Issues and Challenges of Bayesian Inference in PHM: Prior, Data, or Lying

Dr. Zhaojun ‘Steven’ Li

Western New England University, Massachusetts, USA
Outline

• Bayesian vs. Frequentist
• Why Bayesian?
• Where Priors Come from?
• Computational Challenges
• Multiple Priors Integration
• Bayesian Applications in PHM
Modeling and Decision

“Essentially, all models are wrong, but some are useful”

Statistician: George Box

In God we trust, all others bring data
Thomas Bayes

English Theologian [神学家] and Mathematician [数学家]

1702-1761

• Bayesian:
  – One who asks you what you believe before a study in order to tell you what you think afterwards

• Frequentist:
  – One can tell you information for decision making based on assuming the world is known no matter what beliefs you have
Bayesian vs. Frequentist

• Bayesian:
  – Data are observed from the realized sample
  – Parameters are unknown and described probabilistically
  – Data are fixed

• Frequentist:
  – Data are repeatable random sample
  – Underlying parameters is constant in the repeatable process.
  – Parameters are fixed
Why Bayes — a Motivation Example
Binomial Proportion Estimation

• Point estimator
  \[ \hat{p} = \bar{y} = \frac{y}{n} \] (y is # of successes, n is total # of observations)

• Wald Confidence Interval:
  \[ \bar{y} \pm 1.96\sqrt{\frac{\bar{y}(1 - \bar{y})}{n}}, \ SE=\sqrt{\frac{\bar{y}(1 - \bar{y})}{n}} \]

• Under the Null Hypothesis \( p = p_0 \), for LARGE sample size n
  \[ \frac{\bar{y} - p_0}{\sqrt{\frac{\bar{y}(1 - \bar{y})}{n}}} \sim Norm(0, 1) \]

Issues:

  – If \( y=n, \bar{y}=1? \) No matter what confidence level, e.g., 99.99\%, the inference for reliability estimate is always one.
  – **Accuracy suffers** when \( np<5 \) or \( n(1-p)<5 \), which results from the normal approximation.
Why Bayes — a Motivation Example
Clopper-Pearson Interval

• Clopper-Pearson Interval (without assuming Normality)

\[ p_{lb} = B^{-1}\left(\frac{\alpha}{2}; y, n - y + 1\right) \]
\[ p_{ub} = B^{-1}(1 - \frac{\alpha}{2}; y + 1, n - y) \]

where \( B^{-1}(c; a, b) \) is the \( c^{th} \) quantile of the Beta distribution with parameters \( a \) and \( b \)

• Reasonable CI for \( y = 0 \) or \( y = n \), However, can be very broad

• Under the Null Hypothesis \( H_0: p = p_0 \), Beta integral can be used to calculate the cumulative binomial distribution

\[ - \Sigma_{i=y}^{n} \binom{n}{i} p_{lb}^i (1 - p_{lb})^{n-i} = \frac{\alpha}{2} \implies B_{p_{lb}}(y, n - y + 1) = \frac{\alpha}{2} \]
\[ - \Sigma_{i=0}^{y} \binom{n}{i} p_{ub}^i (1 - p_{ub})^{n-i} = \frac{\alpha}{2} \implies B_{p_{ub}}(y + 1, n - y) = 1 - \frac{\alpha}{2} \]

where \( p_{ub} \) and \( p_{lb} \) are the upper bound and lower bound, respectively.
Why Bayes — a Motivation Example
Clopper-Pearson Interval

• F distribution can also be used to calculate the cumulative Binomial probability function
  – Two sided interval:

  \[ p_{lb} = \frac{y}{y + (n - y + 1)F_{2(n-y+1), 2y, 1-\alpha/2}} \]
  \[ p_{ub} = \frac{y + 1}{y + 1 + (n - y)F_{2(n-y), 2(y+1), \alpha/2}} \]

