Chapter 14

Introduction to the Use of Bayesian Methods for Reliability Data

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 26min

Introduction to the Use of Bayesian Methods for Reliability Data
Chapter 14 Objectives

- Describe the use of Bayesian statistical methods to combine **prior** information with data to make inferences.
- Explain the relationship between Bayesian methods and likelihood methods used in earlier chapters.
- Discuss sources of prior information.
- Describe useful computing methods for Bayesian methods.
- Illustrate Bayesian methods for estimating reliability.
- Illustrate Bayesian methods for prediction.
- Compare Bayesian and likelihood methods under different assumptions about prior information.
- Explain the dangers of using wishful thinking or expectations as prior information.

Introduction

- Bayes methods augment likelihood with **prior** information.

- A probability distribution is used to describe our **prior** beliefs about a parameter or set of parameters.

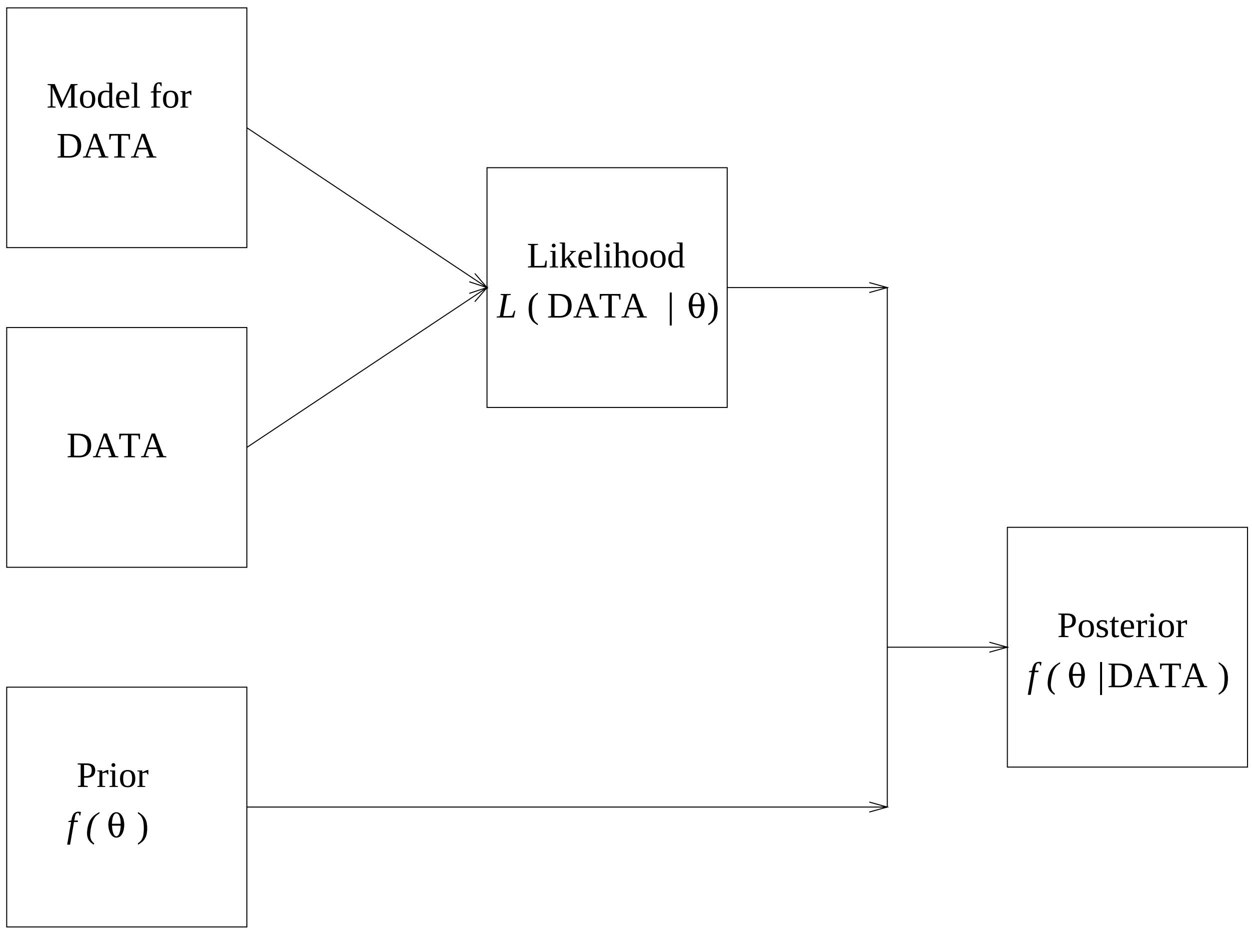
- Sources of prior information:

  Subjective Bayes: prior information subjective.

  Empirical Bayes: prior information from past data.

- Bayesian methods are closely related to likelihood methods.

Bayes Method for Inference



Updating Prior Information Using Bayes Theorem

Bayes Theorem provides a mechanism for combining *prior* information with sample data to make inferences on model parameters.

For a vector parameter $\boldsymbol{\theta}$ the procedure is as follows:

- Prior information on $\boldsymbol{\theta}$ is expressed in terms of a pdf $f(\boldsymbol{\theta})$.
- We observe some data which for the specified model has likelihood $L(\text{DATA}|\boldsymbol{\theta}) \equiv L(\boldsymbol{\theta}; \text{DATA})$.
- Using Bayes Theorem, the conditional distribution of $\boldsymbol{\theta}$ given the data (also known as the **posterior** of $\boldsymbol{\theta}$) is

$$f(\boldsymbol{\theta}|\text{DATA}) = \frac{L(\text{DATA}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int L(\text{DATA}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{R(\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int R(\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

  where $R(\boldsymbol{\theta}) = L(\boldsymbol{\theta})/L(\widehat{\boldsymbol{\theta}})$ is the relative likelihood and the multiple integral is computed over the region $f(\boldsymbol{\theta}) > 0$.

Some Comments on on Posterior Distributions

- The posterior $f(\boldsymbol{\theta}|\text{DATA})$ is function of the prior, the model, and the data.

- In general, it is impossible to compute the multiple integral $\int L(\text{DATA}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$ in closed form.

- New statistical and numerical methods that take advantage of modern computing power are facilitating the computation of the posterior.

Differences Between Bayesian and Frequentist Inference

- Nuisance parameters
  - Bayes methods use marginals.
  - Large-sample likelihood theory suggest maximization.
- There are not important differences in large samples.
- Interpretation
  - Bayes methods justified in terms of probabilities.
  - Frequentist methods justified on repeated sampling and asymptotic theory.

Sources of Prior Information

- Informative
  - Past data
  - Expert knowledge
- Non-informative (or approximately non-informative)
  - Uniform over range of parameter (or function of parameter)
  - Other vague or diffuse priors

Proper Prior Distributions

Any positive function defined on the parameter space that integrates to a finite value (usually 1).

- **Uniform prior:** $f(\theta) = 1/(b - a)$ for $a \leq \theta \leq b$.
  This prior does not express strong preference for specific values of θ in the interval.

- **Examples of non-uniform prior distributions:**
  - Normal with mean at a and and standard deviation b.
  - Beta between specified a and b with specified shape parameters (allows for a more general shape).
  - Isosceles triangle with base (range between) a and b.

For a positive parameter θ, may want to specify the prior in terms of $\log(\theta)$.

Improper Prior Distributions

Positive function $f(\theta)$ over parameter space for which

$$\int f(\theta) d\theta = \infty,$$

- **Uniform** in an interval of infinite length: $f(\theta) = c$ for all θ.

- For a positive parameter θ the corresponding choice is $f[\log(\theta)] = c$ and $f(\theta) = (c/\theta)$, $\theta > 0$.

To use an improper prior, one must have

$$\int f(\theta) L(\theta|\text{DATA}) d\theta < \infty$$

(a condition on the form of the likelihood and the DATA).

- These prior distributions can be made to be proper by specification of a finite interval for θ and choosing c such that the total probability is 1.

Effect of Using Vague (or Diffuse) Prior Distributions

- For a uniform prior $f(\theta)$ (possibly improper) across all possible values of θ

$$f(\boldsymbol{\theta}|\text{DATA}) = \frac{R(\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int R(\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{R(\boldsymbol{\theta})}{\int R(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

  which indicates that the posterior $f(\boldsymbol{\theta}|\text{DATA})$ is proportional to the likelihood.

- The posterior is approximately proportional to the likelihood for a proper (finite range) uniform if the range is large enough so that $R(\boldsymbol{\theta}) \approx 0$ where $f(\boldsymbol{\theta}) = 0$.

