Bayesian Modeling of Population Variability: Practical Guidance and Pitfalls

PSAM-9
May 2008

Dana Kelly
Idaho National Laboratory
Dana.Kelly@inl.gov

Corwin Atwood
Statwood Consulting
cory@statwoodconsulting.com

Idaho National Laboratory

Outline

- Overview of hierarchical Bayes for population variability
- Convergence problems
  - Diagnosing problems
  - Reparameterizing to avoid problems
- Sensitivity to choice of first-stage prior
  - Problems with conjugate priors when variability is large
  - Use of nonconjugate first-stage prior
  - Choosing hyperpriors
- Conclusions

Modeling Population Variability via Hierarchical Bayes

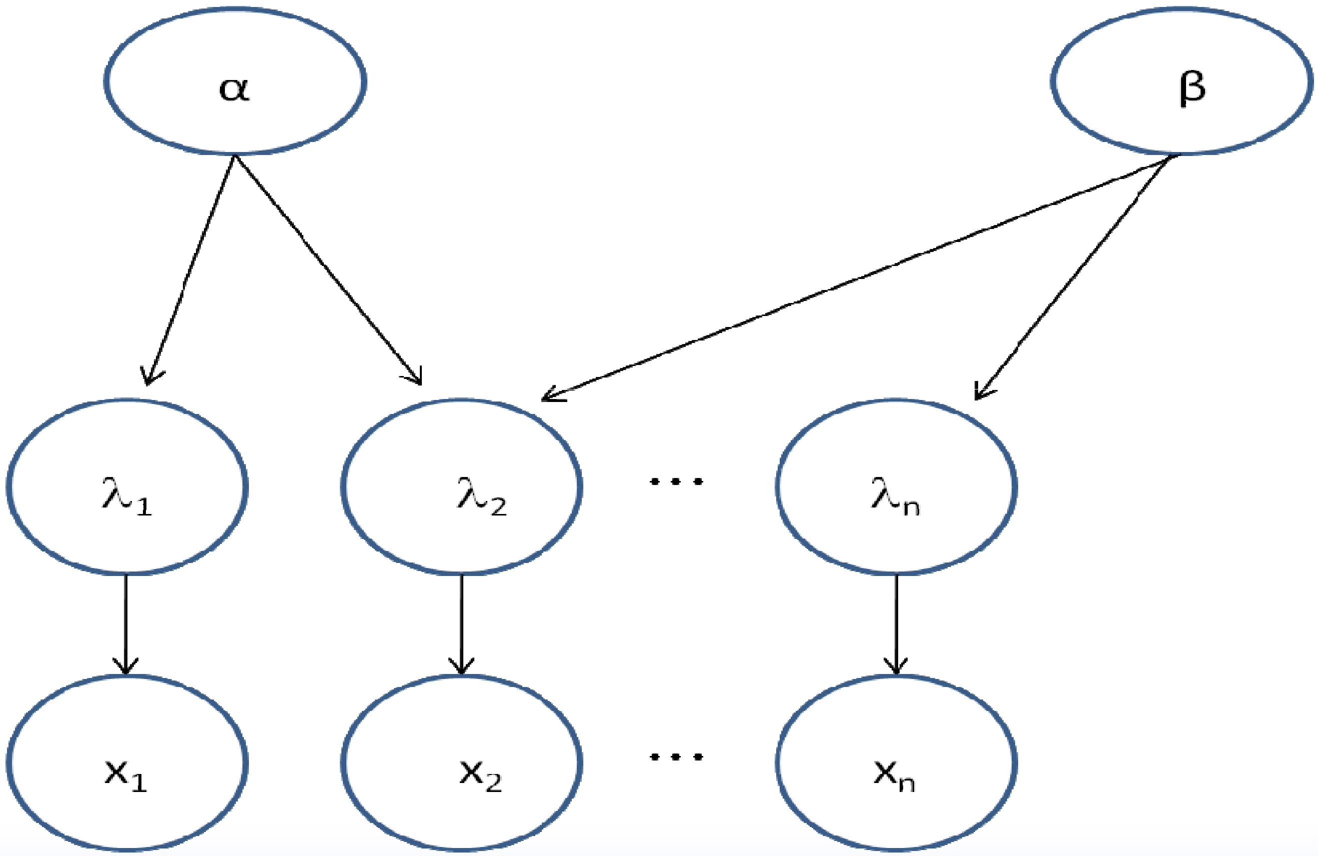
- *Want to use information from more than one source to estimate parameters, such as p or λ*
- *It may be possible that we cannot pool information as estimates from disparate sources might differ significantly*
- *Use hierarchical Bayes analysis to develop population variability curve (PVC)*
  - *Represents source-to-source variability in parameters of interest*
  - *Uses hierarchical prior, specified typically in two stages*