• For Example, \( \alpha = 0.05 \), with small sample size of \( n = 5 \)

  \[ y = 0: CI = [0, 0.53] \]
  \[ y = n: CI = [0.48, 1] \]
Why Bayes — a Motivation Example
Wilson Score Interval

– Using the Standard Deviation under \( H_0: p = p_0 \) (not estimated SE)

– Solving the equations

\[
\frac{\bar{y} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \pm z_{\alpha/2} \text{ for } p_0.
\]

\[
\left( \bar{y} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{[\bar{y}(1-\bar{y}) + \frac{z_{\alpha/2}^2}{4n}]}{n}} \right) / \left( 1 + \frac{z_{\alpha/2}^2}{n} \right)
\]

– **More robust** than Wald interval (reduce the impact from rare events)

– **The coverage probability is closer** to nominal confidence level \( 1 - \alpha \), e.g., 95% at \( \alpha = 0.05 \)
Why Bayes — a Motivation Example
Wilson Score Interval

• Rather than using $\bar{y}$ in Wald CI, the mid point is

$$\bar{y} \left( \frac{n}{n + z_{\alpha/2}^2} \right) + \frac{1}{2} \left( \frac{z_{\alpha/2}^2}{n + z_{\alpha/2}^2} \right)$$

weighted average of estimated parameter $\bar{y}$ and $\frac{1}{2}$.

• The variance of sample proportion

$$\frac{1}{n + z_{\alpha/2}^2} \left[ \bar{y}(1 - \bar{y}) \left( \frac{n}{n + z_{\alpha/2}^2} \right) + \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) \left( \frac{z_{\alpha/2}^2}{n + z_{\alpha/2}^2} \right) \right]$$

weighted average of the variance of a sample proportion of $p = \bar{y}$ and the variance of a sample proportion of $p = 1/2$, using $n + z_{\alpha/2}^2$ in place of the original sample size of $n$. 
Why Bayes — a Motivation Example
Adjusted Wald Interval

• Adjusted Wald Interval
  – Borrowed the mid-point value from Wilson Score by setting $z_{\alpha/2}^2 = 1.96^2 \approx 4$
  – Confidence interval
    • $\bar{y}_{AC} \pm z_{\alpha/2} s_{AC}$
      
      where $s_{AC}^2 = \frac{\bar{y}_{AC}(1-\bar{y}_{AC})}{n+4}$
      
      and $\bar{y}_{AC} = \frac{y + 2}{n+4} = \frac{n}{4+n} \bar{y} + \frac{4}{4+n} \frac{1}{2}$, which is the Wilson score interval’s midpoint
  – Which is also called Agresti-Coull Interval

Agresti and Coull (1998)
Bayesian Inference

**Bayesian Theorem**

\[
p(\theta | y) = \frac{p(y | \theta) p(\theta)}{\int p(y | \tilde{\theta}) p(\tilde{\theta}) d\tilde{\theta}}
\]

- \(\theta\): is the parameter of interest
- \(p(\theta)\): is the **prior** distribution
- \(p(y | \theta)\): the **sampling model**
- \(p(\theta | y)\): is the **posterior**
Bayesian Inference under Conjugate Priors

Example: Beta prior is conjugate for the Binomial model

- Posterior distribution

\[
f(p | y_1, ..., y_n) = \frac{f(y_1, ..., y_n | p)f(p)}{f(y_1, ..., y_n)}
\]

\[
= \frac{1}{p(y_1, ..., y_n)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \left( \sum_{i=1}^{n} y_i \right)
\times p^\sum_{i=1}^{y_i} (1-p)^{n-\sum_{i=1}^{y_i}}
\]

\[
= \frac{1}{p(y_1, ..., y_n)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left( \sum_{i=1}^{n} y_i \right) p^{a+\sum_{i=1}^{y_i}-1}(1-p)^{b+n-\sum_{i=1}^{y_i}-1}
\]

\[
= c(n, y, a, b)p^{a+\sum_{i=1}^{y_i}-1}(1-p)^{b+n-\sum_{i=1}^{y_i}-1}
\]

\[\sim beta(a + \sum_{i=1}^{y_i} - 1, b + n - \sum_{i=1}^{y_i} - 1)\]

- Posterior distribution of \(p\) is \(Beta(2 + y, 2 + n - y)\).