- Other diffuse priors also result in a posterior that is approximately proportional to the likelihood if $R(\boldsymbol{\theta})$ is large relative to $f(\boldsymbol{\theta})$.

Eliciting or Specifying a Prior Distribution

- The elicitation of a meaningful joint prior distribution for vector parameters may be difficult
  - ▶ The marginals may not completely determine the joint distribution.
  - ▶ Difficult to express/elicit dependences among parameters through a joint distribution.
  - ▶ The standard parameterization may not have practical meaning.

- General approach: choose an appropriate parameterization in which the priors for the parameters are approximately independent.

Expert Opinion and Eliciting Prior Information

- Identify parameters that, from past experience (or data), can be specified approximately independently (e.g., for high reliability applications a small quantile and the Weibull shape parameter).

- Determine for which parameters there is useful informative prior information.

- For parameters for which there in **no** useful informative prior information, determine the form and range of the vague prior (e.g., uniform over a wide interval).

- For parameters for which there is useful informative prior information, specify the form and range of the distribution (e.g., lognormal with 99.7% content between two specified points).

Example of Eliciting Prior Information: Bearing-Cage Time to Fracture Distribution

With appropriate questioning, engineers provided the following information:

- Time to fracture data can often be described by a Weibull distribution.

- From previous similar studies involving heavily censored data, (μ, σ) tend to be correlated (making it difficult to specify a joint prior for them).

- For small p (near the proportion failing in previous studies), (t_p, σ) are approximately independent (which allows for specification of approximately independent priors).

Example of Eliciting Prior Information: Bearing-Cage Fracture Field Data (Continued)

- Based on experience with previous products of the same material and knowledge of the failure mechanism, there is strong prior information about the Weibull shape parameter.

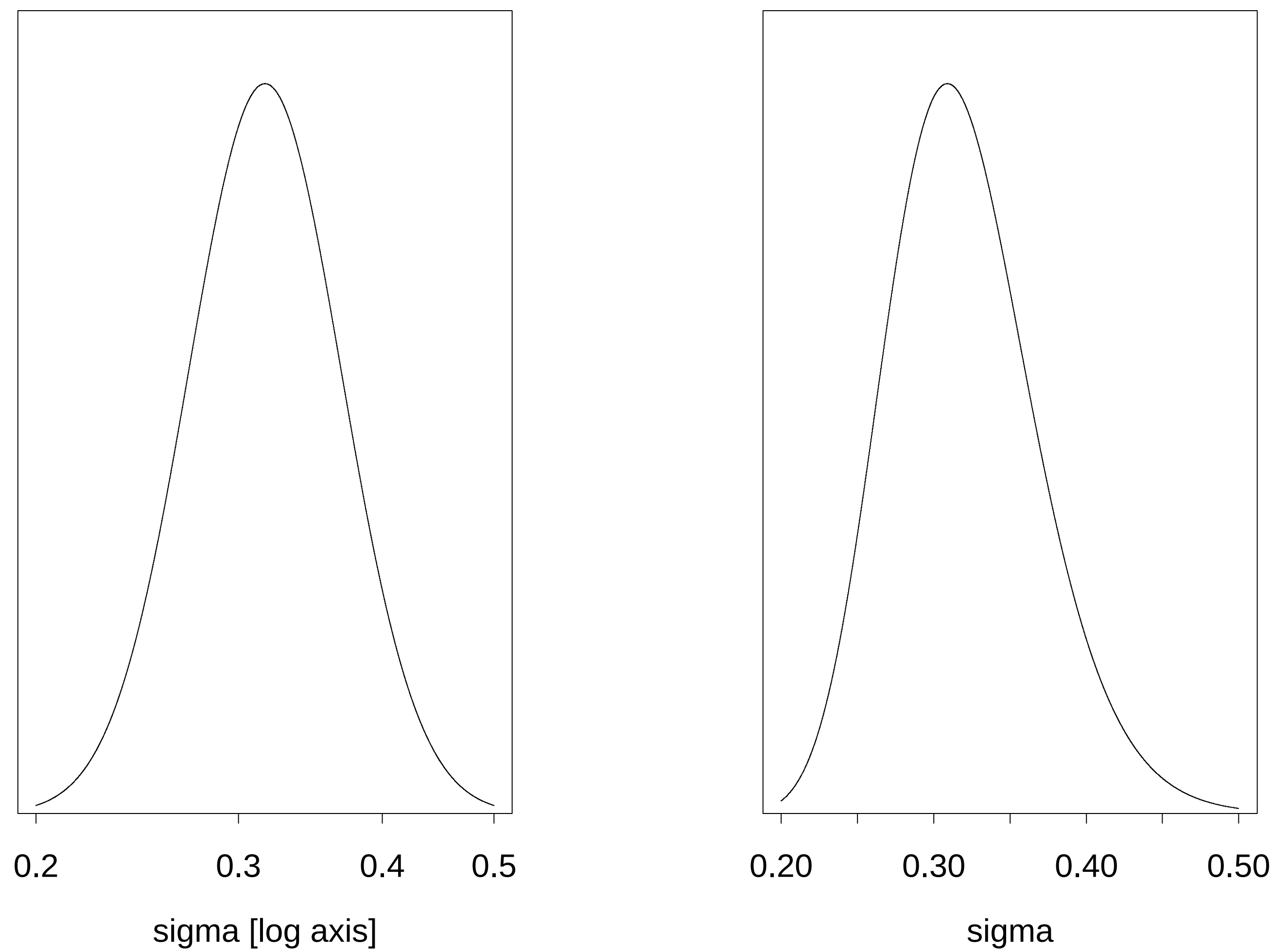
- The engineers did not have strong prior information on possible values for the distribution quantiles.

- For the Weibull shape parameter $\log(\sigma) \sim \text{NOR}(a_0, b_0)$, where a_0 and b_0 are obtained from the specification of two quantiles $\sigma_{\gamma/2}$ and $\sigma_{(1-\gamma/2)}$ of the prior distribution for σ. Then

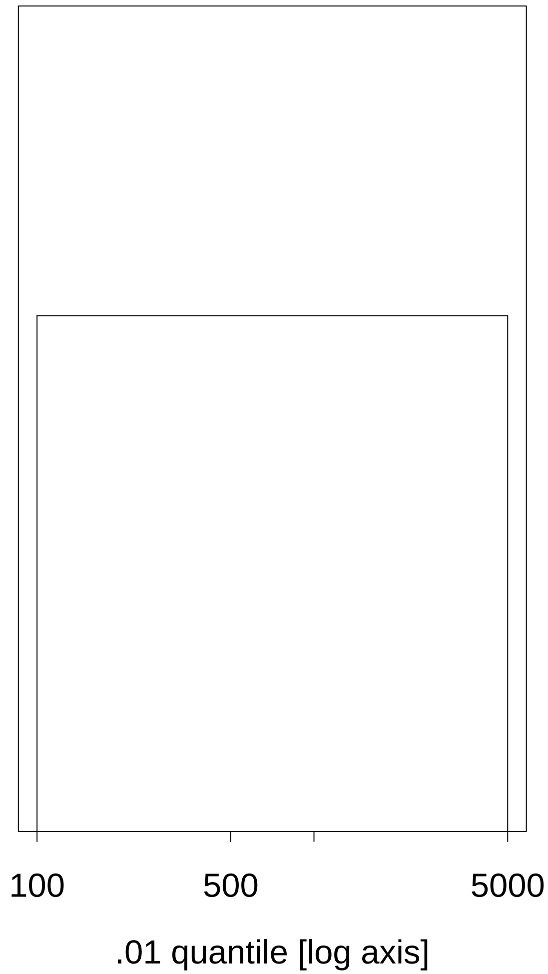
$$a_0 = \log\left[\sqrt{\sigma_{\gamma/2} \times \sigma_{(1-\gamma/2)}}\right], \quad b_0 = \log\left[\sqrt{\sigma_{(1-\gamma/2)}/\sigma_{\gamma/2}}\right]/z_{(1-\gamma/2)}$$

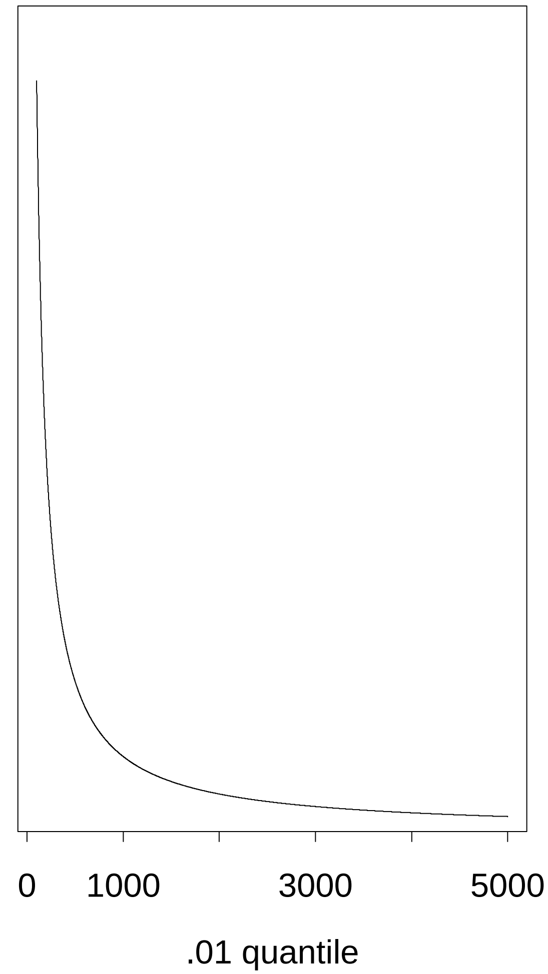
- Uncertainty in the Weibull .01 quantile will be described by $\text{UNIFORM}[\log(a_1), \log(b_1)]$ distribution where $a_1 = 100$ and $b_1 = 5000$ (wide range—not very informative).

