Chapter 15

Prediction of Failures Times and the Number of Field Failures

W. Q. Meeker, L. A. Escobar, and F. G. Pascual
Iowa State University, Louisiana State University, and Washington State University.

Based on Meeker, Escobar, and Pascual (2021): *Statistical Methods for Reliability Data, Second Edition*, John Wiley & Sons Inc.

May 24, 2021
11h 1min

Chapter 15
Prediction of Failures Times and the Number of Field Failures

Topics discussed in this chapter are:

- Prediction applications.
- New-Sample prediction and probability prediction.
- Coverage probabilities concepts and plug-in statistical prediction intervals.
- Calibrating statistical prediction intervals and predictive distributions.
- Within-sample prediction: prediction of the number of future field failures:
  - Single cohort.
  - Multiple cohorts (staggered entry).
- Bayesian predictive distributions.
- Choosing a distribution for prediction and alternative models and methods.

Chapter 15

Segment 1

Prediction Applications

What is Needed to do Prediction?

New-Sample Prediction Basic Ideas

Probability Prediction

Prediction Applications

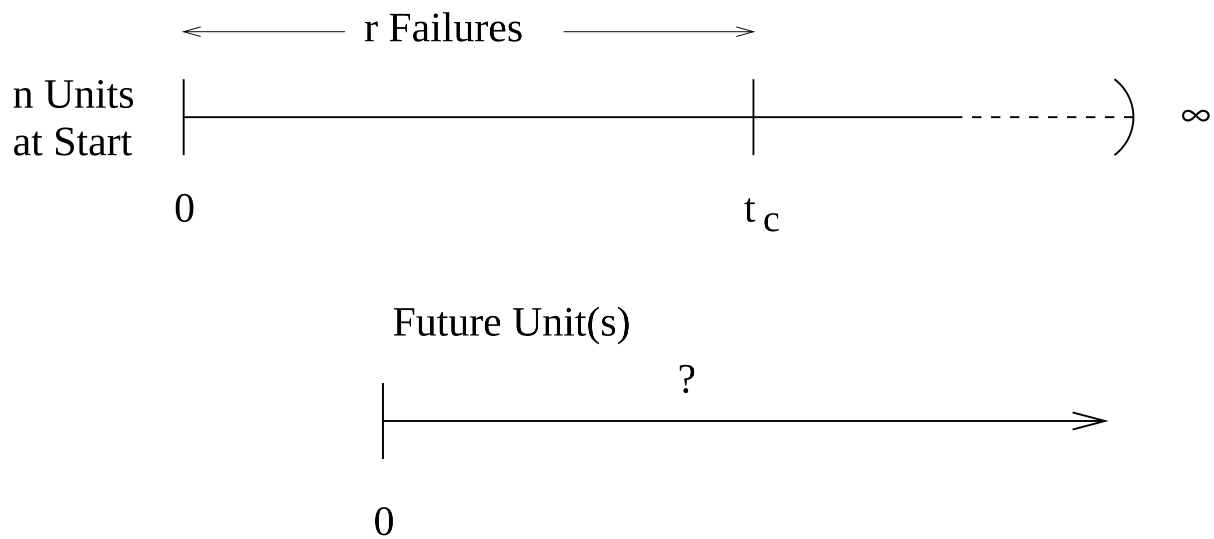
Motivation: Prediction problems are of interest to consumers, managers, engineers, and scientists.

- A consumer would like to **predict** the failure time of a product to be purchased (especially the lower bound on lifetime).
- Finance managers want to **predict** future warranty costs.
- A reliability engineer needs to **predict** the length of a life test of ten units where the test will be terminated after four units fail.
- Managers want to **predict** the **number** of future failures for capital-budget planning.
- Regulators need to **predict** the **number** of future failures to decide whether a product recall is warranted or not.

New-Sample Prediction

Based on previous (possibly censored) life test or field data, one could be interested in:

- **Time** to failure of a **new** item.



Needed for Prediction

In general to predict one needs:

- A probability **distribution or model** to describe random variable of interest (e.g., a failure time having a Weibull distribution). This model depends on parameters in $\boldsymbol{\theta}$.

- **Information** about the parameters in $\boldsymbol{\theta}$. This information could come from:
  - Laboratory test data.
  - Field data.
  - Previous experience or expert opinion.

- Nonparametric new-sample prediction is also possible (e.g., Chapter 5 of Meeker, Hahn, and Escobar, 2017).

Probability Prediction Interval
(θ Given)

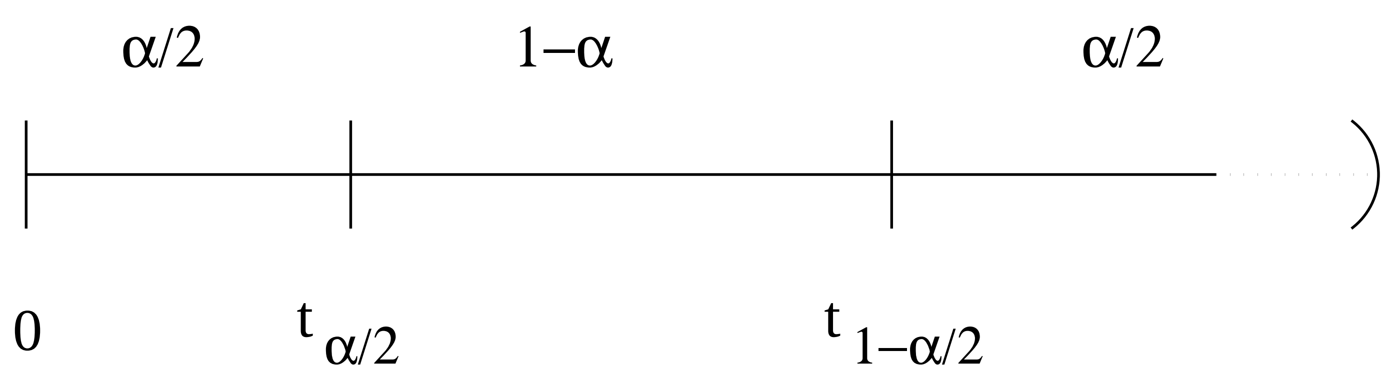
- An **exact** $100(1-\alpha)\%$ probability prediction interval is defined by appropriate quantiles of the distribution:

$$PI(1-\alpha) = [\underset{\sim}{T},\ \widetilde{T}] = [t_{\alpha/2},\ t_{1-\alpha/2}],$$

where $t_p = t_p(\boldsymbol{\theta})$ is the p quantile of T.

- By the definition of the distribution quantiles, the **coverage probability** is

$$\begin{aligned}\Pr[T \in PI(1-\alpha)] &= \Pr(\underset{\sim}{T} \le T \le \widetilde{T}) \\ &= \Pr(t_{\alpha/2} \le T \le t_{1-\alpha/2}) = 1-\alpha.\end{aligned}$$



Example 1: Probability Prediction for the Failure Time of a Single Future Unit Based on Given Parameters

- Suppose that the cycles to failure has a **lognormal** distribution with **given** parameters $\mu = 4.098, \sigma = 0.4761$
- A 90% probability prediction interval is

$$
\begin{aligned}
PI(1-\alpha) &= [\underset{\sim}{T},\ \widetilde{T}] \\
&= [t_{\alpha/2},\ t_{1-\alpha/2}] = [t_{0.05},\ t_{0.95}] \\
&= [\exp(4.098 - 1.645 \times 0.4761), \exp(4.098 + 1.645 \times 0.4761)] \\
&= [26.1,\ 157.1].
\end{aligned}
$$

- Then $\Pr(\underset{\sim}{T} \le T \le \widetilde{T}) = \Pr(26.1 \le T \le 157.1) = 0.90.$
- With misspecified parameters, the coverage probability may not be 0.90.
- Note that with given parameters, a $100(1-\alpha)\%$ probability prediction interval is also a $100(1-\alpha)\%$ tolerance interval.

Chapter 15

Segment 2

Statistical Prediction

Coverage Probabilities Concepts

The Pivotal Method

Statistical Prediction Interval (θ Estimated)

Objective: Want to predict the random quantity T based on a **learning** sample information (DATA).

- The random DATA leads to a parameter estimate $\widehat{\boldsymbol{\theta}}$ and prediction interval $PI(1-\alpha) = [\underset{\sim}{T}(\widehat{\boldsymbol{\theta}}),\ \widetilde{T}(\widehat{\boldsymbol{\theta}})]$.
- Thus $[\underset{\sim}{T}(\widehat{\boldsymbol{\theta}}),\ \widetilde{T}(\widehat{\boldsymbol{\theta}})]$ and T have a joint distribution that may depend on a parameter $\boldsymbol{\theta}$.
- $PI(1-\alpha)$ is an **exact** $100(1-\alpha)\%$ prediction interval procedure if the **coverage probability** is

$$\Pr[T \in PI(1-\alpha)] = \Pr[\underset{\sim}{T}(\widehat{\boldsymbol{\theta}}) \leq T \leq \widetilde{T}(\widehat{\boldsymbol{\theta}})] = 1-\alpha.$$

- First we consider **evaluation of the coverage probability** of the $PI(1-\alpha)$ procedure, then **specification of the procedure**.

Coverage Probabilities Concepts

- **Conditional** coverage probability for **an interval**:
For fixed DATA (and thus fixed $\widehat{\boldsymbol{\theta}}$ and $[\underset{\sim}{T}, \ \widetilde{T}]$):

$$\begin{aligned}\text{CP}[PI(1-\alpha) \mid \widehat{\boldsymbol{\theta}}; \boldsymbol{\theta}] &= \Pr(\underset{\sim}{T} \leq T \leq \widetilde{T} \mid \widehat{\boldsymbol{\theta}}; \boldsymbol{\theta}) \\ &= F(\widetilde{T}; \boldsymbol{\theta}) - F(\underset{\sim}{T}; \boldsymbol{\theta})\end{aligned}$$

Random because $[\underset{\sim}{T}, \ \widetilde{T}]$ depends on $\widehat{\boldsymbol{\theta}}$.
Unknown because $F(t; \boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$.

- **Unconditional** coverage probability for **the procedure**:

$$\begin{aligned}\text{CP}[PI(1-\alpha); \boldsymbol{\theta}] &= \Pr(\underset{\sim}{T} \leq T \leq \widetilde{T}; \boldsymbol{\theta}) \\ &= \text{E}_{\widehat{\boldsymbol{\theta}}}\Big\{\text{CP}[PI(1-\alpha) \mid \widehat{\boldsymbol{\theta}}; \boldsymbol{\theta}]\Big\}.\end{aligned}$$

In general $\text{CP}[PI(1-\alpha); \boldsymbol{\theta}] \neq 1-\alpha$.

- When $\text{CP}[PI(1-\alpha); \boldsymbol{\theta}]$ does not depend on $\boldsymbol{\theta}$, $\text{CP}[PI(1-\alpha); \boldsymbol{\theta}] = 1-\alpha$ and $PI(1-\alpha)$ is an **exact** prediction procedure.

