Chapter 3

Nonparametric Estimation

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 3
Nonparametric Estimation
Objectives

- Show the use of the binomial distribution to estimate $F(t)$ from interval and singly right censored data, without assumptions on $F(t)$. This is called **nonparametric** estimation.
- Explain and illustrate how to compute standard error for $\widehat{F}(t)$ and approximate confidence intervals for $F(t)$.
- Show how to extend nonparametric estimation to allow for multiply right-censored data.
- Illustrate the **Kaplan-Meier** nonparametric estimator for data with observations reported as exact failures.
- Describe and illustrate a generalization that provides a nonparametric estimator of $F(t)$ with arbitrary censoring.

Data for Plant 1 of the Heat Exchanger Tube Crack Data

	Cracked tubes			
100 tubes at start	Year 1	Year 2	Year 3	Uncracked tubes
Plant 1	1	2	2	95
Unconditional Failure Probability	π_1	π_2	π_3	π_4

$$\text{Likelihood:} \quad L(\boldsymbol{\pi}) = \mathcal{C} \times [\pi_1]^1 \times [\pi_2]^2 \times [\pi_3]^2 \times [\pi_4]^{95}$$

$$\sum_{i=1}^{4} \pi_i = 1.$$

A Nonparametric Estimator of $F(t_i)$ Based on Binomial Theory for Interval Singly-Censored Data

We consider the nonparametric estimate of $F(t_i)$ for data situations as illustrate by Plant 1 of the Heat Exchanger Tube Crack:

- The data are:

  n : sample size

  d_i : # of failures (deaths) in the ith interval

- Simple binomial theory gives

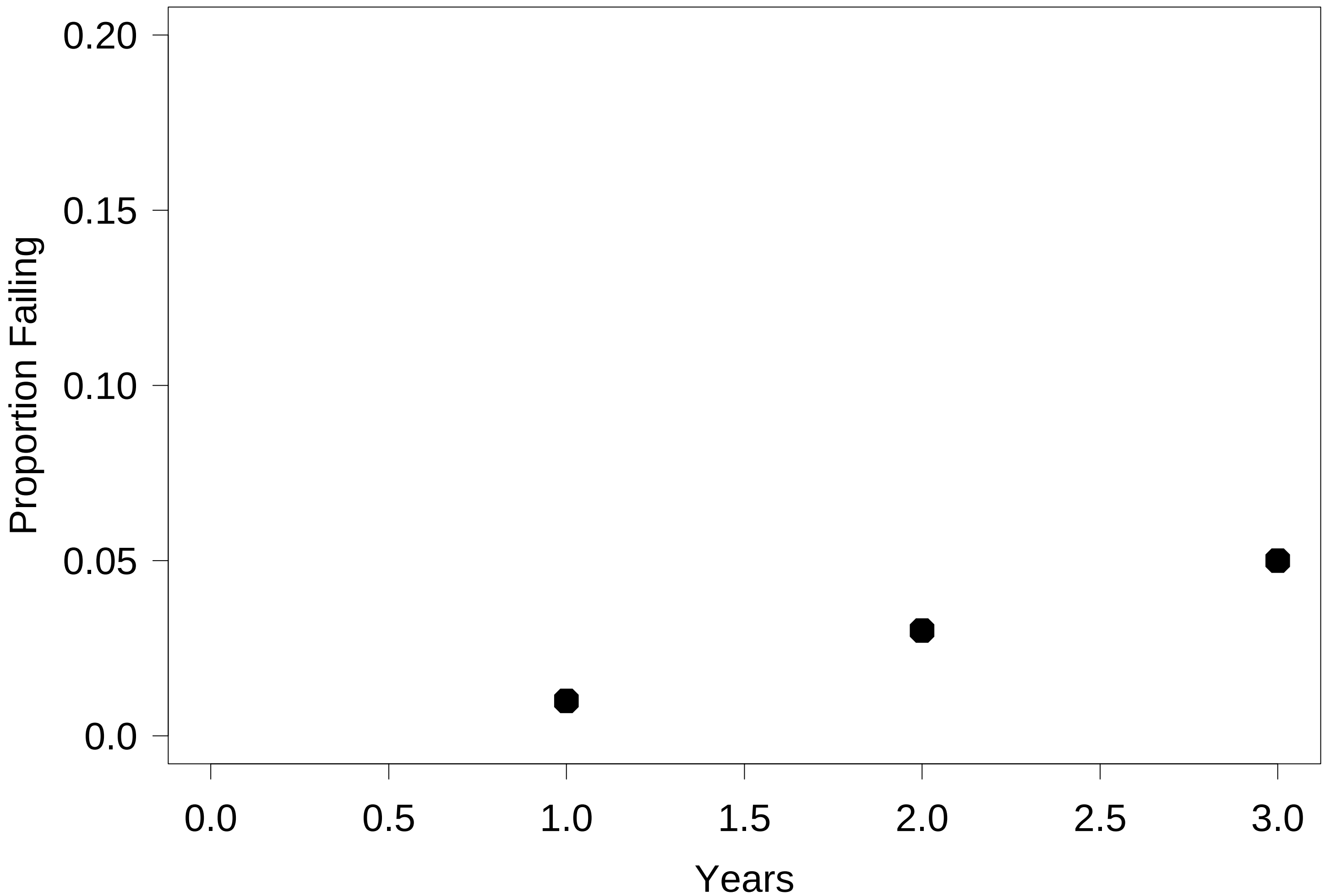
$$\widehat{F}(t_i) = \frac{\text{\# of failures up to time } t_i}{n} = \frac{\sum_{j=1}^{i} d_j}{n}$$

$$\widehat{\text{se}}_{\widehat{F}} = \sqrt{\frac{\widehat{F}(t_i)\left[1 - \widehat{F}(t_i)\right]}{n}}.$$

- For Plant 1 ($n = 100, d_1 = 1, d_2 = 2, d_3 = 2$), one gets:

$$\widehat{F}(1) = 1/100, \quad \widehat{F}(2) = 3/100, \quad \widehat{F}(3) = 5/100.$$

Nonparametric Estimate for Plant 1 from the Heat Exchanger Tube Crack Data



Comments on the Nonparametric Estimate of $F(t_i)$

- $\widehat{F}(t)$ is only defined at the upper ends of the intervals $(t_{i-1}, t_i]$.

- $\widehat{F}(t_i)$ is the ML estimator of $F(t_i)$.

- The increase in $\widehat{F}$ at each value of t_i is

$$\widehat{F}(t_i) - \widehat{F}(t_{i-1}) = d_i/n.$$

Confidence Intervals

A point estimate can be misleading. It is important to quantify uncertainty in point estimates.

- Confidence intervals are very useful in quantifying uncertainty in point estimates due to sampling error arising from limited sample sizes.

- In general, confidence intervals do not quantify possible deviations arising from incorrectly specified model or model assumptions.

Some Characteristic Features of Confidence Intervals

- The level of **confidence** expresses one's confidence (not probability) that a specific interval contains the quantity of interest.

- The actual **coverage** probability is the probability that the procedure will result in an interval containing the quantity of interest.

- A confidence interval is **approximate** if the specified level of confidence is not equal to the actual coverage probability.

- With censored data most confidence intervals are approximate. Better approximations generally require more computations.

Pointwise Binomial-Based Confidence Interval for $F(t_i)$

- A $100(1-\alpha)\%$ conservative confidence interval for $F(t_i)$ based on binomial sampling (see Chapter 6 of Hahn and Meeker, 1991) is

$$
\begin{aligned}
\underset{\sim}{F}(t_i) &= \left\{ 1 + \frac{(n - n\widehat{F} + 1)\mathcal{F}_{(1-\alpha/2;2n-2n\widehat{F}+2,2n\widehat{F})}}{n\widehat{F}} \right\}^{-1} \\
\widetilde{F}(t_i) &= \left\{ 1 + \frac{n - n\widehat{F}}{(n\widehat{F}+1)\mathcal{F}_{(1-\alpha/2;2n\widehat{F}+2,2n-2n\widehat{F})}} \right\}^{-1}
\end{aligned}
$$

  where $\widehat{F} = \widehat{F}(t_i)$ and $\mathcal{F}_{(1-\alpha/2;\nu_1,\nu_2)}$ is the $100(1-\alpha/2)$ quantile of the $\mathcal{F}$ distribution with (ν_1, ν_2) degrees of freedom.

- This confidence interval is conservative in the sense that the actual coverage probability is at least equal to $1-\alpha$.