- The mean of posterior is

\[
\frac{y+2}{n+4} = \frac{n}{4+n} \bar{y} + \frac{4}{4+n} \frac{1}{2}.
\]

the weighted average of \(\bar{y}\) and \(\frac{1}{2}\)
Comparison Between Bayesian VS Frequentist

• The midpoint \( \bar{y} \left( \frac{n}{n+z^2} \right) + \frac{1}{2} \left( \frac{z^2}{n+z^2} \right) \) [Wilson Score]

• \( \bar{y} = \frac{y+2}{n+4} \) [Adjusted Wald]

\[ \sqrt{\frac{y}{n} \left( \frac{n}{n+z^2} \right) + \frac{1}{2} \left( \frac{z^2}{n+z^2} \right)} = \frac{y+1}{2} \frac{z^2}{n+z^2}, \quad z^2 = 1.96^2 \approx 4 \]

Identical to Bayes Posterior mean estimate with Beta(2,2) as the prior distribution.

– Variance Comparison

\[ \frac{(y+2)(n−y+2)}{(n+4)^3} \text{ (Adj. Wald)} \approx \frac{(y+2)(n−y+2)}{(n+4)^2(n+5)} \text{ (Bayes)} \]
Bayesian Inference under Semi-Conjugate Prior

Assuming given and known $\sigma^2$ in Normal model

- **Bayes**
  - $p(u|\sigma^2) \sim \text{Norm}(u_0, \tau_0^2)$ [semi-conjugate prior]
  - Assume $x_1, \ldots, x_n$ are i.i.d. from $\text{Norm}(u, \sigma^2)$, and then posterior $p(u|x_1, \ldots, x_n, \sigma^2)$ is normal.
  - $p(u|x_1, \ldots, x_n, \sigma^2) \sim \text{Norm}(u_n, \tau_n^2)$ [posterior dist.]

  $$u_n = \frac{1}{\tau_0^2} u_0 + \frac{n}{\sigma^2 \tau_0^2} \bar{y}; \quad \tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$ [precision]

- **Frequentist**
  - With given $\sigma$
    - $\mu = \bar{X} = \frac{x_1 + \cdots + x_n}{n}$
  - Confidence interval
    - $\bar{x} \pm z_{\alpha/2} \sigma / \sqrt{n}$

- Credible Interval
  - $u_n \pm z_{\alpha/2} \tau_n$
Summary: Key Components in Bayes

• **Prior**
  – information about the known parameters available from sources independent from observed data.
  – what you know about parameter excluding the information in data

• **Sample distribution/model**
  – the distribution of the observed data conditional on its parameters.
  – **likelihood function**

• **Posterior**
  – the distribution of the parameters after taking into account the observed data.
Prior Elicitation

Types of Prior knowledge

- Non-informative
  - Objective, vague, and diffuse
  - “flat” relative to the likelihood function, such as uniform distribution

- Informative
  - Subjective and decisive
  - Description of elicited expert opinion, historical data, previous studied, etc.