One-Sided Prediction Bounds and Two-Sided Prediction Intervals

- Combining lower and upper $100(1-\alpha/2)\%$ prediction bounds gives an equal-probability two-sided $100(1-\alpha)\%$ prediction interval. **Desire equal probability in each tail.**

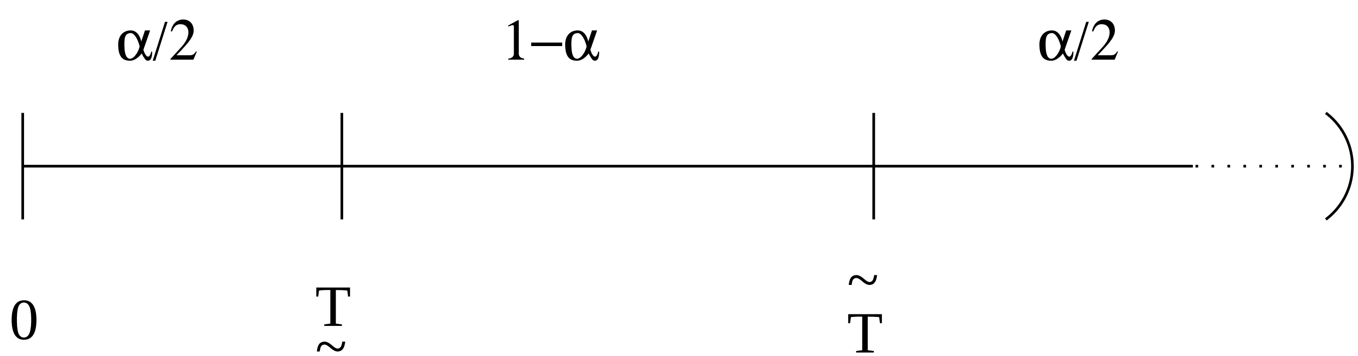
- If

$$\Pr(\underset{\sim}{T} \le T < \infty) = 1 - \alpha/2 \quad \text{and}$$

$$\Pr(0 < T \le \widetilde{T}) = 1 - \alpha/2,$$

then

$$\Pr(\underset{\sim}{T} \le T \le \widetilde{T}) = 1 - \alpha.$$



Prediction Based on a Pivotal Quantity

- With complete data or failure (type 2) censoring,

$$Z_{\log(T)} = \frac{\log(T) - \widehat{\mu}}{\widehat{\sigma}}$$

  is a pivotal quantity, with respect to the joint distribution of T, $\widehat{\mu}$, and $\widehat{\sigma}$. That is, $Z_{\log(T)}$ has a distribution with no unknown parameters.

- One can then write

$$\Pr\left[\widehat{\mu} + z_{\log(T)_{(\alpha/2)}} \times \widehat{\sigma} < \log(T) \leq \widehat{\mu} + z_{\log(T)_{(1-\alpha/2)}} \times \widehat{\sigma}\right] = 1 - \alpha,$$

  where $z_{\log(T)_{(\alpha)}}$ is the α quantile of $Z_{\log(T)}$.

- This leads to the **exact** prediction interval procedure

$$[\underset{\sim}{T}, \ \widetilde{T}] = \left[\exp(\widehat{\mu} + z_{\log(T)_{(\alpha/2)}} \times \widehat{\sigma}), \ \exp(\widehat{\mu} + z_{\log(T)_{(1-\alpha/2)}} \times \widehat{\sigma})\right].$$

  The quantiles $z_{\log(T)_{(\alpha/2)}}$ and $z_{\log(T)_{(1-\alpha/2)}}$ can be obtained by simulating B realizations of $Z_{\log(T)}$.

Chapter 15

Segment 3

Plug-In Statistical Prediction Intervals

Calibrating Plug-In Prediction Bounds and Intervals

Plug-In Statistical Prediction Intervals

- When $\boldsymbol{\theta}$ is **unknown**, a plug-in approximate $100(1-\alpha)\%$ prediction interval is obtained by simply substituting the ML estimates for the parameters:

$$PI(1-\alpha) = [\underset{\sim}{T},\ \widetilde{T}] = [\widehat{t}_{\alpha/2},\ \widehat{t}_{1-\alpha/2}]$$

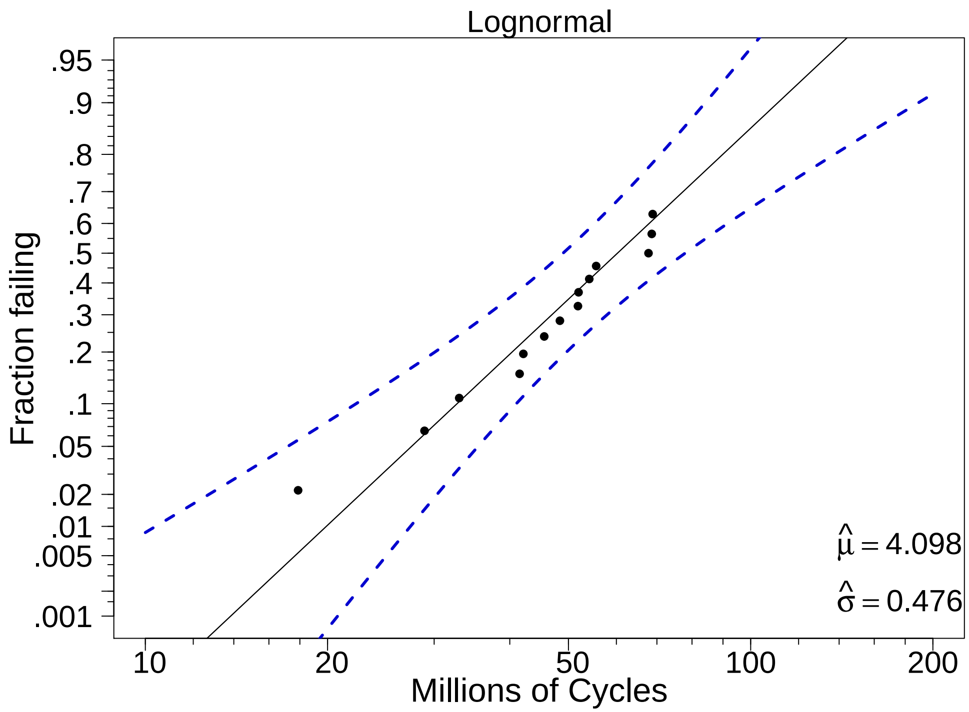
  where $\widehat{t}_p = t_p(\widehat{\boldsymbol{\theta}})$ is the ML estimate of the p quantile of T.

- Usually plug-in intervals are too narrow (coverage probability is too small).

- Coverage probability may be **far** from nominal $(1-\alpha)$, especially with small samples (small number of failures).

Example 2: Prediction Interval for a Single Independent Future Ball Bearing Lifetime T Based on Failure-Censored (Type 2) Censored Data

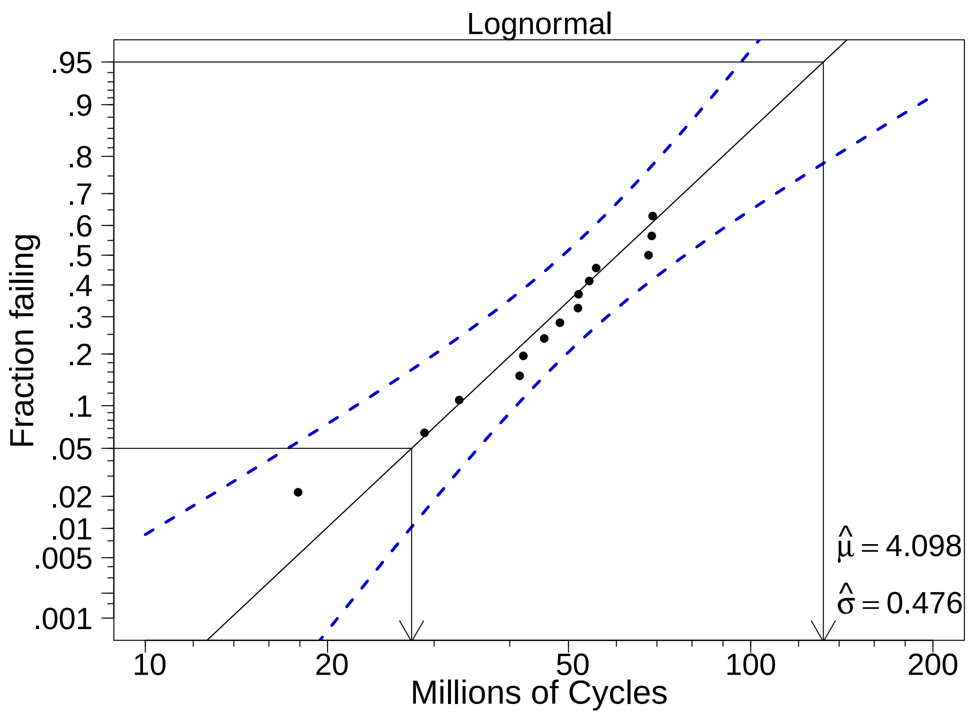
- A life test was run until 15 of 23 ball bearings failed. ML estimates of the lognormal parameters are: $\widehat{\mu} = 4.098$, $\widehat{\sigma} = 0.4761$.



- Need to predict the lifetime of a single future ball bearing.

Example 2: Prediction Interval for a Single Independent Future Ball Bearing Lifetime T Based on Failure-Censored (Type 2) Censored Data

- A life test was run until 15 of 23 ball bearings failed. ML estimates of the lognormal parameters are: $\widehat{\mu} = 4.098$, $\widehat{\sigma} = 0.4761$.



- Need to predict the lifetime of a single future ball bearing.

Example 2: Finding the Plug-In Prediction Interval

- The **plug-in** one-sided **lower** approximate 95% lognormal prediction bound (assuming no sampling error) is:
$$\underset{\sim}{T} = \widehat{t}_{0.05} = \exp[\widehat{\mu} + \Phi^{-1}_{\text{norm}}(0.05)\widehat{\sigma}]$$
$$= \exp[4.0985 + (-1.645)(0.4761)] = 27.53.$$

- The **plug-in** one-sided **upper** approximate 95% lognormal prediction bound (assuming no sampling error) is:
$$\widetilde{T} = \widehat{t}_{0.95} = \exp[\widehat{\mu} + \Phi^{-1}_{\text{norm}}(0.95)\widehat{\sigma}]$$
$$= \exp[4.0985 + 1.645(0.4761)] = 131.84.$$

- Thus a two-sided plug-in approximate 90% prediction interval is $[\underset{\sim}{T},\ \widetilde{T}] = [27.53,\ 131.84]$.

- It is possible to calibrate the plug-in interval to correct for uncertainty in the parameter estimates.

Calibrating Plug-In One-Sided Prediction Bounds

- **Basic idea**: Because the coverage probability of the plug-in method is too small, we should ask for a higher level of confidence to get the desired level.

- To calibrate the lower prediction bound, find α_c such that

$$\begin{aligned}\text{CP}[PI(1-\alpha_c);\widehat{\boldsymbol{\theta}}] &= \Pr\left(\underset{\sim}{T} \le T \le \infty; \widehat{\boldsymbol{\theta}}\right) \\ &= \Pr\left(\widehat{t}_{\alpha_c} \le T \le \infty; \widehat{\boldsymbol{\theta}}\right) = 1-\alpha.\end{aligned}$$

  where $\underset{\sim}{T} = \widehat{t}_{\alpha_c}$ is the ML estimator of the t_{α_c} quantile of T.

- Can do this by using simulation results.

- When for arbitrary α, $\text{CP}[PI(1-\alpha);\boldsymbol{\theta}]$ does not depend on $\boldsymbol{\theta}$, $\text{CP}[PI(1-\alpha_c);\boldsymbol{\theta}] = 1-\alpha$ and the **calibrated** $PI(1-\alpha_c)$ procedure is **exact**.

- For a two-sided prediction interval, calibrate the lower and upper prediction bounds separately and combine.

Simulation of the Sampling/Prediction Process

- Generate bootstrap samples DATA$_j^*$, $j = 1, \dots, B$ for a large number B (e.g., $B =$ 4,000 or $B =$ 10,000).
  - ▶ Simulate from the fitted model (fully parametric bootstrap). Requires specification of the model for censoring and/or truncation. Needed for **exact** procedures.
  - ▶ Resampling (integer-random-weight) bootstrap. May fail if censoring is heavy or if there are other conditions that would limit estimability with bootstrap samples.
  - ▶ Fractional-random-weight bootstrap. Avoids estimability problems that may arise with resampling.
- For each simulated bootstrap sample, compute ML **estimates** $\widehat{\boldsymbol{\theta}}_j^*$ from simulated DATA$_j^*$ [e.g., $\widehat{\boldsymbol{\theta}}_j^* = (\widehat{\mu}_j^*, \widehat{\sigma}_j^*)$ for a (log-)location-scale distribution], $j = 1, \dots, B$.
- Use the bootstrap estimates to compute calibration curves or a predictive distribution.

