Hierarchical Bayesian Reliability Analysis of Complex Dynamical Systems

GABRIELA TONTĂ, LUIGE VLĂDĂREANU, MIHAI STELIAN MUNTEANU, DAN GEORGE TONTĂ

1Department of Electrical Engineering, Measurements and Electric Power Use, Faculty of Electrical Engineering and Information Technology
University of Oradea
Universității st., no. 1, zip code 410087, Oradea, ROMANIA, gtont@uoradea.ro, dtont@uoradea.ro, http://www.uoradea.ro
2Institute of Solid Mechanics of Romanian Academy
C-tin Mile 15, Bucharest 1, 010141, ROMANIA, luigiv@arexim.ro, http://www.acad.ro
3Faculty of Electrical Engineering, Technical University Cluj Napoca,
Constantin Daicoviciu st. no 15, 400020 Cluj - Napoca, ROMANIA, mihai.munteanu@et.utcluj.ro, http://www.utcluj.ro/

Abstract: - The Bayesian methods provide additional information about the meaningful parameters in a statistical analysis obtained by combining the prior and sampling distributions to form the posterior distribution of the parameters. The desired inferences are obtained from this joint posterior. An estimation strategy for hierarchical models, where the resulting joint distribution of the associated model parameters cannot be evaluated analytically, is to use sampling algorithms, known as Markov Chain Monte Carlo (MCMC) methods, from which approximate solutions can be obtained. Both serial and parallel configurations of subcomponents are permitted. Components of the system are assumed to be linked through a reliability block diagram and the manner of failure data collected at the component or subcomponent level can be included into the posterior distribution permit the extension of failure information across similar subcomponents within the same or related systems. An effective and flexible event-based model for assessing the reliability of complex systems including multiple components that illustrates the Bayesian approach is presented.

Key-Words: - Bayesian methods, hierarchical models, complex system, configurations.

1 Introduction
Global modeling of systems' reliability leads to adoption of a repartition law for operating repair time of a high generality, obtained by combination or successions of exponential repairs. Complete specification of the model involves specifying the distribution parameters, namely the parameter $\lambda$ of the exponential distribution. Looking closely estimating parameter exponential distribution was shown that achieving a reasonable accuracy of the estimate assumes a high volume of experimental results difficult to achieve in the current applications in which both sample test subject and duration are of limited economic considerations. One way of increasing the precision of estimation through the enrichment of the experimental material without increasing the volume is using the accelerated requests test. Another way of raising the estimates’ precision is based on the idea that, before conducting a reliability test on a system, there are some information on the system’ reliability, information which, if are not neglected, would contribute to more accurate characterization of the system. The Bayesian methods are assessment methods that take into account the available information on the reliability of a system, whether it is or not of experimental nature. Especially appropriate where the volume of experimental results is low, the Bayesian estimation develops in the context of reliability theory, which is a fertile field for interpretation and suggestions.

2 Literature review
These methods occupy an important place in the modern statistical mathematics, which is especially appropriate where the volume of experimental results is low. The fact made the Bayesian
estimation to develop especially in the context of reliability theory which is a fertile field for interpretation and suggestions for developing Bayesian statistics. To provide context, it is useful to begin with a review of related research in Bayesian system reliability. Most relevant model considered here are the papers by Martz, Waller and Fickas [1] and Martz and Waller [2], where complex systems, comprised of series and parallel subcomponents, were modeled using beta priors and binomial likelihoods at component, subsystem and system levels. Within this framework, an induced higher-level prior was obtained by propagating lower-level posteriors up through the system fault diagram, and combining these posteriors with “native” higher-level priors to obtain an induced prior at the next system level. These “induced” priors were approximated by beta distributions using a methods-of-moments type procedure. The combination of native priors and posterior distributions obtained from lower-level system data, both of which were expressed as beta distributions, was accomplished by expressing the resulting induced priors as a beta distributions with parameters representing a weighted average of the constituent beta densities. This process was propagated through higher and higher system levels until an approximation to the joint posterior distribution on the total system reliability was obtained. Many reliability models do not consider prior expert opinion and data at multiple system levels. Springer and Thompson [3, 4], and Tang [5] provide exact or approximated system reliability distributions obtained by propagating the component posteriors through the system structure. Winterbottom [6] use approximations for exponential lifetimes rather than binomial data. Others propose methods for evaluating or bounding moments of the system reliability posterior distribution Soman and Misra [7]. These moments can also be used in the beta approximations employed by Martz, Waller and Fickas [1], Soman and Misra [7]. These “induced” priors were approximated by beta distributions using a methods-of-moments type procedure. The combination of native priors and posterior distributions obtained from lower-level system data, both of which were expressed as beta distributions, was accomplished by expressing the resulting induced priors as a beta distributions with parameters representing a weighted average of the constituent beta densities. This process was propagated through higher and higher system levels until an approximation to the joint posterior distribution on the total system reliability was obtained. Many reliability models do not consider prior expert opinion and data at multiple system levels. Springer and Thompson [3, 4], and Tang [5] provide exact or approximated system reliability distributions obtained by propagating the component posteriors through the system structure. Winterbottom [6] use approximations for exponential lifetimes rather than binomial data. Others propose methods for evaluating or bounding moments of the system reliability posterior distribution Soman and Misra [7]. These moments can also be used in the beta approximations employed by Martz, Waller and Fickas [1], Soman and Misra [7].

