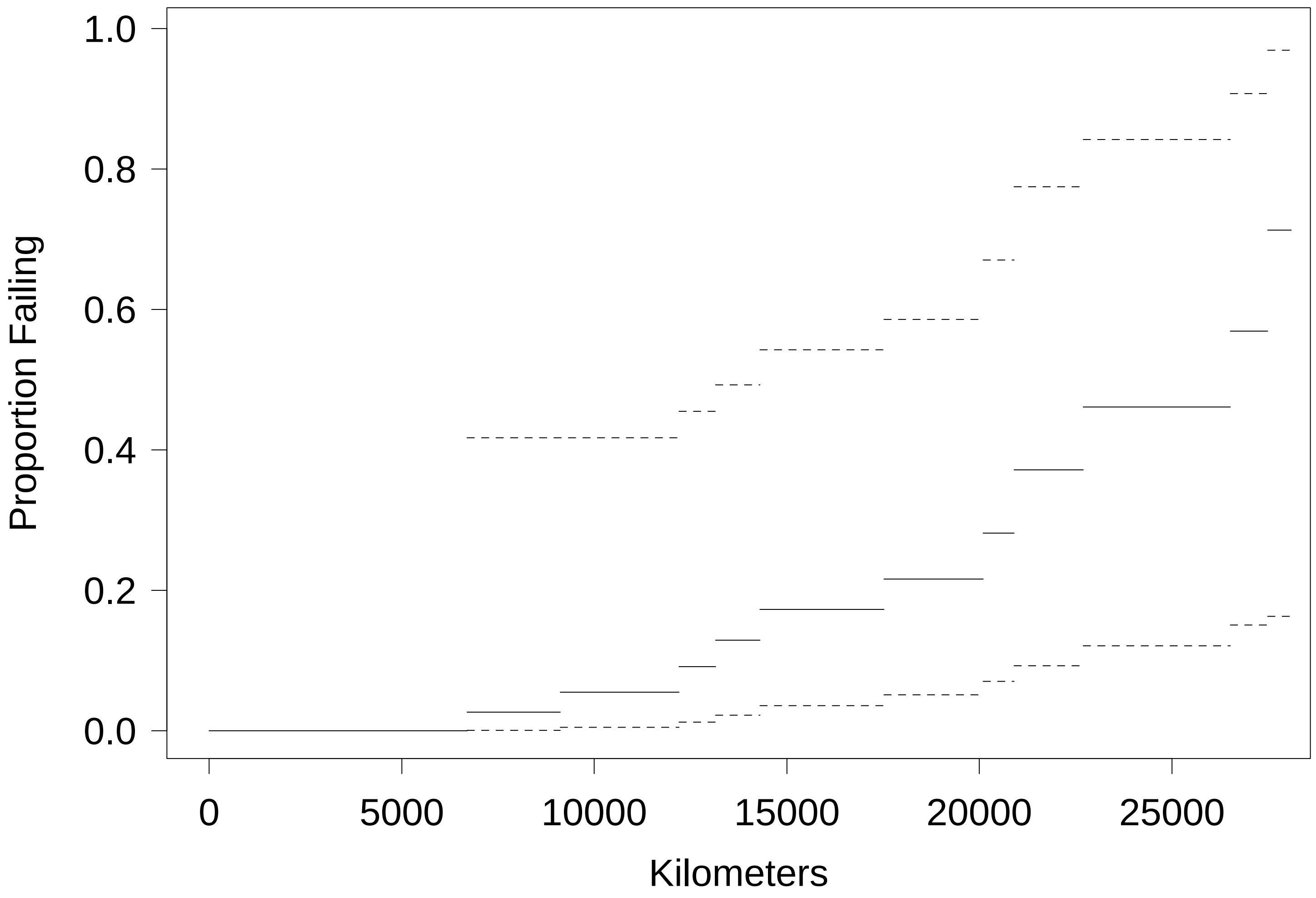
$\widehat{F}(t)$ is a step function until the last reported failure time, but the step increases may be different than $1/n$.

Plotting Positions: $\{\, t_{(i)}$ versus $p_i\}$ with

$$p_i = \frac{1}{2}\left\{\widehat{F}\left[t_{(i)} + \Delta\right] + \widehat{F}\left[t_{(i)} - \Delta\right]\right\}.$$

Justification: This is consistent with the definition for single censoring.

Nonparametric Estimate of $F(t)$ for the Shock Absorbers. Simultaneous Approximate 95% Confidence Bands for $F(t)$



Plotting Positions: Continuous Inspection Data and Single Censoring

Let $t_{(1)}, t_{(2)}, \ldots$ be the ordered failure times. When there is not ties, $\widehat{F}(t)$ is a step function increasing by an amount $1/n$ until the last reported failure time.

Plotting Positions: $\left\{ t_i \text{ versus } \frac{i-.5}{n} \right\}$.

- **Justification:**

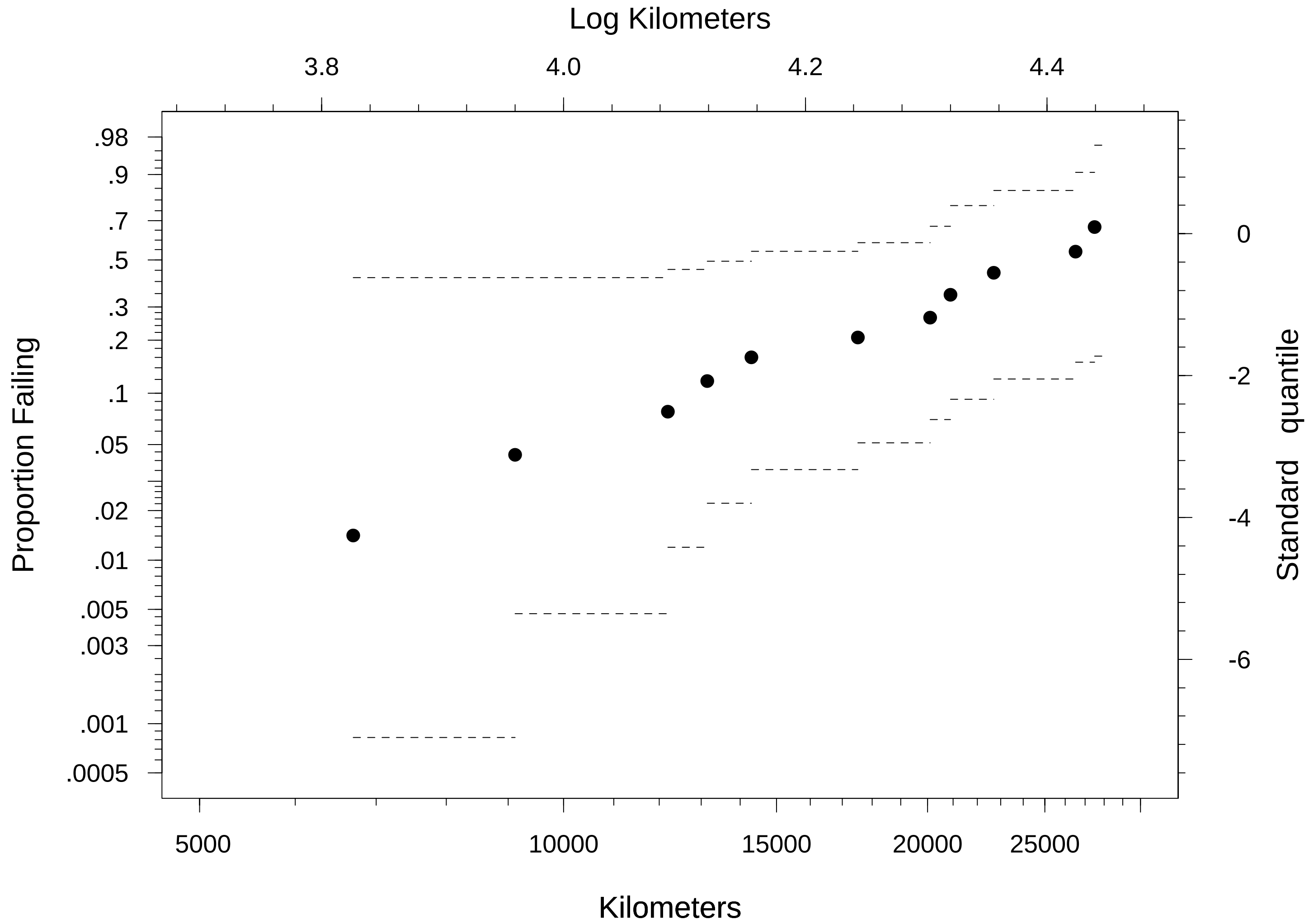
$$\begin{aligned} \frac{i-.5}{n} &= \frac{1}{2}\left\{\widehat{F}\left[t_{(i)}+\Delta\right]+\widehat{F}\left[t_{(i)}-\Delta\right]\right\} \\ \mathrm{E}\left[t_{(i)}\right] &\approx F^{-1}\left(\frac{i-.5}{n}\right). \end{aligned}$$

  where Δ is positive and small.