Finding and Using a Calibration Curve

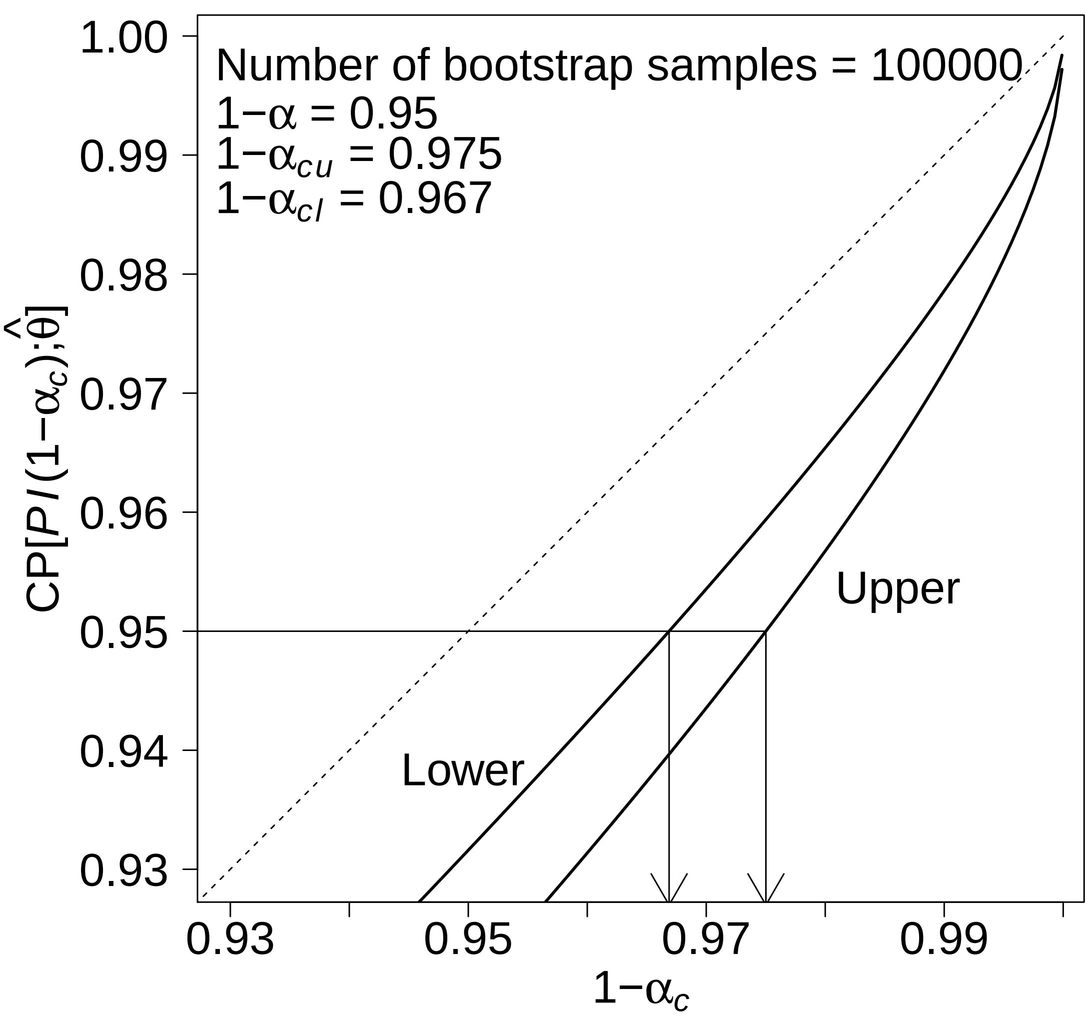
- For each value of $1-\alpha$ in the given range, compute the sample **mean of the conditional coverage probabilities** over all of the B values of $\widehat{\boldsymbol{\theta}}_j^*$, giving the calibration curve.

- Note that the $PI(1-\alpha)$ prediction interval endpoints $\underset{\sim}{T} = \underset{\sim}{T}(\widehat{\boldsymbol{\theta}})$ and $\widetilde{T} = \widetilde{T}(\widehat{\boldsymbol{\theta}})$ depend on the nominal $(1-\alpha)$ and also the ML estimates $\widehat{\boldsymbol{\theta}}$ through the sample data. Then the unconditional coverage probability (which may depend on the unknown true $\boldsymbol{\theta}$) can be computed from the bootstrap sample estimates as

$$\begin{aligned}
\mathsf{CP}[PI(1-\alpha);\boldsymbol{\theta}] &= \mathsf{E}_{\widehat{\boldsymbol{\theta}}}\Big\{\mathsf{CP}[PI(1-\alpha) \mid \widehat{\boldsymbol{\theta}};\boldsymbol{\theta}]\Big\} \\
&= \frac{1}{B}\sum_{j=1}^{B}\Big\{F[\widetilde{T}(\widehat{\boldsymbol{\theta}}_j^*);\boldsymbol{\theta}] - F[\underset{\sim}{T}(\widehat{\boldsymbol{\theta}}_j^*);\boldsymbol{\theta}]\Big\} \\
&= \frac{1}{B}\sum_{j=1}^{B}\Big\{F[\widetilde{T}(\widehat{\boldsymbol{\theta}}_j^*);\boldsymbol{\theta}]\Big\} - \frac{1}{B}\sum_{j=1}^{B}\Big\{F[\underset{\sim}{T}(\widehat{\boldsymbol{\theta}}_j^*);\boldsymbol{\theta}]\Big\}.
\end{aligned}$$

- To obtain a PI with a coverage probability of $100(1-\alpha)\%$, find α_c such that $\mathsf{CP}[PI(1-\alpha_c);\widehat{\boldsymbol{\theta}}] = (1-\alpha)$.

Prediction Interval Calibration Function for the Bearing Life Test Data Censored After 80 Million Cycles, Lognormal Model



Example 3: Finding the Calibrated Prediction Interval

- The **calibrated** one-sided **lower** exact 95% lognormal prediction bound is:

$$\underset{\sim}{T} = \widehat{t}_{(1-0.967)} = \exp[\widehat{\mu} + \Phi^{-1}_{\text{norm}}(1 - 0.967)\widehat{\sigma}]$$
$$= \exp[4.0985 + (-1.8384)(0.4761)] = 25.11.$$

.

- The **calibrated** one-sided **upper** exact 95% lognormal prediction bound is:

$$\widetilde{T} = \widehat{t}_{0.975} = \exp[\widehat{\mu} + \Phi^{-1}_{\text{norm}}(0.975)\widehat{\sigma}]$$
$$= \exp[4.0985 + 1.960(0.4761)] = 153.18$$

.

- Thus a two-sided exact 90% prediction interval is $[\underset{\sim}{T},\ \widetilde{T}] = [25.11,\ 153.18]$
- Extrapolation into the upper tail, however, casts some doubt on the veracity of the upper endpoint of this interval.

Chapter 15

Segment 4

Computing and Using a Predictive Distribution to Find Prediction Intervals

Finding a Predictive Distribution

Alternative Methods of Finding a Predictive Distribution

Technical Results Related to Prediction Methods

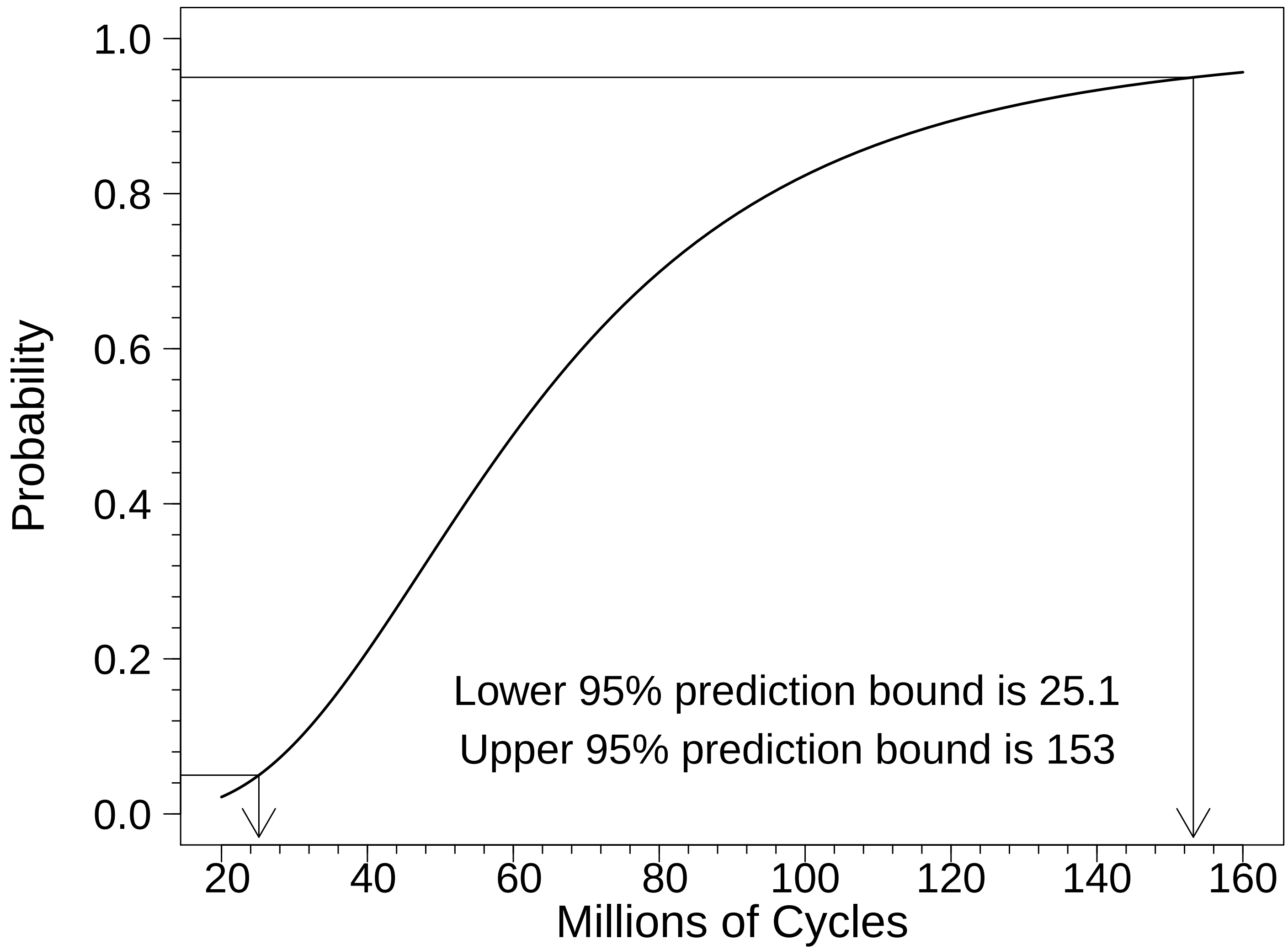
Computing and Using a Predictive Distribution

- Let $G(t;\boldsymbol{\theta})$ denote the cdf of the random variable T to be predicted.
- Prediction intervals can be obtained from a predictive distribution computed from bootstrap simulation results using

$$G_p(t) = \frac{1}{B}\sum_{j=1}^{B} G\left\{G^{-1}\left[G\left(t;\widehat{\boldsymbol{\theta}}\right);\widehat{\boldsymbol{\theta}}_j^{*}\right];\widehat{\boldsymbol{\theta}}\right\}. \qquad (1)$$

- A $100(1-\alpha)\%$ prediction interval is obtained from the $\alpha/2$ and $1-\alpha/2$ quantiles of the predictive distribution $G_p(t)$.
- A one-sided lower (upper) $100(1-\alpha)\%$ prediction bound is obtained from the α $(1-\alpha)$ quantile of the predictive distribution.
- For continuous distributions, using (1) to obtain prediction intervals is equivalent to the calibration method and is exact when $G(T;\widehat{\boldsymbol{\theta}})$ is a pivotal quantity (i.e., the distribution of the random variable $G(T;\widehat{\boldsymbol{\theta}})$ does not depend on $\boldsymbol{\theta}$).