Prior distribution: describes prior information in the form of probability distribution.
Prior Elicitation

Example: Bayesian Reliability for Binary Data

• Inference on the unknown reliability of $r$
• Sampling model
  \[ Y | r \sim \text{binomial}(n = 6, r) \]
• Prior distribution
  – Non-informative prior
    \[ p(r) \sim U(0,1) \]
    which is also $beta(1,1) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} r^{1-1}(1 - r)^{1-1}$
  – Informative prior
    \[ p(r) \sim beta(a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} r^{a-1}(1 - r)^{b-1} \]
Posterior Reliability of $r$ with Beta priors

- **Posterior distribution**

$$p(r|y_1,\ldots,y_n) = \frac{p(y_1,\ldots,y_n|r)p(r)}{p(y_1,\ldots,y_n)}$$

$$= \frac{1}{p(y_1,\ldots,y_n)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} r^{a-1}(1-r)^{b-1} \left(\sum_{i=1}^{n} y_i\right)$$

$$\times r^{\sum y_i}(1-r)^{n-\sum y_i}$$

$$= \frac{1}{p(y_1,\ldots,y_n)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\sum_{i=1}^{n} y_i\right)^{a+\sum y_i-1}(1-r)^{b+n-\sum y_i-1}$$

$$= c(n, y, a, b) r^{a+\sum y_i-1}(1-r)^{b+n-\sum y_i-1}$$

$$\sim beta(a + \sum y_i - 1, b + n - \sum y_i - 1)$$

**Posterior is also Beta distribution**
Non-informative Prior and its Posterior

The same shape of the likelihood function and posterior distribution under uniform prior distribution

Remark: Uniform Dist. is Beta (1, 1)  
Likelihood function: \( L(r) = \prod_{i=1}^{n} f(x_i; r) \)
Posterior of $r$ with Various Priors

The prior and the likelihood function jointly determine the Posterior Distribution.
Sensitivity Analysis

If \( n \) observations are all successes, i.e., \( Y=n \), \( w \) indicates the confidence degree for a specific prior \( r_0 = \frac{a}{a+b} \).

\[
E(r|Y=n) = \frac{n}{a+b+n} \bar{y} + \frac{a+b}{a+b+n} \frac{a}{a+b}
\]

If \( n \) observations are all successes, i.e., \( Y=n \), \( w \) indicates the confidence degree for a specific prior \( r_0 = \frac{a}{a+b} \).

where \( w = a + b \)
Priors Aggregation

- **Averaging:** assign *equal* weights on opinions
  - \( T(p_1, p_2, ..., p_k) = \sum_1^{k} \frac{1}{k} p_i \)
  - *Simple* and ignorant of variability of different opinions

- **Pooling:** assign *different* weights to opinions
  - Linear pooling:
    \( T(p_1, p_2, ..., p_k) = \sum_1^{k} \alpha_i p_i \)
  - **Geometric pooling:**
    \( T(p_1, p_2, ..., p_k) = \prod_1^{k} p_i^{\alpha_i} \)

Geometric pooling is commonly used due to its desirable properties: relative propensity consistency (RPC) and external Bayesianity.
Priors Aggregation

• Choosing the weights in pooling methods becomes a challenge.

\[ T(p_1, p_2, \ldots, p_k) = \prod_{1}^{k} p_i^{\alpha_i} \]

\[ \alpha_i = ?, \ i = 1,2, \ldots, k. \]

• Optimality Criteria
  – Maximize the entropy of pooled distribution
  – Minimize Kullback-Lieble divergence between the pooled prior and individual prior.

De Carolla et al. 2015
System and Subsystem level Prior Aggregation

- **Linear Pooling:**
  \[ q(\phi) = \alpha q_1^*(\phi) + (1 - \alpha)q_2(\phi) \]

- **Geometric Pooling:**
  \[ q(\phi) = q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha} \]

- \( q_1(\theta) \) is the prior of the subsystem level parameter \( \theta \)
- \( q_2(\phi) \) is the **natural prior** of the system level parameter \( \phi \)
- \( \phi = M(\theta) \) – system structure function
- \( q_1^*(\phi) = q_1(M^{-1}(\phi))|J(\phi)| \) - **induced prior**
- \( q(\phi) \) is the **pooled prior** of \( \phi \), called the natural prior
- \( \alpha \) is the pooling parameter
System and Subsystem Level Prior Aggregation
A Two Components Series System Reliability

Priors and the system configuration:

Prior: \( \text{beta}(pC, C(1-p)) \)