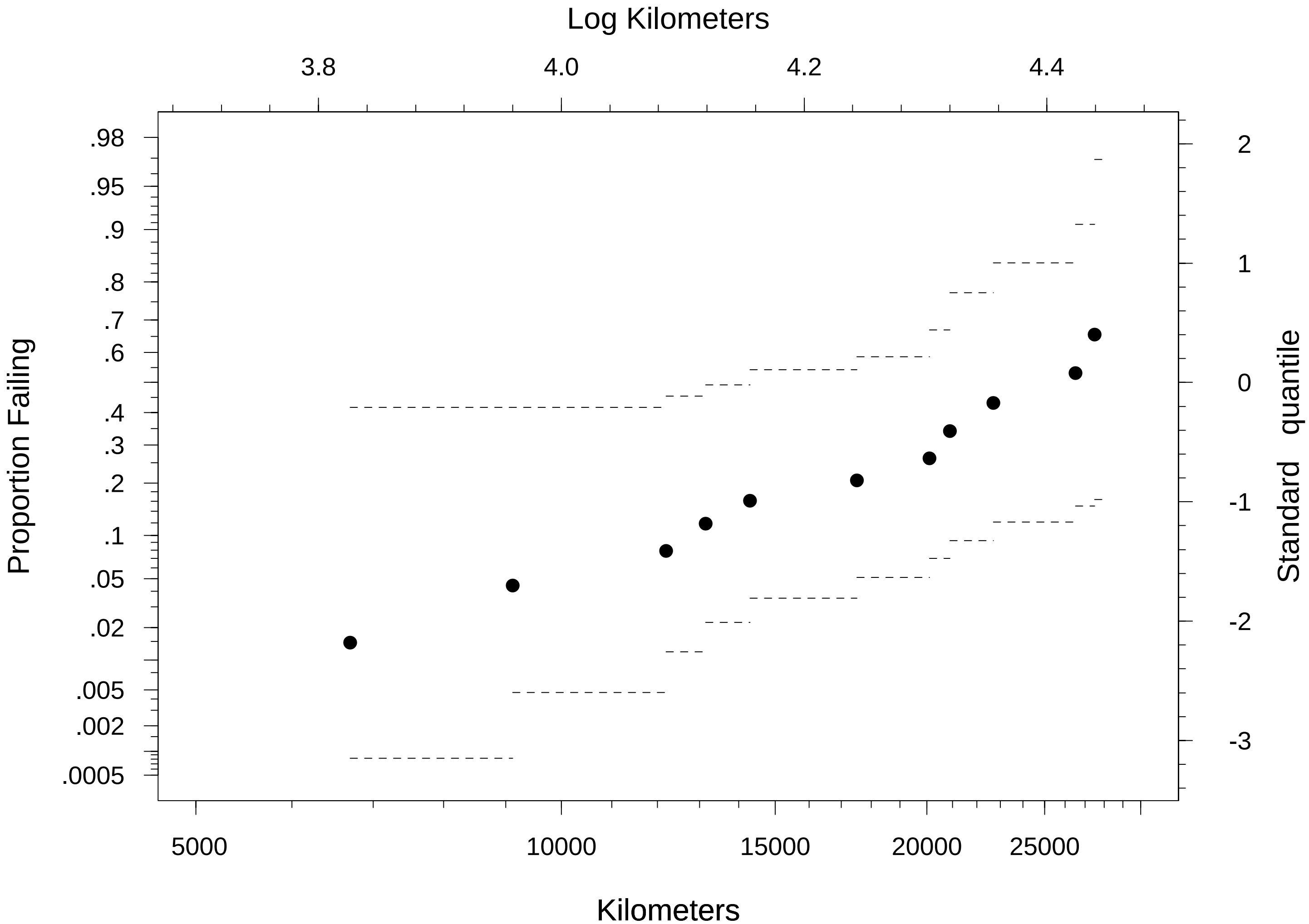
- When the model fits well, the ML line approximately goes through the points.

- Need to adjust these plotting positions when there are ties.

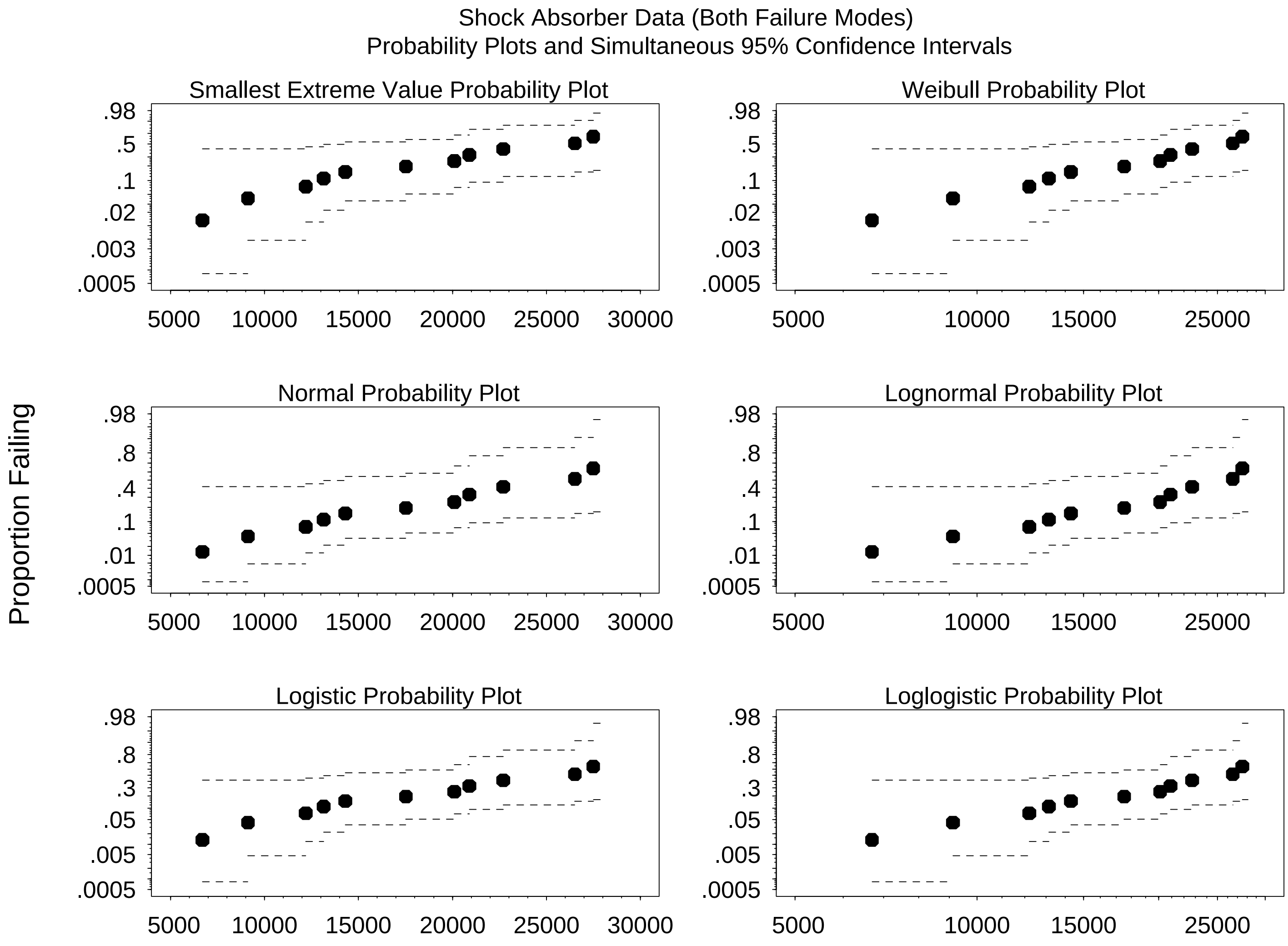
Weibull Probability Plot of the Shock Absorber Data. Also Shown are Simultaneous Approximate 95% Confidence Bands for $F(t)$



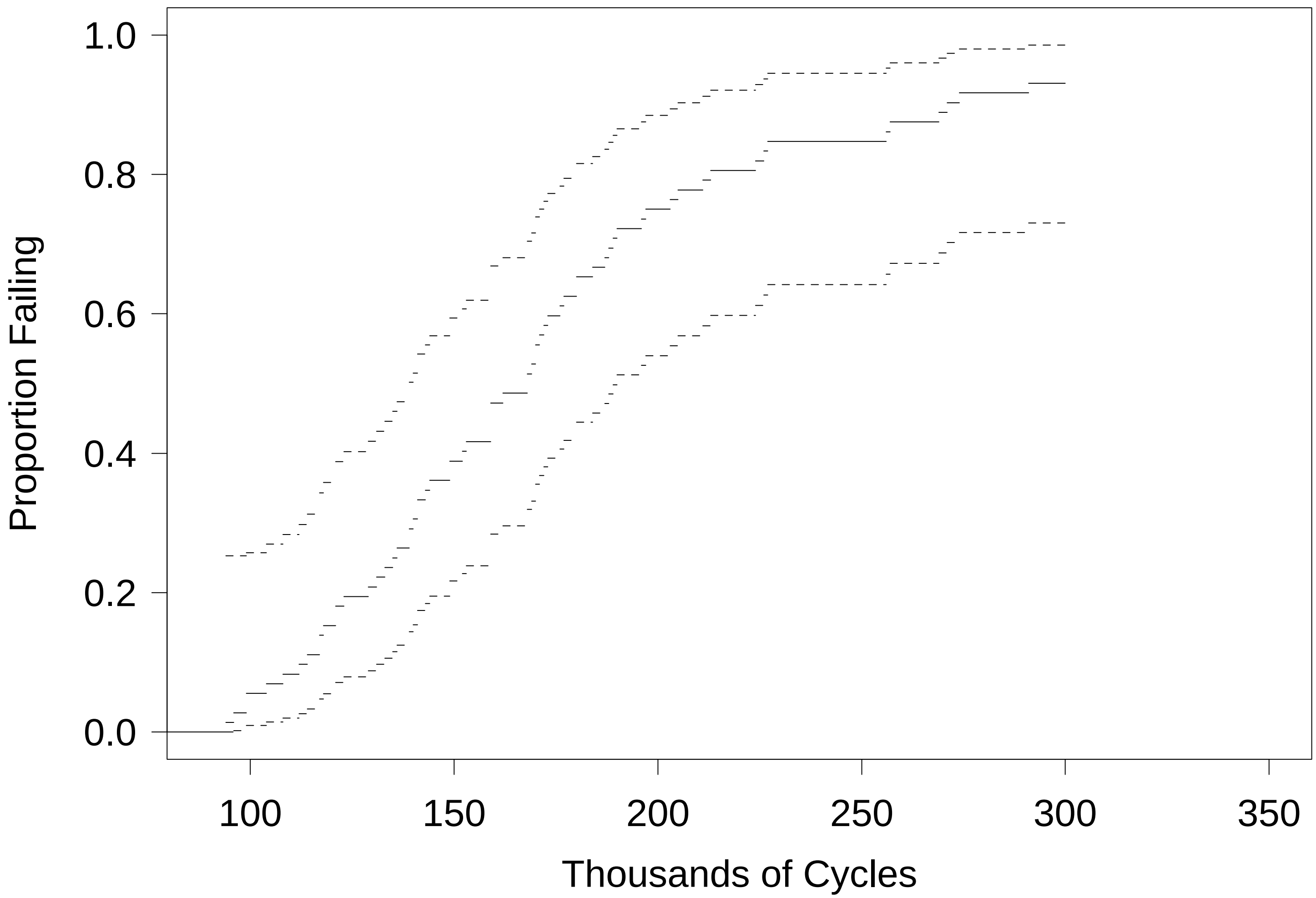
Lognormal Probability Plot of the Shock Absorber Data. Also Shown are Simultaneous Approximate 95% Confidence Bands for $F(t)$