Hierarchical Priors

- *Bayesian approach is to specify prior in stages (hierarchies)*
  - *First stage is gamma(α, β) prior for λ_i (or other functional form)*
  - *Second stage is joint prior $\pi(\alpha, \beta)$*
    - *Called hyperprior*
    - *α, β called hyperparameters*
    - *Often use diffuse (noninformative) independent priors for hyperparameters*
  - *Two stages typical, but can model three or more*

$$\underbrace{\pi(\lambda)}_{\text{Overall prior}} = \iint \underbrace{\pi(\lambda \mid \alpha, \beta)}_{\text{First-stage prior}} \underbrace{\pi(\alpha, \beta)}_{\text{Second-stage prior (hyperprior)}} d\alpha d\beta$$

Bayesian Network Formulation of Problem



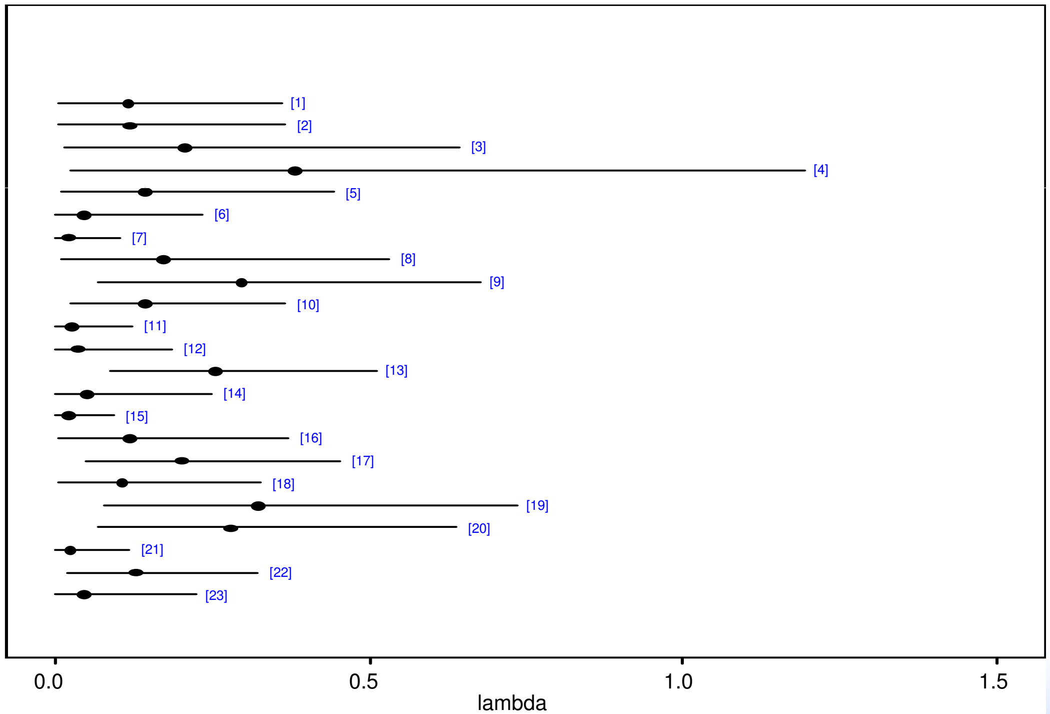
First Example: Loss of Offsite AC Power

- *Data taken from NUREG/CR-5496*

Events	Exposure time (yr)	Events	Exposure Time (yr)
1	13.054	5	21.5
1	12.77	0	10.075
1	7.22	0	26.32
1	3.944	1	12.54
1	10.548	3	17.5
0	10.704	1	14.3
0	24	3	10.89
1	8.76	3	12.5
3	11.79	0	21.38
2	17.5	2	19.65
0	20.03	0	11.34
0	13.39		

