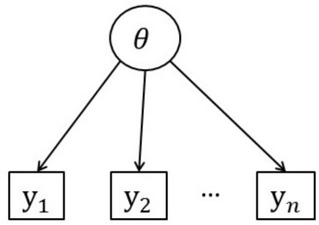
5 Week 5

5.1 Directed acyclic graphs

a)-c) Write down 1) the joint probability distribution of all the parameteters and observation variables y in the directed acyclic graphs (DAG) shown in Figure 1 and 2) a Stan pseudo code that tells how a model corresponding to that DAG would be written. You can assume that all variables get values in real numbers. An example of a model answer is provided on the right



Joint distribution

$$p(y_1, \dots y_n, \theta) = p(\theta) \prod_{i=1}^{n} p(y_i|\theta)$$

Stan pseudo-code:

```
data{
   int<lower=0> n;
   real y[n];
}
parameters{
   real theta;
}
model{
   theta ~ p();
   for( i in 1 : n ) {
      y[i] ~ p(theta);
   }
}
```

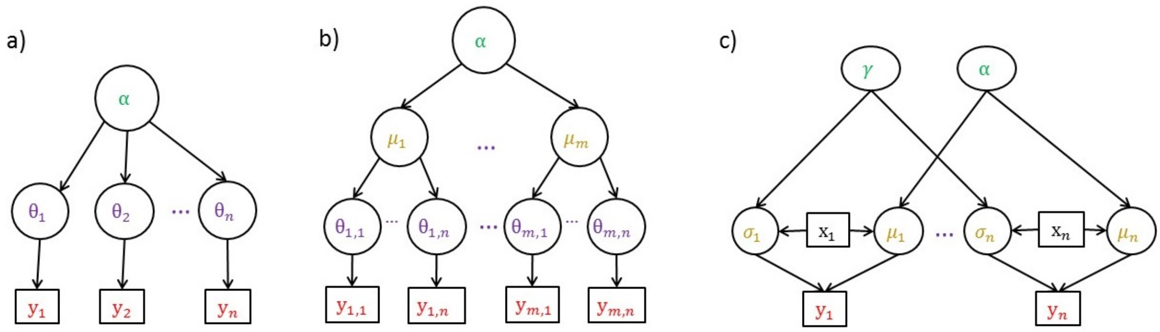


Figure 1: The DAGs for which the joint distribution and pseudo code have to be defined. Note variables denoted by x should be treated as covariates.

d) Draw a Directed acyclic graph (DAG) and write a Stan pseudo code of the following model

$$\begin{aligned}
y_{i,j} &\sim N(\mu_j, \sigma_j^2), i = 1, \dots, n, j = 1, \dots, J \\
\mu_j &\sim N(\mu_0, \phi) \\
\mu_0 &\sim N(0, 10^6) \\
\phi &\sim \text{Inv-}\chi^2(\nu_1, s_1^2) \\
\sigma_j^2 &\sim \text{Inv-}\chi^2(\nu_2, s_2^2)
\end{aligned}$$

See exercise 5.1 for an example on the needed accuracy for the pseudocode.

GRADING: a)-c) Two points for correct joint density function and 3 points for correct pseudo-code. d) 3 points for correct DAG and 2 points for correct pseudo-code.

5.2 Hierarchical model for Rat tumor experiments

R-template `ex_ratTumor.R`.

This exercise is modified from the example in Section 5.1 and 5.3 in BDA3. Construct the following hierarchical model for the 71 experiments summarized in the data file

$$\begin{aligned} y_i &\sim \text{Binom}(\theta_i, N_i) \\ \theta_i &\sim \text{Beta}(\mu s, s - \mu s) \\ \mu &\sim \text{Unif}(0, 1) \\ s &\sim \text{log-}N(4, 4). \end{aligned}$$

Here, μ is the prior mean of θ_i and s governs the uncertainty about it. The parametrization of log-Gaussian distribution $s \sim \text{log-}N(m, \sigma^2)$ is such that $E[\log(s)] = m$ and $Var[\log(s)] = \sigma^2$

1. Implement the model in Stan and sample from the posterior for parameters. Check for convergence for all parameters, and examine autocorrelation for s, μ and few θ_i. visualize posterior of μ, s and $\theta_i, i = 1, \dots, 71$.

2. what is the interpretation of posterior of μ? what is the interpretation of posterior of θ_i? How does this differ from the interpretation of μ.

3. calculate the posterior predictive distribution of outcome $\tilde{y}_{71}$ of a new experiment with $\tilde{N}_{71} = 20$ new test animals in laboratory 71.

4. calculate the posterior predictive distribution of θ_{72} and $\tilde{y}_{72}$ with $\tilde{N}_{72} = 20$ new test animals in a new laboratory of number 72 (a laboratory from where we don't have data yet) that is similar to the existing 71 laboratories

5. Sample from the posterior distribution of the so called pooled estimate of θ (you can do this either with Stan or directly in R). This corresponds to a model

$$\begin{aligned} y_i &\sim Binom(\theta, N_i) \\ \theta &\sim Beta(1, 1) \end{aligned}$$

   In this model all laboratories are thought to have the same parameter θ. Note that this model is equivalent to model

$$\begin{aligned} \sum_{i=1}^{71} y_i &\sim Binom(\theta, \sum_{i=1}^{71} N_i) \\ \theta &\sim Beta(1, 1). \end{aligned}$$

   Discuss how does the posterior of the pooled θ differ from the population mean, μ, and from the individual θ_i in the hierarchical model?

Grading: Each of the steps provides 4 points from correct answer and 2 points from an answer that is towards the right direction but includes minor mistake (e.g. a bug or typo)