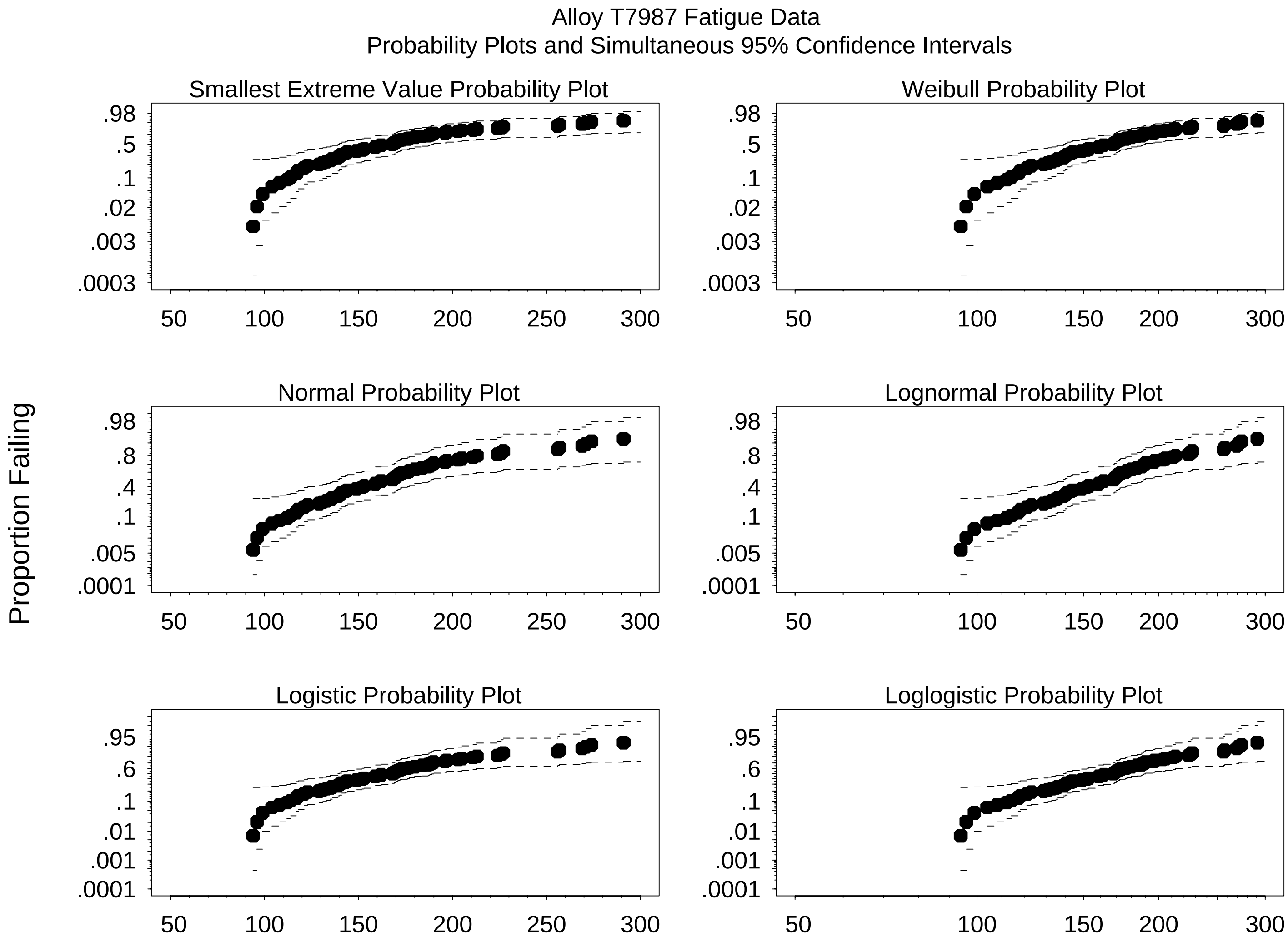
Six-Distribution Probability Plots of the Shock Absorber Data



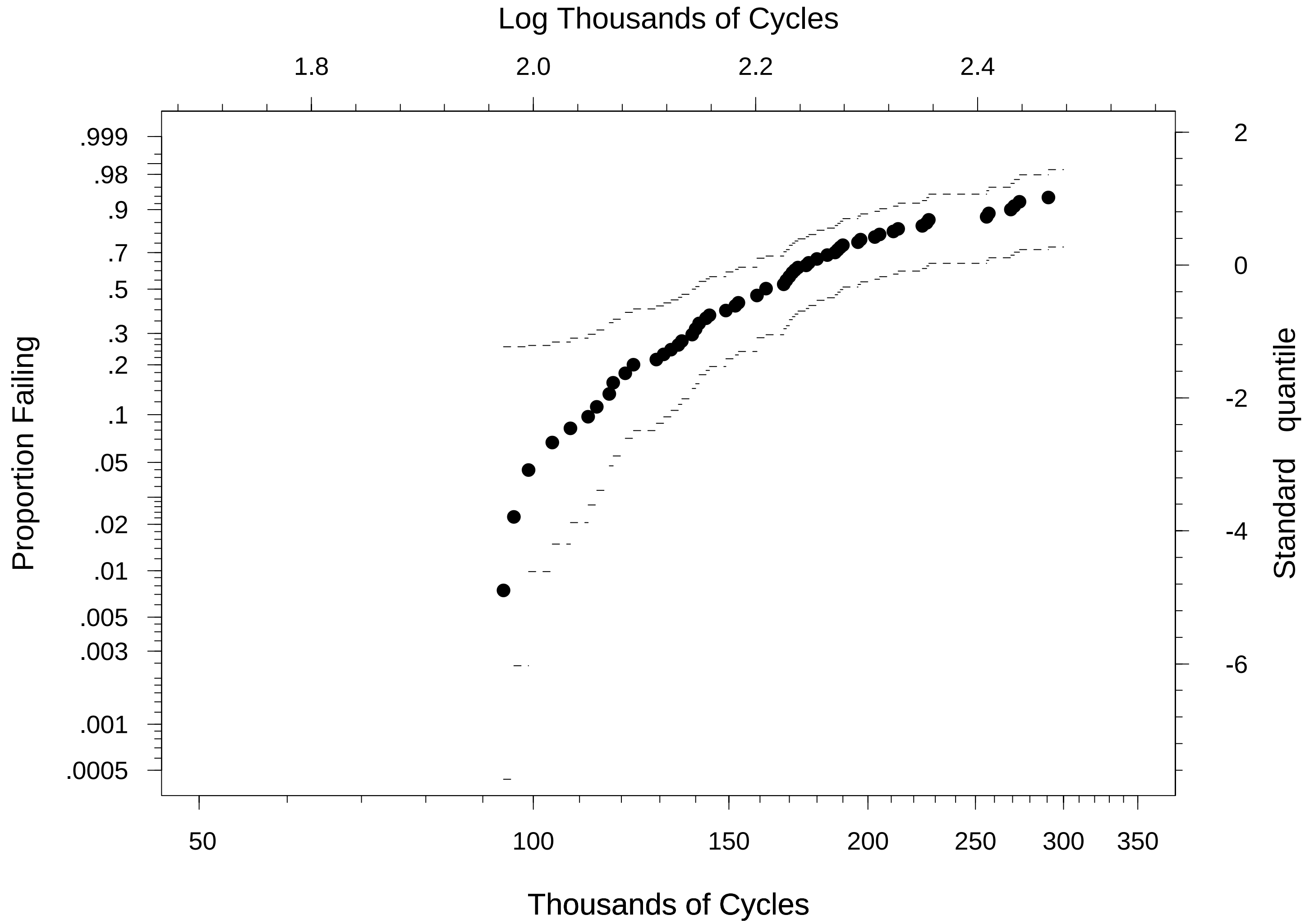
Plot of Nonparametric Estimate of $F(t)$ for the Alloy T7987 Fatigue Life and Simultaneous Approximate 95% Confidence Bands for $F(t)$



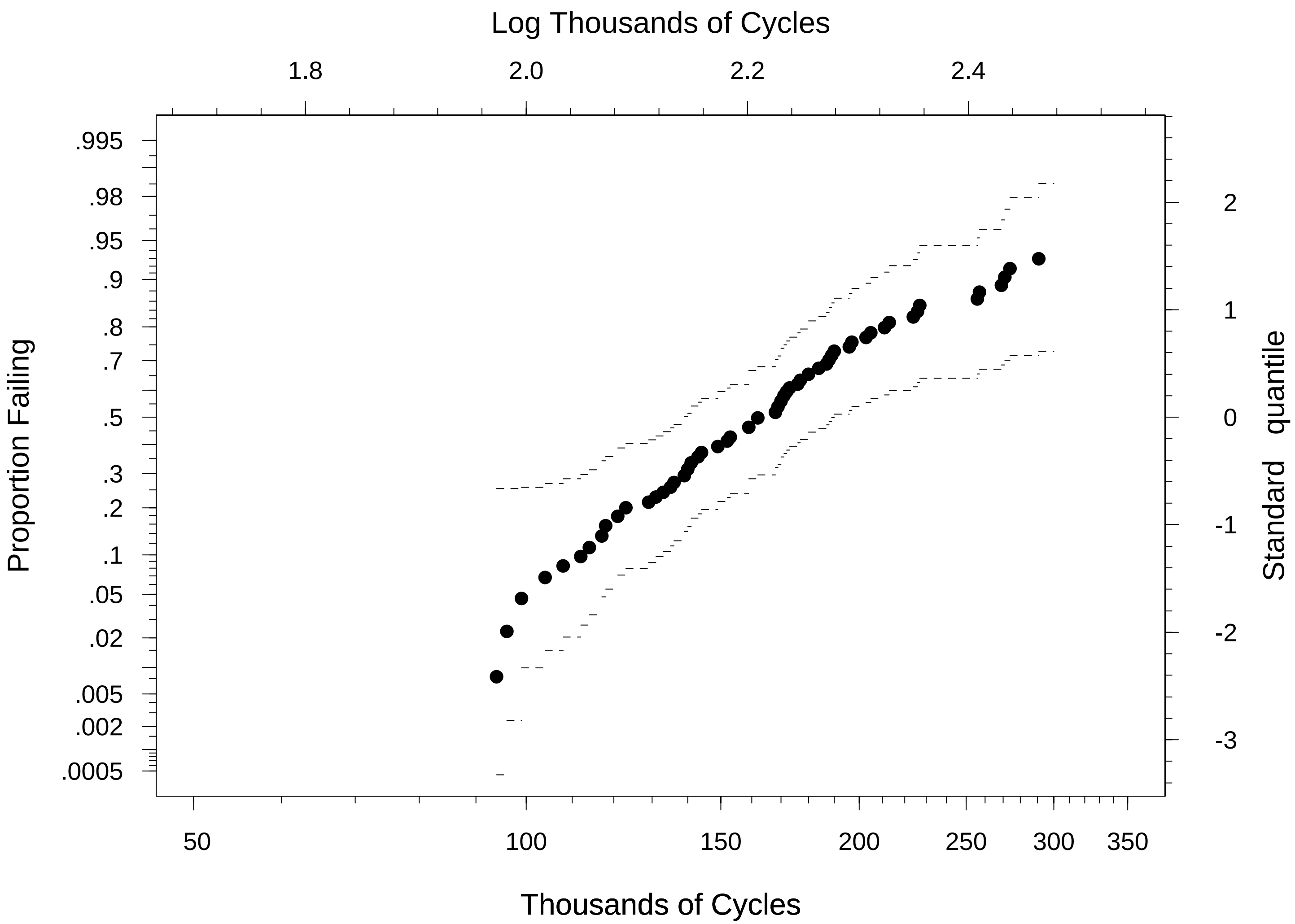
Six-Distribution Probability Plots
Alloy T7987 Fatigue Life