Side-by-Side Interval Plot Illustrates Plant-to-Plant Variability

- *95% credible intervals from update of Jeffreys prior for each plant*

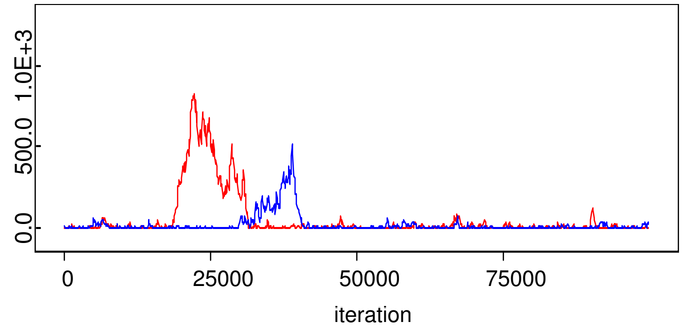


Hierarchical Bayes Model for LOSP Data

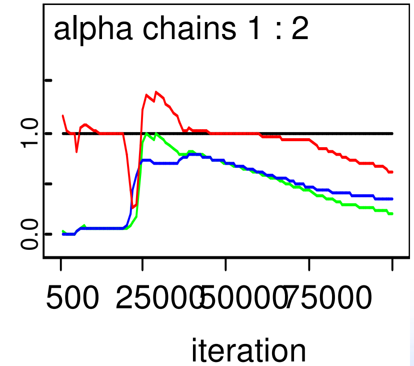
- *Will use gamma first-stage prior*
- *Independent diffuse hyperpriors on first-stage gamma parameters*
- *Will run two MCMC chains*
  - *Initial values selected by finding empirical Bayes estimates of gamma parameters*
    - *Starting values dispersed around EB estimates to obtain good coverage of joint posterior distribution*

Illustration of Convergence Problems

- *Plot of first 100,000 iterations shows poor mixing of chains*



- *Brooks-Gelman-Rubin (BGR) convergence diagnostic confirms lack of convergence*
  - *Red line should be near 1.0*
  - *Blue/green lines not stable*



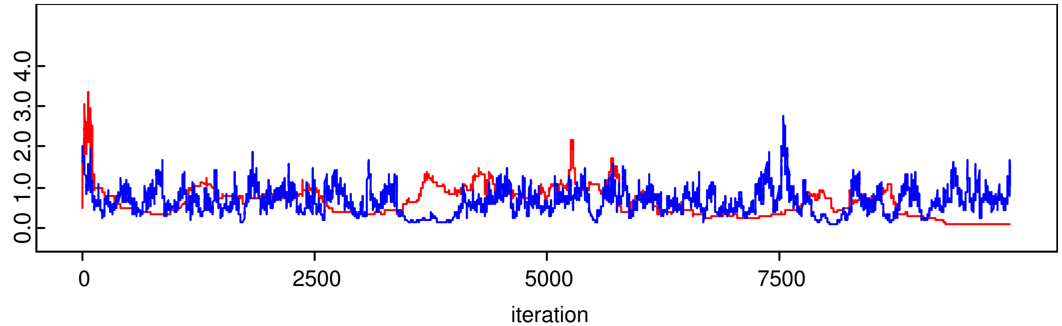
Convergence Problems Can Arise from Highly Correlated Parameters

- *Rank correlation coefficient for gamma parameters is 0.98*
- *Reparameterize gamma first-stage prior in terms of "independent" parameters*
  - *Use mean = α/β and coefficient of variation = std.dev./mean = $\alpha^{-0.5}$*
  - *Use independent diffuse hyperpriors on mean and CV*

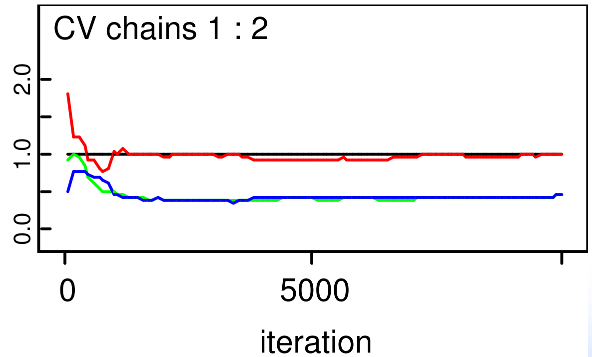
Idaho National Laboratory

Convergence Results with Reparameterized Model

- *History for first 10,000 iterations shows chains well mixed*

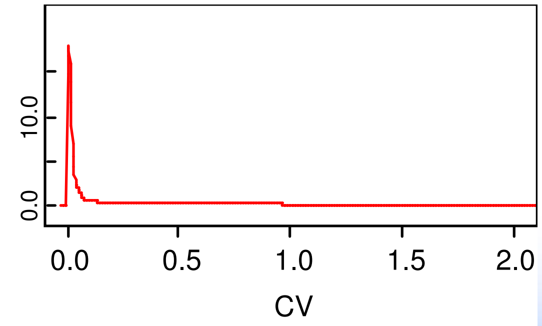


- *BGR diagnostic shows no problems*



Results for Reparameterized Model

- *Mean is 0.09/yr*
- *90% credible interval is (0.02, 0.20)*
- *Numerically close to EB results*
  - *Expected as variability is not too large*
  - *Illustrated by marginal posterior distribution for CV, which is peaked at small values*