Prior pdfs for $\log(\sigma)$ and σ when $\sigma_{.005} = .2, \sigma_{.995} = .5$



Prior pdfs for $\log(t_{.01})$ **and** $t_{.01}$ **when** $a_1 = 100, b_1 = 5000$





Joint Lognormal-Uniform Prior Distributions

- The prior for $\log(\sigma)$ is normal

$$f[\log(\sigma)] = \frac{1}{b_0}\phi_{\text{nor}}\left[\frac{\log(\sigma)-a_0}{b_0}\right], \quad \sigma > 0.$$

  The corresponding density for σ is $f(\sigma) = (1/\sigma) f[\log(\sigma)]$.

- The prior for $\log(t_p)$ is uniform

$$f[\log(t_p)] = \frac{1}{\log(b_1/a_1)}, \quad a_1 \le t_p \le b_1.$$

  The corresponding density for t_p is $f(t_p) = (1/t_p) f[\log(t_p)]$.

- Consequently, the joint prior distribution for (t_p, σ) is

$$f(t_p, \sigma) = \frac{f[\log(t_p)]}{t_p}\,\frac{f[\log(\sigma)]}{\sigma} \quad a_1 \le t_p \le b_1,\ \sigma > 0.$$

Joint Prior Distribution for (μ, σ)

- The transformation $\mu = \log(t_p) - \Phi^{-1}_{\text{sev}}(p)\sigma, \sigma = \sigma$ yields the prior for (μ, σ)

$$
\begin{aligned}
f(\mu, \sigma) &= \frac{f\left[\log(t_p)\right]}{t_p} \times \frac{f[\log(\sigma)]}{\sigma} \times t_p \\
&= f\left[\log(t_p)\right] \times \frac{f[\log(\sigma)]}{\sigma} \\
&= \frac{1}{\log(b_1/a_1)} \times \frac{\phi_{\text{nor}}\left\{\left[\log(\sigma) - a_0\right]/b_0\right\}}{\sigma b_0}
\end{aligned}
$$

where $\log(a_1) - \Phi^{-1}_{\text{sev}}(p)\sigma \le \mu \le \log(b_1) - \Phi^{-1}_{\text{sev}}(p)\sigma$, $\sigma > 0$.

- The region in which $f(\mu, \sigma) > 0$ is South-West to North-East oriented because $\text{Cov}(\mu, \sigma) = -\Phi^{-1}_{\text{sev}}(p)\text{Var}(\sigma) > 0$.

Joint Posterior Distribution for (μ, σ)

- The likelihood is

$$L(\mu,\sigma) = \prod_{i=1}^{2003} \left\{ \frac{1}{\sigma t_i} \phi_{\text{sev}} \left[\frac{\log(t_i) - \mu}{\sigma} \right] \right\}^{\delta_i} \times \left\{ 1 - \Phi_{\text{sev}} \left[\frac{\log(t_i) - \mu}{\sigma} \right] \right\}^{1-\delta_i}$$

where δ_i indicates whether the observation i is a failure or a right censored observation.

- The posterior distribution is

$$f(\mu,\sigma|\text{DATA}) = \frac{L(\mu,\sigma) f(\mu,\sigma)}{\int\int L(v,w) f(v,w) dv dw} = \frac{R(\mu,\sigma) f(\mu,\sigma)}{\int\int R(v,w) f(v,w) dv dw}.$$

Methods to Compute the Posterior

- **Numerical integration**: to obtain the posterior, one needs to evaluate the integral $f(\boldsymbol{\theta}|\text{DATA}) = \int R(\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$ over the region on which $f(\boldsymbol{\theta}) > 0$.

  In general there is not a closed form for the integral and the computation has to done numerically using fixed quadrature or adaptive integration algorithms.

- **Simulation methods**: the posterior can be approximated using Monte Carlo simulation resampling methods.

Computing the Posterior Using Simulation

Using simulation, one can draw a sample from the posterior using only the likelihood and the prior. The procedure for a general parameter $\boldsymbol{\theta}$ and prior distribution $f(\boldsymbol{\theta})$ is as follows:

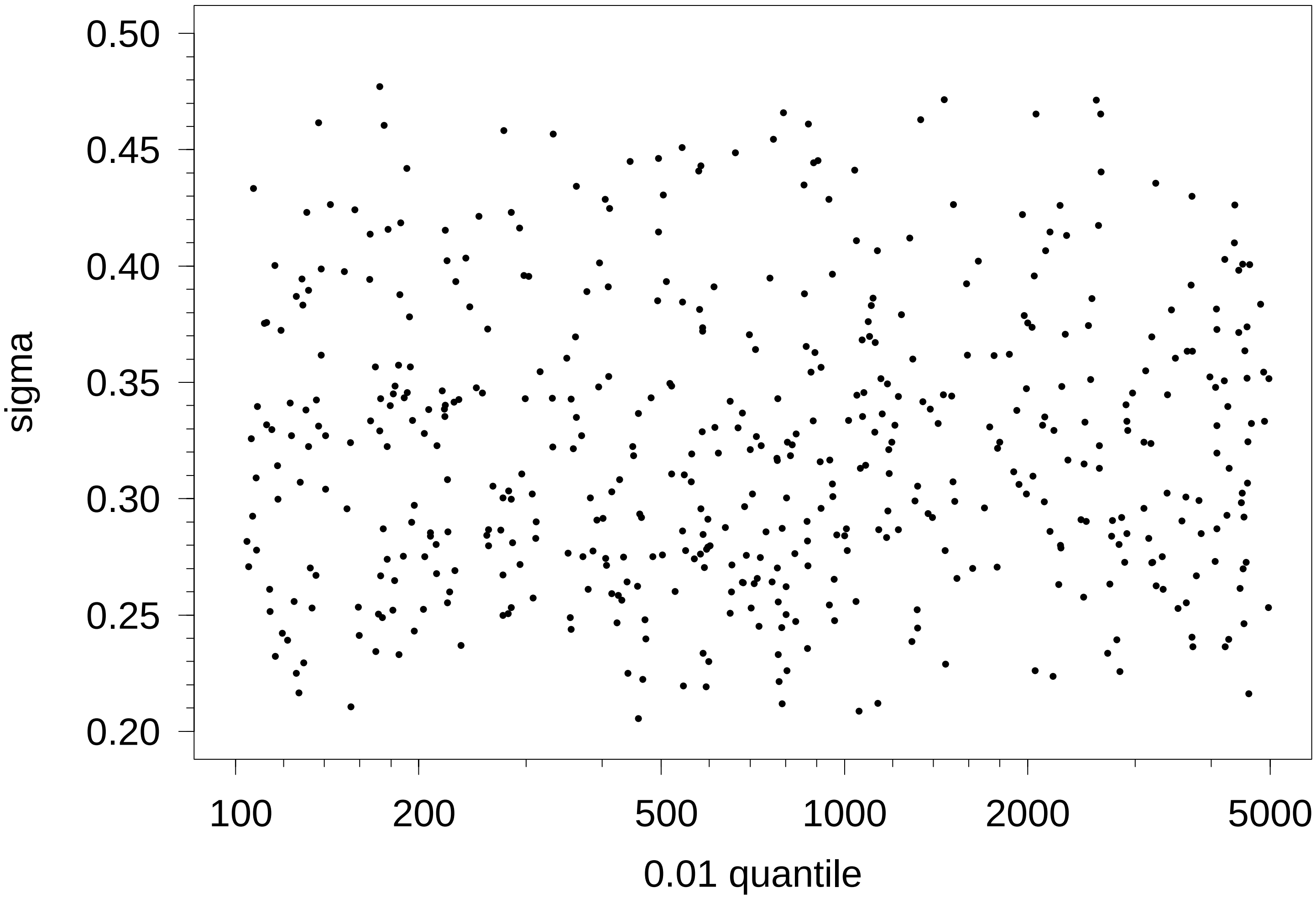
- Let $\boldsymbol{\theta}_i$, $i = 1, \ldots, M$ be a random sample from $f(\boldsymbol{\theta})$.