Predictive Distribution for a Ball Bearing Lifetime Giving an Exact Prediction Interval



Alternative Method of Computing Prediction Intervals Using Calibration and an Extra Layer of Simulation

A $100(1-\alpha)\%$ prediction interval for T can be obtained by doing the following:

- Simulate T_j^* from the distribution $G(t;\widehat{\boldsymbol{\theta}})$, $j = 1,\ldots,B$, where $\widehat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$ from the original data.
- Compute $\nu_j^* = G(T_j^*;\boldsymbol{\theta}_j^*)$ for $j = 1,\ldots,B$.
- Compute $\nu_{\alpha/2}$ and $\nu_{1-\alpha/2}$, the $\alpha/2$ and $1-\alpha/2$ quantiles of the empirical distribution of the B ν_j^* values.
- Solve for $\underset{\sim}{T}$ and $\widetilde{T}$ in

$$G(\underset{\sim}{T};\widehat{\boldsymbol{\theta}}) = \nu_{\alpha/2}$$
$$G(\widetilde{T};\widehat{\boldsymbol{\theta}}) = \nu_{1-\alpha/2}$$

  to give the $100(1-\alpha)\%$ prediction interval for T.
- Useful if $G^{-1}(p;\boldsymbol{\theta})$ is difficult to compute. But the extra layer of simulation requires a larger value of B.

A Direct Method of Computing a Predictive Distribution

- The Bayesian predictive distribution for a random variable T is defined as

$$G_p(t) = \int_{\boldsymbol{\theta}} G(t;\boldsymbol{\theta}) f(\boldsymbol{\theta}|\text{DATA}) d\boldsymbol{\theta} \tag{2}$$

  where $f(\boldsymbol{\theta}|\text{DATA})$ is a joint posterior distribution for the parameter vector $\boldsymbol{\theta}$ and the integration is over the entire parameter space of $\boldsymbol{\theta}$.

- A non-Bayesian predictive distribution can be obtained by defining $f(\boldsymbol{\theta}|\text{DATA})$ to be a confidence distribution for $\boldsymbol{\theta}$.

- Usually $f(\boldsymbol{\theta}|\text{DATA})$ will be represented by draws from the distribution obtained from a simulation method (GPQ, bootstrap, or MCMC), to be mapped into draws from the predictive distribution $G_p(t)$.

Implementing the Direct Method of Computing a Predictive Distribution

There are two convenient ways to evaluate the integral in (2).

- View the integral in (2) as the expectation of $G(t;\boldsymbol{\theta})$ with respect to the posterior (or confidence) distribution $f(\boldsymbol{\theta}|\text{DATA})$ that can be evaluated by

$$G_p(t) = \frac{1}{B}\sum_{j=1}^{B} G(t;\widehat{\boldsymbol{\theta}}_j^*),$$

  where $\widehat{\boldsymbol{\theta}}_j^*, j = 1, \ldots, B$ are draws from $f(\boldsymbol{\theta}|\text{DATA})$.

- Generate T_j^* from the distribution $G(t;\widehat{\boldsymbol{\theta}}_j^*)$ for $j = 1, \ldots, B$, giving draws from the predictive distribution $G_p(t)$.

  Using this extra-layer-of-simulation method will require a larger value of B, but is useful when it is easy to simulate random variables from $G(t;\boldsymbol{\theta})$, but not easy to compute $G(t;\boldsymbol{\theta})$.

The GPQ (Fiducial) Method to Compute a Predictive Distribution for (Log-)Location-Scale Distributions

Failure time T has (log-)location-scale distribution with cdf $\Pr(T \leq t) = G(t; \mu, \sigma)$ with location parameter μ and scale parameter σ.

- Compute the GPQs for μ and σ.

$$\widehat{\mu}_j^{**} = \widehat{\mu} + \left(\frac{\widehat{\mu} - \widehat{\mu}_j^*}{\widehat{\sigma}_j^*}\right)\widehat{\sigma}$$

$$\widehat{\sigma}_j^{**} = \left(\frac{\widehat{\sigma}}{\widehat{\sigma}_j^*}\right)\widehat{\sigma}, \quad \text{for } j = 1, \ldots, B.$$

- Then the predictive distribution $G_p(t)$ can be computed as

$$G_p(t) = \frac{1}{B}\sum_{j=1}^{B} G(t; \widehat{\mu}_j^{**}, \widehat{\sigma}_j^{**}).$$

- This method can also be used for non-(log-)location-scale distributions if a GPQ is available.

Computation of a Predictive Distribution Using an Extra Layer of Simulation

Failure time T has a (log-)location-scale distribution with cdf $\Pr(T \le t) = G(t; \mu, \sigma)$ with location parameter μ and scale parameter σ.

- Compute the GPQs for μ and σ.

$$\widehat{\mu}_j^{**} = \widehat{\mu} + \left(\frac{\widehat{\mu} - \widehat{\mu}_j^*}{\widehat{\sigma}_j^*}\right)\widehat{\sigma}, \quad \sigma_j^{**} = \left(\frac{\widehat{\sigma}}{\widehat{\sigma}_j^*}\right)\widehat{\sigma},$$

  for $j = 1, \ldots, B$.

- Simulate T_j^* from the distribution $G(t; \widehat{\mu}_j^{**}, \sigma_j^{**})$, $j = 1, \ldots, B$. The empirical distribution of the T_j^* values provides a predictive distribution $G_p(t)$ that, if B is large enough, will agree with the GPQ method predictive distribution.

- This method is useful when it is difficult to compute $G(t; \mu, \sigma)$, but easy to simulate values of T for given values of μ and σ, but a larger value of B will be needed.

Prediction Intervals for a Future Ball Bearing Lifetime

- The exact 90% prediction interval can be obtained by using calibration or the 0.05 and 0.95 quantiles of the predictive distribution.

- Comparison of approximate 90% prediction intervals for a future bearing lifetime

Method	Interval Endpoints Lower	Upper
Plug-In	[27.5,	131.8]
Calibration using (1)	[25.1,	153.2]
Direct method using (2)	[25.1,	153.2]

The confidence distribution $f(\boldsymbol{\theta}|\text{DATA})$ used in (2) corresponds to the GPQ (fiducial) method of constructing a joint confidence region for μ and σ.

An Alternative Method for Computing Bootstrap Samples

- Computing bootstrap estimates can be computationally intensive, especially for a complicated model or when data sets are large.

- An alternative is to draw samples from the large-sample approximate distribution of the ML estimators: a multivariate normal distribution.

- This method will perform well when there is a large amount of information in the data about the parameters (large sample or a large number of failures when there is censoring).

Some Technical Results

- If a pivotal prediction interval method exists
  - The calibration method
  - The predictive distribution method(s), and
  - The pivotal method

  all give the same **exact** prediction interval so that:

  $$\text{CP}[PI(1-\alpha);\boldsymbol{\theta}] = (1-\alpha).$$

- The coverage probability of the plug-in method is

  $$\text{CP}[PI(1-\alpha);\boldsymbol{\theta}] = (1-\alpha) + O_p\Big(n^{-1}\Big).$$

- When there is no pivotal quantity, the coverage probability of the calibrated method is

  $$\text{CP}[PI(1-\alpha);\boldsymbol{\theta}] = (1-\alpha) + O_p\Big(n^{-2}\Big).$$

Chapter 15

Segment 5

Within-Sample Prediction

Distribution of the Number of Failures

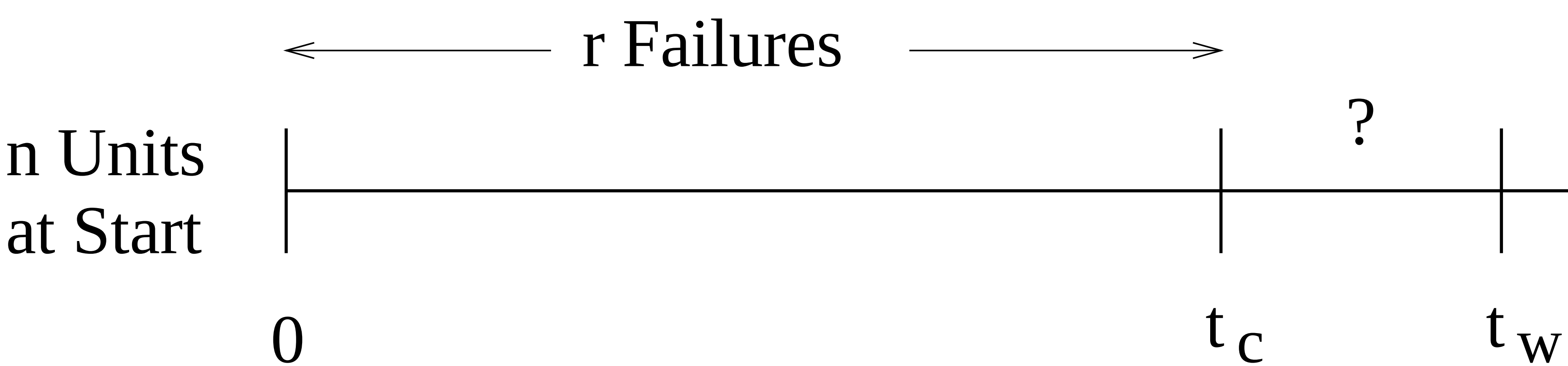
Plug-In Prediction Bound

Computing the Predictive Distribution

Within-Sample Prediction

Predict **future number of failures**, conditional on **early** data from the field.

- Suppose n units are in service until t_c and r failures were observed.
- The failure-time distribution is $\Pr(T \leq t) = F(t; \boldsymbol{\theta})$.
- The DATA are the first r failure times from a sample of size n: $t_{(1)} < \ldots < t_{(r)} \leq t_c$.
- There are $(n - r)$ units at risk to fail in the future.
- Want a prediction interval for K, the **number** of additional failures in interval $[t_c, t_w)$, conditional on the data up to t_c.



Distribution of K and Plug-In Prediction Bounds

- Conditional on DATA, the number of failures K in $(t_c, t_w]$ is distributed as

$$K \sim \text{BINOM}(n - r, \rho)$$

where, from the distribution of remaining life,

$$\rho = \frac{\Pr(t_c < T \le t_w)}{\Pr(T > t_c)} = \frac{F(t_w; \boldsymbol{\theta}) - F(t_c; \boldsymbol{\theta})}{1 - F(t_c; \boldsymbol{\theta})}. \qquad (3)$$

- $G(k) = \Pr(K \le k) = \texttt{pbinom}(k, n - r, \rho)$.

- Obtain $\widehat{\rho}$ by evaluating (3) at $\widehat{\boldsymbol{\theta}}$.

- The **plug-in** $100(1-\alpha)\%$ **lower** and **upper** prediction bounds for K are the α and $1-\alpha$ quantiles of the distribution of K:

$$\underset{\sim}{K} = \max(\texttt{qbinom}(\alpha, n - r, \widehat{\rho}) - 1, 0)$$
$$\widetilde{K} = \texttt{qbinom}(1 - \alpha, n - r, \widehat{\rho}),$$

which depend on the data through $\widehat{\boldsymbol{\theta}}$ and $\widehat{\rho}$.