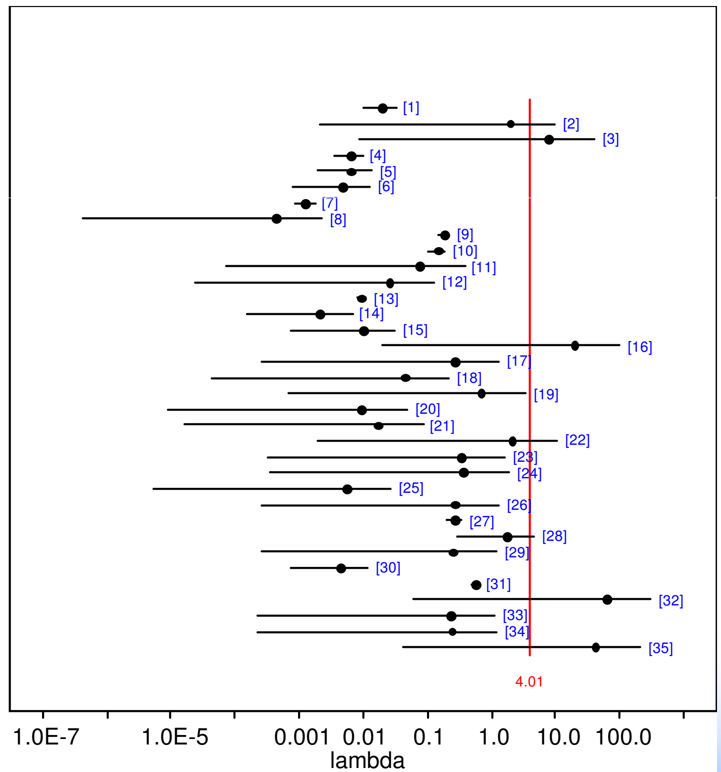
Sensitivity to Choice of First-Stage Prior

- *Re-analyze first example with lognormal first-stage prior*
  - *Use independent diffuse hyperpriors on lognormal parameters*
- *Mean is 0.10/yr*
- *90% credible interval is (0.02, 0.24)*
- *Little sensitivity to choice of first-stage prior for this example*
  - *Expected as variability is not too large*

Second Example: Digital I&C Failure Data

- *35 data sources, assumed to be Poisson-distributed*
- *Side-by-side interval plot illustrates extreme variability in Poisson rate*

Data taken from Yue, Meng and Chu, Tsong-Lun. *Estimation of Failure Rates of Digital Components Using a Hierarchical Bayesian Method.* New Orleans : 2006. International Conference on Probabilistic Safety Assessment and Management.

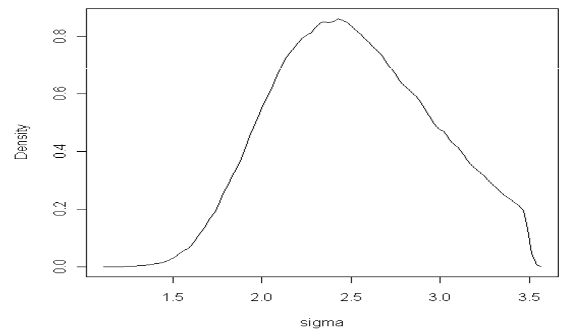


Results with Gamma First-Stage Prior

- *Mean is 0.09/yr*
  - *EB mean is 0.07/yr*
  - *Median is 0.01/yr*
- *90% credible interval is (6.7E-8, 0.4)*
- *Posterior mean of α is 0.24*
  - *EB estimates α = 0.24*
- *Conjugate first-stage prior can only capture large variability by having small value of α*
  - *Gives vertical asymptote at 0*
    - *Unrealistically small lower percentiles*

Lognormal First-Stage Prior

- *Lognormal density goes to 0 at 0*
  - *No vertical asymptote*
- *Must avoid overly restrictive hyperpriors, especially on* σ
  - *Data-based unif(1, 3.5) hyperprior causes truncation of upper tail of posterior density for* σ
  - *Leads to low estimate of mean*



  - *Mean depends strongly on* σ
  - *Used flat hyperprior on* μ *and uniform(0, 5) hyperprior on* σ
    - σ *= 1.4 corresponds to error factor of 10*