Pointwise Normal-Approximation Confidence Interval for $F(t_i)$

- For a specified value of t_i, an approximate $100(1-\alpha)\%$ confidence interval for $F(t_i)$ is

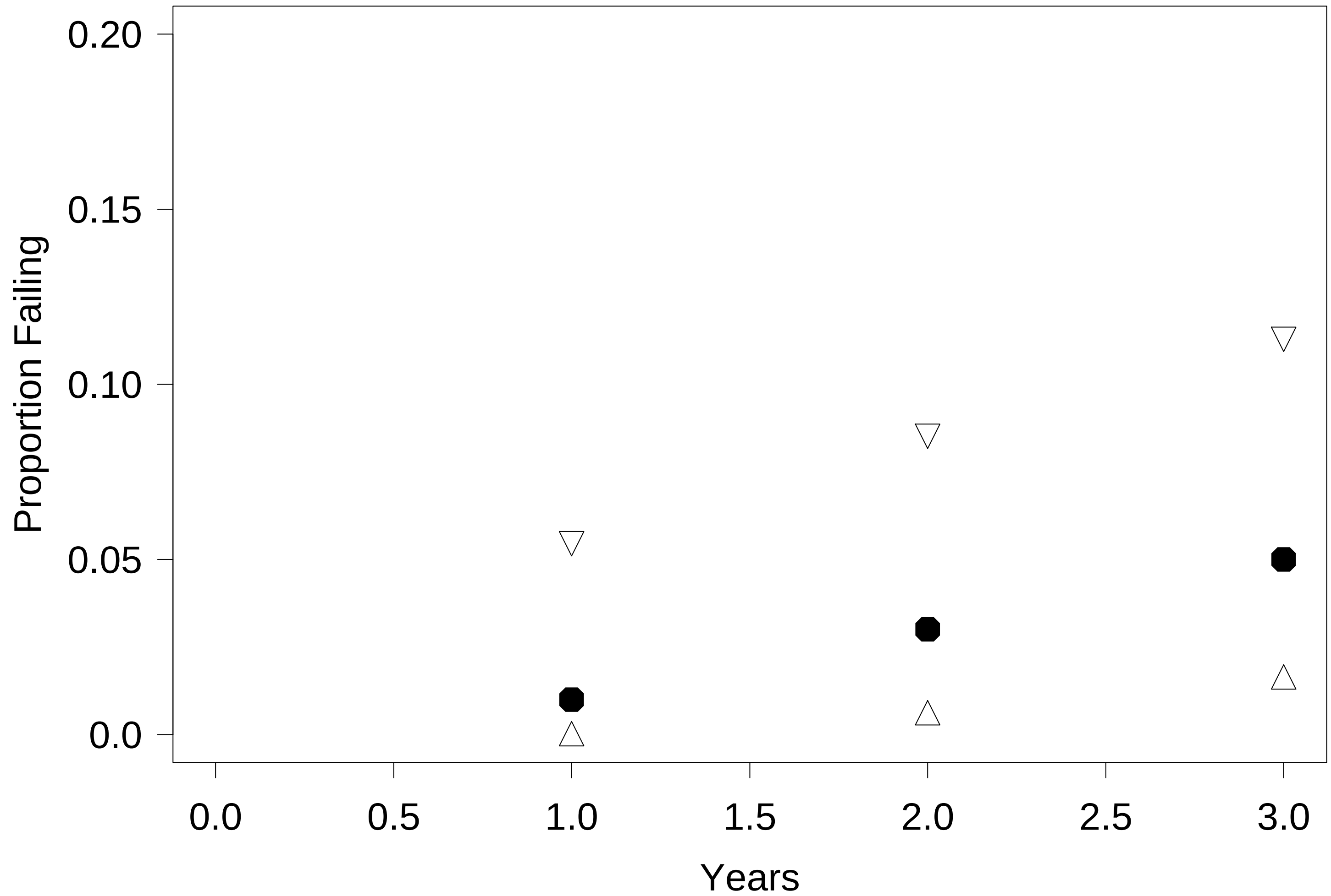
$$[\underset{\sim}{F}(t_i), \quad \tilde{F}(t_i)] = \widehat{F}(t_i) \pm z_{(1-\alpha/2)} \widehat{\text{se}}_{\widehat{F}}.$$

where $z_{(1-\alpha/2)}$ is the $1-\alpha/2$ quantile of the standard normal distribution and $\widehat{\text{se}}_{\widehat{F}} = \sqrt{\widehat{F}(t_i)\left[1-\widehat{F}(t_i)\right]/n}$ is an estimate of the standard error of $\widehat{F}(t_i)$.

- This confidence interval is based on

$$Z_{\widehat{F}} = \frac{\widehat{F}(t_i) - F(t_i)}{\widehat{\text{se}}_{\widehat{F}}} \mathrel{\dot\sim} \text{NOR}(0,1).$$

Plant 1 Heat Exchanger Tube Crack Nonparametric Estimate with Conservative Pointwise 95% Confidence Intervals Based on Binomial Theory



Calculations of the Nonparametric Estimate of $F(t_i)$ for Plant 1 from the Heat Exchanger Tube Crack Data

Year	t_i	d_i	$\widehat{F}(t_i)$	$\widehat{\text{se}}_{\widehat{F}}$	Pointwise Confidence Interval $\underset{\sim}{F}(t_i)$	$\widetilde{F}(t_i)$
$(0-1]$	1	1	0.01	.00995		
95% Confidence Intervals for $F(1)$						
Based on	Binomial Theory				[.0003,	.0545]
Based on	$Z_{\widehat{F}} \dot{\sim}$ NOR$(0,1)$				[−.0095,	.0295]
$(1-2]$	2	2	0.03	.01706		
95% Confidence Intervals for $F(2)$						
Based on	Binomial Theory				[.0062,	.0852]
Based on	$Z_{\widehat{F}} \dot{\sim}$ NOR$(0,1)$				[−.0034,	.0634]
$(2-3]$	3	2	0.05	0.02179		
95% Confidence Intervals for $F(3)$						
Based on	Binomial Theory				[.0164,	.1128]
Based on	$Z_{\widehat{F}} \dot{\sim}$ NOR$(0,1)$				[.0073,	.0927]

Integrated Circuit (IC) Failure Times in Hours
Data from Meeker (1987)

.10	.10	.15	.60	.80	.80
1.20	2.50	3.00	4.00	4.00	6.00
10.00	10.00	12.50	20.00	20.00	43.00
43.00	48.00	48.00	54.00	74.00	84.00
94.00	168.00	263.00	593.00		

When the test ended at 1370 hours, there were 28 observed failures and 4128 unfailed units.

Note: Ties in the data. Reason?

Nonparametric Estimator of $F(t)$ Based on Binomial Theory for Exact Failures and Singly Right Censored Data

When the number of inspections increases the width of the intervals $(t_{i-1}, t_i]$ approaches zero and the failure times are exact.

- For the **integrated circuit life test data**, we have: $n = 4156$ with 28 exact failures in 1370 hours.

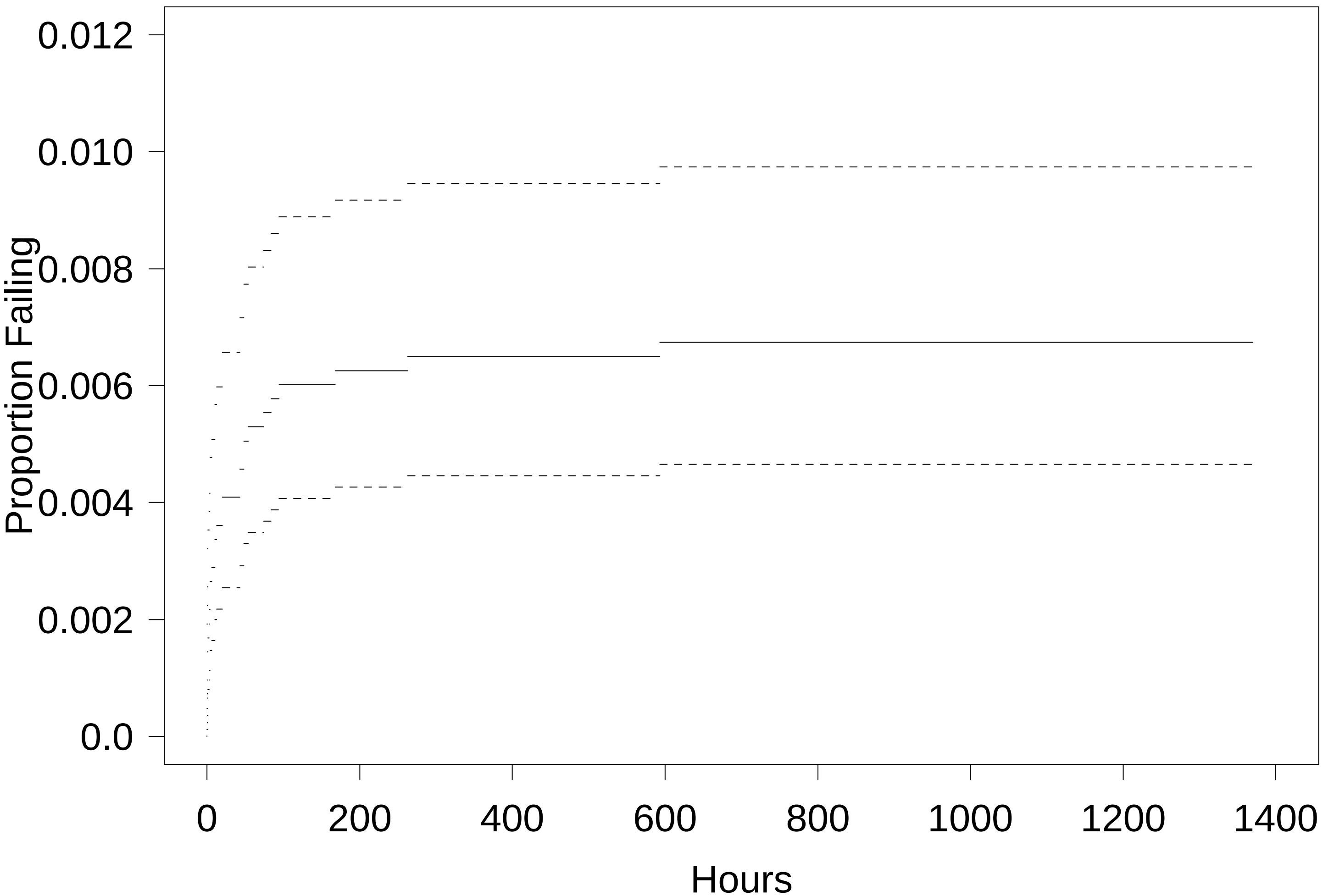
  For any particular t_e, $0 < t_e \leq 1370$, simple binomial theory gives

$$\widehat{F}(t_e) = \frac{\text{\# of failures up to time } t_e}{n}$$

$$\widehat{\text{se}}_{\widehat{F}} = \sqrt{\frac{\widehat{F}(t_e)\left[1 - \widehat{F}(t_e)\right]}{n}}.$$

- Methods to obtain confidence intervals for $F(t_e)$ are the same as the methods described for the interval data.