Example 4: Prediction of the Number of Future Failures for Product A

- $n = 10{,}000$ units put into service; 80 failures in 48 months. The number units at risk is

$$n - r = 10000 - 80 = 9920 \text{ units.}$$

- Weibull time to failure distribution assumed with ML estimates: $\widehat{\eta} = 1152$, $\widehat{\beta} = 1.518$. The probability of failing between month 48 and month 60 is

$$\widehat{\rho} = \frac{\widehat{F}(60) - \widehat{F}(48)}{1 - \widehat{F}(48)} = 0.003233.$$

- Point prediction for the number failing between 48 and 60 months is

$$\widehat{K} = (n - r) \times \widehat{\rho} = 9920 \times 0.003233 = 32.07.$$

- The plug-in lower and upper 95% prediction bounds for the number failing between 48 and 60 months are

$$\underset{\sim}{K} = \texttt{qbinom}(0.05, 9920, 0.003233) - 1 = 22$$
$$\widetilde{K} = \texttt{qbinom}(0.95, 9920, 0.003233) = 42.$$

The GPQ (Fiducial) Method to Compute a Predictive Distribution for the Number Failing Between t_c and t_w

- The distribution of the number of failures K has a binomial distribution with cdf $\Pr(K \le k) = G(k; n-r, \rho)$.
- Compute the GPQs for μ and σ.

$$\widehat{\mu}_j^{**} = \widehat{\mu} + \left(\frac{\widehat{\mu} - \widehat{\mu}_j^{*}}{\widehat{\sigma}_j^{*}}\right)\widehat{\sigma}$$

$$\sigma_j^{**} = \left(\frac{\widehat{\sigma}}{\widehat{\sigma}_j^{*}}\right)\widehat{\sigma}, \quad j = 1, \ldots, B.$$

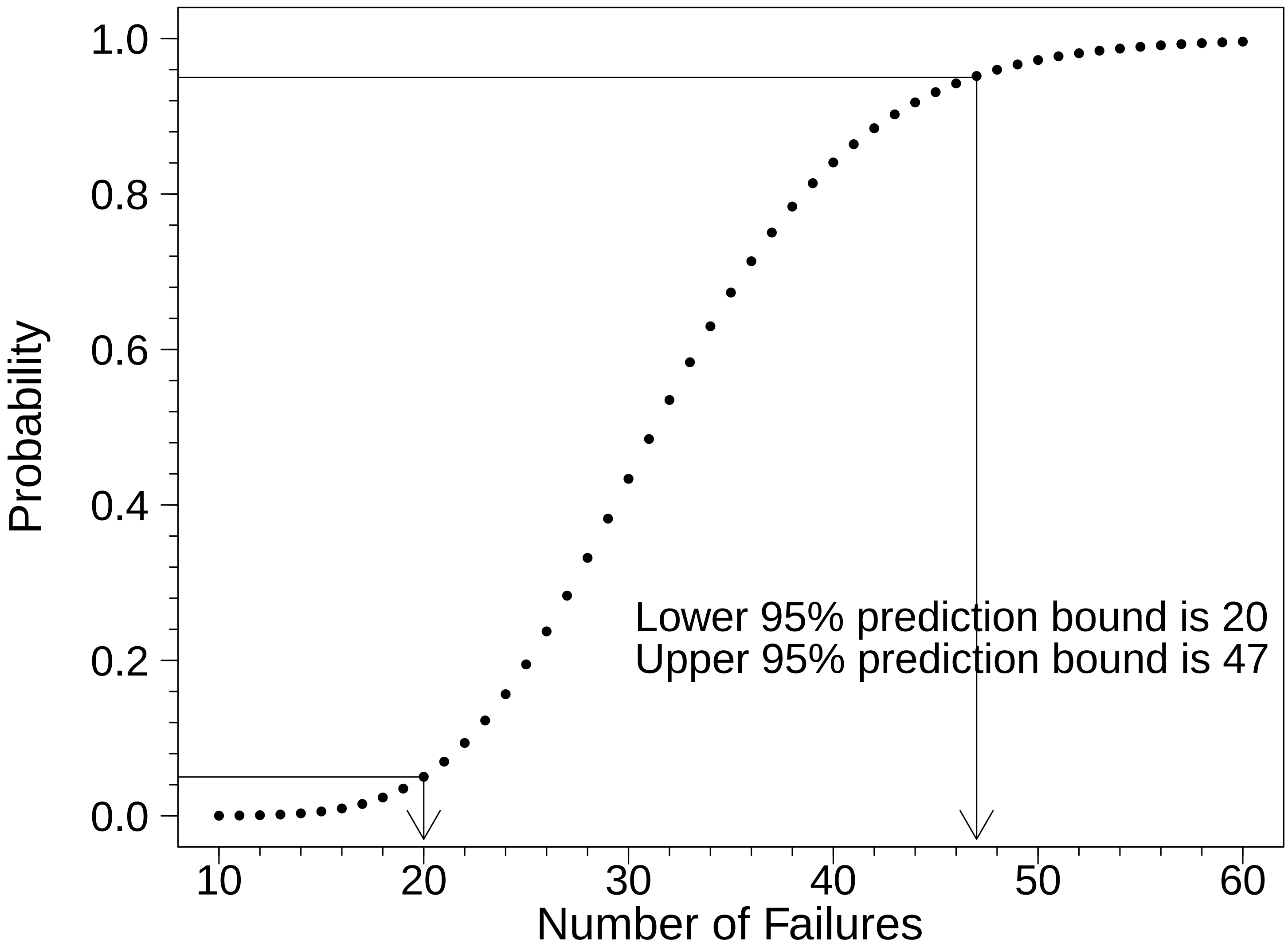
- Then the predictive distribution $G_p(t)$ can be computed as

$$G_p(k) = \frac{1}{B}\sum_{j=1}^{B} G(k; n-r, \widehat{\rho}_j^{**})$$

where

$$\widehat{\rho}_j^{**} = \frac{F(t_w; \widehat{\mu}_j^{**}, \sigma_j^{**}) - F(t_c; \widehat{\mu}_j^{**}, \sigma_j^{**})}{1 - F(t_c; \widehat{\mu}_j^{**}, \sigma_j^{**})}$$

Example 4. The Predictive Distribution and Upper and Lower Prediction Bounds for the Number of Future Field Failures for Product A



Example 4 Comparison

Comparison of approximate 90% prediction intervals for the number of Product A failures in the next 12 months:

Method	Interval Endpoints Lower	Upper
Plug-In	[22,	42]
Calibration using (1)	[18,	45]
Direct method using (2)	[19,	47]

The confidence distribution $f(\boldsymbol{\theta}|\text{DATA})$ used in (2) corresponds to the GPQ (fiducial) method of constructing a joint confidence region for μ and σ.

Chapter 15

Segment 6

Staggered Entry Within-Sample Prediction

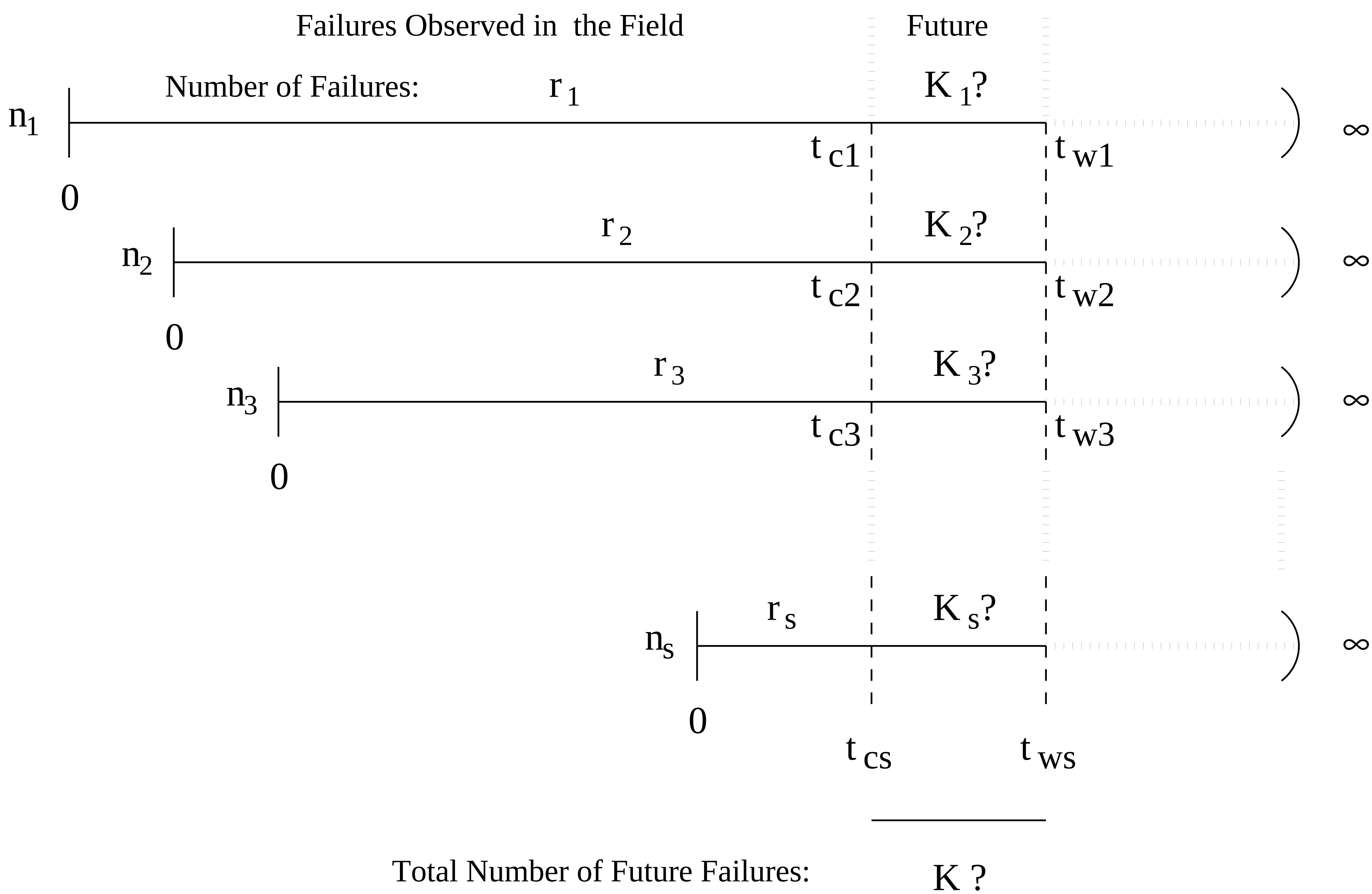
Distribution of the Number of Failures

Plug-In Prediction Bound

The Poisson-Binomial Distribution

Computing the Predictive Distribution

Staggered Entry Prediction Problem



Example 5: Bearing-Cage Field-Failure Data (from Abernethy et al. 1983)

- A total of 1703 units were introduced into service over a period of eight years (about 1600 in the past three years).

- Time measured in hours of service.

- Six out of 1703 units had failed by the data-freeze date.

- Unexpected failures early in life suggested the need for a design change.

- For the current fleet a prediction is needed on how many failures will occur in the next year (point prediction and upper prediction bound requested), assuming 300 hours of service for each aircraft.