- The ith observation, $\boldsymbol{\theta}_i$, is retained with probability $R(\boldsymbol{\theta}_i)$.

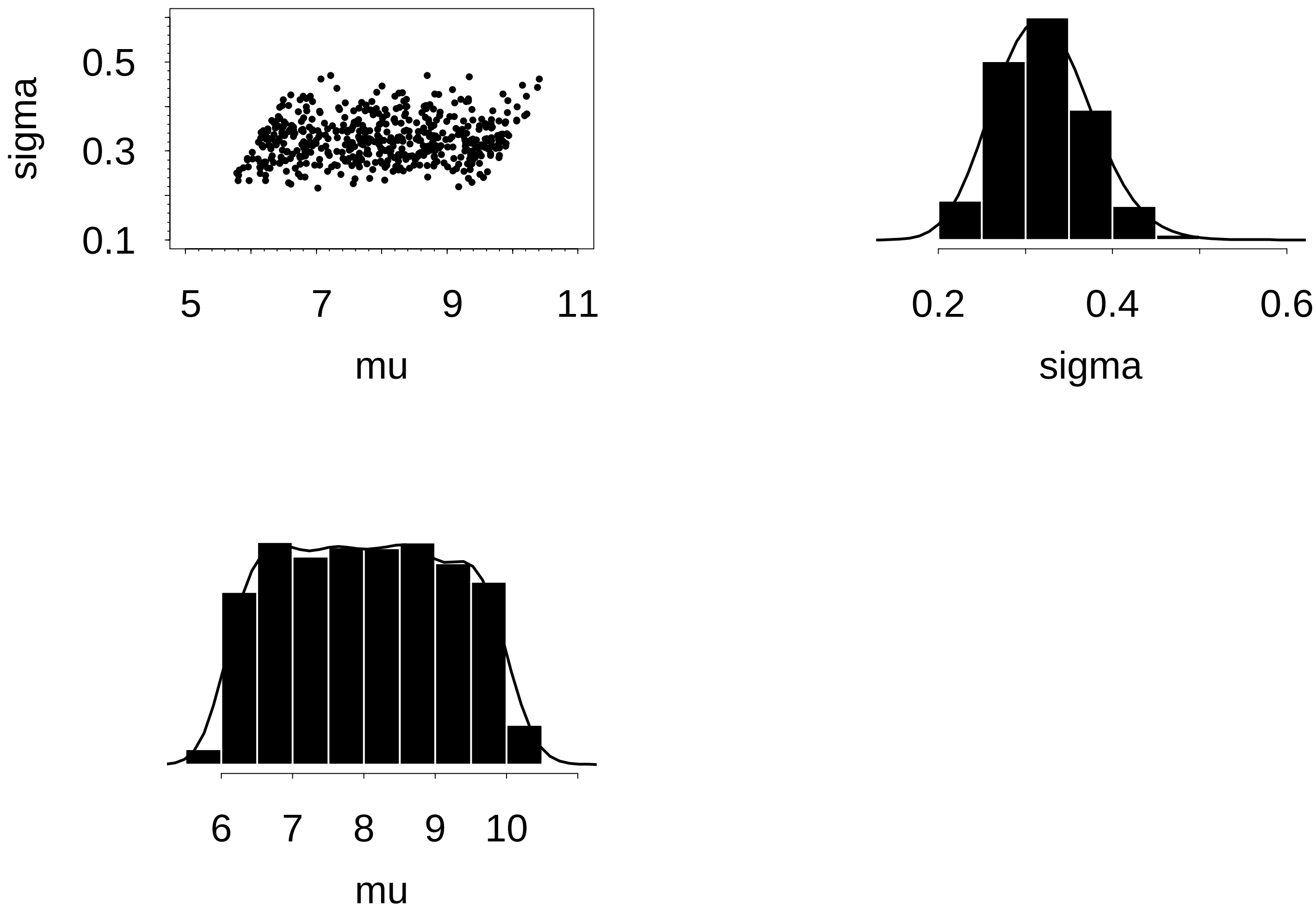
  Then if U_i is a random observation from a uniform $(0, 1)$, $\boldsymbol{\theta}_i$ is retained if
  $$U_i \le R(\boldsymbol{\theta}_i).$$

- It can be shown that the retained observations, say $\boldsymbol{\theta}_1^\star, \ldots \boldsymbol{\theta}_{M^\star}^\star$ $(M^\star \le M)$ are observations from the posterior $f(\boldsymbol{\theta}|\text{DATA})$.

Simulated Joint Prior for $t_{.01}$ and σ



Simulated Joint and Marginal Prior Distributions for μ and σ



Sampling from the Prior

The joint prior for $\boldsymbol{\theta} = (\mu, \sigma)$, is generated as follows:

- Use the inverse cdf method (see Chapter 4) to obtain a pseudorandom sample for t_p, say

$$(t_p)_i = a_1 \times b_1^{U_{1i}}, \quad i = 1, \ldots, M$$

  where $U_{11}, \ldots, U_{1M}$ are a pseudorandom sample from a uniform $(0, 1)$.

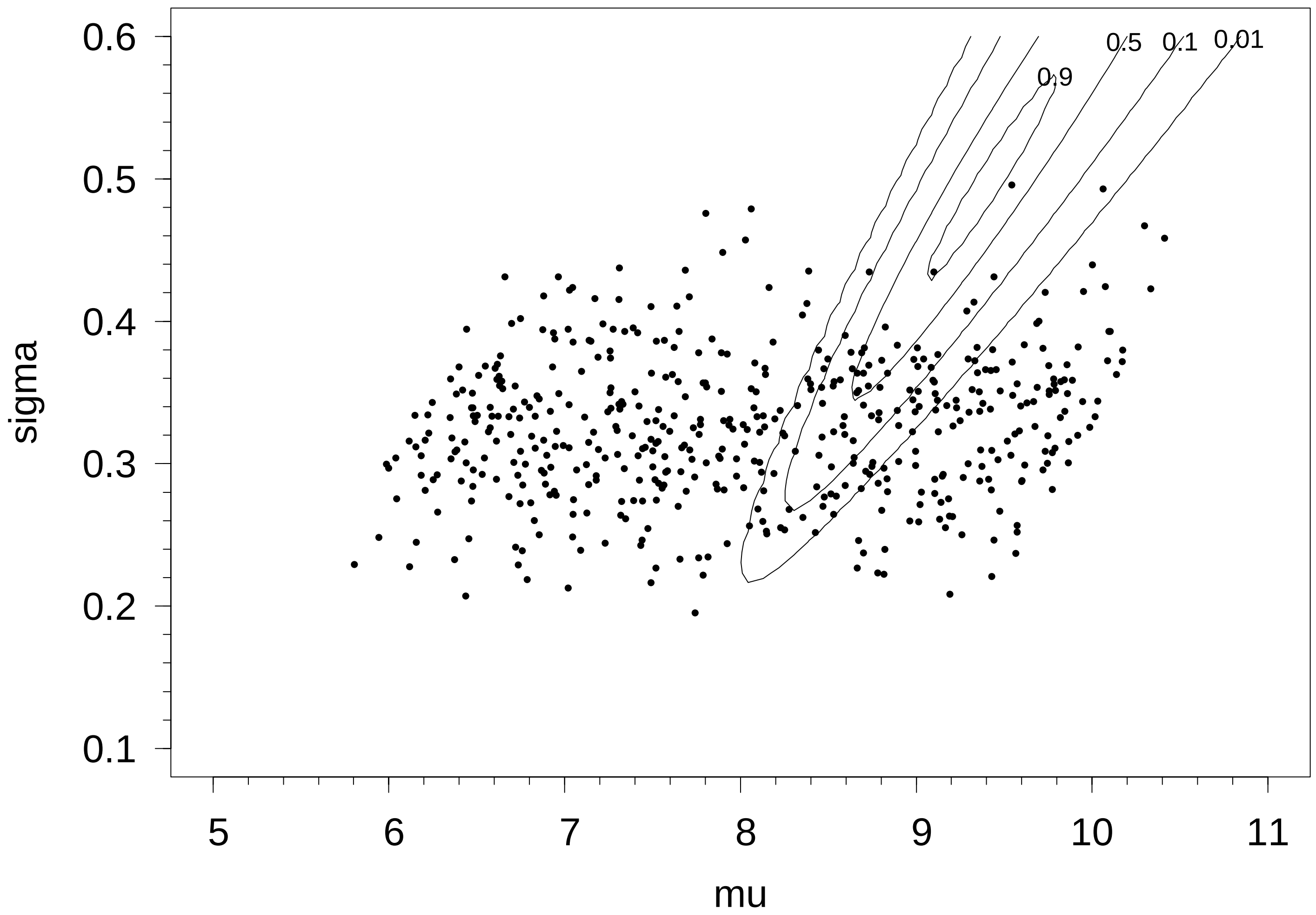
- Similarly, obtain a pseudorandom sample for σ, say