Nonparametric Estimate for the IC Data with Normal Approximation Pointwise 95% Confidence Intervals Based on $Z_{\mathrm{logit}(\widehat{F})}$

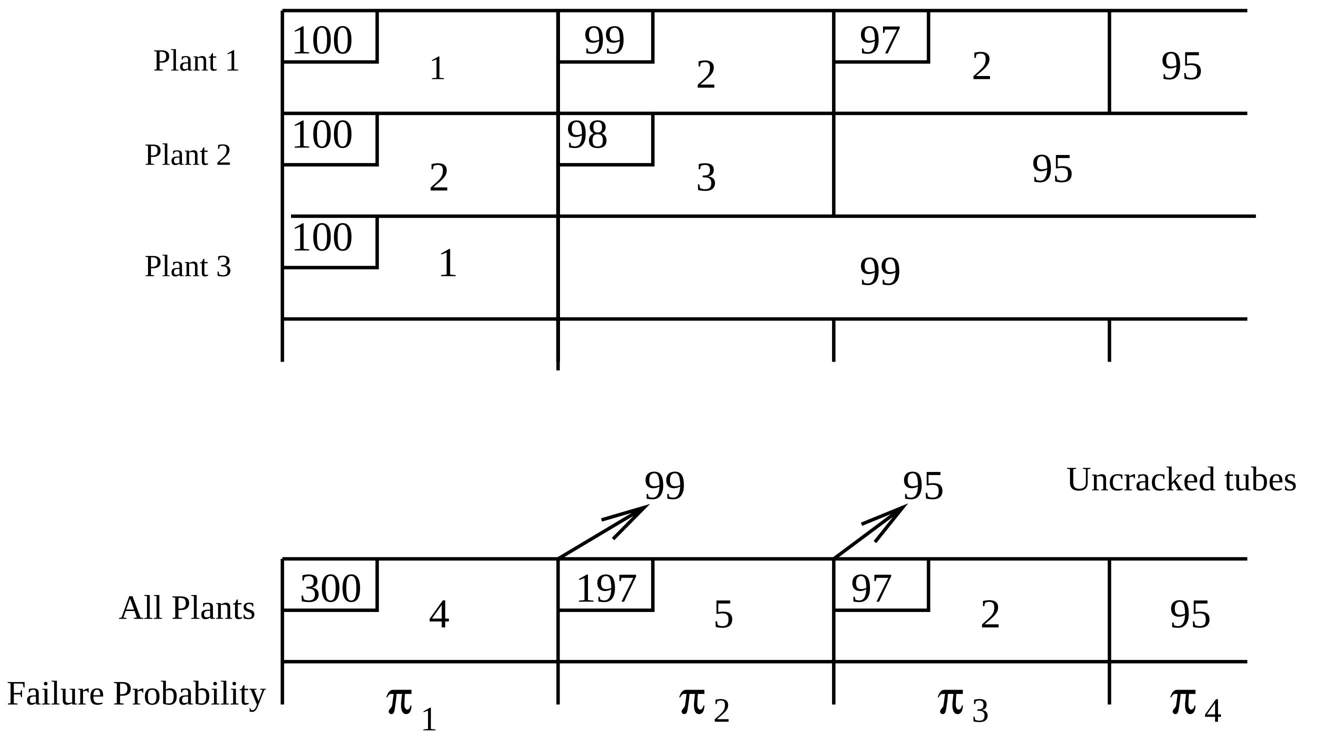


Comments on the Nonparametric Estimate of $F(t)$

- $\widehat{F}(t)$ is defined for all t in the interval $(0, t_c]$ where t_c is the singly censoring time.
- $\widehat{F}(t)$ is the ML estimator of $F(t)$.
- The estimate $\widehat{F}(t)$ is a step up function with a step of size $1/n$ at each exact failure time.

  Sometimes the step size is a multiple of $1/n$ because there are ties on the failure times.
- When there is no censoring, $\widehat{F}(t)$ is the well known empirical cdf.

Pooling of the Heat Exchanger Tube Crack Data



Likelihood: $L(\underline{\pi}) = c\ [\pi_1]^4[\pi_2]^5[\pi_3]^2[\pi_4]^{95}[\pi_3+\pi_4]^{95}[\pi_2+\pi_3+\pi_4]^{99}$

A Nonparametric Estimator of $F(t_i)$ Based on Interval Data and Multiple Censoring

The combined data from the heat exchanger tube crack are multiply censored and the simple binomial method to estimate $F(t_i)$ cannot be used.

Here we describe a more general method to compute a nonparametric estimator of $F(t_i)$.

$$\widehat{F}(t_i) = 1 - \widehat{S}(t_i)$$

$$\text{where} \quad \widehat{S}(t_i) = \prod_{j=1}^{i} \left[1 - \widehat{p}_j\right] \quad \text{with} \quad \widehat{p}_j = \frac{d_j}{n_j}$$

n : sample size

d_i : # of failures (deaths) in the ith interval

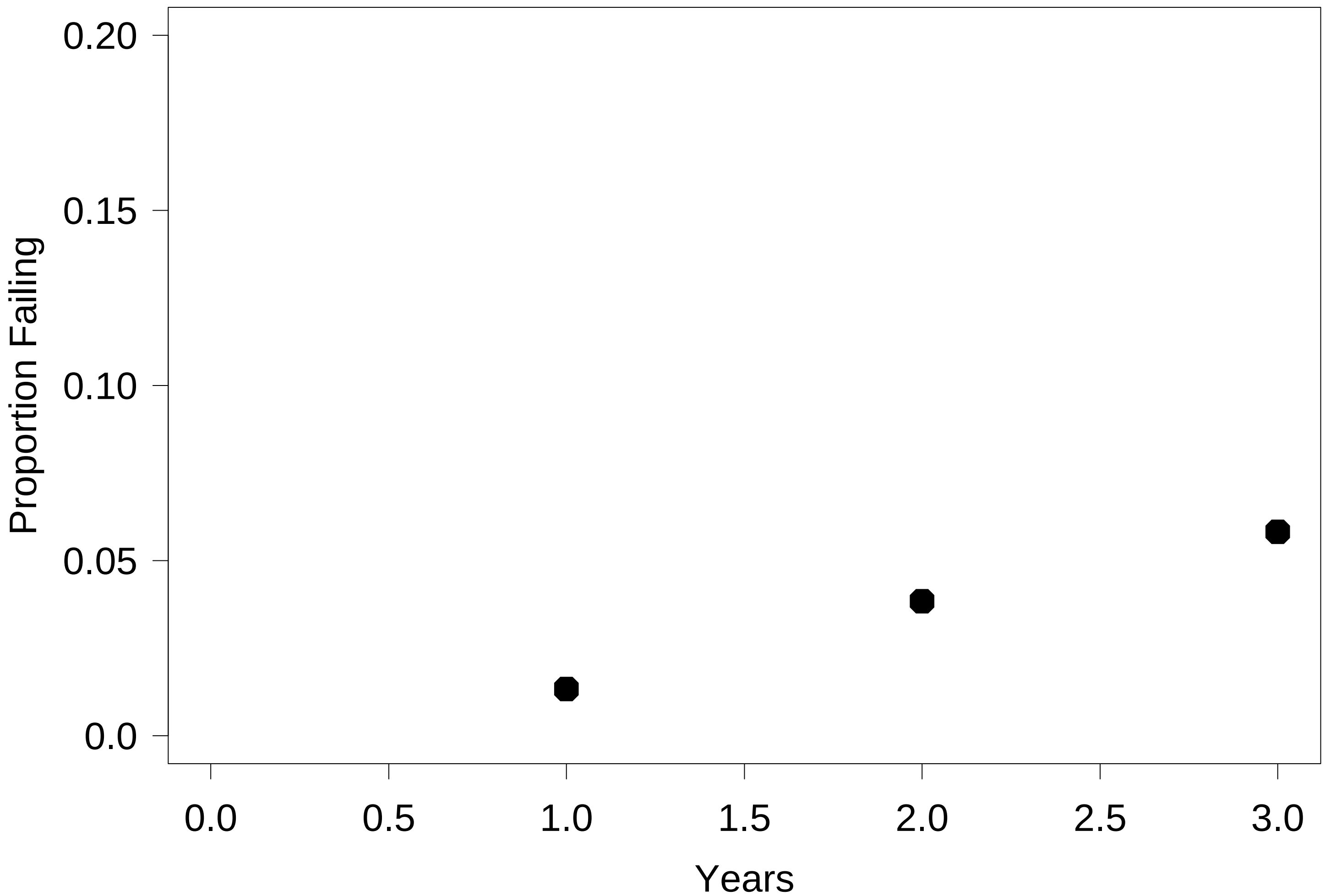
$n_i = n - \sum_{j=0}^{i-1} d_j - \sum_{j=0}^{i-1} r_j$, the risk set at t_{i-1}

r_i : # of right censored obs at t_i

Calculations of the Nonparametric Estimate of $F(t_i)$ for the Heat Exchanger Tube Crack Data

Year	t_i	n_i	d_i	r_i	$\widehat{p}_i$	$1 - \widehat{p}_i$	$\widehat{S}(t_i)$	$\widehat{F}(t_i)$
$(0 - 1]$	1	300	4	99	4/300	296/300	.9867	.0133
$(1 - 2]$	2	197	5	95	5/197	192/197	.9616	.0384
$(2 - 3]$	3	97	2	95	2/97	95/97	.9418	.0582

Nonparametric Estimate for the Heat Exchanger Tube Crack Data



Approximate Variance of $\widehat{F}(t_i)$

- Recall, $\widehat{F}(t_i) = 1 - \widehat{S}(t_i)$ and $\widehat{S}(t_i) = \prod_{j=1}^{i}\left[1 - \widehat{p}_j\right]$.
- Then $\text{Var}\left[\widehat{F}(t_i)\right] = \text{Var}\left[\widehat{S}(t_i)\right]$.
- A Taylor series first-order approximation of $\widehat{S}(t_i)$ is

$$\widehat{S}(t_i) \approx S(t_i) + \sum_{j=1}^{i} \left.\frac{\partial S}{\partial q_j}\right|_{q_j} \left(\widehat{q}_j - q_j\right)$$

where $q_j = 1 - p_j$.