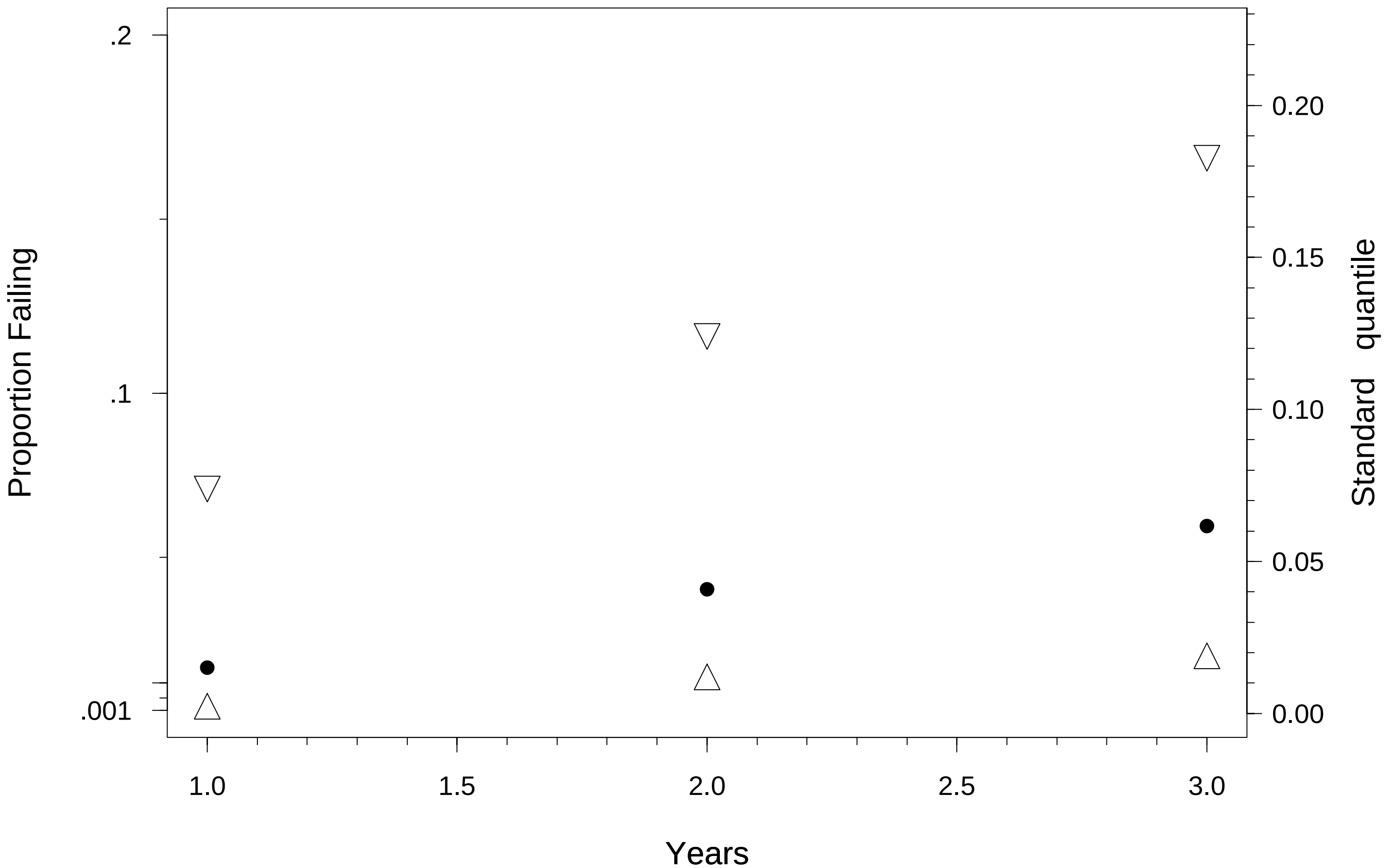
Weibull Probability Plot for the Alloy T7987 Fatigue Life and Simultaneous Approximate 95% Confidence Bands for $F(t)$



Lognormal Probability Plot for the Alloy T7987 Fatigue Life and Simultaneous Approximate 95% Confidence Bands for $F(t)$



Exponential Distribution Probability Plot of the Heat-Exchanger Tube Crack Data and Simultaneous Approximate 95% Confidence Bands for $F(t)$



Plotting Positions: Interval Censored Inspection Data

Let $(t_0, t_1], \ldots, (t_{m-1}, t_m]$ be the inspection times.

The upper endpoints of the inspection intervals $t_i, i = 1, 2, \ldots$, are convenient plotting times.

Plotting Positions: $\{\, t_i$ versus $p_i \,\}$, with

$$p_i = \widehat{F}(t_i)$$

When there are no censored observations beyond t_m, $F(t_m) = 1$ and this point cannot be plotted on probability paper.

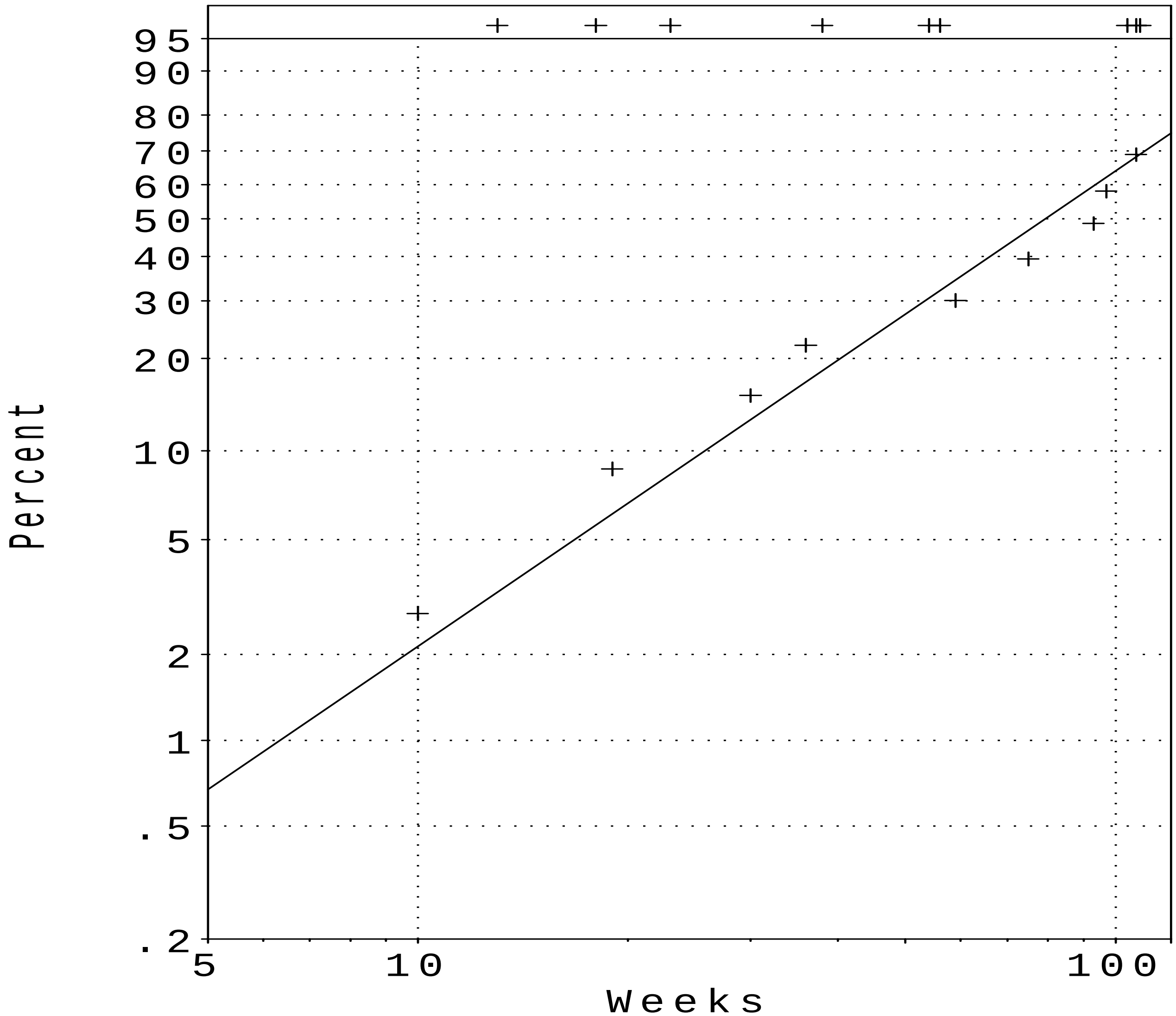
Justification: with no losses, from standard binomial theory,