Bearing Cage Data and Future-Failure Risk Analysis

Group i	Hours in Service	n_i	Failed r_i	At Risk $(n_i - r_i)$	$\widehat{\rho}_i$	$(n_i - r_i) \times \widehat{\rho}_i$
1	50	288	0	288	0.000763	0.2196
2	150	148	0	148	0.001158	0.1714
3	250	125	1	124	0.001558	0.1932
4	350	112	1	111	0.001962	0.2178
5	450	107	1	106	0.002369	0.2511
6	550	99	0	99	0.002778	0.2750
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
17	1650	6	0	6	0.007368	0.0442
18	1750	0	0	0	0.007791	0.0000
19	1850	1	0	1	0.008214	0.0082
20	1950	0	0	0	0.008638	0.0000
21	2050	2	0	2	0.009062	0.0181
Total		$n = 1703$	$r = 6$	$n - r = 1697$		$\widehat{K} = 5.058$

Distribution of the Number of Future Failures with Staggered Entry

- Conditional on DATA$_i$, the number of additional failures K_i in group i during interval $(t_{cj}, t_{wi}]$ (where $t_{wi} = t_{cj} + \Delta t$) is distributed as $K_i \sim \text{BINOM}(n_i - r_i, \rho_i)$ with

$$\rho_i = \frac{\Pr(t_{cj} < T \leq t_{wi})}{\Pr(T > t_{cj})} = \frac{F(t_{wi}; \boldsymbol{\theta}) - F(t_{cj}; \boldsymbol{\theta})}{1 - F(t_{cj}; \boldsymbol{\theta})}, \quad i = 1, \ldots, s.$$

- Obtain $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_s)$ by evaluating $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_s)$ at $\widehat{\boldsymbol{\theta}}$.

- Want to predict the total number of additional failures $K = \sum_{i=1}^{s} K_i$ over Δt. Conditional on the DATA (and the fixed censoring times) $K \sim \text{POIBIN}(k; \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho})$ a sum of s independent but non-identically distributed binomial random variables with parameters $\boldsymbol{n} - \boldsymbol{r} = (n_1 - r_1, \ldots, n_s - r_s)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_s)$. K has a Poisson-binomial distribution.

- The **plug-in** $100(1-\alpha)\%$ one-sided prediction bounds are

$$\begin{aligned} \underset{\sim}{K} &= \max(\texttt{qpoibin}(\alpha, \boldsymbol{n} - \boldsymbol{r}, \widehat{\boldsymbol{\rho}}) - 1, 0) \\ \widetilde{K} &= \texttt{qpoibin}(1 - \alpha, \boldsymbol{n} - \boldsymbol{r}, \widehat{\boldsymbol{\rho}}). \end{aligned}$$

The Poisson-Binomial Distribution

- The sum of independent, but not identically distributed Bernoulli random variables has a Poisson-binomial distribution.
- The R package `poibin` can be used to compute Poisson-binomial probabilities and quantiles.
- With large n_i values and a large number of groups, computing Poisson-binomial probabilities and especially quantiles can be computationally intensive. Fortunately, good approximations are available.
- For large n_i values and small ρ_i values (typical in many applications), the Poisson approximation provides an excellent approximation, where the mean of the Poisson distribution is taken to be $\mu = \sum_{i=1}^{s} n_i \rho_i$.
- For large n_i values and ρ_i not too small, the normal distribution approximation can be used with the same mean but standard deviation $\sigma = \sqrt{\sum_{i=1}^{s} n_i \rho_i (1 - \rho_i)}$.

Example 5–Computations

- The **plug-in** 95% **lower** prediction bound on K is

$$\underset{\sim}{K} = \widehat{K}_{0.05} - 1 = \texttt{qpoibin}(0.05, \widehat{\boldsymbol{\rho}}, \boldsymbol{n} - \boldsymbol{r}) - 1 = 1.$$

- The **plug-in** 95% **upper** prediction bound on K is

$$\widetilde{K} = \widehat{K}_{0.95} = \texttt{qpoibin}(0.95, \widehat{\boldsymbol{\rho}}, \boldsymbol{n} - \boldsymbol{r}) = 9.$$

- A plug-in approximate 90% prediction interval is

$$[\underset{\sim}{K}, \ \ \widetilde{K}] = [1, \ \ 9].$$

- The plug-in interval can be improved by using a procedure that accounts for uncertainty in the parameter estimates.

The GPQ (Fiducial) Method to Compute a Predictive Distribution for the Number Failing Between t_c and t_w with Multiple Cohorts

- The distribution of the number of failures K has a Poisson-binomial distribution with cdf $\Pr(K \le k) = G(k; \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho})$.
- Compute the GPQs for μ and σ.

$$\widehat{\mu}_j^{**} = \widehat{\mu} + \left(\frac{\widehat{\mu} - \widehat{\mu}_j^{*}}{\widehat{\sigma}_j^{*}}\right)\widehat{\sigma}$$

$$\sigma_j^{**} = \left(\frac{\widehat{\sigma}}{\widehat{\sigma}_j^{*}}\right)\widehat{\sigma}, \quad j = 1, \ldots, B.$$

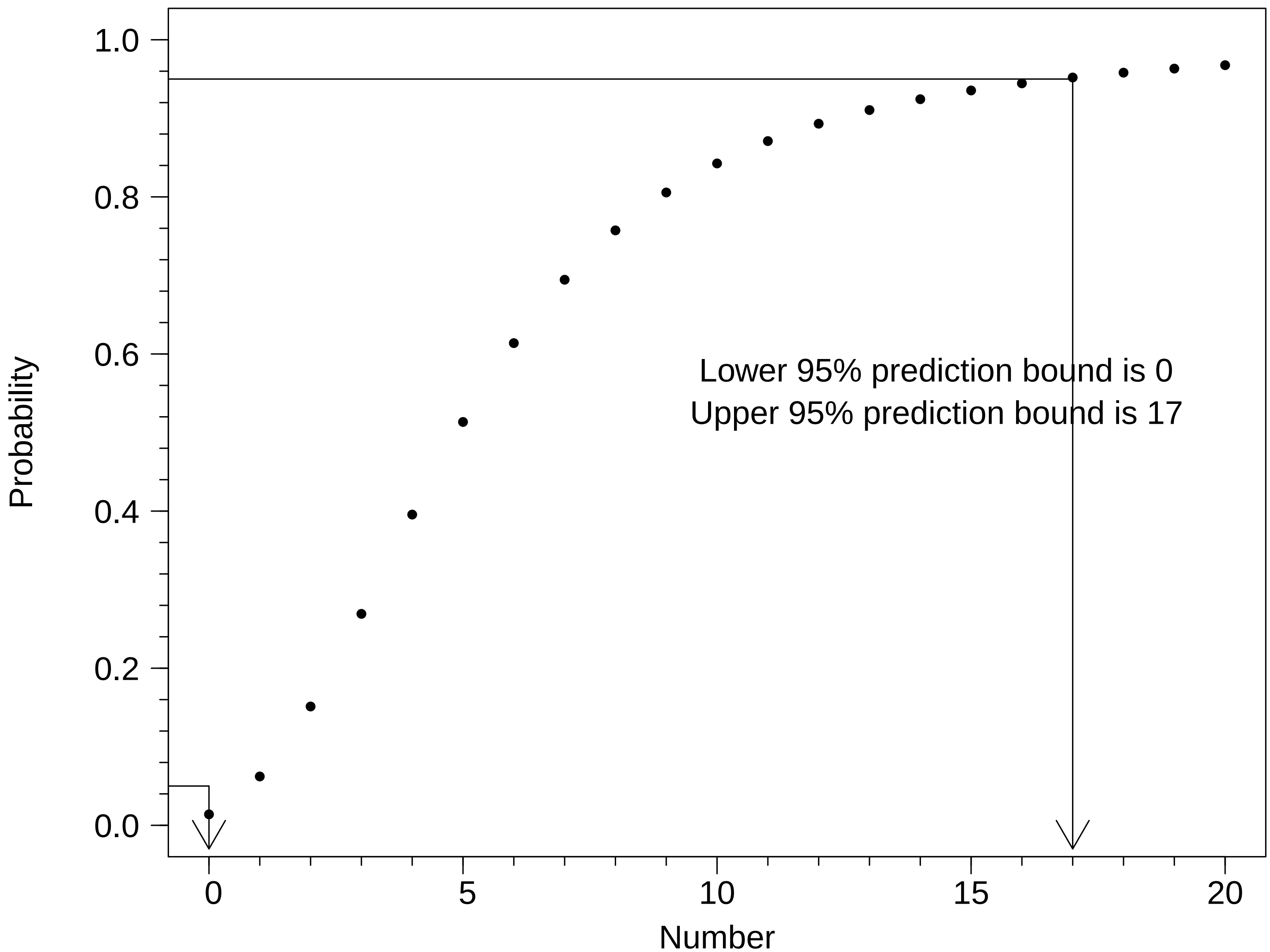
- Then a predictive distribution $G_p(k)$ can be computed as

$$G_p(k) = \frac{1}{B}\sum_{j=1}^{B} G(k; ; \boldsymbol{n} - \boldsymbol{r}, \widehat{\boldsymbol{\rho}}_j^{**}), \quad k = 0, \ldots n - r$$

$$\widehat{\boldsymbol{\rho}}_j^{**} = (\widehat{\rho}_{j,1}^{**}, \ldots, \widehat{\rho}_{j,s}^{**})$$

$$\widehat{\rho}_{j,i}^{**} = \frac{F(t_{w,i}; \widehat{\mu}_j^{**}, \sigma_j^{**}) - F(t_{c,i}; \widehat{\mu}_j^{**}, \sigma_j^{**})}{1 - F(t_{c,i}; \widehat{\mu}_j^{**}, \sigma_j^{**})}, \quad i = 1, \ldots, s.$$

Example 5: Fiducial/GPQ Predictive Distribution Giving the Upper and Lower Prediction Bounds on the Number of Future Field Failures with Staggered Entry



Examples 5 and 6–Comparisons

Comparison of approximate 90% prediction intervals for the number of failures in the next year (assuming 300 hours of operation per aircraft):

Method	Interval Endpoints Lower	Upper
Plug-In	[1,	9]
Calibration using (1)	[0,	12]
Direct method using (2) with GPQs	[0,	17]
Bayesian weakly informative	[0,	10]
Bayesian informative for β	[0,	10]

Chapter 15

Segment 7

Bayesian Prediction Procedures

Alternative Models and Methods

Bayesian Prediction Motivation

- Bayesian prediction methods are important and recommended when one or more of the following hold:
  - ▶ There is a small amount of information in the data so that the adequacy of large-sample theory is in question.
  - ▶ In complicated models involving random effects, where Bayesian estimation is easier to do.
  - ▶ There is informative prior information that should be used.
  - ▶ If there is no informative prior information available, then "weakly informative" of diffuse prior distributions can be used.
- Bayesian prediction methods are relatively easy to apply once draws from the joint posterior distribution of the model parameters are available.