- Then it follows that

$$\text{Var}\left[\widehat{S}(t_i)\right] \approx S^2(t_i) \sum_{j=1}^{i} \frac{p_j}{n_j(1 - p_j)}.$$

Estimating the Standard Error of $\widehat{F}(t_i)$

- Using the variance formula, one gets

$$\widehat{\mathsf{Var}}\left[\widehat{F}(t_i)\right] = \widehat{\mathsf{Var}}\left[\widehat{S}(t_i)\right] = \widehat{S}^2(t_i) \sum_{j=1}^{i} \frac{\widehat{p}_j}{n_j(1-\widehat{p}_j)}$$

  which is known as **Greenwood's** formula.

- An estimate of the standard error, $\mathsf{se}_{\widehat{F}}$, is

$$\widehat{\mathsf{se}}_{\widehat{F}} = \sqrt{\widehat{\mathsf{Var}}\left[\widehat{F}(t_i)\right]} = \widehat{S}(t_i)\sqrt{\sum_{j=1}^{i} \frac{\widehat{p}_j}{n_j(1-\widehat{p}_j)}}.$$

Pointwise Normal-Approximation Confidence Interval for $F(t_i)$-Based on Logit Transformation

- Generally better confidence intervals can be obtained by using the logit transformation ($\text{logit}(p) = \log[p/(1-p)]$) and basing the confidence intervals on

$$Z_{\text{logit}(\widehat{F})} = \frac{\text{logit}[\widehat{F}(t_i)] - \text{logit}[F(t_i)]}{\widehat{\text{se}}_{\text{logit}(\widehat{F})}} \;\dot{\sim}\; \text{NOR}(0,1).$$

- A pointwise normal-approximation $100(1-\alpha)\%$ confidence interval for $\text{logit}[F(t_i)]$ is

$$\begin{aligned}\left[\underset{\sim}{\text{logit}}(\widehat{F}), \quad \widetilde{\text{logit}}(\widehat{F})\right] &= \text{logit}(\widehat{F}) \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{\text{logit}(\widehat{F})} \\ &= \text{logit}(\widehat{F}) \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{F}}/[\widehat{F}(1-\widehat{F})]\end{aligned}$$

since $\widehat{\text{se}}_{\text{logit}(\widehat{F})} = \widehat{\text{se}}_{\widehat{F}}/[\widehat{F}(1-\widehat{F})]$.

Pointwise Normal-Approximation Confidence Interval for $F(t_i)$-Based on Logit Transformation

- The confidence interval for $F(t_i)$ is obtained from the interval for logit(F) and using the inverse logit transformation

$$\text{logit}^{-1}(v) = \frac{1}{1+\exp(-v)}$$

- Then

$$\begin{aligned}[\underset{\sim}{F}(t_i), \quad \tilde{F}(t_i)] &= \text{logit}^{-1}\left[\text{logit}(\widehat{F}) \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{\text{logit}(\widehat{F})}\right] \\ &= \frac{1}{1+\exp\left[-\text{logit}(\widehat{F}) \mp z_{(1-\alpha/2)}\widehat{\text{se}}_{\text{logit}(\widehat{F})}\right]} \\ &= \left[\frac{\widehat{F}}{\widehat{F}+(1-\widehat{F})\times w}, \quad \frac{\widehat{F}}{\widehat{F}+(1-\widehat{F})/w}\right]\end{aligned}$$

where $w = \exp\{z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{F}}/[\widehat{F}(1-\widehat{F})]\}$.

- The endpoints $\underset{\sim}{F}(t_i)$ and $\tilde{F}(t_i)$ will always lie between 0 and 1.

Normal-Approximation Pointwise Confidence Intervals for the Heat Exchanger Tube Crack Data

- Computation of standard errors

$$\widehat{\mathrm{Var}}\left[\widehat{F}(t_i)\right] = \widehat{S}^2(t_i)\sum_{j=1}^{i}\frac{\widehat{p}_j}{n_j(1-\widehat{p}_j)}$$

$$\widehat{\mathrm{Var}}\left[\widehat{F}(t_1)\right] = (.9867)^2\left[\frac{.0133}{300(.9867)}\right] = .0000438$$

$$\widehat{\mathrm{se}}_{\widehat{F}(t_1)} = \sqrt{.0000438} = .00662$$

$$\widehat{\mathrm{Var}}\left[\widehat{F}(t_2)\right] = (.9616)^2\left[\frac{.0133}{300(.9867)} + \frac{.0254}{197(.9746)}\right] = .0001639$$

$$\widehat{\mathrm{se}}_{\widehat{F}(t_2)} = \sqrt{.0001639} = .0128$$

Normal-Approximation Pointwise Confidence Intervals for the Heat Exchanger Tube Crack Data

Computation of approximate 95% confidence intervals:

- For $F(1)$ with $\widehat{F}(t_1) = .0133$, $\widehat{\text{se}}_{\widehat{F}(t_1)} = \sqrt{.0000438} = .00662$

  Based on: $Z_{\widehat{F}} = [\widehat{F}(t_1) - F(t_1)]/\widehat{\text{se}}_{\widehat{F}} \,\dot{\sim}\, \text{NOR}(0,1)$.

  $$[\underset{\sim}{F}(t_1),\ \tilde{F}(t_1)] = .0133 \pm 1.96(.00662) = [.0003,\ .0263].$$

  Based on: $Z_{\text{logit}(\widehat{F})} = [\text{logit}[\widehat{F}(t_1)] - \text{logit}[F(t_1)]/\widehat{\text{se}}_{\text{logit}(\widehat{F})} \,\dot{\sim}\, \text{NOR}(0,1)$.

  $$[\underset{\sim}{F}(t_1),\ \tilde{F}(t_1)] = \left[\frac{.0133}{.0133 + (1 - .0133) \times w},\ \frac{.0133}{.0133 + (1 - .0133)/w}\right] = [.0050,\ .0350]\,.$$

  $w = \exp\{1.96(.00662)/[.0133(1 - .0133)]\} = 2.687816$.

- For $F(2)$ with $\widehat{F}(t_2) = .0384$, $\widehat{\text{se}}_{\widehat{F}(t_2)} = \sqrt{.0001639} = .0128$

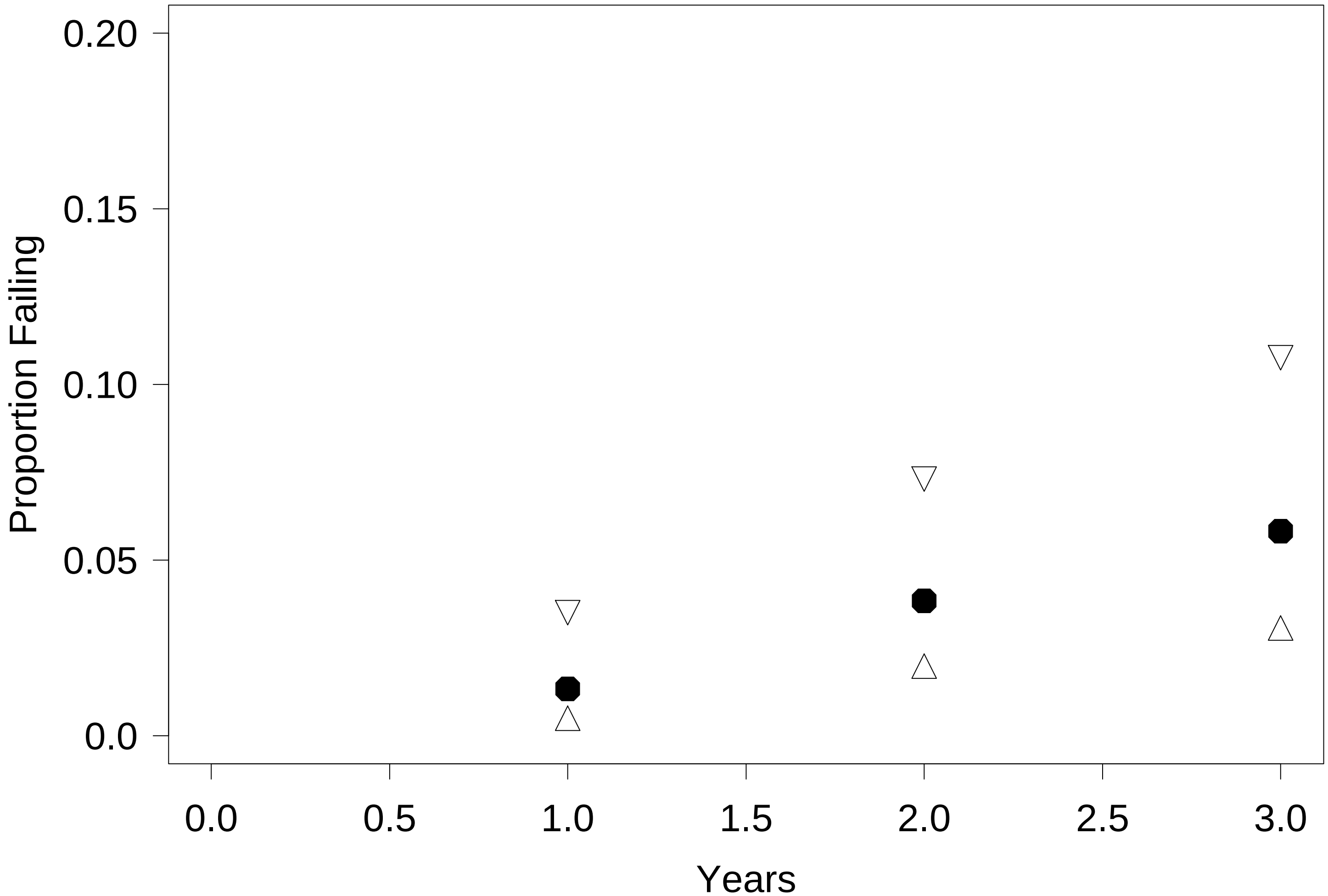
  Based on: $Z_{\widehat{F}}$, $\quad [\underset{\sim}{F}(t_2),\ \tilde{F}(t_2)] = [.0133,\ .0635]$.