<table>
<thead>
<tr>
<th>Subsystem</th>
<th>Mode (p)</th>
<th>Confidence level (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub 1</td>
<td>0.90</td>
<td>10</td>
</tr>
<tr>
<td>Sub 2</td>
<td>0.90</td>
<td>10</td>
</tr>
<tr>
<td>System</td>
<td>0.90</td>
<td>10</td>
</tr>
</tbody>
</table>

\[ M(\theta) = \theta_1 \theta_2 \]

Comparison among various system priors

The pooled prior “approaches” to the induced prior when \( \alpha \) increases.
Effect of Pooling Parameter $\alpha$ on the Posterior Inference

Data: $rbinom(n = 10,000, s = 100, p = 0.96)$

Observations:

- Posterior modes under both **Geometric pooling** and **linear pooling** decrease as $\alpha$ increases due to the higher weight on the induced prior.
- Posterior modes of **linear pooling** are slightly greater than those of **Geometric pooling**.

```latex
\text{Linear Pooling:}\quad q(\phi) = \alpha q_1^*(\phi) + (1 - \alpha)q_2(\phi)

\text{Geometric Pooling:}\quad q(\phi) = q_1^*(\phi)^{\alpha}q_2(\phi)^{1-\alpha}
```

$n$: the number of runs; $s$: sampling size; $p$: the assumed true reliability
### Computational Issues in Bayes

**When do we need simulation for posterior inference?**

**Which simulation technique to use?**

<table>
<thead>
<tr>
<th>Prior</th>
<th>Posterior</th>
<th>Computation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjugate/Semi-conjugate</td>
<td>Closed form solution</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>Non-conjugate</td>
<td>No closed form solution</td>
<td>Full conditionals exist</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No full conditionals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gibbs Sampler</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metropolis- Hastings</td>
</tr>
</tbody>
</table>

*A prior distribution is conjugate when the posterior is in the same family as the prior distribution.*
## Conjugate Priors

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Model parameters</th>
<th>Prior</th>
<th>Posterior [closed form sol.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>$p$ probability</td>
<td>$\text{beta}(\alpha, \beta)$</td>
<td>$\text{beta}(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i)$</td>
</tr>
<tr>
<td>Binomial</td>
<td>$p$ probability</td>
<td>$\text{beta}(\alpha, \beta)$</td>
<td>$\text{beta}(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i)$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$\lambda$ rate</td>
<td>$\Gamma(\alpha, \beta)$</td>
<td>$\Gamma(\alpha + \sum_{i=1}^{n} x_i, \beta + n)$</td>
</tr>
<tr>
<td>Normal</td>
<td>$\mu$ mean</td>
<td>$N(\mu_0, \sigma_0^2)$</td>
<td>$N\left(\frac{\mu_0 + \sum_{i=1}^{n} x_i}{\sigma_0^2 + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\right)$</td>
</tr>
<tr>
<td></td>
<td>(with known $\sigma$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Examples of conjugate priors within the exponential family distributions.
Simulation on Bayesian Inference

• Challenge: how to draw samples from \( p(\theta | y) \)
  – No closed form posterior
  – High dimensional super-parameters

• MC on Bayesian Posterior Inference:
  – Sample \( \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(S)} \) from \( p(\theta | y) \)
  – Approximate posterior summaries of interest through these samples (e.g. the posterior mean and variance).
Numerical Simulation Methods

- **Monte Carlo (MC) methods**: estimate the expected value based on sample mean

  \[ I = \int_{a}^{b} h(y)dy = (b - a) \int_{a}^{b} h(y) \frac{1}{b - a} dy = (b - a)E(h(y)) \]

  - **Importance Sampling**: \( f(y) \) is difficult to sample from; \( g(y) \) is the proposal distribution

    \[ I = \int_{a}^{b} h(y)f(y)dy = \int_{a}^{b} h(y) \frac{f(y)}{g(y)} g(y)dy \]

- **MCMC**: (generate a Markov chain whose limiting distribution is equal to the desired distribution.)