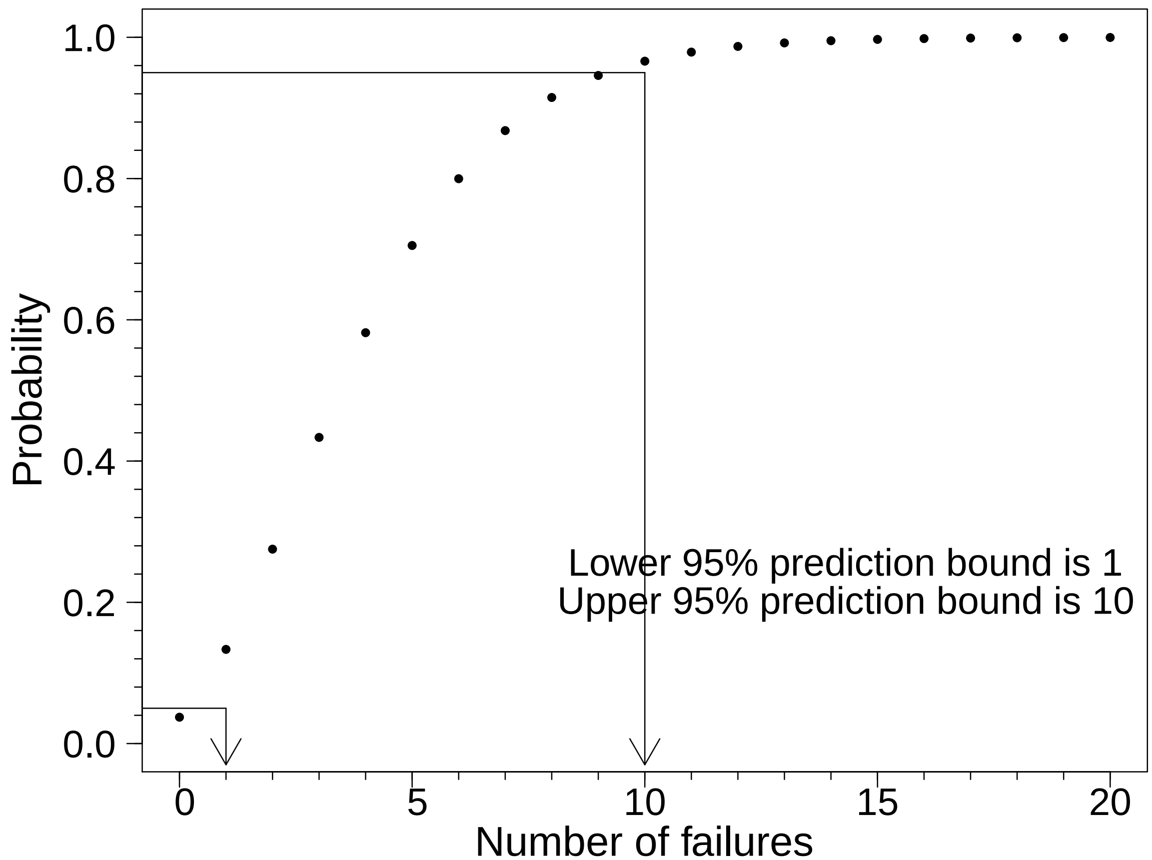
Bayesian Prediction Methods

- As with the non-Bayesian prediction methods, there are two alternative approaches. Both are similar to the GPQ methods described earlier, except that draws from the joint posterior distribution of the parameters are used instead of the draws from the joint GPQ (fiducial) distributions.

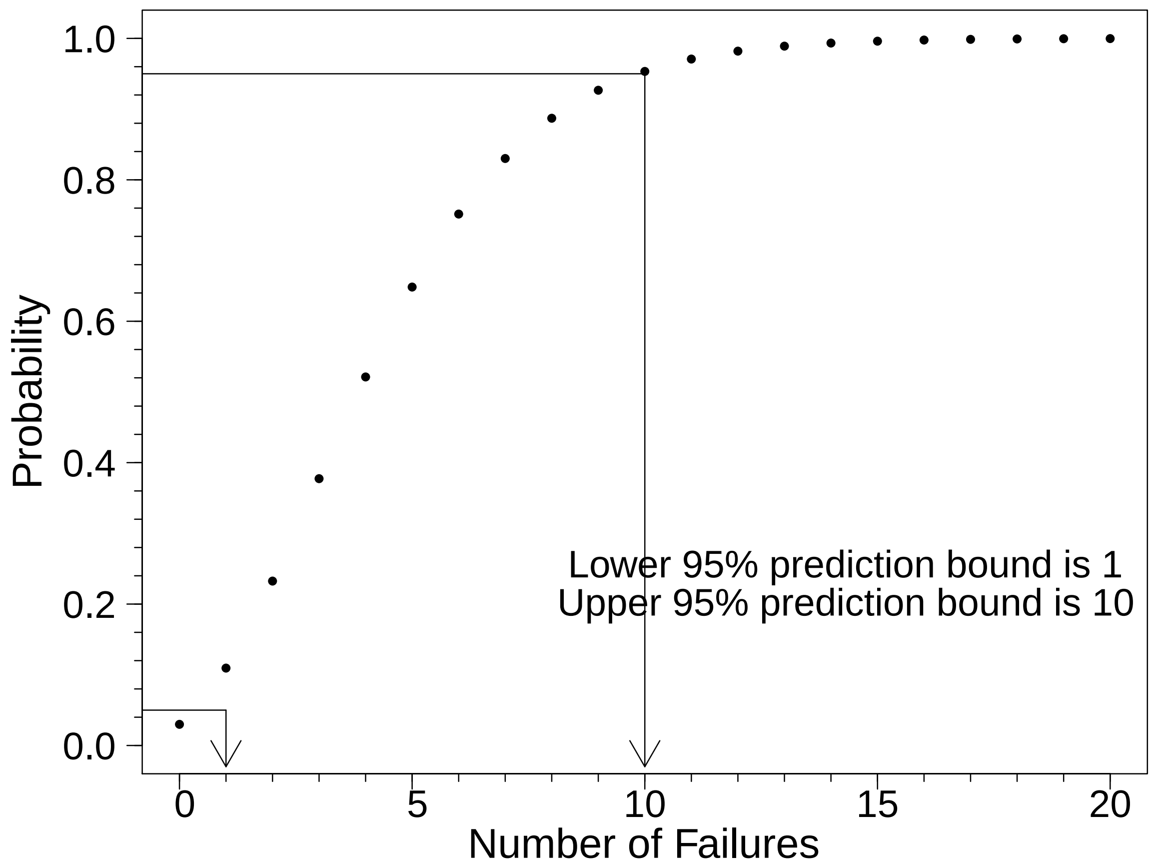
  - Direct computation of the predictive distribution. This method works well if one can easily compute the cdf of the predictand.
  - Extra layer of simulation. This method works well as long as one can simulate values of the predictand, given draws from the joint posterior.

  Similar to the non-Bayesian methods, to obtain the same precision as the direct method, (i.e., reduce Monte Carlo error), the number of draws from the joint posterior distribution has to be larger for the "extra layer of simulation" method.

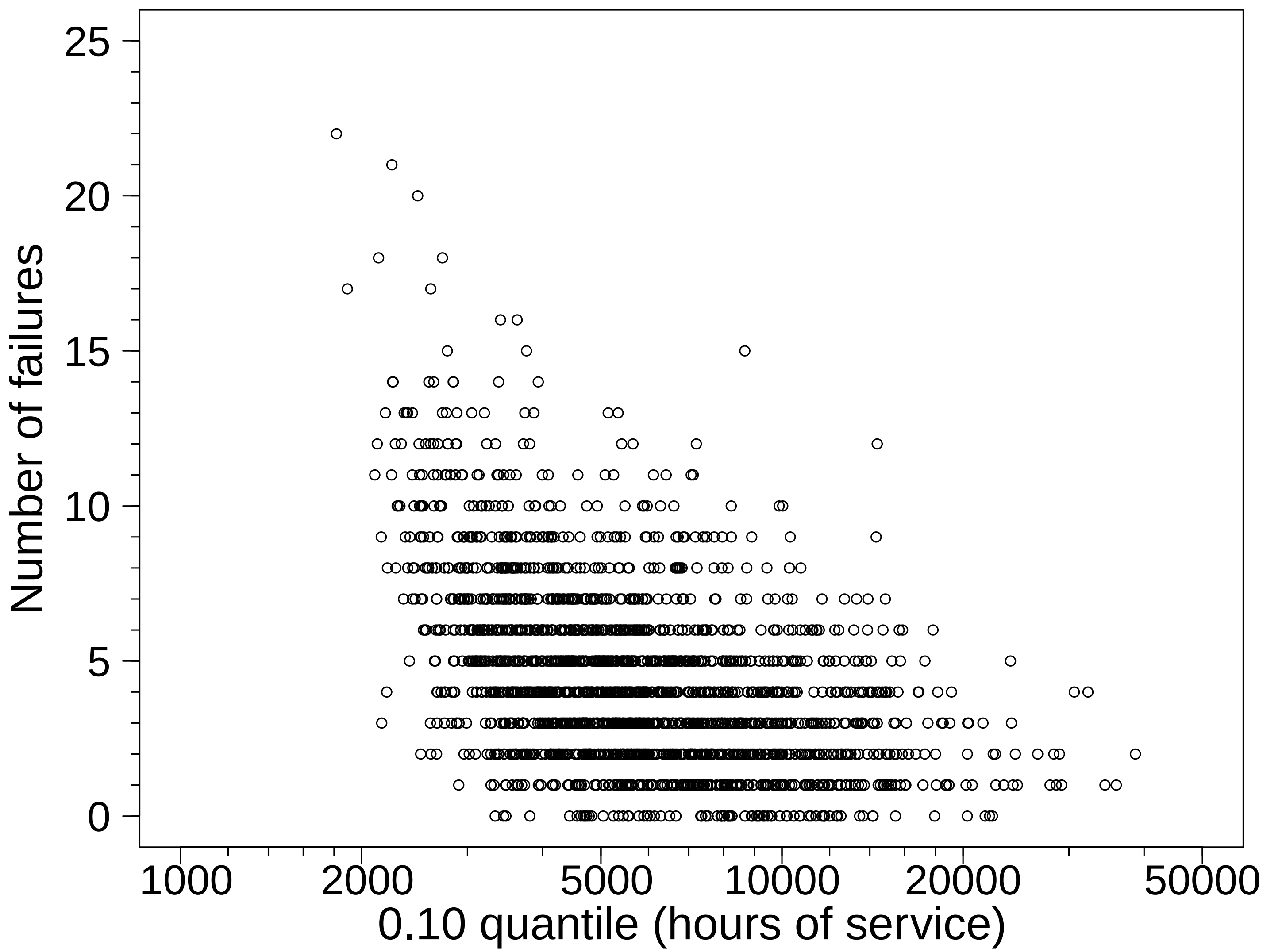
Example 6: Bayesian Weakly Informative Prior Predictive Distribution Giving the Upper and Lower Prediction Bounds on the Number of Future Field Failures with Staggered Entry