  Based on: $Z_{\text{logit}(\widehat{F})}$, $\quad [\underset{\sim}{F}(t_2),\ \tilde{F}(t_2)] = [.0198,\ .0730]\,.$

Results of Calculations for Nonparametric Pointwise Confidence Intervals for $F(t_i)$ for the Heat Exchanger Tube Crack Data

Year	t_i	$\widehat{F}(t_i)$	$\widehat{\text{se}}_{\widehat{F}}$	Pointwise Confidence Intervals
$(0-1]$	1	.0133	.00662	
95% Confidence Intervals for $F(1)$				
Based on $Z_{\text{logit}(\widehat{F})} \dot{\sim} \text{NOR}(0,1)$				[.0050, .0350]
Based on $Z_{\widehat{F}} \dot{\sim} \text{NOR}(0,1)$				[.0003, .0263]
$(1-2]$	2	.0384	.0128	
95% Confidence Intervals for $F(2)$				
Based on $Z_{\text{logit}(\widehat{F})} \dot{\sim} \text{NOR}(0,1)$				[.0198, .0730]
Based on $Z_{\widehat{F}} \dot{\sim} \text{NOR}(0,1)$				[.0133, .0635]
$(2-3]$	3	.0582	.0187	
95% Confidence Intervals for $F(3)$				
Based on $Z_{\text{logit}(\widehat{F})} \dot{\sim} \text{NOR}(0,1)$				[.0307, .1076]
Based on $Z_{\widehat{F}} \dot{\sim} \text{NOR}(0,1)$				[.0216, .0949]

Heat Exchanger Tube Crack Nonparametric Estimate with Pointwise 95% Confidence Intervals Based on $Z_{\mathrm{logit}(\widehat{F})}$



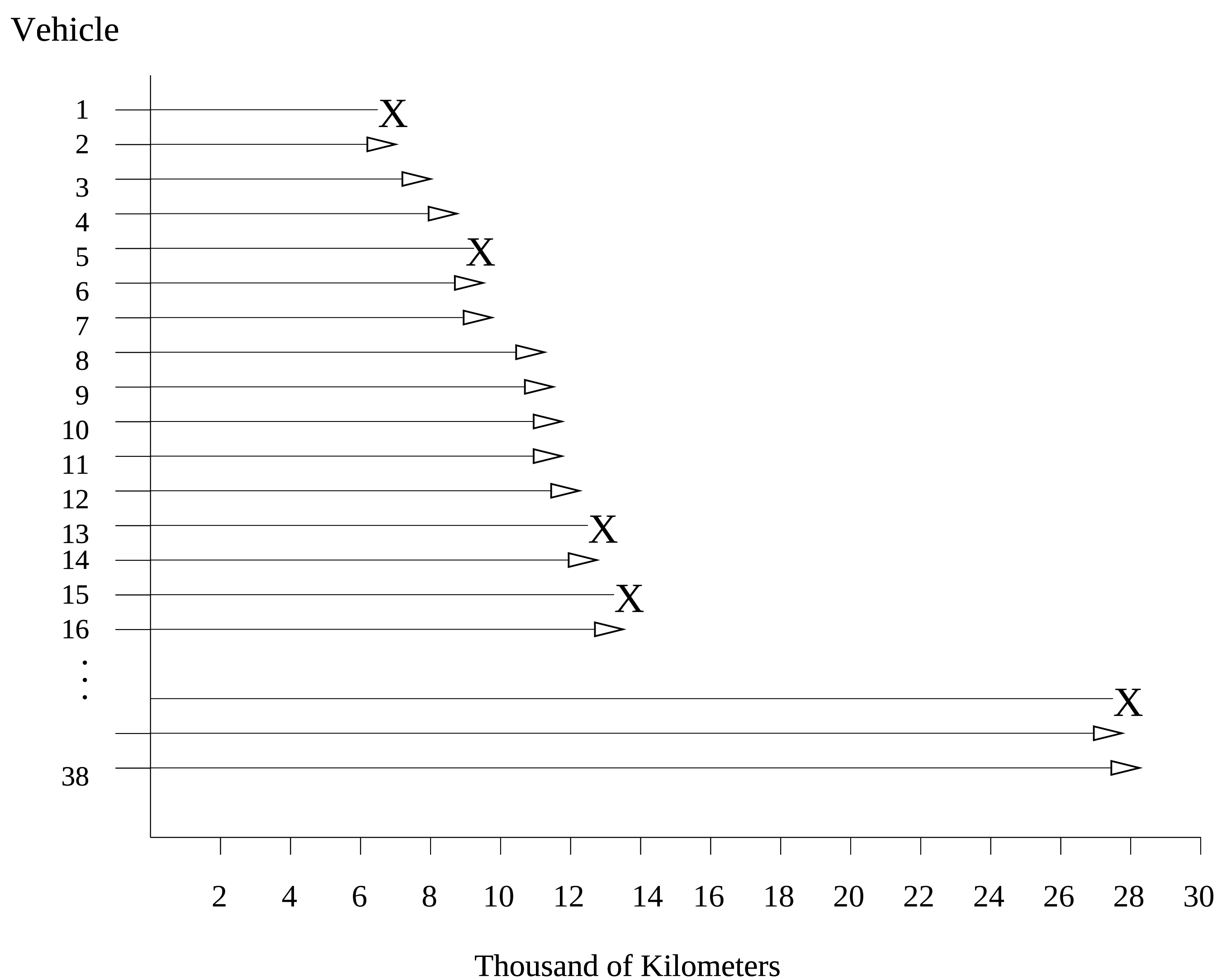
Shock Absorber Failure Data

First reported in O'Connor (1985).

- Failure times, in number of kilometers of use, of vehicle shock absorbers.
- Two failure modes, denoted by M1 and M2.
- One might be interested in the distribution of time to failure for mode M1, mode M2, or in the overall failure-time distribution of the part.

  Here we do not differentiate between modes M1 and M2. We will estimate the distribution of time to failure by either mode M1 or M2.

Failure Pattern in the Shock Absorber Data
Failure Mode Ignored
(O'Connor 1985)



Nonparametric Estimation of $F(t)$ with Exact Failures (Kaplan-Meier) Estimator

In the limit, as the number of inspections increases and the width of the inspection intervals approaches zero, we get the **product-limit** or **Kaplan-Meier** estimator:

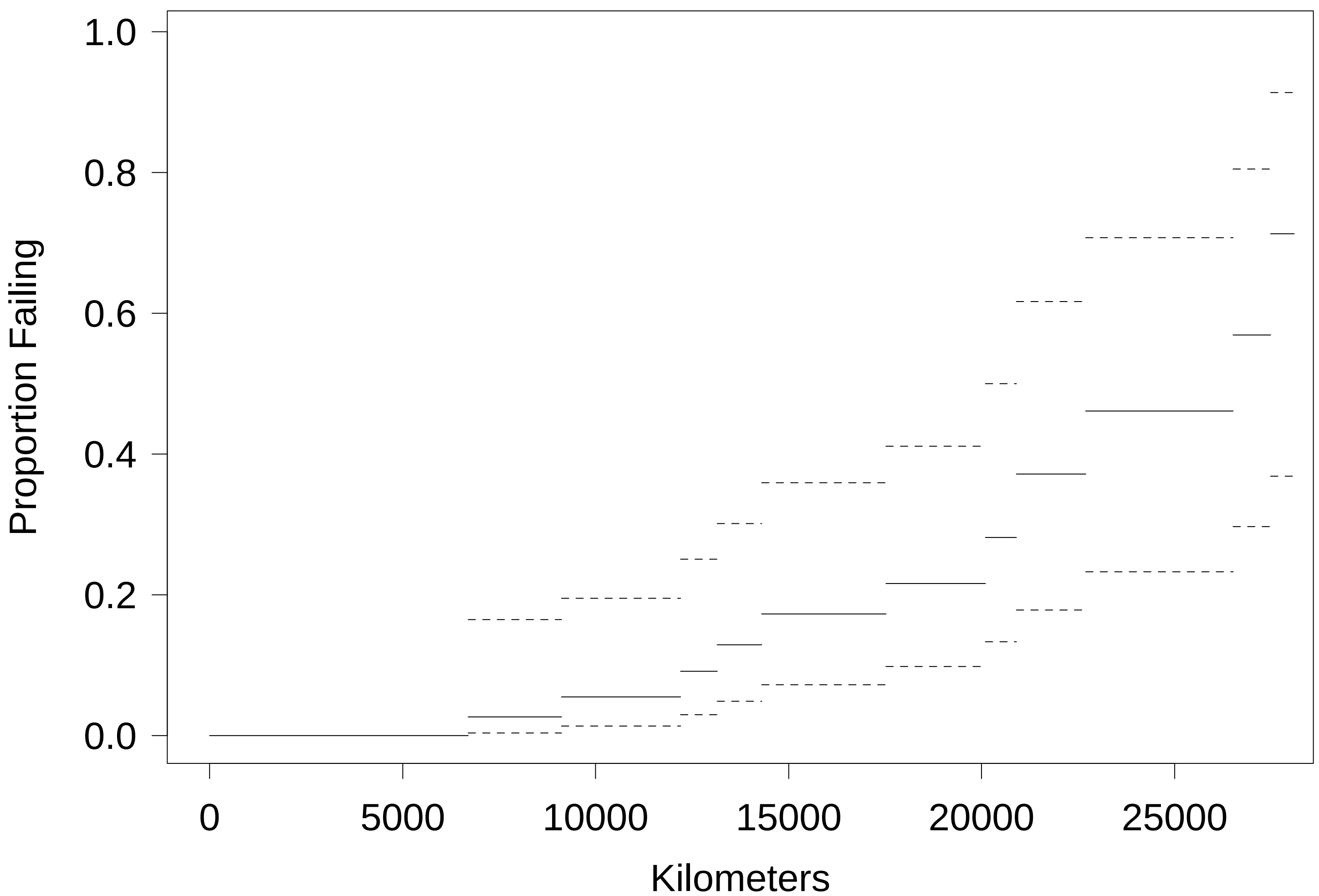
- Failures are concentrated in a small number of intervals of infinitesimal length.
- $\widehat{F}(t)$ will be **constant** over all intervals that have no failures.
- $\widehat{F}(t)$ is a step function with **jumps** at each reported failure time.

Note: The binomial estimator for exact failures and singly right censored data is a special case of the Kaplan-Meier estimate.

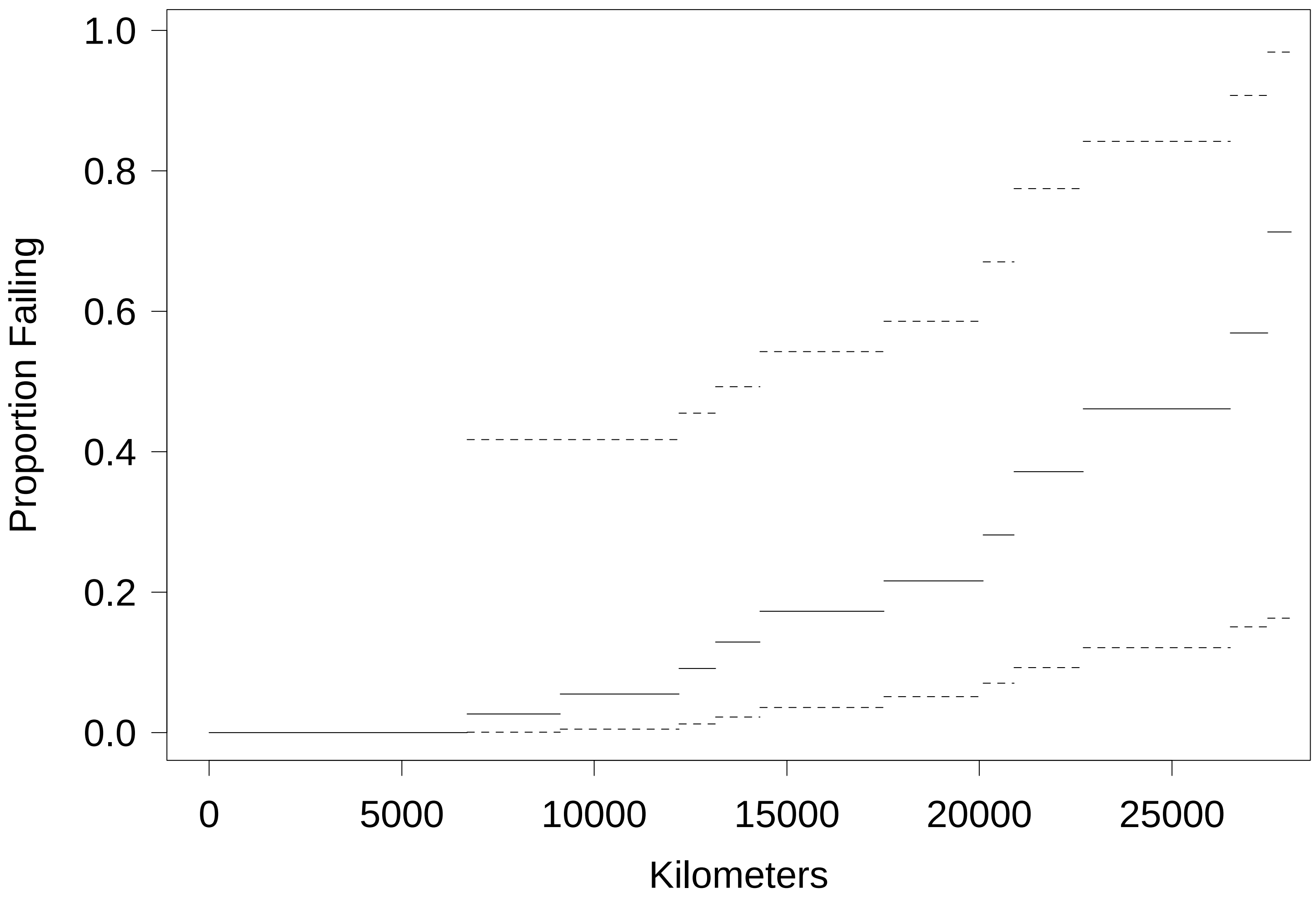
Nonparametric Estimates for the Shock Absorber Data up to 12,220 km

				Conditional		Unconditional	
t_j (km)	n_j	d_j	r_j	$\widehat{p}_j$	$1-\widehat{p}_j$	$\widehat{S}(t_j)$	$\widehat{F}(t_j)$
6,700	38	1	0	1/38	37/38	0.97368	0.02632
6,950	37	0	1				
7,820	36	0	1				
8,790	35	0	1				
9,120	34	1	0	1/34	33/34	0.94505	0.05495
9,660	33	0	1				
9,820	32	0	1				
11,310	31	0	1				
11,690	30	0	1				
11,850	29	0	1				
11,880	28	0	1				
12,140	27	0	1				
12,200	26	1	0	1/26	25/26	0.90870	0.09130
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Nonparametric Estimate for Shock Absorber Data with Pointwise 95% Confidence Intervals Based on $Z_{\mathrm{logit}(\widehat{F})}$



Nonparametric Estimate for Shock Absorber Data with Simultaneous 95% Confidence Bands Based on $Z_{\text{logit}(\widehat{F})}$



Need for Nonparametric Simultaneous Confidence Bands for $F(t)$

- **Pointwise confidence intervals** for $F(t)$ are useful for making a statement about $F(t)$ at one particular value of t.

- **Simultaneous confidence bands** for $F(t)$ are necessary to quantify the sampling uncertainty over a range of values of t.

Nonparametric Simultaneous Confidence Bands for $F(t)$

Approximate $100(1-\alpha)\%$ simultaneous confidence bands for F can be obtained from

$$\left[\underset{\sim}{F}(t),\ \tilde{F}(t)\right] = \widehat{F}(t) \pm e_{(a,b,1-\alpha/2)} \widehat{\mathrm{se}}_{\widehat{F}}(t) \quad \text{for all } t \in [t_L(a), t_U(b)]$$

where $[t_L(a), t_U(b)]$ is a complicated function of the censoring pattern in the data.