  - Gibbs sampler
  - Metropolis- Hastings
Importance Sampling on Bayesian

\[ p(\theta | y) = \frac{p(y|\theta) p(\theta)}{\int p(y|\tilde{\theta}) p(\tilde{\theta}) d\tilde{\theta}} \propto p(y|\theta) p(\theta) \]

- **Sampling Importance Re-sampling (SIR)**
- \( p(\theta) \) as the proposal distribution
  - Sample \( \theta_j, j = 1, ..., k \), from \( p(\theta) \)
  - Calculate the normalized weights
    \[ w_j = \frac{L(\theta_j)}{\sum_{j=1}^{k} L(\theta_j)}, L(\theta_j) = \prod_{i=1}^{n} p(y_i|\theta_j) \]
  - Sample \( \theta_j^* \) from \( \theta_j, j = 1, ..., k \), with replacement using the weight \( w_j \) as sampling probabilities.

Rubin (1987, 1988)
Adaptive SIR for Better Dealing with Extreme Weights

- Sample $\theta_j, j = 1, \ldots, k$, from $p(\theta)$
- Calculate the normalized weights

$$w_j = \frac{L(\theta_j)}{\sum_{j=1}^{k} L(\theta_j)}, L(\theta_j) = \prod_{i=1}^{n} p(y_i|\theta_j)$$

- Reject extreme weights
  1. Select $M$ nodes from $w_j$ randomly and sort the selected nodes descendingly;
  2. Track the cumulative sum of the sorted nodes, and assign the threshold weight $t_w = \sum_{i=1}^{\lfloor \alpha \times M \rfloor} w_j$, reject $w_{\text{reject}}$, if $w_{\text{reject}} > t_w$, where $\alpha$ is denoted as the threshold percentage, $\lfloor p \times M \rfloor$ stands for the floor value of $p \times M$;
  3. Split the rejected weight into $k = \left\lfloor \frac{w_{\text{reject}}}{t_w} + 1 \right\rfloor$ smaller weighs which are $w_{\text{reject}} \div k$;
  4. Repeat 2-4 until the pre-set requirement is reached, e.g. a pre-determined variance of weights.

Ref: Yuan, C., and Druzdzel M. (2009) Improving Importance Sampling by Adaptive Split-Rejection Control in Bayesian Networks
MCMC - Gibbs sampler

• Assumption: full conditionals exists.
• Suppose we have a vector of parameters \( \phi = \{ \phi_1, ..., \phi_p \} \).
  The distribution about \( \phi \) is in the form of \( p(\phi) = p(\phi_1, ..., \phi_p) \).

Procedure:

• Give a starting point \( \phi^{(0)} = \{ \phi_1^{(0)}, \phi_2^{(0)}, ..., \phi_p^{(0)} \} \)

For \( s=1, ..., n \)

  – Sample \( \phi_1^{(s)} \sim p \left( \phi_1 \mid \phi_2^{(s-1)}, \phi_3^{(s-1)}, ..., \phi_p^{(s-1)} \right) \)
  – Sample \( \phi_2^{(s)} \sim p(\phi_2 \mid \phi_1^{(s)}, \phi_3^{(s-1)}, ..., \phi_p^{(s-1)}) \)
  – ...
  – Sample \( \phi_p^{(s)} \sim p(\phi_p \mid \phi_1^{(s)}, \phi_2^{(s)}, ..., \phi_{p-1}^{(s)}) \)

• \( \phi^{(s)} = \{ \phi_1^{(s)}, \phi_2^{(s)}, ..., \phi_p^{(s)} \} \)
Metropolis-Hasting

- Parameter of interest is $\theta$
- A proposal value $\theta^*$ nearby the current value $\theta^{(s)}$ is sampled from a symmetric proposal distribution $J(\theta^* | \theta^{(s)})$.