3 Bayesian analysis framework

As in the classical or frequent approach to inference, the Bayesian approach is developed in the presence of observations \( x \) whose value is initially uncertain and described through a probability distribution with density or probability function \( f(x|\theta) \). The quantity \( \theta \) serves as an index of the family constants. The likelihood function is:

\[
\ln(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{x_i - \theta}{\sigma} \right)^2 \right\} \alpha \exp \left\{ -\frac{n}{2\sigma^2} \left( \bar{x} - \theta \right)^2 \right\}
\] (1)

where \( \bar{x} \) is the arithmetic average of the \( x_i \). Therefore, the posterior density is:

\[
\pi(\theta) \propto \exp \left\{ -\frac{1}{2} \left( \frac{\bar{x} - \theta}{\sigma} \right)^2 \right\} \exp \left\{ -\frac{1}{2} \left( \frac{\theta - \mu}{\tau^2} \right)^2 \right\}
\]

\[
\pi(\theta) \propto \exp \left\{ -\frac{1}{2} \left( \frac{\theta - \mu}{\tau^2} \right)^2 \right\}
\]

where:

\[
\tau^2 = n\sigma^2 + \tau^2 \quad \text{and} \quad \mu = \tau^2 \left( n\sigma^2 \bar{x} + \sigma^2 \mu \right).
\]

The last passage is obtained using that if \( x, a_1, a_2, b_1 \) and \( b_2 \) are scalars then:

\[
\frac{1}{b_1} + \frac{1}{b_2} = \frac{1}{b_1 + b_2}
\]

where \( b_1^{-1} = b_1 + b_2 \) and \( a_1^{-1} + a_2^{-1} \) with \( z = \theta \), \( a_1 = \bar{x} \), \( a_2 = \mu \), \( b_1 = \sigma^2 / n \) and \( b_2 = \tau^2 \). Incorporating the multiplicative term that does not depend on \( \theta \) to the proportionality constant gives:

\[
\pi(\theta) \propto \exp \left\{ -\frac{1}{2} \left( \frac{\theta - \mu}{\tau^2} \right)^2 \right\}
\]

(4)

In other words, the posterior distribution of \( \theta \) is:

\[\pi(\theta) = N\left(\mu + \tau^2 \bar{x}\right)\]

We may note that by increasing the value of \( \tau^2 \), the information contained in the prior is reduced and so is its influence on the analysis. In the limit when \( \tau^2 \rightarrow \infty \) the non-informative prior \( p(\theta) \alpha k \) is obtained and \( \pi(\theta) = N\left(\bar{x}, \sigma^2 / n\right) \). There is plenty of controversy among Bayesians about the specification of non-informative prior distributions.
disagreement is due to an inherent anomaly of these distributions. Often, this specification leads to improper distributions. These are distributions that do not integrate to 1 as prescribed by the theory of probability, when \( \tau^2 \to \infty \), \( \int \rho(\theta) \theta \neq 1 \). There are many different definitions of non-informative prior distributions, especially in multivariate cases. One of the most commonly accepted definitions is Jeffreys’ prior, given by:

\[
p(\theta) \propto \left| I(\theta) \right|^{1/2}
\]

(5)

where \( I(\theta) = \mathbb{E} \left[ -\frac{\partial^2 \log f(x|\theta)}{\partial \theta \partial \theta} \right] \) is the expected Fisher information.