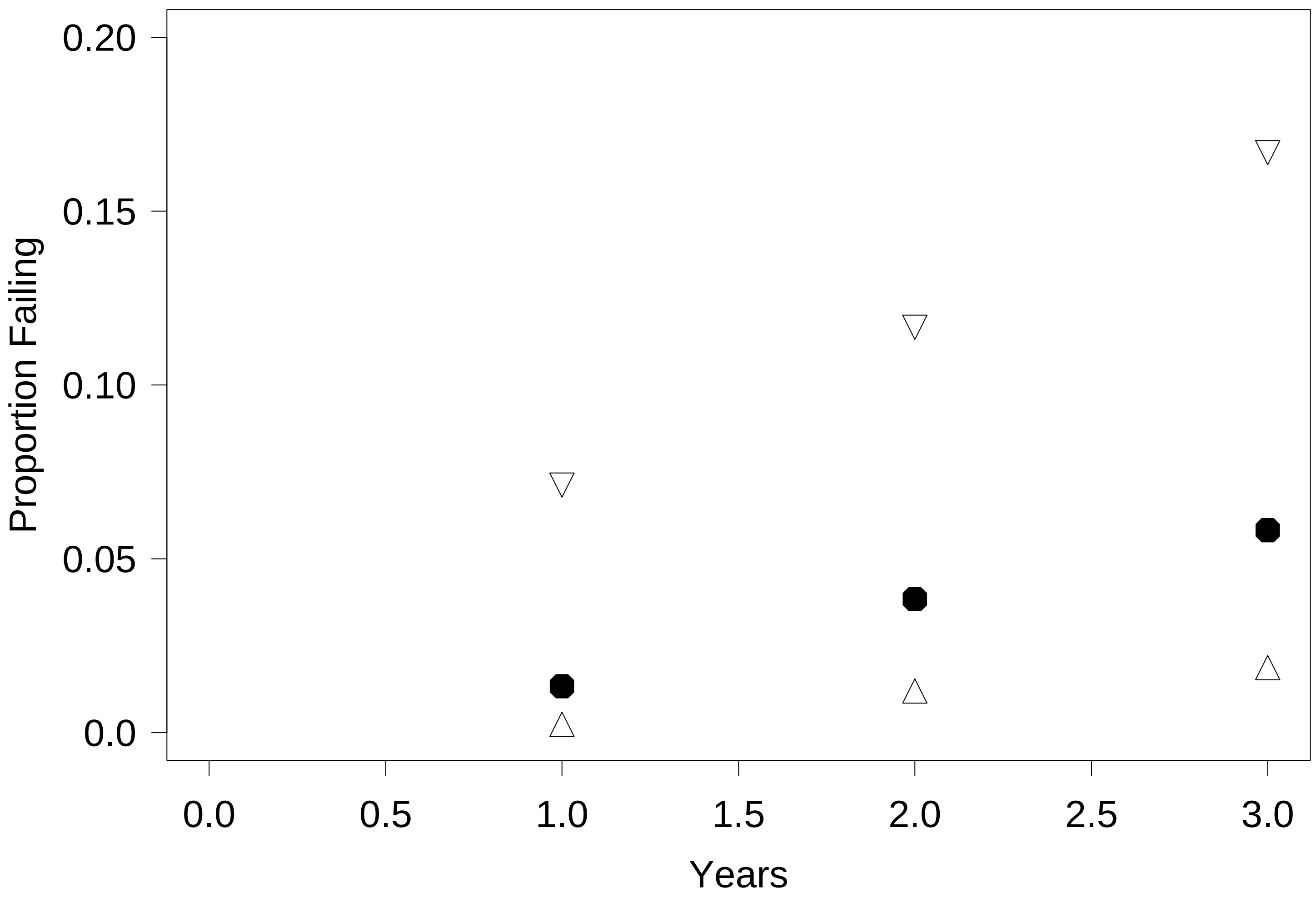
Comments:

- The approximate factors $e_{(a,b,1-\alpha/2)}$ can be computed from a large-sample approximation given in Nair (1984).
- $e_{(a,b,1-\alpha/2)}$ is the same for all values of t.
- The factors $e_{(a,b,1-\alpha/2)}$ are greater than the corresponding $z_{(1-\alpha/2)}$.

Factors $e_{(a,b,1-\alpha/2)}$ for Computing the EP Nonparametric Simultaneous Approximate Confidence Bands

Limits		Confidence Level			
a	b	.80	.90	.95	.99
.005	.999	2.92	3.17	3.41	3.88
.01	.999	2.90	3.15	3.39	3.87
.05	.999	2.84	3.10	3.34	3.82
.001	.995	2.92	3.17	3.41	3.88
.005	.995	2.86	3.12	3.36	3.85
.01	.995	2.84	3.10	3.34	3.83
.05	.995	2.76	3.03	3.28	3.77
.001	.99	2.90	3.15	3.39	3.87
.005	.99	2.84	3.10	3.34	3.83
.01	.99	2.81	3.07	3.31	3.81
.05	.99	2.73	3.00	3.25	3.75
.001	.95	2.84	3.10	3.34	3.82
.005	.95	2.76	3.03	3.28	3.77
.01	.95	2.73	3.00	3.25	3.75
.05	.95	2.62	2.91	3.16	3.68
.001	.9	2.80	3.07	3.31	3.80
.005	.9	2.72	3.00	3.25	3.75
.01	.9	2.68	2.96	3.21	3.72
.05	.9	2.56	2.85	3.11	3.64

Nonparametric Estimate Heat Exchanger Tube Crack Data with Simultaneous 95% Confidence Bands Based on $Z_{\max \text{logit}(\widehat{F})}$



Better Nonparametric Simultaneous Confidence Bands for $F(t)$

- The approximate 100(1−α)% simultaneous confidence bands

$$\left[\underset{\sim}{F}(t),\ \tilde{F}(t)\right] = \widehat{F}(t) \pm e_{(a,b,1-\alpha/2)}\widehat{\text{se}}_{\widehat{F}}(t) \quad \text{for all } t \in [t_L(a), t_U(b)]$$

  are based on the the approximate distribution of

$$Z_{\text{max}\widehat{F}} = \max_{t \in [t_L(a), t_U(b)]} \left[\frac{\widehat{F}(t) - F(t)}{\widehat{\text{se}}_{\widehat{F}(t)}}\right].$$

- It is generally better to compute the simultaneous confidence bands based on the logit transformation of $\widehat{F}$. This gives

$$[\underset{\sim}{F}(t),\ \tilde{F}(t)] = \left[\frac{\widehat{F}(t)}{\widehat{F}(t) + [1 - \widehat{F}(t)] \times w}, \frac{\widehat{F}(t)}{\widehat{F}(t) + [1 - \widehat{F}(t)]/w}\right]$$

  where $w = \exp\{e_{(a,b,1-\alpha/2)}\widehat{\text{se}}_{\widehat{F}}/[\widehat{F}(1 - \widehat{F})]\}$.

  These are based on the approximate distribution of

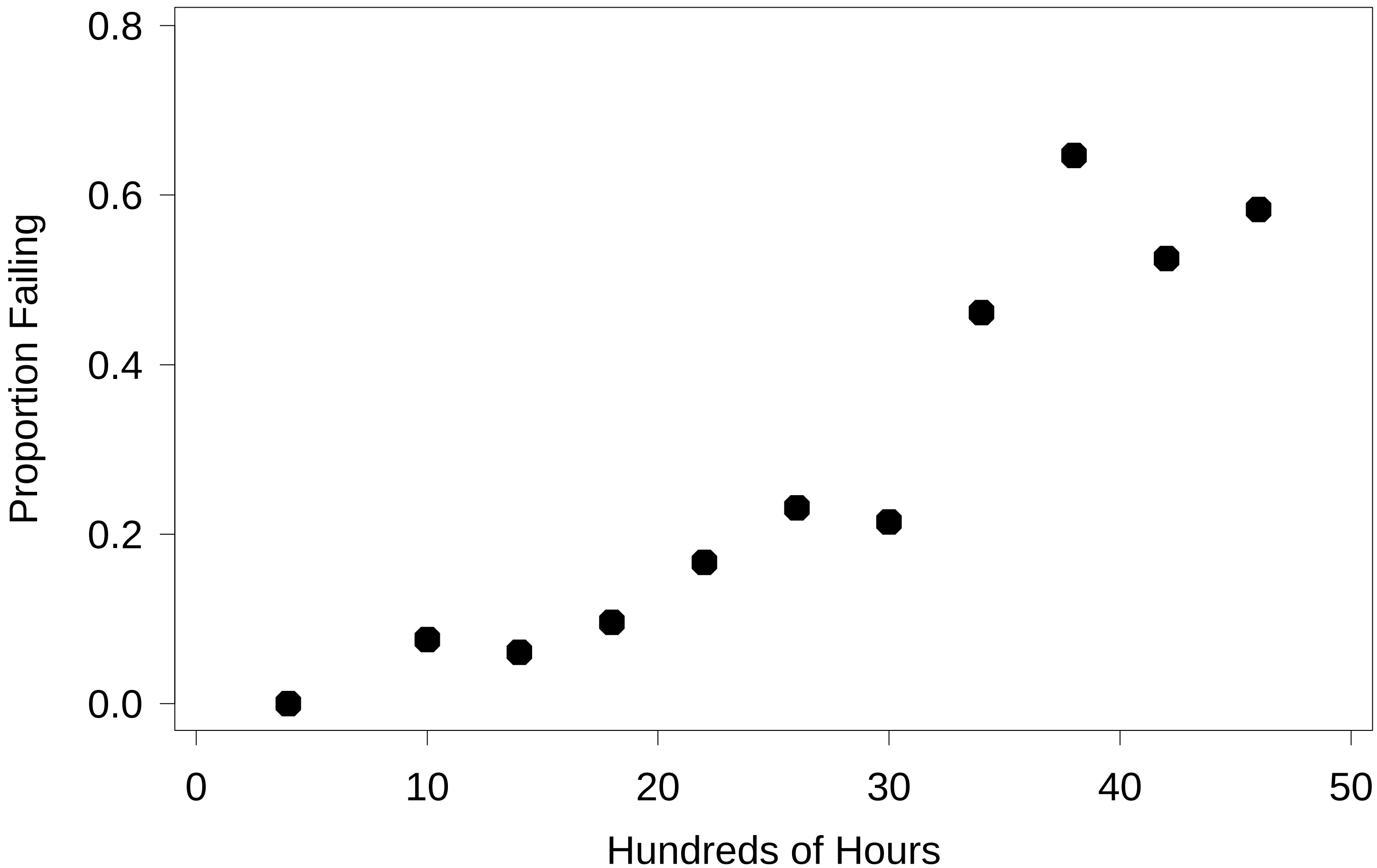
$$Z_{\text{max}\,\text{logit}(\widehat{F})} = \max_{t \in [t_L(a), t_U(b)]} \left[\frac{\text{logit}[\widehat{F}(t)] - \text{logit}[F(t)]}{\widehat{\text{se}}_{\text{logit}[\widehat{F}(t)]}}\right].$$

Nonparametric Estimation of $F(t_i)$ with Arbitrary Censoring

- The methods described so far works only for some kinds of censoring patterns (multiple right censoring, interval censoring with intervals that do not overlap, and some other very special censoring patterns.)