$$\sigma_i = \exp\left[a_0 + b_0 \Phi_{\text{nor}}^{-1}(U_{2i})\right]$$

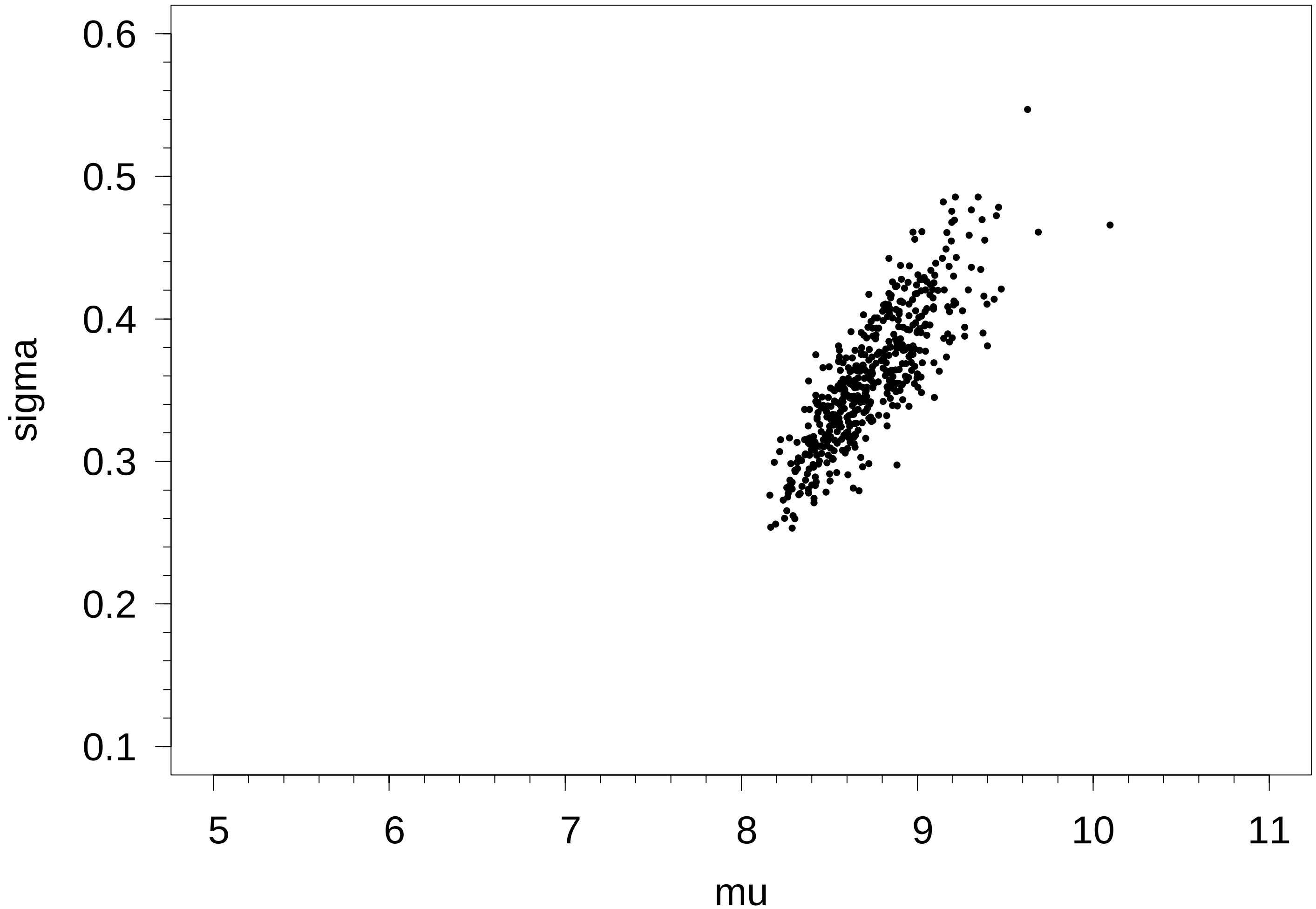
  where $U_{21}, \ldots, U_{2M}$ are another independent pseudorandom sample from a uniform $(0, 1)$.

- Then $\boldsymbol{\theta}_i = (\mu_i, \sigma_i)$ with $\mu_i = \log\left[(t_p)_i\right] - \Phi_{\text{sev}}^{-1}(p)\sigma_i$ is a pseudorandom sample from the (μ, σ) prior.

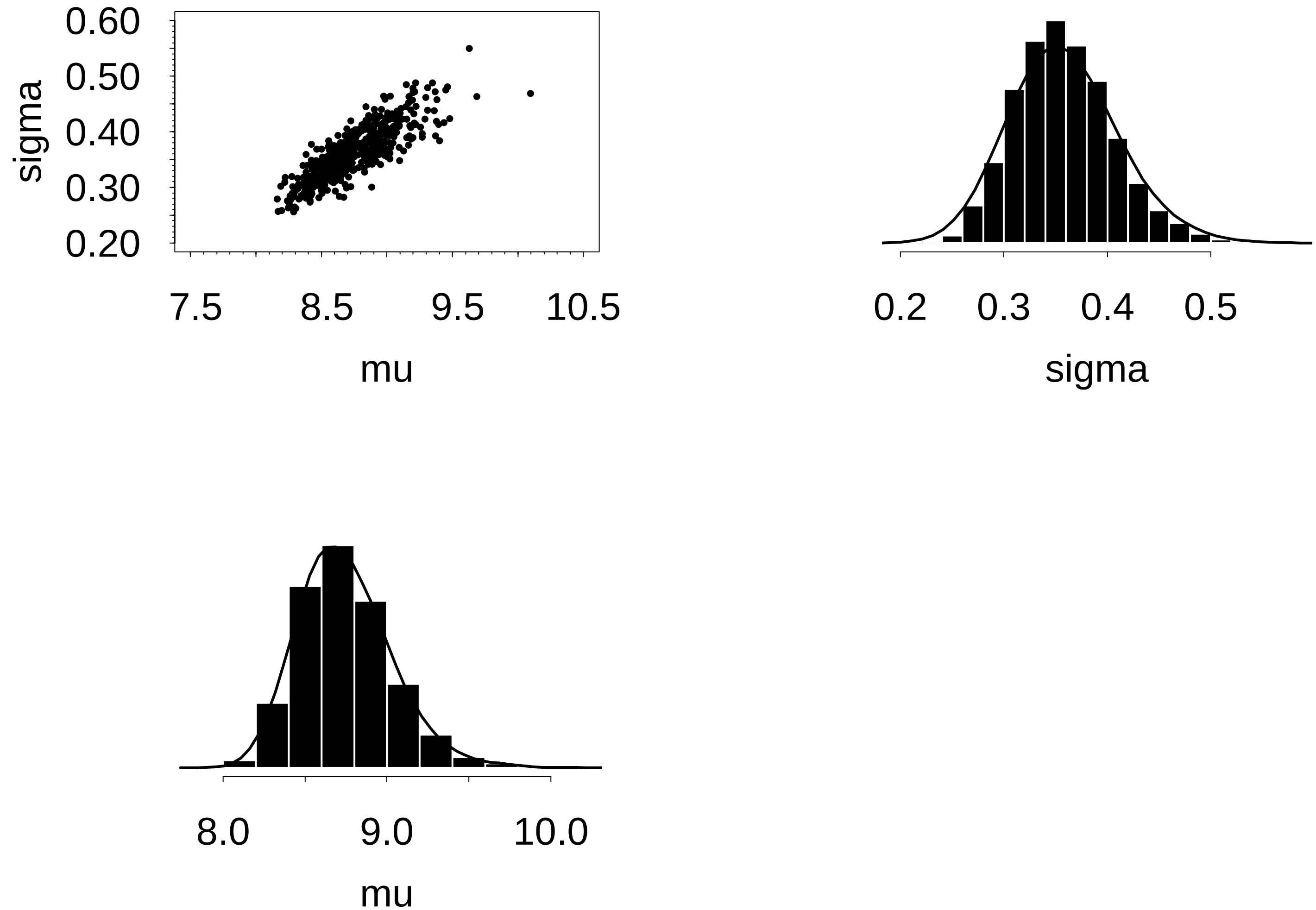
Simulated Joint Prior Distribution with μ and σ Relative Likelihood



Joint Posterior for μ and σ



Joint Posterior and Marginals for μ and σ for the Bearing Cage Data



Comments on Computing Posteriors Using Resampling

The number of observations $M^\star$ from the posterior is random with an expected value of

$$\mathsf{E}(M^\star) = M \int f(\boldsymbol{\theta}) R(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

Consequently,

- When the prior and the data do not agree well, $M^\star << M$ otherwise and a larger prior sample will be required.