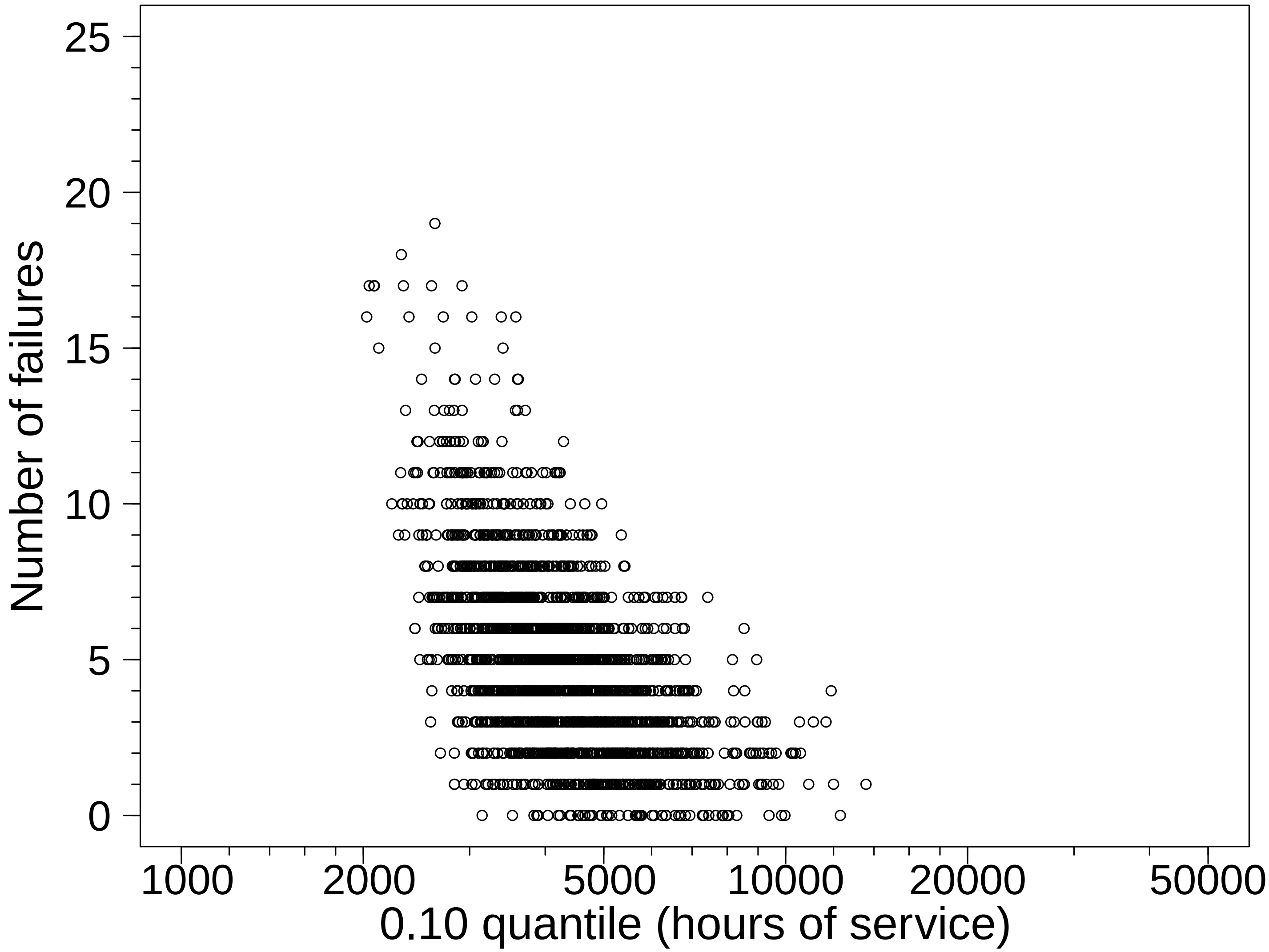
Example 6: Bayesian Informative Prior on β Predictive Distribution Giving the Upper and Lower Prediction Bounds on the Number of Future Bearing Cage Field Failures



Example 6: Bayesian Weakly Informative Prior Joint Posterior Distribution of $t_{0.10}$ and K



Example 6: Bayesian Informative Prior on β
Joint Posterior Distribution of $t_{0.10}$ and K



Chapter 15

Segment 8

Choosing a Distribution for Prediction

Alternative Models and Methods

Prediction and Extrapolation

- Extrapolation is usually required when predicting the number of failures based on an ongoing time-to-failure process.
- Example: Predict the number of returns in a three-year warranty period based on field data for the first year of operation.
- When extrapolation is required, predictions can be strongly dependent on the distribution choice.
- In some applications where there has been staggered entry over a long period of time and the failure-time distribution has not changed importantly over that time, there may be less extrapolation.

Choosing a Distribution for Prediction

- In most applications, especially with heavy censoring, there is little or no useful information about the failure-time distribution in the data

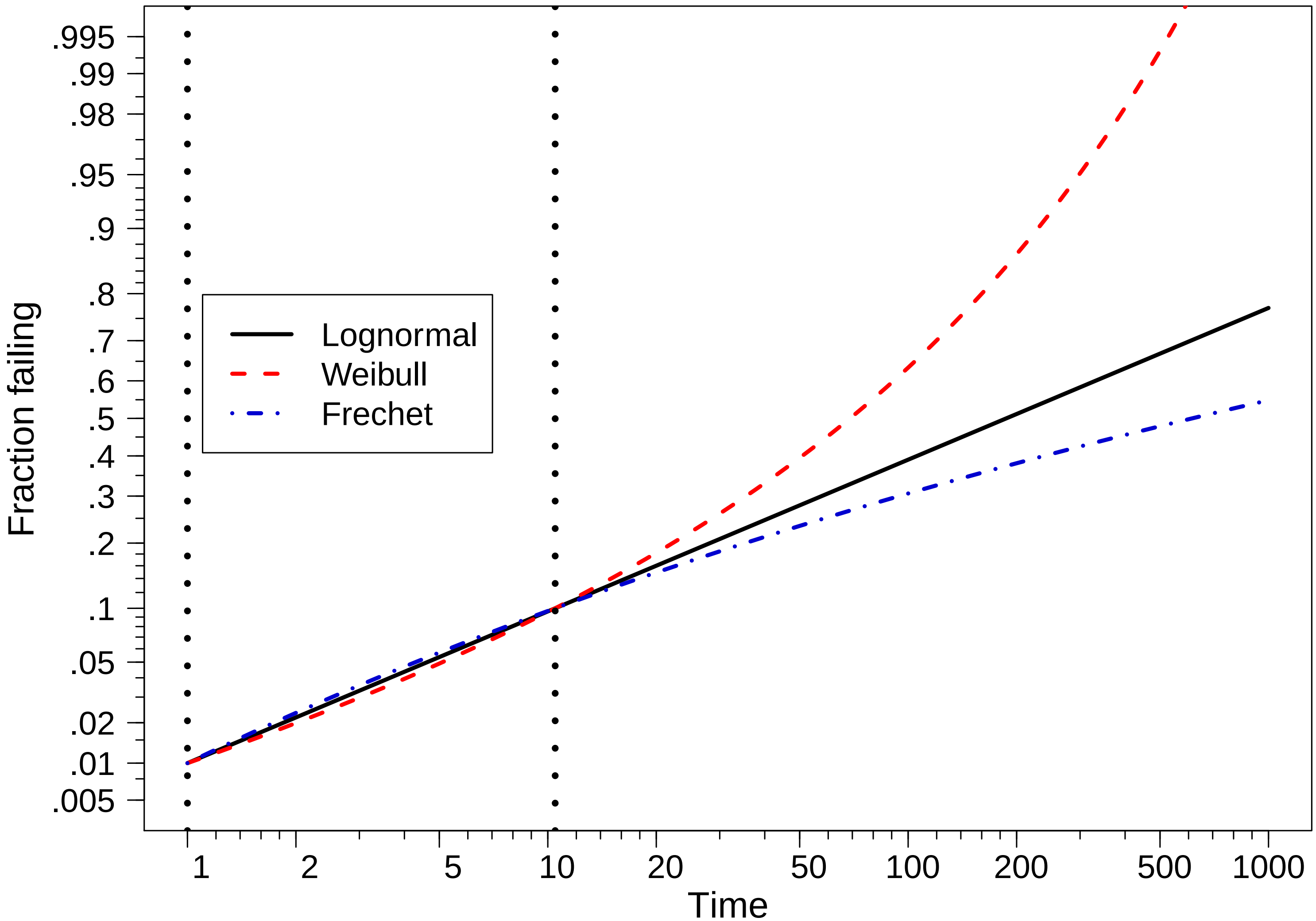
- It is best to choose a failure-time distribution based on knowledge of the failure mechanism and the related physics/chemistry of failure.

- When there is no information available to choose a distribution, use sensitivity analyses, comparing different distributions.

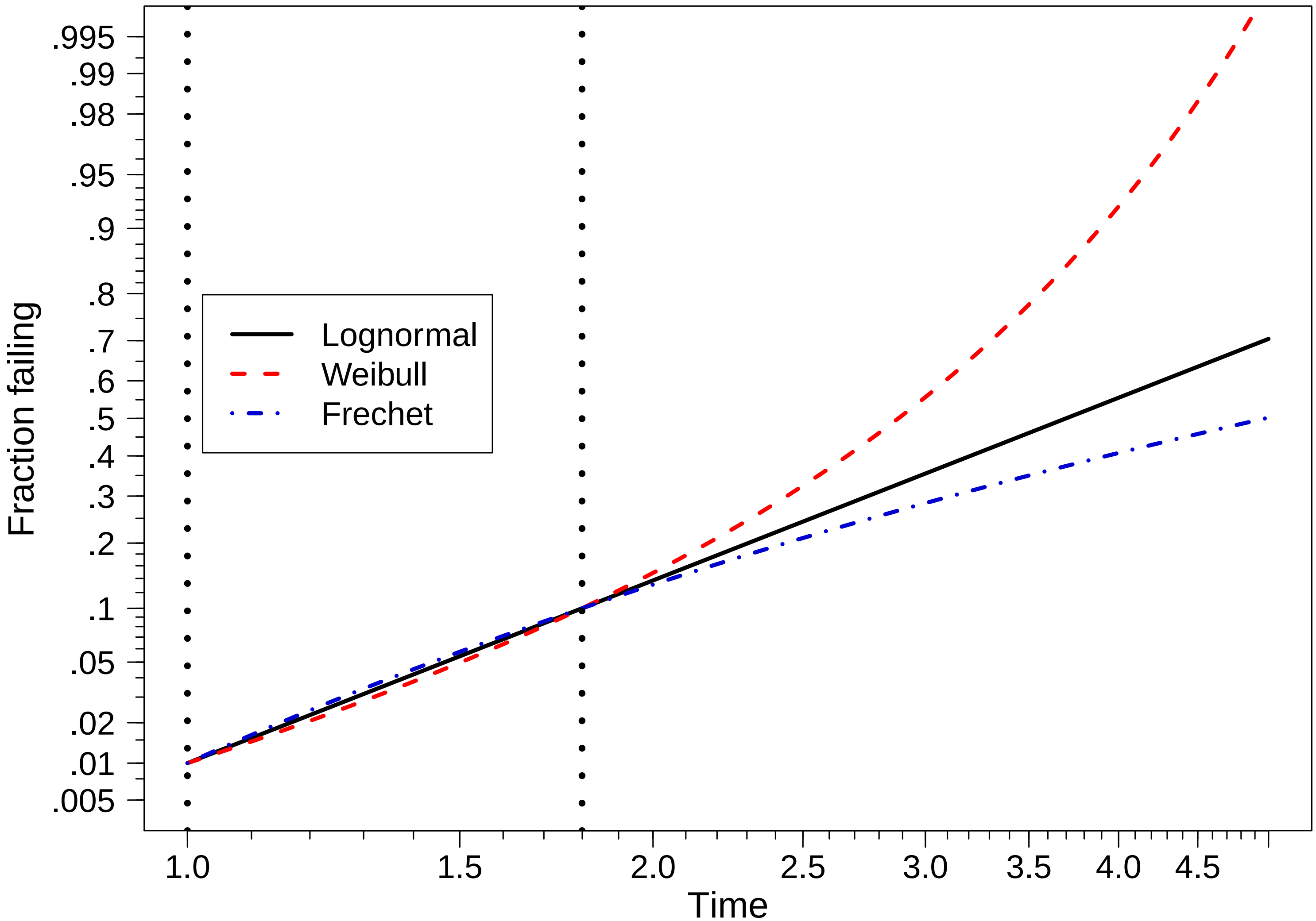
  - ▶ The Weibull distribution is always more pessimistic (conservative) than the lognormal.

  - ▶ The Fréchet distribution is always more optimistic than the lognormal.

Comparison of Weibull, lognormal, and Fréchet cdfs for a Weibull shape parameter $\beta = 1$



Comparison of Weibull, lognormal, and Fréchet cdfs for a Weibull shape parameter $\beta = 4$



Alternative Models and Methods Involving Prediction

This prediction methodology described here has been or could be extended to:

- Staggered entry with differences in warranty period.
- Limited failure population (defective sub-population) model.
- Making separate predictions for different failure modes.
- Time-constant covariates such as different use rates.
- Allowing for a retirement process for the at-risk units.
- Dynamic (time-varying) covariates like weather.
- Modeling of spatial and temporal variability in environmental factors like UV radiation, acid rain, temperature, and humidity.

References

Meeker, W. Q., L. A. Escobar, and F. G. Pascual (2021). *Statistical Methods for Reliability Data* (Second Edition). Wiley. [1]