Introduction to Hierarchical Models

22S:138 Bayesian Statistics

Lecture 11
Oct 6, 2004

Kate Cowles, Ph.D.

Hierarchical models

- Bayesian models with more than two levels or "stages"
- may arise for several reasons
  - we have insufficient knowledge to specify the parameters of priors
  - we wish to model data or parameters that cannot be considered exchangeable but that are related

Example: Pump failure data

- A hierarchical model is fit to data on failure rates of the pump at each of 10 power plants. The number of failures for the i-th pump is assumed to follow a Poisson distribution:

  $$x_i \sim Poisson(\theta_i t_i), \quad i = 1, \ldots, 10$$

  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).
- Important point: we do *not* assume that all the pumps have the same failure rate. In fact, one of the questions of interest is to estimate the rates for the individual pumps.
- We do not consider the x_i, t_i pairs *exchangeable*.

- Write the likelihood of the data:

- Recall that the definition of exchangeable observations is their likelihood is invariant to permutations of the indices.
- If we exchanged the subscripts on two x_i, t_i pairs, and did not change the indices of the corresponding θ_is, the evaluation of the likelihood would change.
- The *first stage* of a hierarchical model is the sampling distribution of the observed data, or the likelihood.

The second stage

- The second stage gives priors on the parameters that appeared in the first stage.
- In the pump failures example, a conjugate gamma prior distribution is adopted for the failure rates:

$$\theta_i \sim Gamma(\alpha, \beta), \quad i = 1\ldots, 10$$

- This says that, although the failure rates for the individual pumps are not the same, they are related. They are all drawn from a common distribution.
- We do not know enough about failure rates of pumps in nuclear power plants to be able to specify fixed numbers for the prior parameters α and β. In fact, we want the data to inform us about these values.
- Consequently, we will make α and β additional *unknown* paramters in the model.

  – "hyperparameters"

- At the third stage of the hierarchical model for pump failures, the following priors are specified for the hyperparameters α and β:

$$\alpha \sim Exponential(1.0)$$
$$\beta \sim Gamma(0.10, 1.0)$$

WinBUGS program to fit "Pump" model

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

  alpha ~ dexp(1.0);
  beta ~ dgamma(0.1,1.0);

}
```


Data and initial values

```
list(t = c(94.3,15.7,62.9,126.0,5.24,31.4,1.05,1.05,2.1,
     10.5),
     x = c(5,1,5,14,3,19,1,1,4,22), N=10)

list(alpha = 1.0, beta = 1.0,
     theta = c(0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1))
```

Results

node	mean	sd	MC error	2.5%	med	97.5%	start	sam
alpha	0.70010	0.26990	.0047060	0.28510	0.66340	1.3380	1001	100
beta	0.92900	0.53250	.0097800	0.19380	0.83150	2.2050	1001	100
theta[1]	0.05980	0.02542	.0002680	0.02128	0.05627	0.1195	1001	100
theta[2]	0.10080	0.07855	.0008177	0.00838	0.08181	0.3023	1001	100
theta[3]	0.08927	0.03759	.0003702	0.03160	0.08469	0.1762	1001	100
theta[4]	0.11600	0.03048	.0003170	0.06363	0.11320	0.1825	1001	100
theta[5]	0.60560	0.31500	.0030870	0.15290	0.55290	1.3590	1001	100
theta[6]	0.61050	0.13930	.0014000	0.36680	0.59960	0.9096	1001	100
theta[7]	0.90250	0.72520	.0079370	0.07559	0.71670	2.7510	1001	100
theta[8]	0.89640	0.72500	.0082620	0.07614	0.70980	2.7850	1001	100
theta[9]	1.59000	0.77670	.0090040	0.48280	1.45200	3.4520	1001	100
theta[10	1.99300	0.42510	.0049150	1.26400	1.95800	2.9160	1001	100

Compare to maximum likelihood estimates for individual pumps

hours	failures	mle	theta
94.30	5	.0530	.0598
15.70	1	.0637	.1008
62.90	5	.0795	.0893
126.00	14	.1111	.1160
5.24	3	.5725	.6056
31.40	19	.6051	.6105
1.05	1	.9528	.9025
1.05	1	.9524	.8964
2.10	4	1.9048	1.5900
10.50	22	2.0952	1.9309

- individual estimates are “shrunk” away from mle toward a common mean
- individual estimates “borrow strength” from the rest of the data
- thetas for observations with large ”sample size” (time observed) are shrunk less

than thetas for other observations

- thetas far from the common mean are shrunk more than those near it