$$\mathsf{E}[\widehat{F}(t_i)] = F(t_i).$$

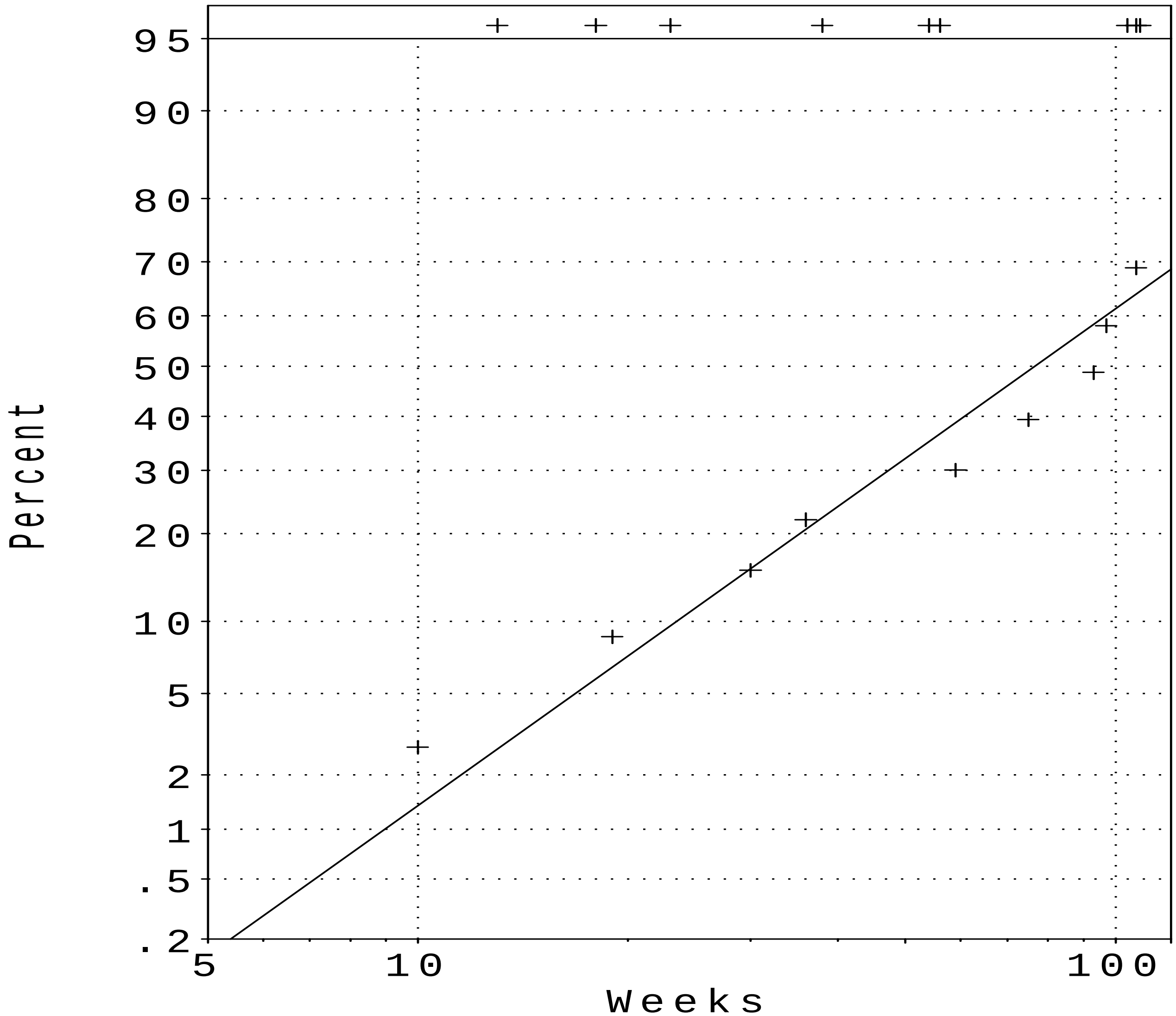
Biomedical Examples

Here we show some probability plots for the IUD data

Nonparametric Estimate for IUD Data
(Weibull probability plot)



Nonparametric Estimate for IUD Data
(Lognormal Probability Plot)

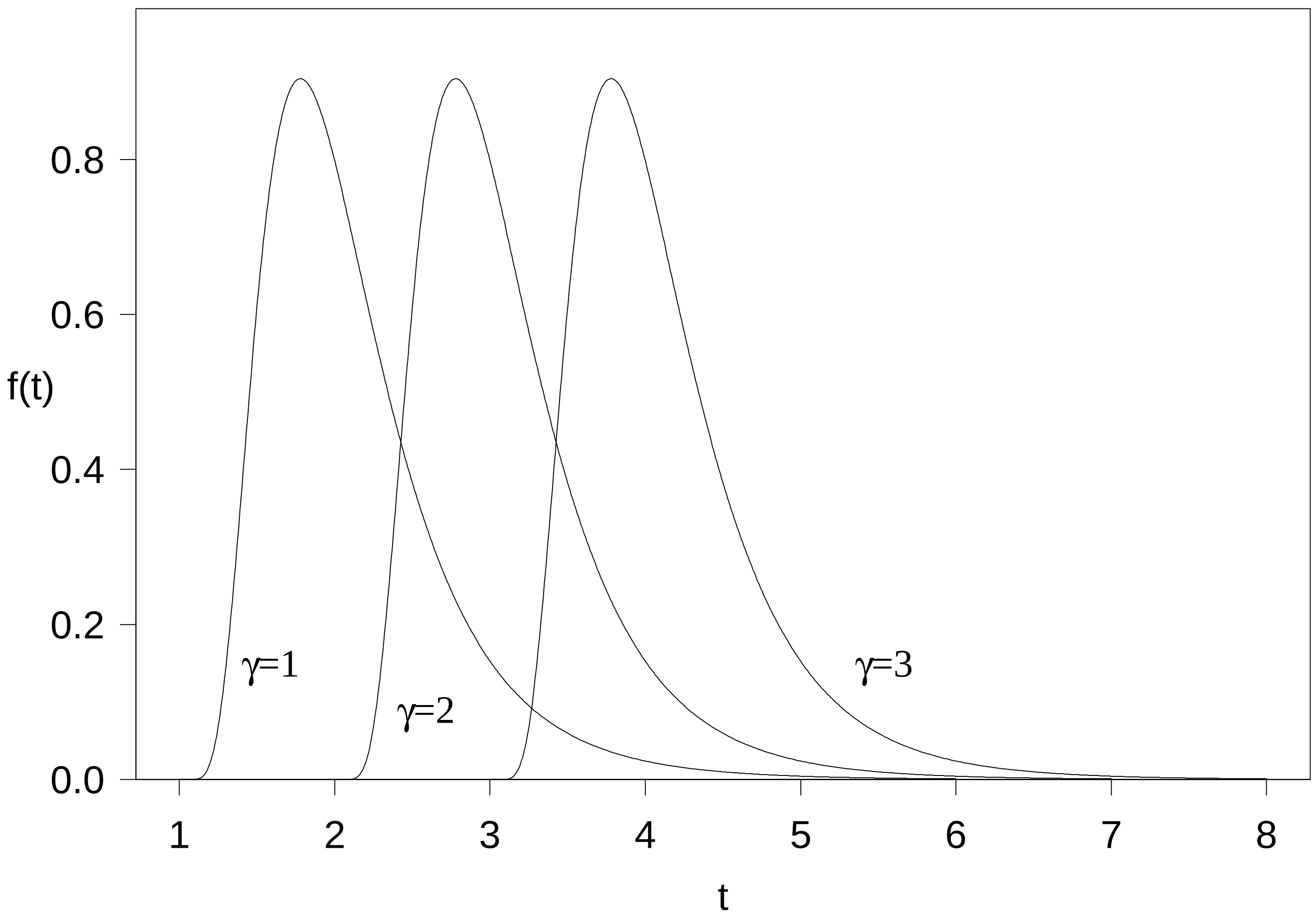


Probability Plots with Specified Shape Parameters

The probability plotting techniques can be extended to construct probability plots for:

- Distributions that are not members of the location-scale family.
- To help identify, graphically, the need for non-zero threshold parameter.
- Estimate graphically a shape parameter.

Pdf for three-parameter lognormal distributions for $\mu = 0$ and $\sigma = .5$ with $\gamma =$ 1,2,3



Distributions with a Threshold Parameter

- The lognormal, Weibull, gamma, and other similar distributions can be generalized by the addition of a **threshold** parameter, γ, to shift the beginning of the distribution away from 0.

- These distributions are particularly useful for fitting skewed distributions that are shifted far to the right of 0.

- For example, the cdf and quantiles of the 3-parameter lognormal distribution can be expressed as

$$p = F(t;\mu,\sigma,\gamma) = \Phi_{\text{nor}}\left[\frac{\log(t-\gamma)-\mu}{\sigma}\right], \quad t > \gamma$$

Linearizing the 3-Parameter Gamma CDF

CDF: $\quad p = F(t; \theta, \kappa, \gamma) = \Gamma_{\text{I}}\left(\frac{t-\gamma}{\theta}; \kappa\right), \quad t > \gamma.$

Quantiles : $\quad t_p = \gamma + \Gamma_{\text{I}}^{-1}(p; \kappa)\theta.$

where $\Gamma_{\text{I}}(z; \kappa) = \int_0^z x^{\kappa-1}e^{-x}dx/\Gamma(\kappa)$ and $\Gamma(\kappa) = \int_0^\infty x^{\kappa-1}e^{-x}dx$.

Conclusion:

$\{\ t_p$ versus $\Gamma_{\text{I}}^{-1}(p; \kappa)\ \}$ will plot as a straight line.

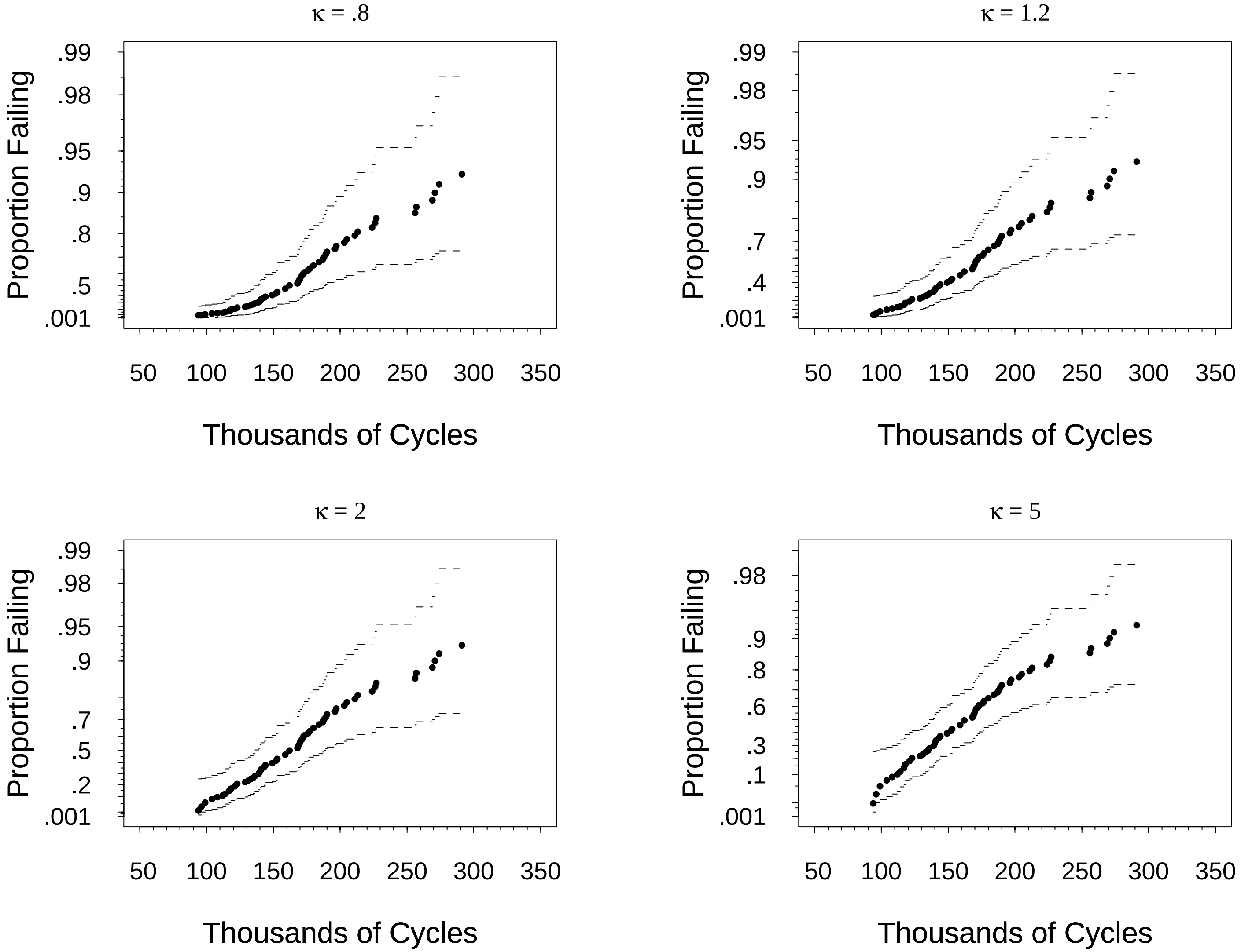
The probability axis **depends** on specification of the shape parameter κ.