- Can add to the posterior by sequentially filtering groups of prior points until a sufficient number is available in the posterior.

Posterior and Marginal Posterior Distributions for the Model Parameters

- Inferences on individual parameters are obtained by using the marginal posterior distribution of the parameter of interest. The marginal posterior of θ_j is

$$f[\boldsymbol{\theta}_j|\text{DATA}] = \int f(\boldsymbol{\theta}|\text{DATA})d\boldsymbol{\theta}'.$$

  where θ' is the subset of the parameters excluding θ_j.

- Using the general resampling method described above, one gets a sample for the posterior for $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_i^\star = (\mu_i^\star, \sigma_i^\star)$, $i = 1, \ldots, M^\star$.

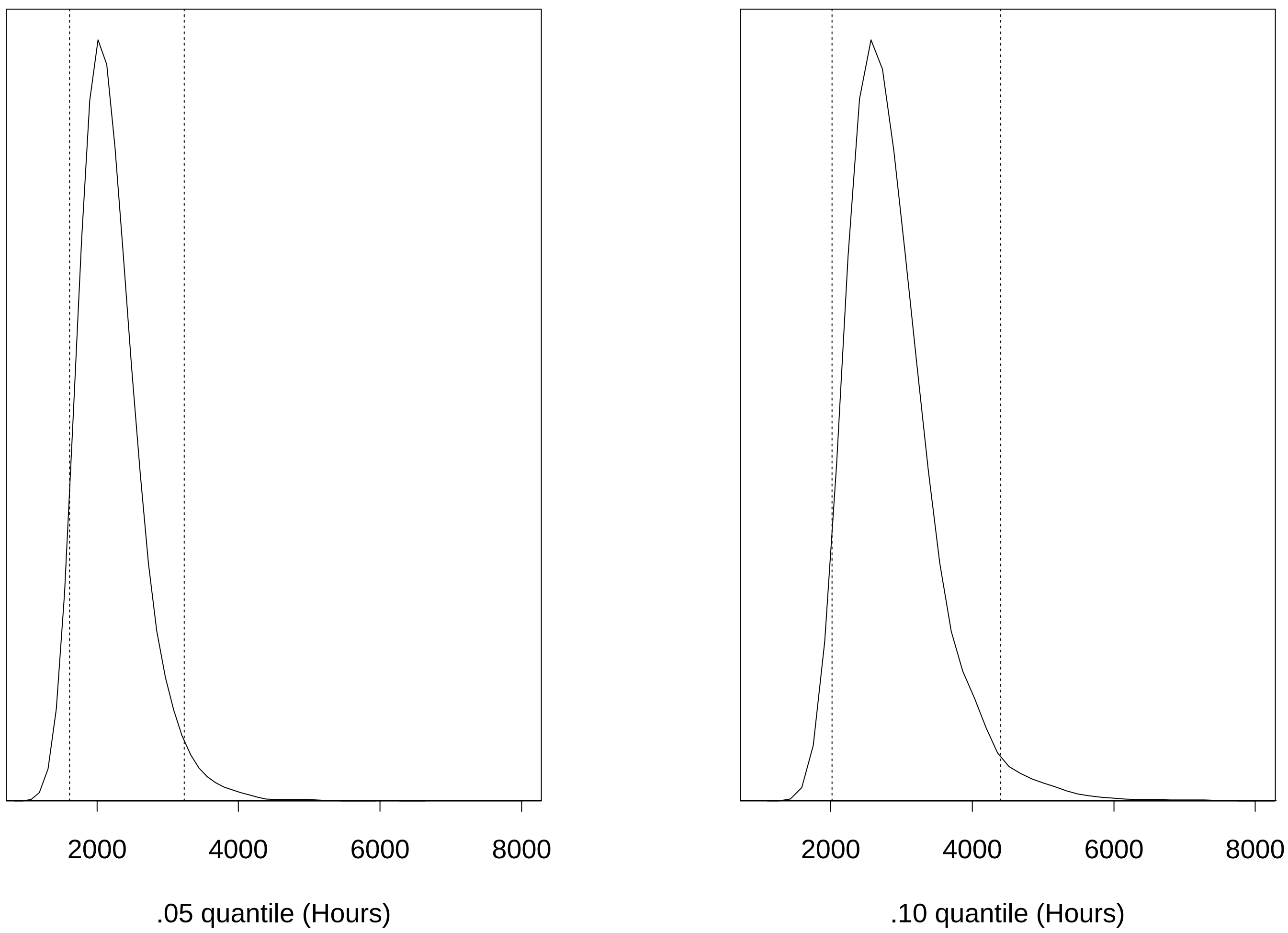
- Inferences for μ or σ alone are based on the corresponding **marginal** distributions $\mu_i^\star$ and $\sigma_i^\star$, respectively.

Posterior and Marginal Posterior Distributions for the Functions of Model Parameters

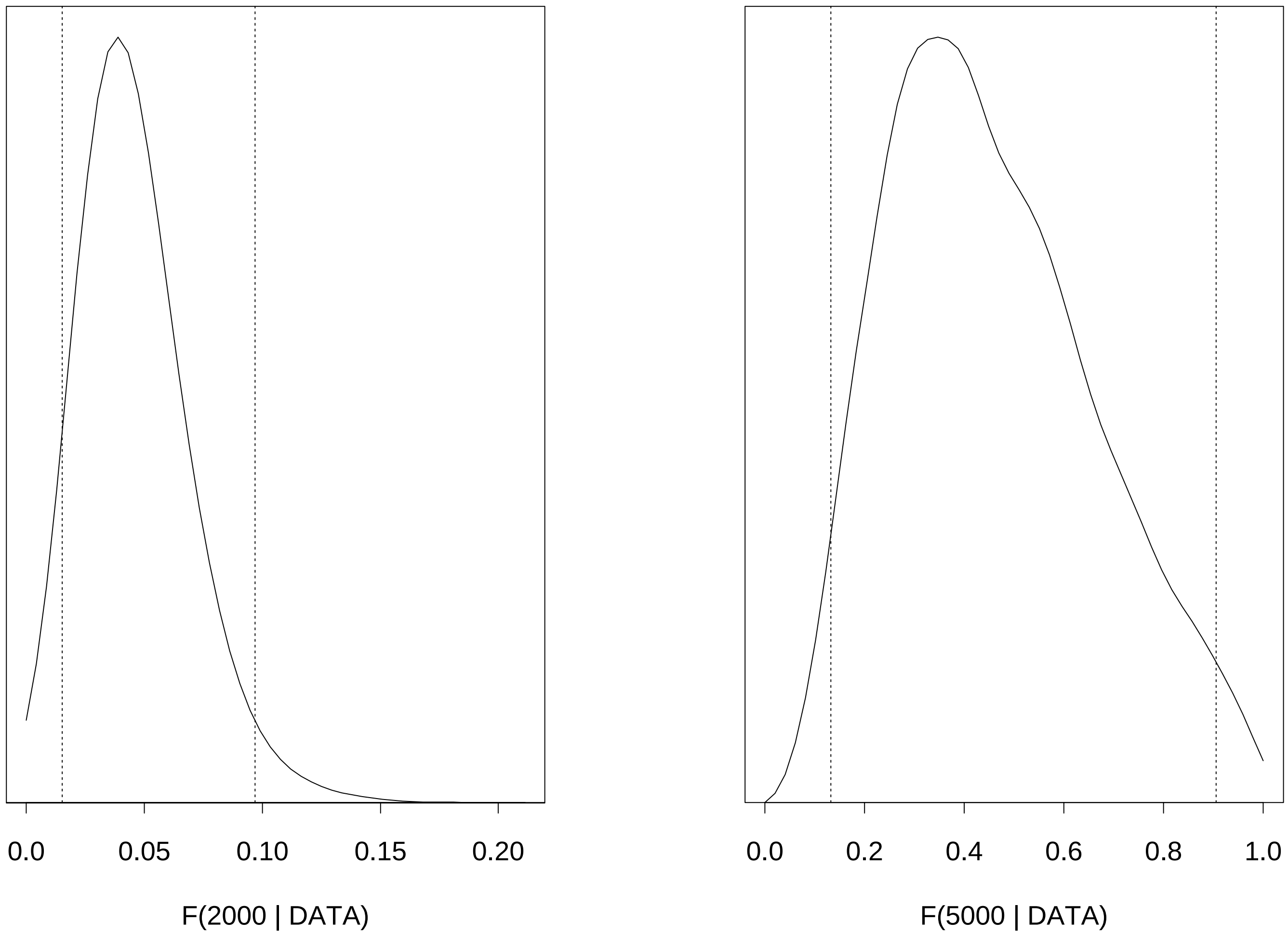
- Inferences on a scalar function of the parameters $g(\boldsymbol{\theta})$ are obtained by using the marginal posterior distribution of the functions of the parameters of interest, $f[g(\boldsymbol{\theta})|\text{DATA}]$.

- Using the simulation method, inferences are based on the simulated posterior marginal distributions. For example:
  - ▶ The marginal posterior distribution of $f(t_p|\text{DATA})$ for inference on quantiles is obtained from the empirical distribution of $\mu_i^\star + \Phi_{\text{sev}}^{-1}(p)\sigma_i^\star$.
  - ▶ The marginal posterior distribution of $f[F(t_e)|\text{DATA}]$ for inference for failure probabilities at t_e is obtained from the empirical distribution of $\Phi_{\text{sev}}\left[\frac{\log(t_e)-\mu_i^\star}{\sigma_i^\star}\right]$.

Simulated Marginal Posterior Distributions for $t_{.05}$ and $t_{.10}$



Simulated Marginal Posterior Distributions for $F(2000)$ and $F(5000)$



Bayes Point Estimation

Bayesian inference for $\boldsymbol{\theta}$ and functions of the parameters $\boldsymbol{g}(\boldsymbol{\theta})$ are entirely based on their posterior distributions $f(\boldsymbol{\theta}|\text{DATA})$ and $f[\boldsymbol{g}(\boldsymbol{\theta})|\text{DATA}]$.

Point Estimation:

- If $g(\boldsymbol{\theta})$ is a scalar, a common Bayesian estimate of $g(\boldsymbol{\theta})$ is its posterior mean, which is given by

$$\widehat{g}(\boldsymbol{\theta}) = \mathrm{E}[g(\boldsymbol{\theta})|\text{DATA}] = \int g(\boldsymbol{\theta}) f(\boldsymbol{\theta}|\text{DATA}) d\boldsymbol{\theta}.$$

  In particular, for the ith component of $\boldsymbol{\theta}$, $\widehat{\theta}_i$ is the posterior mean of θ_i. This estimate is the the Bayes estimate that minimizes the square error loss.

- Other possible choices to estimate $g(\boldsymbol{\theta})$ include (a) the posterior mode, which is very similar to the ML estimate and (b) the posterior median.

One-Sided Bayes Confidence Bounds

- A $100(1-\alpha)\%$ Bayes lower confidence bound (or credible bound) for a scalar function $g(\boldsymbol{\theta})$ is value $\underset{\sim}{g}$ satisfying

$$\int_{\underset{\sim}{g}}^{\infty} f[g(\boldsymbol{\theta})|\text{DATA}]dg(\boldsymbol{\theta}) = 1-\alpha$$

- A $100(1-\alpha)\%$ Bayes upper confidence bound (or credible bound) for a scalar function $g(\boldsymbol{\theta})$ is value $\tilde{g}$ satisfying

$$\int_{-\infty}^{\tilde{g}} f[g(\boldsymbol{\theta})|\text{DATA}]dg(\boldsymbol{\theta}) = 1-\alpha$$

Two-Sided Bayes Confidence Intervals

- A $100(1-\alpha)\%$ Bayes confidence interval (or credible interval) for a scalar function $g(\boldsymbol{\theta})$ is any interval $[\underset{\sim}{g}, \quad \tilde{g}]$ satisfying

$$\int_{\underset{\sim}{g}}^{\tilde{g}} f[g(\boldsymbol{\theta})|\text{DATA}]dg(\boldsymbol{\theta}) = 1-\alpha \tag{1}$$

- The interval $[\underset{\sim}{g}, \quad \tilde{g}]$ can be chosen in different ways
  - Combining two $100(1-\alpha/2)\%$ intervals puts equal probability in each tail (preferable when there is more concern for being incorrect in one direction than the other).
  - A $100(1-\alpha)\%$ Highest Posterior Density (HPD) confidence interval chooses $[\underset{\sim}{g}, \quad \tilde{g}]$ to consist of all values of g with $f(g|\text{DATA}) > c$ where c is chosen such that (1) holds. HPD intervals are similar to likelihood-based confidence intervals. Also, when $f[g(\boldsymbol{\theta})|\text{DATA}]$ is unimodal the HPD is the narrowest Bayes interval.

Bayesian Joint Confidence Regions

The same procedure generalizes to confidence regions for vector functions $\boldsymbol{g}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$.

- A $100(1-\alpha)\%$ Bayes confidence region (or credible region) for a vector valued function $\boldsymbol{g}(\boldsymbol{\theta})$ is defined as

$$\mathsf{CR}_{\mathsf{B}} = \{\boldsymbol{g}(\boldsymbol{\theta}) | f[g|\mathsf{DATA}] \geq c\}$$

where c is chosen such that

$$\int_{\mathsf{CR}_{\mathsf{B}}} f[\boldsymbol{g}(\boldsymbol{\theta})|\mathsf{DATA})d\boldsymbol{g}(\boldsymbol{\theta}) = 1 - \alpha$$

- In this case the presentation of the confidence region is difficult when $\boldsymbol{\theta}$ has more than 2 components.

Bayes Versus Likelihood

- Summary table or plots to compare the Likelihood versus the Bayes Methods to compare confidence intervals for μ, σ, and $t_{.1}$ for the Bearing-cage data example.

Prediction of Future Events

- Future events can be predicted by using the Bayes predictive distribution.

- If X [with pdf $f(\cdot|\boldsymbol{\theta})$] represents a future random variable
  - the posterior predictive pdf of X is

$$\begin{aligned} f(x|\text{DATA}) &= \int f(x|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\text{DATA}) d\boldsymbol{\theta} \\ &= \mathsf{E}_{\boldsymbol{\theta}|\text{DATA}} \left[f(x|\boldsymbol{\theta}) \right] \end{aligned}$$

  - the posterior predictive cdf of X is

$$\begin{aligned} F(x|\text{DATA}) &= \int_{-\infty}^{x} f(u|\boldsymbol{\theta}) du = \int F(x|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\text{DATA}) d\boldsymbol{\theta} \\ &= \mathsf{E}_{\boldsymbol{\theta}|\text{DATA}} \left[F(x|\boldsymbol{\theta}) \right] \end{aligned}$$

  where the expectations are computed with respect to the posterior distribution of $\boldsymbol{\theta}$.

Approximating Predictive Distributions

- $f(x|\text{DATA})$ can be approximated by the average of the posterior pdfs $f(x|\boldsymbol{\theta}_i^\star)$. Then

$$f(x|\text{DATA}) \;\approx\; \frac{1}{M^\star}\sum_{i=1}^{M^\star} f(x|\boldsymbol{\theta}_i^\star).$$

- Similarly, $F(x|\text{DATA})$ can be approximated by the average of the the posterior cdfs $F(x|\boldsymbol{\theta}_i^\star)$. Then

$$F(x|\text{DATA}) \;\approx\; \frac{1}{M^\star}\sum_{i=1}^{M^\star} F(x|\boldsymbol{\theta}_i^\star).$$

- A two-sided $100(1-\alpha)\%$ Bayesian prediction interval for a new observation is given by the $\alpha/2$ and $(1-\alpha/2)$ quantiles of $F(x|\text{DATA})$.

Location-Scale Based Prediction Problems

Here we consider prediction problems when $\log(T)$ has a location-scale distribution.

- Predicting a future value of T. In this case, $X = T$ and $x = t$, then

$$f(t|\boldsymbol{\theta}) = \frac{1}{\sigma t}\phi(\zeta), \quad F(t|\boldsymbol{\theta}) = \Phi(\zeta)$$

where $\zeta = [\log(t) - \mu]/\sigma$.