Idaho National Laboratory

Results with Lognormal First-Stage Prior

- *Mean is 1.1 /yr*
  - *Median is 0.007/yr*
- *90% credible interval is (6.3E-5, 0.55)*
- *Recall results with gamma first-stage prior:*
  - *Mean = 0.09/yr, median = 0.01/yr*
  - *90% interval (6.65E-8, 0.43)*
- *Mean is not robust, median and 95% value relatively robust*

Conclusions

- *Convergence can be an issue for hierarchical Bayes*
  - *May need to reparameterize to accelerate convergence*
- *When variability is large, results can be sensitive to choice of first-stage prior*
  - *Conjugate prior requires small shape parameter to represent large variability*
    - *Leads to unrealistically small lower percentiles*
  - *Nonconjugate first-stage prior gives more realistic lower percentiles, but mean may not be representative*

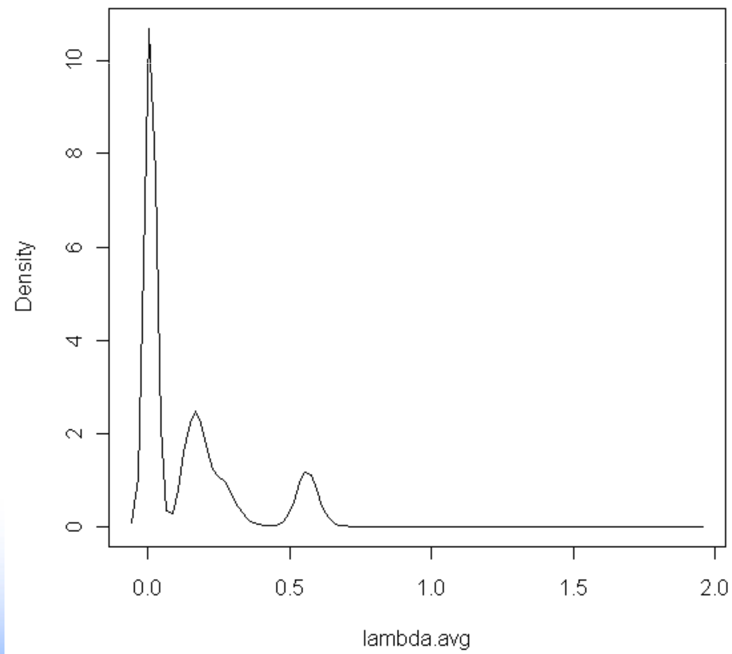
Conclusions

- *In cases of large variability, median is more robust estimate than mean*
- *Recommended first-stage priors when variability is large;*
  - *Poisson data: lognormal prior for* λ
  - *Binomial data: logistic-normal prior for p*
    - *Lognormal prior for p can give values > 1*
    - *Logistic-normal and lognormal approximately same for small p*

Conclusions

- *With extreme source-to-source variability, may want to consider clustering sources and developing mixture prior or eliminating some sources altogether*

Mixture prior with 9 clusters, equally weighted



Backup Slides

Hierarchical Bayes Model for LOSP Data

- *WinBUGS script*

```
model       {
for (i in 1 : N) {
lambda[i] ~ dgamma(alpha, beta) #Model variability in  frequency - gamma first stage
}
lambda.avg ~ dgamma(alpha, beta) #Industry population variability curve – gamma
alpha ~ dgamma(0.0001, 0.0001) #Vague hyperprior for alpha
beta ~ dgamma(0.0001, 0.0001) #Vague hyperprior for beta
}

inits
list(alpha=1, beta=1000)
list(alpha=10, beta=100)
```

WinBUGS Script for Reparameterized Model

```
model	{
for (i in 1 : N) {
lambda[i] ~ dgamma(alpha, beta) #Model variability in  frequency - gamma first stage
}
lambda.avg ~ dgamma(alpha, beta) #Industry population variability curve – gamma
alpha <- pow(CV, -2)
beta <- alpha/mean
mean ~ dgamma(0.0001, 0.0001)
CV ~ dgamma(0.0001, 0.0001)
}

Inits
list(CV=0.5, mean=1)
list(CV=2, mean=0.1)
```

Results with Lognormal First-Stage Prior

- *WinBUGS script*

```
model		{
for (i in 1 : N) {
lambda[i] ~ dlnorm(mu, tau) #Lognormal first-stage prior
}
lambda.avg ~ dlnorm(mu, tau)#Industry population variability curve – lognormal
mu ~ dflat()
tau <- pow(sigma, -2)
sigma ~ dunif(0, 5)
}

inits
list(mu=-3, sigma=2)
list(mu=-1, sigma=1)
```