γ is the intercept on the time axis (because $\Gamma_{\text{I}}^{-1}(p; \kappa) = 0$ when $p = 0$). The slope of the cdf line is equal to $1/\theta$.

Note:

Changing θ changes the slope of the line and changing γ changes the position of the line.

Gamma Probability Plot with $\kappa = .8, 1.2, 2, 5$ for the Alloy T7987 Fatigue Life with Simultaneous Approximate 95% Confidence Bands for $F(t)$



Linearizing the 3-Parameter Weibull CDF Using Linear Time Axis and Specified Shape Parameter

CDF: $$p = F(t; \mu, \sigma) = \Phi_{\text{sev}}\left[\frac{\log(t-\gamma)-\mu}{\sigma}\right], \quad t > \gamma.$$

$$\text{Quantiles}: \; t_p = \gamma + \eta[-\log(1-p)]^{1/\beta},$$

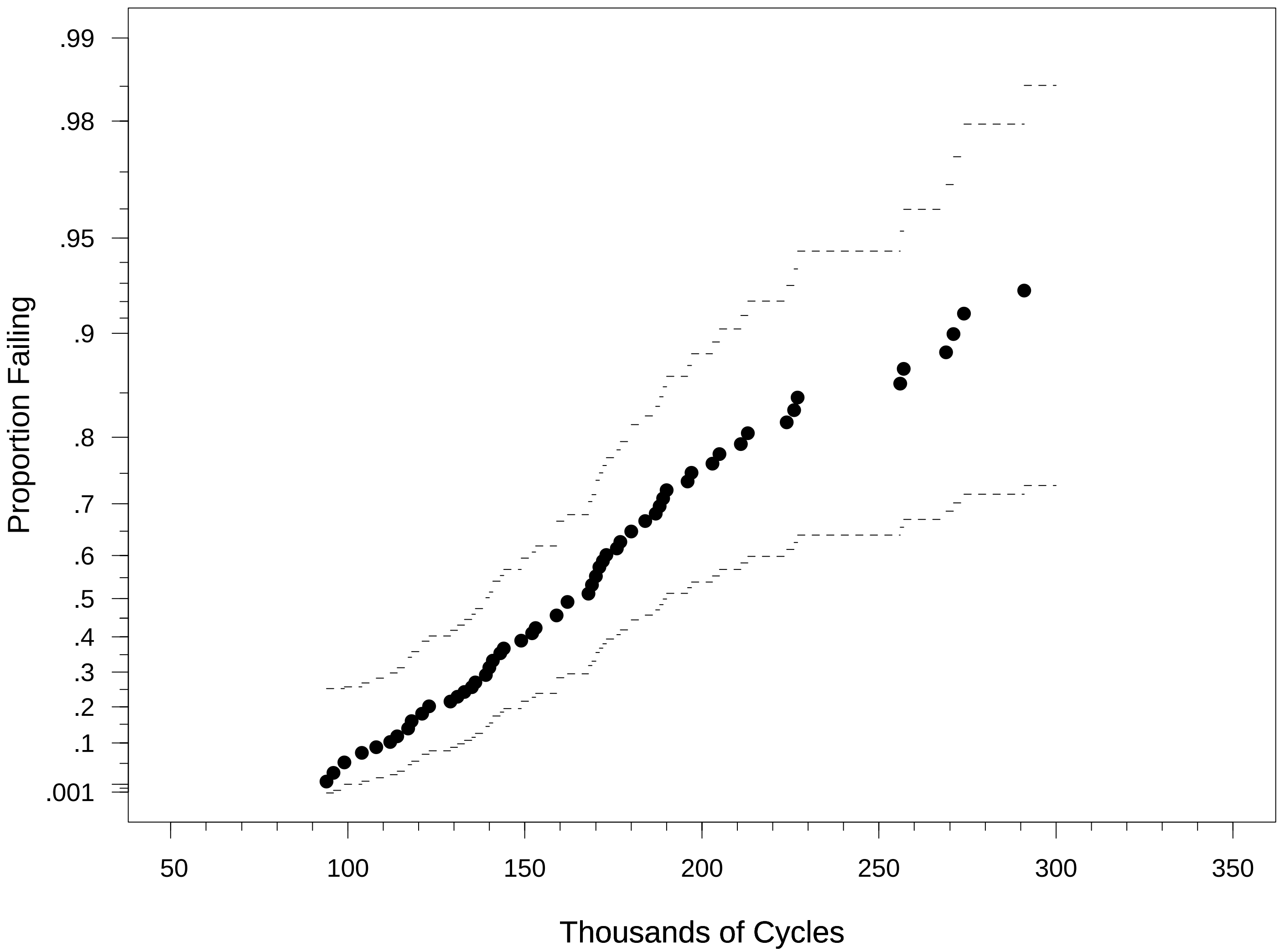
where $\Phi_{\text{sev}}(z) = 1 - \exp[-\exp(z)]$, $\eta = \exp(\mu)$, $\beta = 1/\sigma$.

Conclusion:

$\{\ t_p$ versus $[-\log(1-p)]^{1/\beta}\ \}$ will plot as a straight line.

- The probability axis for this linear-time-axis Weibull probability plot requires specification of the shape parameter β.
- γ is the intercept on the time axis. The slope of the cdf line is equal to $1/\eta$.
- The plot allows graphical estimation the threshold parameter γ.

Linear-Scale Weibull Plot with $\beta = 1.4$ for the Alloy T7987 Fatigue Life with Simultaneous Approximate 95% Confidence Bands for $F(t)$



Linearizing the Generalized Gamma CDF

$$\text{CDF:} \qquad p = F(t;\theta,\beta,\kappa) = \Gamma_{\mathrm{I}}\left[\left(\frac{t}{\theta}\right)^{\beta};\kappa\right].$$

$$\text{Quantiles}: \quad t_p = \theta\left[\Gamma_{\mathrm{I}}^{-1}(p;\kappa)\right]^{1/\beta}.$$

Then $\log(t_p) = \log(\theta) + \log[\Gamma_{\mathrm{I}}^{-1}(p;\kappa)]\,\frac{1}{\beta}$.

Conclusion:

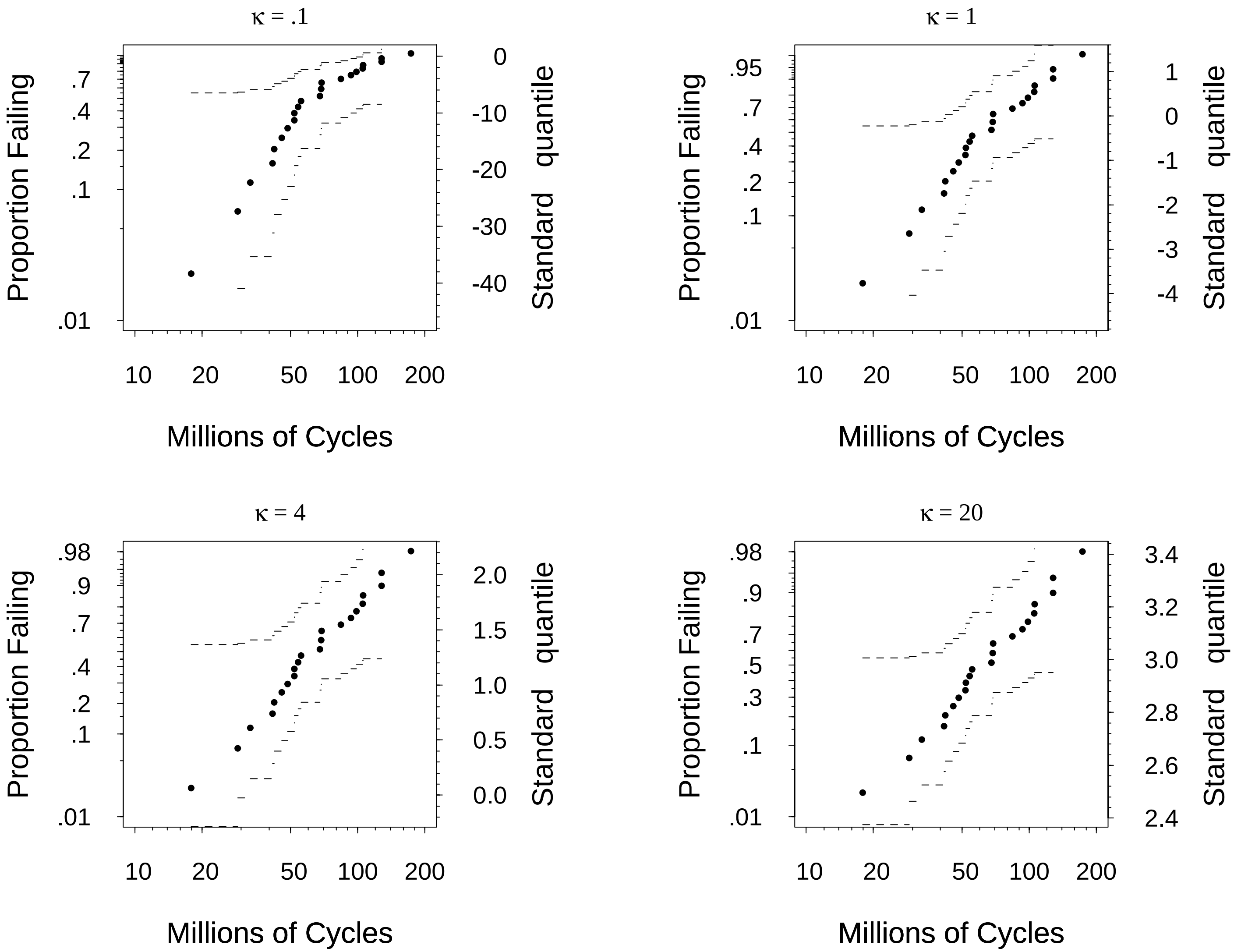
$\{$ $\log(t_p)$ versus $\log[\Gamma_{\mathrm{I}}^{-1}(p;\kappa)]$ $\}$ will plot as a straight line.

