PUMPS EXAMPLE

This problem is based on "Pumps" example in OpenBUGS Examples Volume I. The goal is to informally compare different model formulations (e.g., hierarchical vs. non-hierarchical parameter model). The data are for 10 power plant pumps and their failure rates, where: y_j = the number of failures for pump j that occur over the period t_j, where t_j is the length of operation time of pump j (in 1000s of hours). The likelihood is based on a Poisson distribution such that:

$$y_j \sim Poisson\left(\theta_j \cdot t_j\right)$$

We will specify the following conjugate prior (hierarchical model) for each θ_j, and we will specify the following priors for the root nodes, α and β:

$$\theta_j \sim gamma(\alpha, \beta)$$

$$\alpha \sim expon(1)$$

$$\beta \sim gamma(0.1, 1.0)$$

1.) <u>Model 1 – hierarchical failure rate model</u>: Use JAGS to implement the hierarchical version of the Pumps model (as above). That is, run the model that has already been programmed in OpenBUGS (see "Pumps example OpenBUGS" and associated data, "Pumps data"). Within the JAGS code, compute `Etheta` = $E[\theta_j] = \alpha / \beta$ and the "average" θ across pumps as `theta.ave` = mean(`theta`[]); these quantities should not occur inside of a "for loop."

   a. Using `jags.model()`, initialize the model for 3 chains and allow the JAGS model to generate initial values (i.e., no need to provide initials).

   b. Using `coda.samples()`, monitor `theta, beta, alpha, theta.ave, Etheta,` and `deviance`. Run the model for 10,000 iterations.

   c. Briefly evaluate the MCMC chains for burn-in, convergence, and autocorrelation.

   d. Compute posterior statistics (mean, median, std dev, 95% CI) for `theta, beta, alpha, theta.ave,` and `Etheta`.

   e. Create a box plot or caterpillar plot of the `theta`'s.

   f. You will want to keep this output handy for comparison with the other two models.

2.) <u>Model 2 – non-hierarchical failure rate model</u>: Modify model 1 by assuming independent priors for each θ_j, that is, treat each θ_j as a root node and assign a non-informative conjugate prior to each, e.g., θ_j ~ gamma(0.1, 0.1). Note, you can't compute `Etheta`, but we can still compute `theta.ave` for comparison. Repeat steps 1a – 1e and keep the output for comparison.

3.) <u>Model 3 – common failure rate, non-hierarchical model</u>: Modify model 2 by assuming a common failure rate for all pumps (i.e., do not index θ by j, just assume a common θ); assign a conjugate, non-informative prior to θ (i.e., use the same prior as used in Model 2 for each θ_j). Note, you can't compute `Etheta` or `theta.ave`, but you can compare θ from this model to `Etheta` or `theta.ave` obtained from models 1 and 2 (all quantities represent the "average" failure rate across pumps). Repeat steps 1a – 1e and keep the output for comparison.

4.) <u>Comparison:</u> Compare the results you obtained from the 3 models and address the following:

a. Is the assumption of a common failure across pump appropriate, or do we need separate failure rates for each pump?

b. How is the uncertainty (e.g., CI widths or std. deviations) in the θ_j's affected by the assumption of a hierarchical model (treats the θ_j's as coming from a pop'n of θ_j's) vs. a non-hierarchical model (θ_j treated as completely independent, don't arise from a common distribution)?

c. How does the population estimate (mean, median, CI) of the "common" failure rate (`Etheta` in Model 1 or `theta.ave` in Models 1 and 2 vs. `theta` in model 3) differ between models?

BUGS

Pumps: conjugate gamma-Poisson hierarchical model

George *et al* (1993) discuss Bayesian analysis of hierarchical models where the conjugate prior is adopted at the first level, but for any given prior distribution of the hyperparameters, the joint posterior is not of closed form. The example they consider relates to 10 power plant pumps. The number of failures x_i is assumed to follow a Poisson distribution

$$x_i \sim \text{Poisson}(\theta_i t_i) \quad i = 1,...,10$$

where θ_i is the failure rate for pump *i* and t_i is the length of operation time of the pump (in 1000s of hours). The data are shown below.