- The nonparametric maximum likelihood generalizations provided by the **Peto/Turnbull** estimator can be used for
  - ▶ Arbitrary censoring (e.g., both left and right).
  - ▶ Censoring with overlapping intervals.
  - ▶ Truncated data.

Plot of Proportions Failing Versus Hours of Exposure for the Turbine Wheel Inspection Data



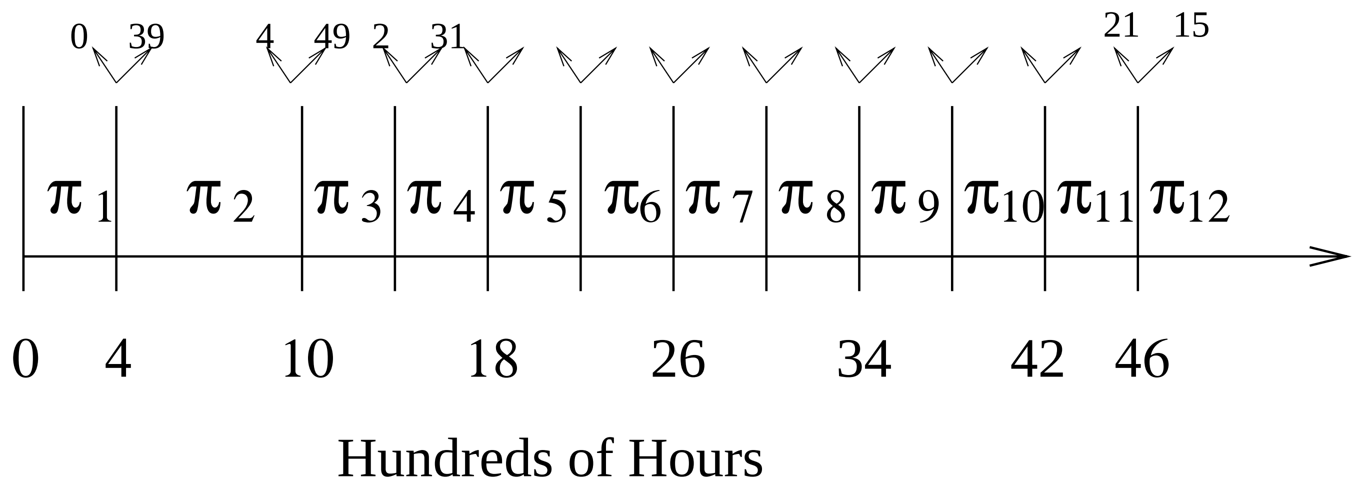
Turbine Wheel Inspection Data Summary

100-hours of Exposure t_i	# Cracked Left Censored	# Not Cracked Right Censored	Proportion Cracked Crude Estimate of $F(t)$
4	0	39	$0/39 = .000$
10	4	49	$4/53 = .075$
14	2	31	$2/33 = .060$
18	7	66	$7/73 = .096$
22	5	25	$5/30 = .167$
26	9	30	$9/39 = .231$
30	9	33	$9/42 = .214$
34	6	7	$6/13 = .462$
38	22	12	$22/34 = .647$
42	21	19	$21/40 = .525$
46	21	15	$21/36 = .583$

Data from Nelson (1982), page 409.

- The analysts did not know the initiation time for any of the wheels.
- All they knew about each wheel was its exposure time and whether a crack had initiated or not. Units grouped by exposure time.

Basic Parameters Used in Computing the Nonparametric ML Estimate of $F(t)$ for the Turbine Wheel Data



Nonparametric Estimation of $F(t)$ with Arbitrary Censoring-General Approach

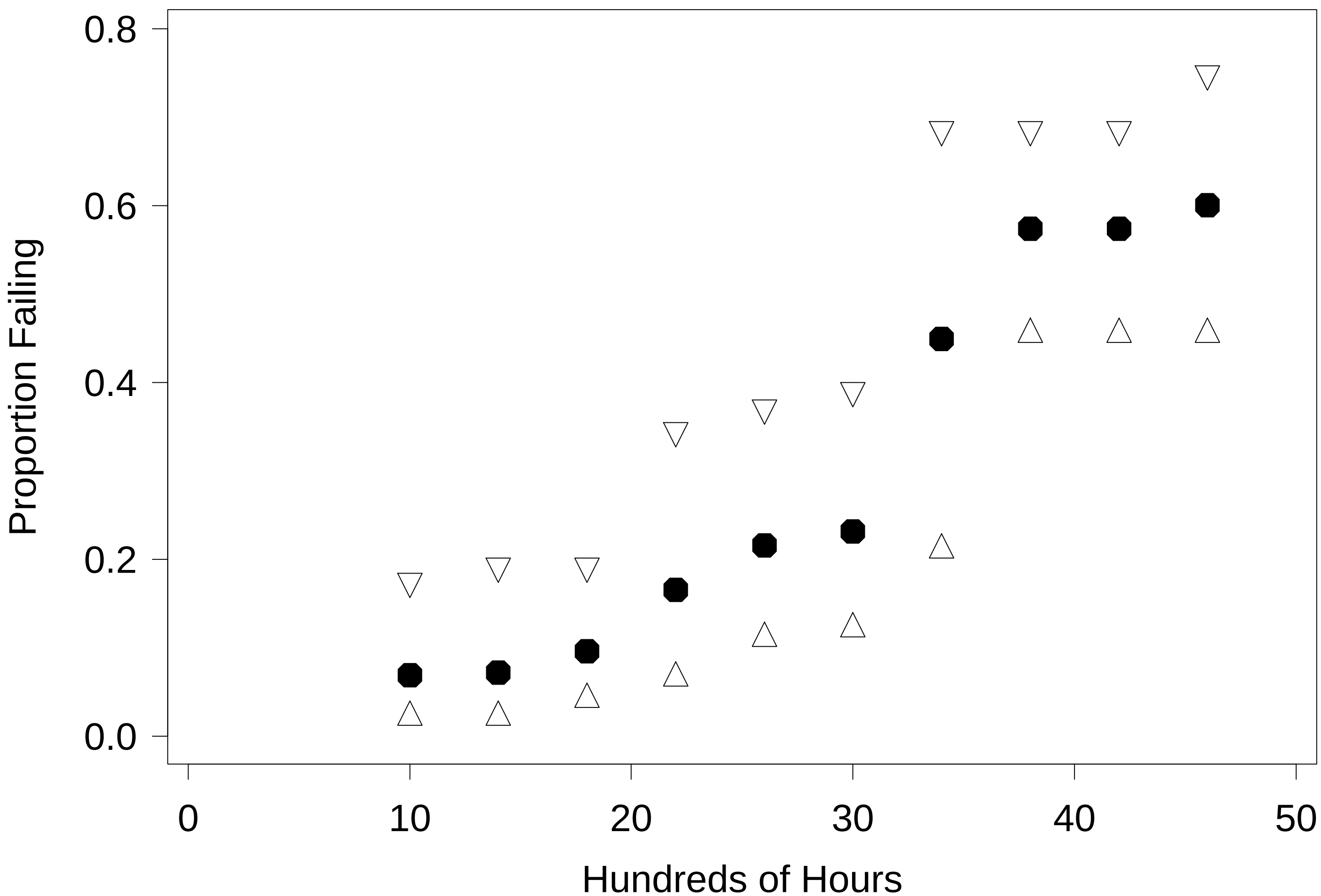
- **Basic idea:** write the likelihood and maximize this likelihood to obtain $\widehat{\boldsymbol{p}}$ or $\widehat{\boldsymbol{\pi}}$ from which one gets $\widehat{F}(t_i)$ (Peto 1973).

- **Illustration:** the likelihood for the turbine wheel inspection data is

$$
\begin{aligned}
L(\boldsymbol{\pi}) = L(\boldsymbol{\pi}; \text{DATA}) \; = \; & \mathcal{C} \times [\pi_1]^0 \times [\pi_2 + \cdots + \pi_{12}]^{39} \times \\
& [\pi_1 + \pi_2]^4 \times [\pi_3 + \cdots + \pi_{12}]^{49} \times \\
& [\pi_1 + \cdots + \pi_3]^2 \times [\pi_4 + \cdots + \pi_{12}]^{31} \times \\
& \vdots \\
& [\pi_1 + \cdots + \pi_{11}]^{21} \times [\pi_{12}]^{15}
\end{aligned}
$$

where $\pi_{12} = 1 - \sum_{i=1}^{11} \pi_i$. The values of $\pi_1, \ldots, \pi_{11}$ that maximize $L(\boldsymbol{\pi})$ gives $\widehat{\boldsymbol{\pi}}$, the ML estimator of $\boldsymbol{\pi}$. Then $\widehat{F}(t_i) = \sum_{j=1}^{i} \widehat{\pi}_j$, $i = 1, \ldots, m$.

Nonparametric ML estimate for the turbine wheel data with 95% Pointwise Confidence Intervals for $F(t_i)$ Based on $Z_{\mathrm{logit}(\widehat{F})}$



Other Topics in Chapter 3

- Maximum likelihood methods to compute nonparametric confidence intervals and confidence bands.

- Uncertain censoring times.