The scale parameter θ is the intercept on the time scale, corresponding to the time where the cdf crosses the horizontal line at $\log[\Gamma_{\mathrm{I}}^{-1}(p;\kappa)] = 0$.

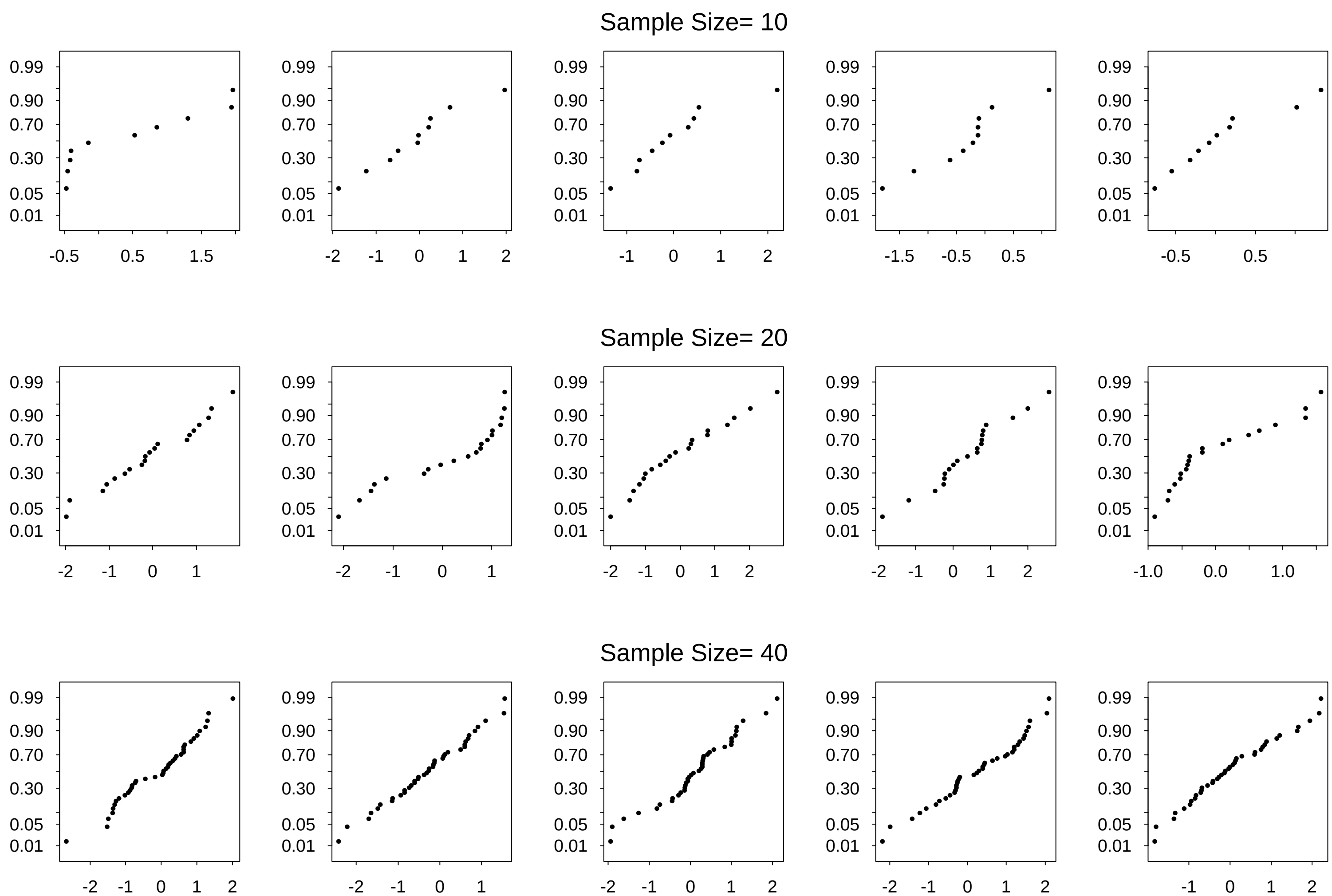
The slope of the line on the graph with time on the horizontal axis is β.

Note: The probability scale for the GENG probability plot requires a given value of the shape parameter κ.

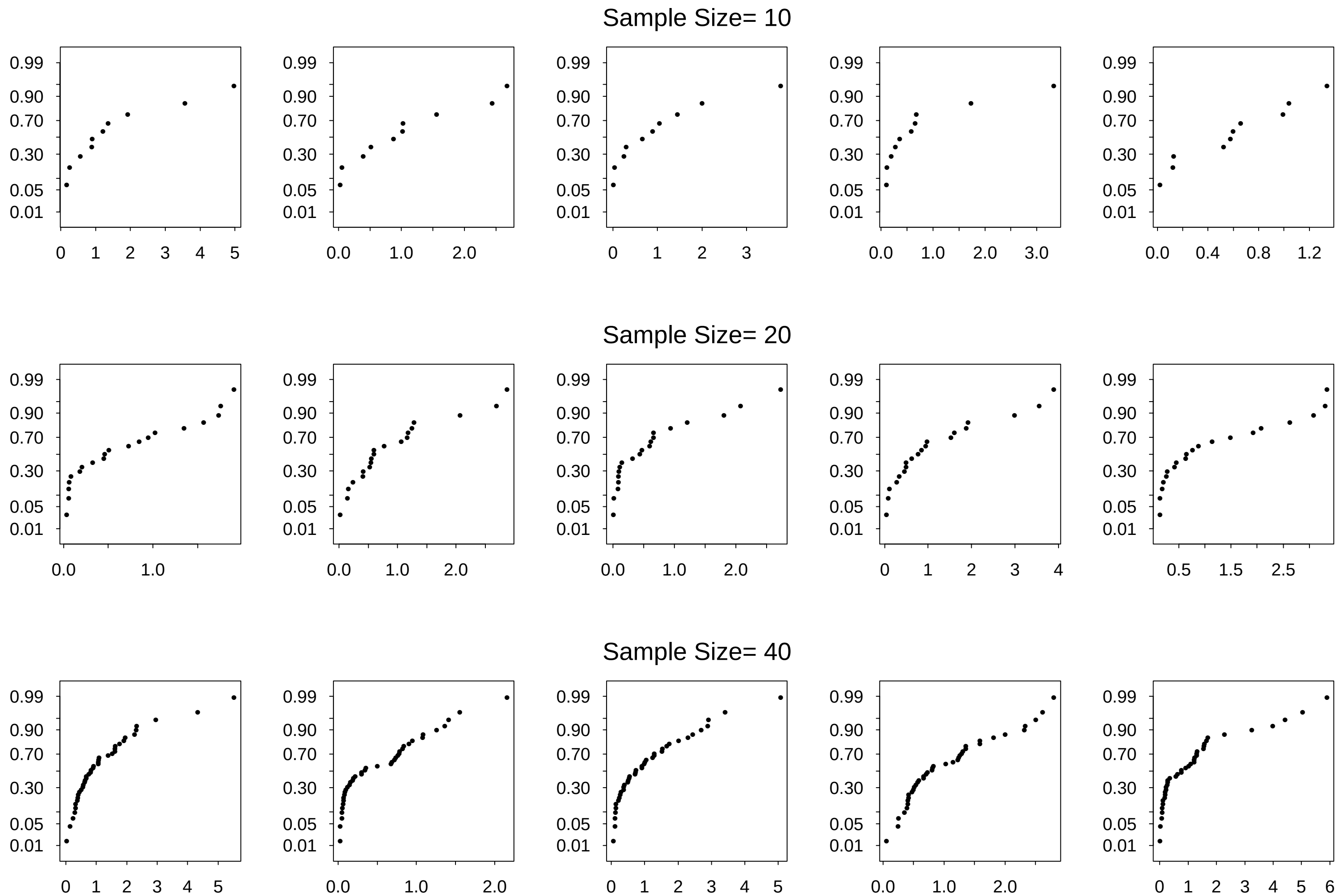
GENG Probability Plots of the Ball Bearing Fatigue Data with Specified κ= .1, 1, 4, and 20



Random Normal Variates Plotted on Normal Probability Plots with Sample Sizes of n=10, 20, and 40. Five Replications of Each Probability Plot