Pump	t_i	x_i
1	94.5	5
2	15.7	1
3	62.9	5
4	126	14
5	5.24	3
6	31.4	19
7	1.05	1
8	1.05	1
9	2.1	4
10	10.5	22

A conjugate gamma prior distribution is adopted for the failure rates:

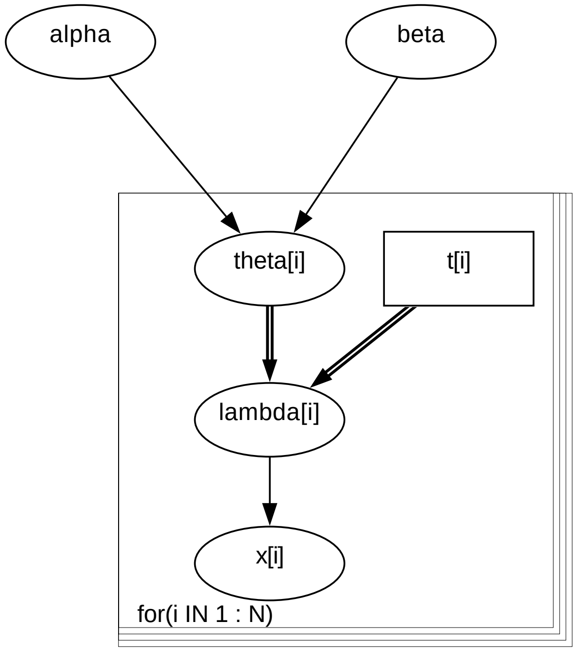
$$\theta_i \sim \text{Gamma}(\alpha, \beta), \; i = 1,...,10$$

George *et al* (1993) assume the following prior specification for the hyperparameters α and β

$$\alpha \sim \text{Exponential}(1.0)$$
$$\beta \sim \text{Gamma}(0.1, 1.0)$$

They show that this gives a posterior for β which is a gamma distribution, but leads to a non-standard posterior for α. Consequently, they use the Gibbs sampler to simulate the required posterior densities.

Graphical model for pump example:



BUGS *language for pump example:*

```
model
{
   for (i in 1 : N) {
      theta[i] ~ dgamma(alpha, beta)
      lambda[i] <- theta[i] * t[i]
      x[i] ~ dpois(lambda[i])
   }
   alpha ~ dexp(1)
   beta ~ dgamma(0.1, 1.0)
}
```

[Data](#) (click to open)

[Inits for chain 1](#)   [Inits for chain 2](#)   (click to open)

[Results](#)

A burn in of 1000 updates followed by a futher 10000 updates gave the parameter estimates:

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
alpha	0.6951	0.2764	0.005396	0.2812	0.6529	1.351	1001	10000
beta	0.9189	0.542	0.01017	0.1775	0.8138	2.265	1001	10000
theta[1]	0.05981	0.02518	2.629E-4	0.02127	0.05621	0.1176	1001	10000
theta[2]	0.1027	0.08174	9.203E-4	0.00808	0.08335	0.3138	1001	10000
theta[3]	0.08916	0.03802	4.144E-4	0.03116	0.08399	0.1798	1001	10000
theta[4]	0.1157	0.0301	3.152E-4	0.06443	0.1128	0.1818	1001	10000
theta[5]	0.5977	0.3124	0.003209	0.1491	0.5426	1.359	1001	10000
theta[6]	0.6104	0.1376	0.00145	0.3726	0.6007	0.9089	1001	10000
theta[7]	0.9035	0.7396	0.007221	0.07521	0.7072	2.844	1001	10000
theta[8]	0.9087	0.7523	0.007056	0.07747	0.7094	2.887	1001	10000
theta[9]	1.583	0.7647	0.007846	0.4667	1.461	3.446	1001	10000
theta[10]	1.984	0.4212	0.004278	1.24	1.953	2.891	1001	10000