**Step 1** Sample $\theta^* \sim J(\theta | \theta^{(s)})$.

**Step 2** Compute the acceptance ratio

$$ r = \frac{p(\theta^* | y)}{p(\theta^{(s)} | y)} = \frac{p(y | \theta^*)p(\theta^*)}{p(y | \theta^{(s)})p(\theta^{(s)})} $$

**Step 3** Let

$$ \theta^{(s+1)} = \begin{cases} 
\theta^* & \text{with probability } \min(r, 1) \\
\theta^{(s)} & \text{with probability } 1 - \min(r, 1)
\end{cases} $$

- **Step 3** can be accomplished by sampling $u \sim \text{uniform}(0, 1)$ and setting $\theta^{(s+1)} = \theta^*$ if $u < r$ and setting $\theta^{(s+1)} = \theta^{(s)}$ otherwise. ($r$ is large enough to make sure the state transits to next state)
Bayesian Reliability Inference with Multilevel information
– An Example

Scenario 1 (S1)  Scenario 2 (S2)  Scenario 3 (S3)

System data + priors  Subsystem data+ priors  System & Subsystem data+ priors

Subsystem prior
↓ +System structure
Induced system prior
↓ +System prior
Pooled system prior
↓ Updated subsystem prior
Updated subsystem prior
↓ +System data
Updated system prior
↓ +System structure
Updated subsystem posterior
↓ +System data
Updated subsystem posterior
↓ +System structure
System/subsystem posterior
Update Subsystem Prior Using Bayesian Melding Method

\[
\tilde{q}_\phi(\phi) \propto q_1^*(\phi)^\alpha \times q_2(\phi)^{1-\alpha} \quad \phi = M(\theta)
\]

\[
\tilde{q}_\theta(\theta) \propto q_1(\theta) \times \left( \frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha}
\]

Update subsystem prior

\[
\tilde{q}_\theta(\theta) \propto q_1(\theta) \times \left( \frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha}
\]

Subsystem prior updating

\[\tilde{q}_\theta(C_3) = \tilde{q}_\phi(C_2)\]
\[q_1(C_3) = q_1^*(C_2)\]

\[\tilde{q}_\theta(C_1) = \tilde{q}_\theta(C_3) \left( \frac{q_1(C_1)}{q_1(C_3)} \right) = \tilde{q}_\phi(C_2) \left( \frac{q_1(C_1)}{q_1^*(C_2)} \right) \propto q_1^*(C_2)^\alpha q_2(C_2)^{1-\alpha} \left( \frac{q_1(C_1)}{q_1^*(C_2)} \right)\]

\[= q_1(C_1) \left( \frac{q_1(C_1)}{q_1^*(C_2)} \right)^{1-\alpha}\]

Posterior Inference for System Reliability

Scenario 1: when only system level data are available

Monte Carlo:

Scenario 2 & 3: Adaptive Sampling Important Re-sampling (SIR):
Resampling the initial samples by certain importance weights

\[ w_{i2} = \left( \frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha} L_1(\theta) \]
\[ w_{i3} = \left( \frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha} L_1(\theta)L_2(\theta) \]

\[ \tilde{\pi}_\theta(\theta) \propto q_1(\theta) \left( \frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha} L_1(\theta) \]
\[ \tilde{\pi}_\theta(\theta) \propto q_1(\theta) \left( \frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha} L_1(\theta)L_2(M(\theta)) \]

Posterior Based on Bayesian Melding

- Computation complexity of **MCMC** $O(n \log(n))$, while **SIR** $O(n)$
Mis-specified Priors From the True reliability

An example: Two-component series system

$$M(\theta) = \theta_1 \theta_2$$

The subsystem priors

$$q_1(\theta) \sim \prod_{i=1}^{2} beta(pe_i C_{ei} + 1, (1 - pe_i) C_{ei} + 1)$$