A formalized a theory of Bayesian inference mostly using this prior and justified it on the grounds of invariance under parametric transformations. In general, Bayes’ theorem it leads to prior densities in the form \( p(\theta) \propto k \) for location parameters \( \theta \), \( p(\sigma) \propto \sigma \) for scale parameters \( \sigma \). When a location \( \theta \) and scale \( \sigma \) are present, it is based on expected discrepancy measures of information and under asymptotic normality coincides with Jeffreys’ prior in the univariate case. In the multidimensional case, the reference approach works on splitting the parameter vector into groups and seems to avoid some difficulties of other approaches in the multiparameter case, though at the cost of a more complex derivation of the prior. The impropriety of some vague prior specifications is a nuisance but in general they lead to proper posterior distributions and inference can be made without any difficulty. There are exceptions and in some cases the posterior remains improper. This is a serious problem as in many complex models verification of propriety is far from trivial. For these models, exact inference cannot be performed and the approximations used may lead to a number of inconsistencies. Another important element for Bayesian inference is the predictive or marginal distribution of \( x \) with density \( f(x) \) given by (1). It provides the expected distribution for the observation \( x \) as \( f(x) = \mathbb{E}[f(x|\theta)] \) and the expectation is taken with respect to the prior distribution of \( \theta \). A similar derivation can be applied to the prediction of a future observation \( y \) after observing \( x \). This prediction should be based on the distribution of \( y|x \), that is, on the updated probabilistic description based on the available information. If \( y \) and \( x \) are conditionally independent given \( \theta \) then

\[
f(y|x) = \int f(y|x,\theta) \rho(\theta) \theta = \int f(y|x|\theta) \rho(\theta) \theta
\]

(6)

and again the density is obtained as the expectation of the sampling distribution but this time with respect to the posterior of \( \theta \). Conditional independence between \( x \) and \( y \) is obtained, for example, if \( x = (x_1, \ldots, x_n) \) and \( y = (x_{n+1}, \ldots, x_{n+m}) \) are samples from \( f(x|\theta) \). The predictive distribution is then used to predict future values of this population. Predictive distributions form the basis of the predictive approach to inference. This approach is described, detailed and applied to a variety of problems by Aitchinson and Dunsmore [10]. The main thrust of their argument is that the ultimate test of any inferential procedure is the confrontation against reality.

4. Hierarchical Model

The normal regression model was specified in the previous section with the aim of establishing relations between the response variable \( y \) and a set of explanatory variables \( x_1, \ldots, x_p \) through regression coefficients gathered in the vector \( \beta \). Many times, the problem is structured in a way that qualitative probabilistic statements about \( \beta \) can and should be incorporated into the model. Consider observations \( y_{ij} = N(\beta_i, \sigma^2) \), \( j=1, \ldots, n_i \), \( i=1, \ldots, d \), collected from \( d \) groups with different means \( \beta_i \) but the same dispersion. This model is a special case of a regression model with observation vector \( y=(y_{11}, \ldots, y_{1n}, \ldots, y_{dn}) \) and design matrix \( X = \text{diag}(1_{n_1}, \ldots, 1_{n_d}) \) where \( 1_{n_i} \) is the \( n \)-dimensional vector of \( 1 \). The model is completed with a prior distribution for \( (\beta, \sigma^2) \). One possibility is to assume prior independence between the means \( \beta_i \), \( i=1, \ldots, d \). If the \( d \) groups are similar in some sense, a plausible alternative is to assume that the means are a sample from a population of means. This population may be hypothetical and, to fix ideas, is assumed here to be homogeneous. Assuming a normal population, \( \beta_1, \ldots, \beta_d \) is a sample from a \( N(\mu, \tau^2) \) where \( \mu \) is the mean and \( \tau^2 \) measures the dispersion of the population of means. The model is completed with a prior distribution for \( (\mu, \tau^2) \). The complete prior specification is:

1st level: \( \beta|\mu, \tau^2 \sim N(1_{d1}, \tau^2 1_{d1}) \)

(7)

2nd level: \( \mu \sim N(h_0, B_0) \)

\( \sigma^2 = F_\sigma \) and \( \tau^2 = F_\tau \)

for independent probability distributions \( F_\sigma \) and
The prior density for the model parameters $(\beta, \mu, \sigma^2, \tau^2)$ is:

$$p(\beta, \mu, \sigma^2, \tau^2) = p(\beta|\mu, \sigma^2) p(\mu) p(\sigma^2) p(\tau^2)$$

We note that the prior in the example was specified in two stages. The (two-stage) model of the example can be generalized in many ways. A generalization towards a normal regression model is given by:

$$\beta_2 = N(b, B)$$

The design matrix with the covariates for the response vector $y$ and the regression coefficient were respectively renamed to $X_1$ and $\beta_1$. This is due to the presence of another design matrix containing a further set of explanatory variables $X_2$ and another regression coefficient $\beta_2$. This matrix contains the values that explain the variations of the values of $\beta_1$ and $\beta_2$ contains the coefficients of this explanation.