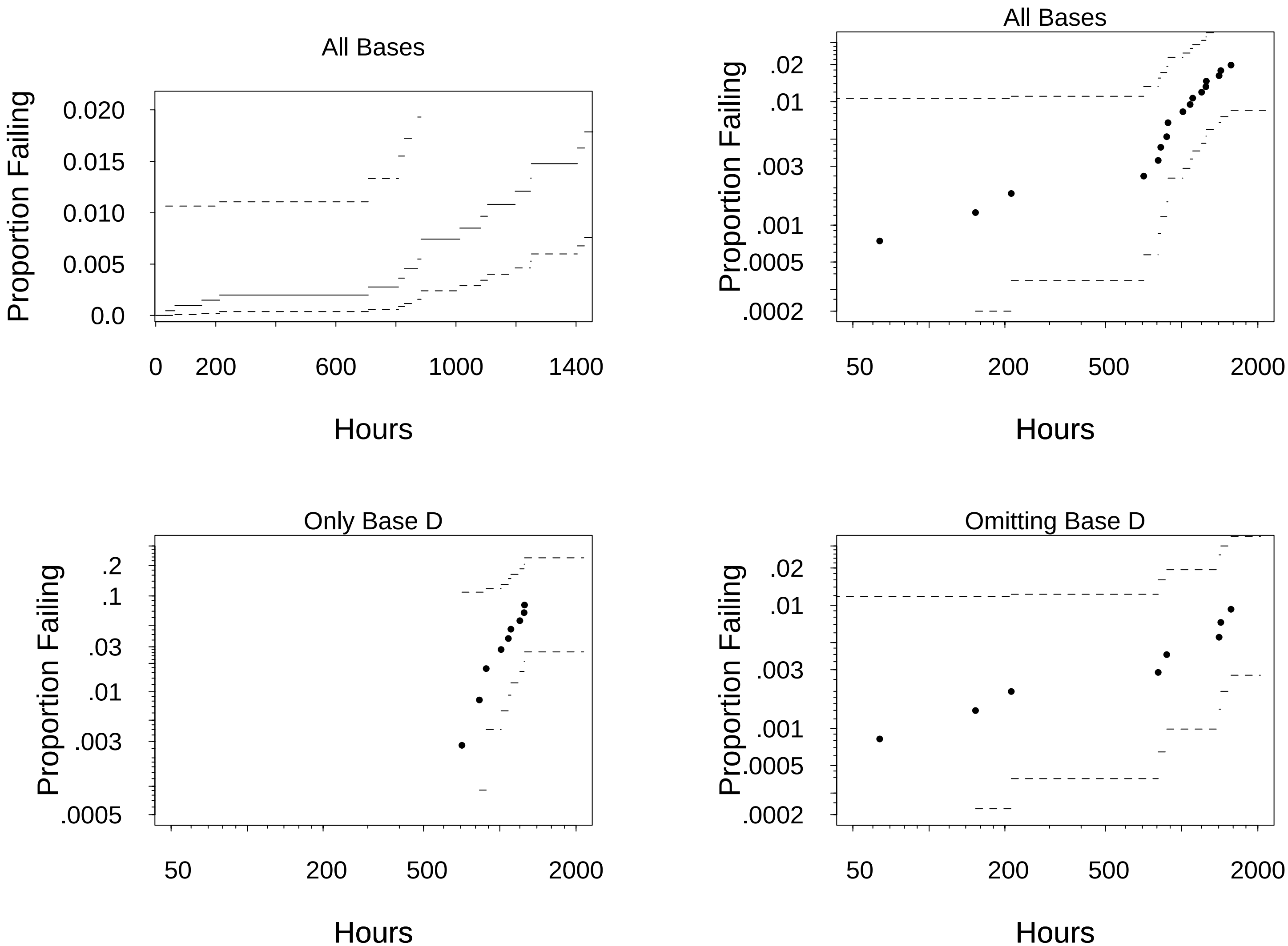
Random Exponential Variates Plotted on Normal Probability Plots with Sample Sizes of n=10, 20, and 40. Five Replications of Each Probability Plot



Notes on the Application of Probability Plotting

- Using simulation to help interpret probability plots
  - ▶ Try different assumed distributions and compare the results.
  - ▶ Assess linearity; allowing for more variability in the tails.
    - ∗ Use simultaneous nonparametric confidence bands.
    - ∗ Use simulation or bootstrap to calibrate.
- Possible reason for a bend in a probability plot
  - ▶ Sharp bend or change in slope generally indicates an abrupt change in a failure process.

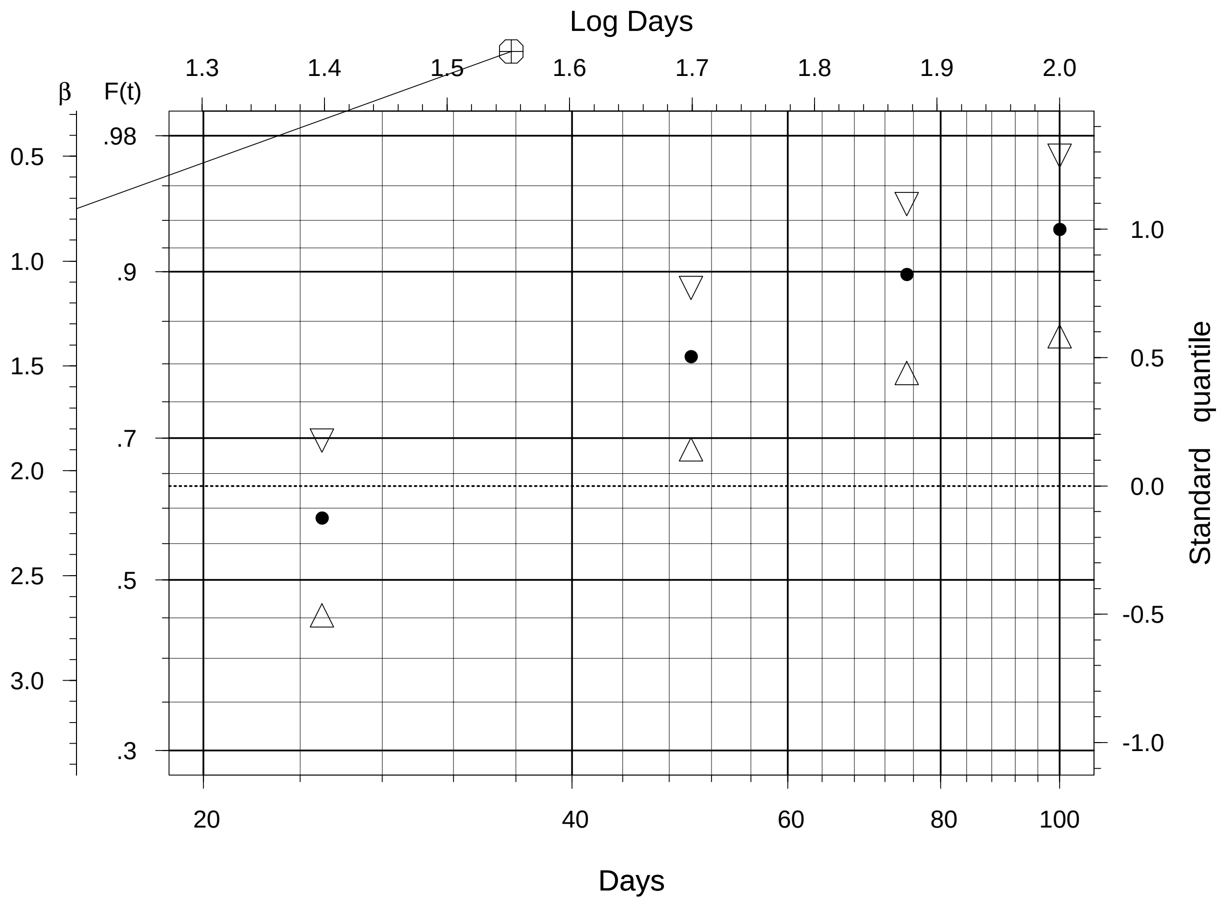
Bleed System
Failure Data Analysis
CDF plot and Weibull Probability Plots



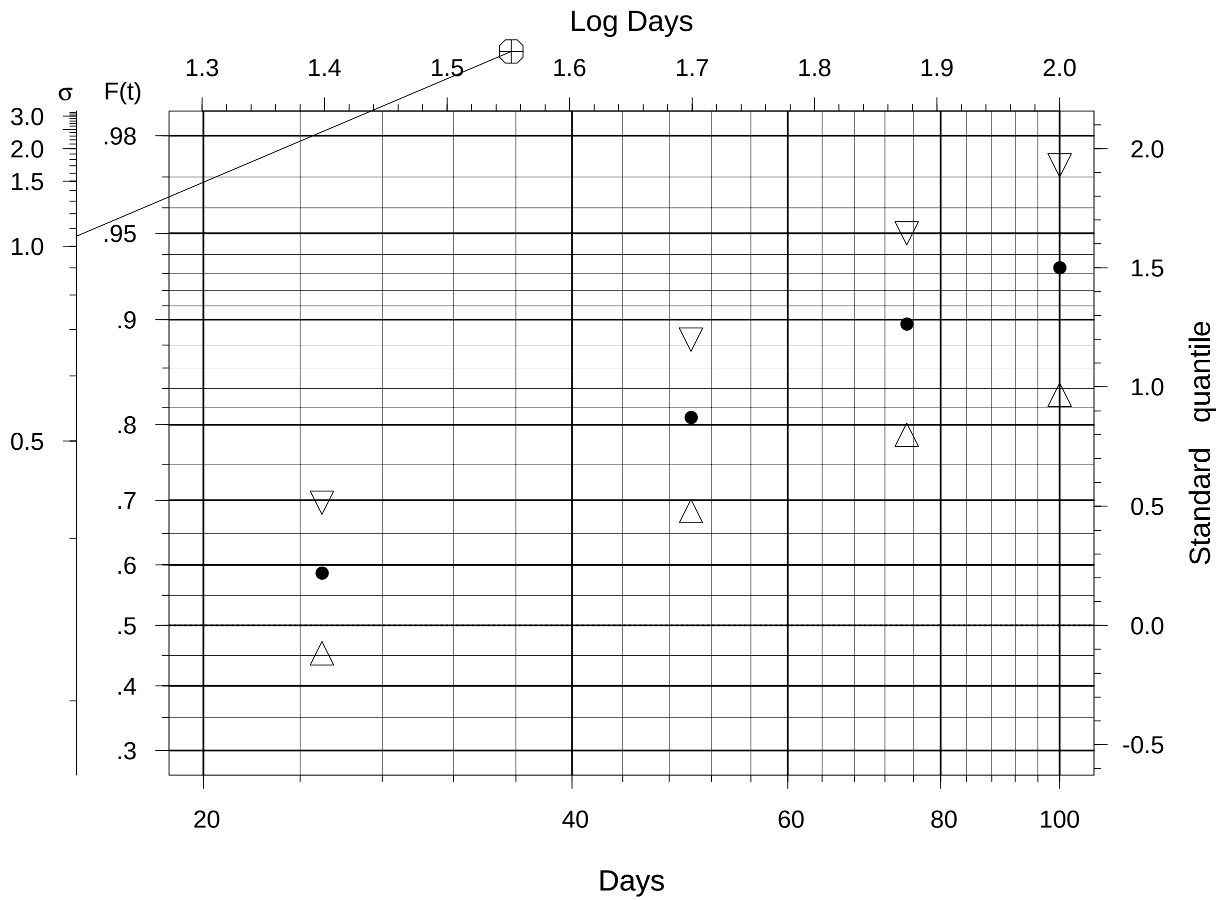
Bleed System Failure Data
(Abernethy, Breneman, Medlin, and Reinman 1983)

- Failure and running times for 2256 bleed systems.
- The Weibull probability plot suggest changes in the failure distribution after 600 hours. The data shows that 9 of the 19 failures had occurred at Base D.
- Separate analyses of the Base D data and the data from the other bases indicated different failure distributions.
- The large slope ($\beta \approx 5$) for Base D indicated strong wearout.
- The relatively small slope for the other bases ($\beta \approx .85$) suggested infant mortality or accidental failures.
- The problem at base D was caused by salt air. A change in maintenance procedures there solved the main part of the reliability problem with the bleed systems.

Weibull Probability Plot of the V7 Transmitter Tube Failure Data with Simultaneous Approximate 95% Confidence Bands for $F(t)$.



Lognormal Probability Plot of the V7 Transmitter Tube Failure Data with Simultaneous Approximate 95% Confidence Bands for $F(t)$.



Other Topics in Chapter 6

Probability plotting for arbitrarily censored data.