<table>
<thead>
<tr>
<th>Subsystem</th>
<th>Mode, $p_e$</th>
<th>Confidence level, $C_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub 1</td>
<td>0.90</td>
<td>10</td>
</tr>
<tr>
<td>Sub 2</td>
<td>0.90</td>
<td>10</td>
</tr>
<tr>
<td>System</td>
<td>0.90</td>
<td>10</td>
</tr>
</tbody>
</table>

Assuming the true reliability of 0.5
S1: Assess the System Posterior Reliability with System Data Available only

**Posterior modes mean and their 95% credible interval**

- Assuming true reliability of 0.5

**Observations:**

- Means of posterior modes are close to each other
- Approach to the assumed true reliability as sample size increases
S2: Assess the subsystem posterior reliability with Subsystem Data only

Posterior modes mean and their 95% credible interval - Assuming true reliability of 0.5

Observations:

• Means of posterior modes are close to each other
• The width of CI of S2 is narrower than that of the traditional Bayes
S3: Assess the subsystem posterior reliability with System Data and Subsystem Data

Posterior modes mean and their 95% credible interval

- Assuming true reliability of 0.5

Observations:

- Means of posterior modes are close to each other
- The mean of posterior modes from S3 is the closer to the assuming true reliability 0.5.
- The width of CI of S3 is narrower with larger small sizes.

Traditional Bayes: Subsystem prior and Subsystem Data

S3: Updated Subsystem Prior and System Data and Subsystem Data
Battery Health Prognostics Under Bayesian

Backgrounds:
- The battery capacity degrades over cycles with the loss of active Lithium Ions
- Batteries are considered as failed when their capacity reaches the threshold capacity, e.g., 80% of rating capacity
- Four $LiCoO_2$ batteries are tested under constant current constant voltage (CCCV)

Methodology: Mixed effects model is used to quantify the variations of unit-to-unit and within unit.

$$y_j = \alpha_j + \beta_j x_j(t) + \varepsilon_j$$

Research Scheme

Battery Capacity Fade Modeling and Prognostics

Step 1: Covariates Selection

- Physics-based
- Data-driven
- Hybrid: Physics-based and Data-driven

Step 2: Model Selection

- Fixed Effect Models
- Random Effect Models
- Bayesian Information Criteria

Step 3: Prognostics

- One-step-ahead Prediction
  - Moving Windows Data Input
  - Cumulative Data Input
- Cycle to Failure Distribution
Mixed Effects Models

\[
M2: y_{j[i]} = \alpha_{j[i]} + \beta_{1j[i]}\sqrt{t_{j[i]}} + \varepsilon_j
\]

\[
M4: y_{j[i]} = \alpha_{j[i]} + \beta_{j[i]}\log t_{j[i]} + \varepsilon_j
\]

Performance:
Both M2 and M4 overestimate the battery capacity in the late stage
Mixed Effects Models

![Graph showing mixed effects models for M5 & M7](image)

**M5:** \( y_{j[i]} = \alpha + \beta_1 \sqrt{t_{j[i]}} + \beta_2 \log t_{j[i]} + \beta_3 \sqrt{t_{j[i]}} \times \log t_{j[i]} + \varepsilon_j \)

**Selected model M7:** \( y_{j[i]} = \alpha_{j[i]} + \beta_1_{j[i]} \sqrt{t_{j[i]}} + \beta_2_{j[i]} \log t_{j[i]} + \beta_3_{j[i]} \sqrt{t_{j[i]} \times \log t_{j[i]}} + \varepsilon_j \)

**Performance:**
Residuals of M7 of random effects are smaller than M5 of fixed effects
Conclusions

- Bayesian inference can well integrate prior knowledge with data
- Multiple prior integration can be challenging
- When only limited data are available, prior misspecification can be misleading for posterior inference
- Knowledge updating is one advantage of Bayesian inference for PHM
Questions & Comments

Thank you.