- Thus, for the Bearing-cage fracture data, approximations of the predictive pdf and cdf for a **new** observation are:

$$\begin{aligned} f(t|\text{DATA}) &\approx \frac{1}{M^\star}\sum_{i=1}^{M^\star}\frac{1}{\sigma_i^\star t}\phi_{\text{sev}}(\zeta_i^\star) \\ F(t|\text{DATA}) &\approx \frac{1}{M^\star}\sum_{i=1}^{M^\star}\Phi_{\text{sev}}(\zeta_i^\star) \end{aligned}$$

where $\zeta_i^\star = [\log(t) - \mu_i^\star]/\sigma_i^\star$.

Prediction of an Order Statistic

Here we consider prediction of the kth order statistic in a future sample of size m from the distribution of T when $\log(T)$ has a location-scale distribution.

- In this case, $X = T_{(k)}$ and $x = t_{(k)}$, then

$$
\begin{aligned}
f[t_{(k)}|\boldsymbol{\theta}] &= \frac{m!}{(k-1)!\,(m-k)!} \times [\Phi(\zeta)]^{k-1} \times \frac{1}{\sigma t_{(k)}} \phi(\zeta) \\
&\quad \times [1-\Phi(\zeta)]^{m-k} \\
F[t_{(k)}|\boldsymbol{\theta}] &= \sum_{j=k}^{m} \frac{m!}{j!\,(m-j)!} [\Phi(\zeta)]^{j} \times [1-\Phi(\zeta)]^{m-j}
\end{aligned}
$$

where $\zeta = [\log(t_{(k)}) - \mu]/\sigma$.

Predicting the 1st Order Statistic

When $k = 1$ (predicting the 1st order statistic), the formulas simplify to

- Predictive pdf

$$f[t_{(1)}|\boldsymbol{\theta}] \;=\; m \times [\Phi(\zeta)]^{m-1} \times \frac{1}{\sigma t_{(1)}} \phi(\zeta) \times [1 - \Phi(\zeta)]^{m-1}$$

- Predictive cdf

$$F[t_{(1)}|\boldsymbol{\theta}] \;=\; 1 - [1 - \Phi(\zeta)]^{m}$$

where $\zeta = [\log(t_{(1)}) - \mu]/\sigma$.

Predicting the 1st Order Statistic for the Bearing-Cage Fracture Data

For the Bearing-cage fracture data:

- An approximation for the predictive pdf for the 1st order statistic is

$$f[t_{(1)}|\text{DATA}] \;\approx\; \frac{1}{M^\star}\sum_{i=1}^{M^\star}\left\{m \times \frac{1}{\sigma_i^\star t}\phi\left(\zeta_i^\star\right) \times \left[1 - \Phi\left(\zeta_i^\star\right)\right]^{m-1}\right\}$$

- The corresponding predictive cdf is

$$F[t_{(k)}|\text{DATA}] \;\approx\; \frac{1}{M^\star}\sum_{i=1}^{M^\star}\left\{1 - \left[1 - \Phi\left(\zeta_i^\star\right)\right]^{m}\right\}$$

where $\zeta_i^\star = [\log(t) - \mu_i^\star]/\sigma_i^\star$.

Predicting a New Observation

- $F(t|\text{DATA})$ can be approximated by the average of the posterior probabilities $F(t|\boldsymbol{\theta}_i^\star), \; i = 1, \ldots, M^\star$.

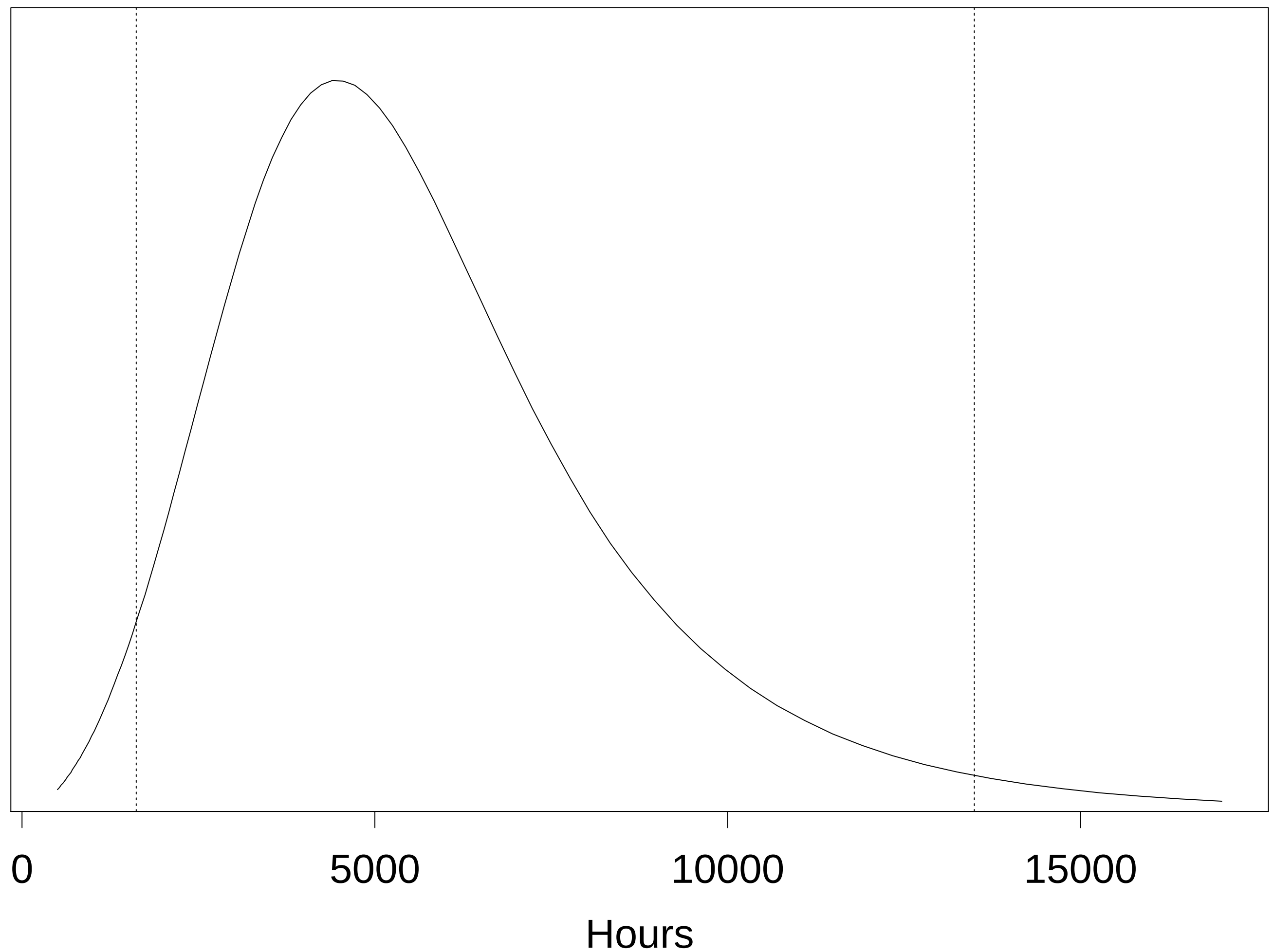
- Similarly, $f(t|\text{DATA})$ can be approximated by the average of the posterior densities $f(t|\boldsymbol{\theta}_i^\star), \; i = 1, \ldots, M^\star$.

- In particular for the Bearing-cage fracture data, an approximation for the predictive pdf and cdf are

$$
\begin{aligned}
f(t|\text{DATA}) &\approx \frac{1}{M^\star} \sum_{i=1}^{M^\star} \frac{1}{\sigma_i^\star t} \phi_{\text{sev}} \left[\frac{\log(t) - \mu_i^\star}{\sigma_i^\star} \right] \\
F(t|\text{DATA}) &\approx \frac{1}{M^\star} \sum_{i=1}^{M^\star} \Phi_{\text{sev}} \left[\frac{\log(t) - \mu_i^\star}{\sigma_i^\star} \right].
\end{aligned}
$$

- A $100(1-\alpha)\%$ Bayesian prediction interval for a new observation is given by the percentiles of this distribution.

Predictive Density and Prediction Intervals for a Future Observation from the Bearing Cage Population



Caution on the Use of Prior Information

- In many applications, engineers really have useful, indisputable prior information. In such cases, the information should be integrated into the analysis.

- We must beware of the use of **wishful thinking** as prior information. The potential for generating seriously misleading conclusions is high.

- As with other inferential methods, when using Bayesian methods, it is important to do sensitivity analyses with respect to uncertain inputs to ones model (including the inputted prior information)