Depending on the problem, more stages may be required for an appropriate description of the model. Its form may remain unchanged with additional equations in the form:

$$\beta_j | \beta_{j+1} = N(X_{j+1} \beta_{j+1}, C_j)$$

In general, the higher the stage, the harder is the specification of the distributions. Rarely, models have more than three stages and it is very common that the prior at the higher stage is set to be non-informative. The matrices $C$ and $B$ are being assumed to be known. This assumption is not reasonable in general and a modification sometimes suggested is the substitution of $C$ and $B$ by $\phi^{-1} C$ and $\phi^{-1} B$ respectively. The variances $C$ and $B$ will then measure prior dispersion relatively to the likelihood dispersion. This dependence on $X$ allows the use of the results about conjugacy. The analysis still remains conditional on the (assumed known) values of the matrices $C$ and $B$. The derivations below concentrate on the two-stage model to simplify the notation even though there is no technical problem in the extension to the $k$-stage models, $k > 2$. The first point to mention is that the structure imposed upon the joint distribution of all the variables in the problem, that is $(y, \beta_1, \beta_2, \phi)$, may be written as follows:

$$p(y, \beta_1, \beta_2, \phi) = p(y|\beta_1, \phi)p(\beta_1)p(\beta_2)p(\phi)$$

The hierarchical character of the model then becomes clear with the successive conditional specifications. Unfortunately, the analysis is not analytically tractable and it is not possible to obtain the marginal posterior distributions of $\beta_1$ and $\phi$. The analysis conditional on knowledge of $\beta_2$ is not new and was performed in the previous section. If $\beta_2$ is known, the prior does not depend on the probabilistic specification of $\beta_2$. Replacement of $b_0$ by $X_2 \beta_2$ in the regression model is necessary. Hence, the full conditional posterior for $\beta_1$ is $N(b_0, B)$ and for $\phi$ is $G(n/2, n1 S_1 / 2)$ where:

$$b_0 = B_1^{-1} X_1 \beta_1 + \phi X_1 y, \quad \phi = C^{-1} + \phi X_1. \quad n_1 = n + n_0$$

In many cases, they are also conditionally conjugate. The hierarchical models allow easy derivation of full conditional posterior distributions not only the normal distribution. In many cases, they are also conditionally conjugate. The hierarchical structures of model that passes through $\beta_1$ all information provided by $y$ to $\beta_2$. More formally, $y$ and $\beta_2$ are conditionally independent given $\beta_1$. This model building strategy based on conditional independence allows easy derivation of full conditional distributions. In many cases, they are also conditionally conjugate. The hierarchical models can also be defined for other observational distributions not only the normal distribution. Extended for observations $y_i$ with density $f(y|\beta_i, \phi)$. The $\beta_i$ are a sample from a population with density $p(\beta|\lambda)$ and this constitutes the first stage of the prior distribution. Again, the model is completed with a second stage prior for $\lambda$ and $\phi$. These models are conditionally independent hierarchical models. A case of particular interest is
when the $\beta_s$ are conditionally conjugate given the values of $\phi$ and $\lambda$. In the context of exponential family distributions, $y_{ij} \approx EF(\mu_i)$, the $\theta_i$ form a sample from a $CP(\alpha, \beta)$ distribution and the model is completed with a prior for $\alpha$ and $\beta$. Then, the full conditional posterior of the $\theta_i$ is still given by a product of independent $CP(\alpha^*, \beta^*)$ distributions. A more general extension of the hierarchical model is considered the two-stage regression model and generalized linear models can be combined to give a generalized linear hierarchical model:

$$y_{ij} \mid \mu_i \approx EF(\mu_i), \ i=1,...,n$$

$$\eta = X_1 \beta_1$$

$$\beta_1 \mid \beta_2 \approx N(X_2 \beta_2, C)$$

$$\beta_2 \approx N(b, B)$$

where $\eta = (\eta_1, \eta_2)$ and $\eta_i = g(\mu_i), i=1,...,\eta$.

When the EF distribution considered is the normal, the above model is obtained. When the regression structure simply classifies observations by groups, the normal prior model for transformed means suggested above. Some basic aspects of the dynamic linear models with time-varying parameters, adequate to the modeling of time series and regression are defined by a pair of equations, called the observation equation and the evolution or system equation, respectively given by:

$$y_t = F_t \beta_t + \epsilon_t, \ \epsilon_t \approx N(0, \sigma^2_t)$$

$$\beta_t = G_t \beta_{t-1} + \omega_t, \ \omega_t \approx N(0, W_t)$$

where $\{y_t\}$ is a sequence of observations through time, conditionally independent given $\beta_t$ and $\sigma^2_t$.

$F_t$ is a vector of explanatory variables as in the previous sections, $\beta_t$ is a $d$-dimensional vector of regression coefficients or state parameters at time $t$ and $G_t$ is a matrix describing the parametric evolution. The errors $\epsilon_t$ and $\omega_t$ are mutually independent and $\sigma^2_t$ and $W_t$ are the error variances respectively associated to the univariate observation and the $d$-dimensional vector of parameters. The model is completed with a prior $\beta_0 \approx N(\alpha, R)$. Dynamic linear models provide another nice example of specification of a prior for a highly dimensional parameter by combination of qualitative and quantitative information. The system equation provides qualitative information about the relation between successive values of the state parameters. Quantitative information is provided by the prior distributions of $\beta_t$ and evolution errors $\omega_t$. The complete expression of the prior distribution results from the combination of these sources of information. Dynamic regression models are defined by $G_t = I_d, \ \forall t$. If, in addition, $W_t = 0, \ \forall t$, the static regression model is obtained. This is equivalent to setting the regression coefficients $\beta_t$ fixed in time. The simplest time series model is the first order model and is given by equations (16), (17):

$$y_t = \beta_t + \epsilon_t, \ \epsilon_t \approx N(0, \sigma^2_t)$$

$$\beta_t = \beta_{t-1} + \omega_t, \ \omega_t \approx N(0, W_t)$$

and $\beta_t$ is scalar. The model can be thought of as a first order Taylor series approximation of a smooth function representing the time trend of the series. This model is useful for stock control, production planning and data analysis. Observational and system variances may evolve in time, offering great scope for modeling the variability of the system. The linear growth model is more elaborate by incorporation of an extra time-varying parameter $\beta_2$ representing the growth of the level of the series. The model becomes:

$$y_{1,t} = \beta_{1,t} + \epsilon_t, \ \epsilon_t \approx N(0, \sigma^2_t)$$

$$\beta_{1,t} = \beta_{1,t-1} + \beta_{2,t} + \omega_{1,t}$$

$$\beta_{2,t} = \beta_{2,t-1} + \omega_{2,t}$$

$$\omega_t = (\omega_{1,t}, \omega_{2,t}) \approx N(0, W_t)$$

This model can be written in the form (20) with $F_t = (1, 0)$ and:

$$G_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \forall t$$

The choice of $F_t$ and $G_t$ depends on the model and the nature of the series one wishes to describe. Complete specification of the model requires full description of the variances of and $W_t$. In general they are assumed to be constant in time with of typically larger than the entries of $W_t$ in applications. Exact inference is only possible when $W_t$ is replaced by of $W_t$. The matrix $W_t$ becomes a matrix of weights relative to the observational variance. Typically it is unknown and must be estimated making the analytical treatment impossible. Nevertheless, consider initially that $\sigma^2_t$ and $W_t$ are known and let $y_t = \{y_t, y^{t-1}\}$ with $y_0$ describing the initial information available, including the values of $F_t$ and $G_t, \forall t$ also assumed known here. Observations are independent conditionally on the state parameters. This structure
matches well with the Bayesian approach by the accommodation of sequential procedures and subjective specifications. A detailed hierarchical model showing the dependencies of each variable on others is presented in figure 1 by the means of Directed Acyclic Graph (DAG).

![Fig. 1 Directed Acyclic Graph (DAG) for a hierarchical model](image)

5 Case study

The method described can be applied to the structure of the hierarchical model exemplified by fig. 2.

![Fig. 2 Block reliability diagram for a hierarchical model](image)

Applying the model we obtained the reliability posterior distributions for each of the components and the precision parameter. The system histogram of reliability posterior distributions with the system data included and system data excluded are shown in Figure 3. We note the agreement between the two posterior distributions:

![Fig. 3 Histogram of reliability posterior distributions](image)

6 Conclusion

From the analysis we ultimately concluded that the model captured well some of the main aspects of the problem. Other outstanding issues include the development of diagnostics to assess the adequacy of the system diagram in describing the functioning of the system, and the introduction of models for dependencies between subcomponents within subsystems.

References:


