STAT 535

MCMC Case-Study

A Dataset

From a case-control study

Cases: babies who died of *sudden infant death syndrome* (SIDS)

Controls: babies healthy at age XXX.

Question: SIDS associated with maternal *anemia*?

Subject-level variables

$Y = 0/1$: control/case (DISEASE)
$V = 0/1$: *actual* no/yes anemia (EXPOSURE)
$V_1^* = 0/1$: interview indicates no/yes anemia
$V_2^* = 0/1$: medical records indicate no/yes anemia

	controls		cases	
	$V_2^* = 0$	$V_2^* = 1$	$V_2^* = 0$	$V_2^* = 1$
$V_1^* = 0$	147	15	125	15
$V_1^* = 1$	34	20	49	24

Modelling Assumptions

Given (V, Y), (V_1^*, V_2^*) are

(i) independent,

and

(ii) have distributions which only depend on V.

Interpretation?

Six Unknown Parameters

Two exposure prevalences:

$$r_0 = Pr(V = 1|Y = 0) \qquad r_1 = Pr(V = 1|Y = 1)$$

($\psi = log(r_1/(1 - r_1)) - log(r_0/(1 - r_0))$ of particular interest.)

Two parameters reflecting performance of V_1^* as a surrogate for V:

$$p_1 = Pr(V_1^* = 1|V = 1) \qquad q_1 = Pr(V_1^* = 0|V = 0),$$

and similarly

$$p_2 = Pr(V_2^* = 1|V = 1) \qquad q_2 = Pr(V_2^* = 0|V = 0).$$

Assign independent uniform priors to the six parameters, i.e. $f(p_1, q_1, p_2, q_2, r_0, r_1) = 1$. Interpretation?

What is the joint distribution of everything???

$$\begin{aligned} f(everything) &= \prod_{i=1}^{n} \{f(v_{1i}^*|v_i, p_1, q_1)\times \\ &\quad f(v_{2i}^*|v_i, p_2, q_2) \times \\ &\quad f(v_i|y_i, r_0, r_1) \times \\ &\quad f(y_i)\} \times \\ &\quad f(p_1, q_1, p_2, q_2, r_0, r_1) \end{aligned}$$

$$\begin{aligned} f(\text{everything}) &\propto \prod_{i=1}^{n} \times \\ &\quad p_1^{V_i V_{1i}^*}(1-p_1)^{V_i(1-V_{1i}^*)} \times \\ &\quad q_1^{(1-V_i)(1-V_{1i}^*)}(1-q_1)^{(1-V_i)V_{1i}^*} \times \\ &\quad p_2^{V_i V_{2i}^*}(1-p_2)^{V_i(1-V_{2i}^*)} \times \\ &\quad q_2^{(1-V_i)(1-V_{2i}^*)}(1-q_2)^{(1-V_i)V_{2i}^*} \times \\ &\quad r_0^{V_i(1-Y_i)}(1-r_0)^{(1-V_i)(1-Y_i)} \times \\ &\quad r_1^{V_i Y_i}(1-r_1)^{(1-V_i)Y_i} \end{aligned}$$

Bizzare looking! Structure?

Full Conditionals - Beta distributions!

$$(p_1|e.e.) \sim Beta\left\{1 + \sum_{i:V_i=1} V_{1i}^*, 1 + \sum_{i:V_i=1} (1 - V_{1i}^*)\right\}$$

Interpretation? Similarly,

$$(q_1|e.e.) \sim Beta\left\{1 + \sum_{i:V_i=0} (1 - V_{1i}^*), 1 + \sum_{i:V_i=0} V_{1i}^*\right\}$$

$$(p_2|e.e.) \sim Beta\left\{1 + \sum_{i:V_i=1} V_{2i}^*, 1 + \sum_{i:V_i=1} (1 - V_{2i}^*)\right\}$$

$$(q_2|e.e.) \sim Beta\left\{1 + \sum_{i:V_i=0} (1 - V_{2i}^*), 1 + \sum_{i:V_i=0} V_{2i}^*\right\}$$

$$(r_0|e.e.) \sim Beta\left\{1+\sum_{i:Y_i=0} V_i, 1+\sum_{i:Y_i=0}(1-V_i)\right\}$$

$$(r_1|e.e.) \sim Beta\left\{1+\sum_{i:Y_i=1} V_i, 1+\sum_{i:Y_i=1}(1-V_i)\right\}$$

Also, $(V_i|\text{e.e.}) \sim$ Bernoulli.

For instance,

$$Pr(V_i = 1|V_{1i}^* = 1, V_{2i}^* = 1, Y = 0) = \frac{p_1 p_2 r_0}{p_1 p_2 r_0 + (1-q_1)(1-q_2)(1-r_0)}$$

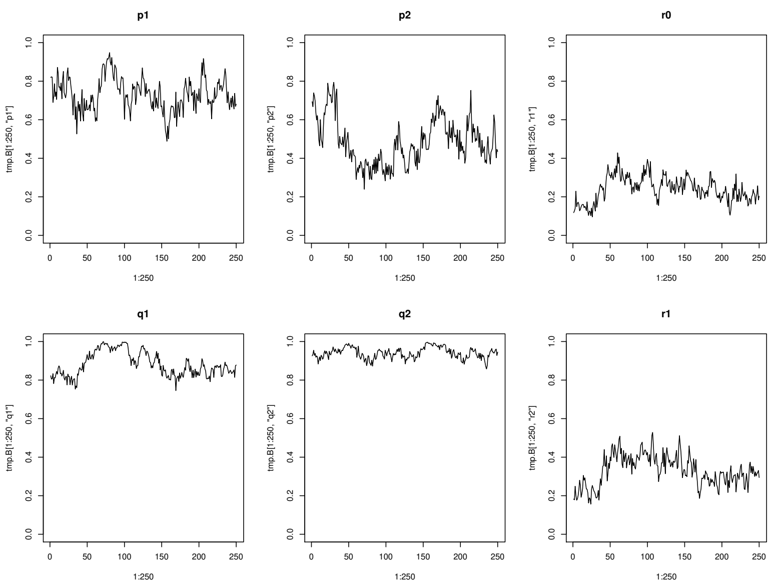
or

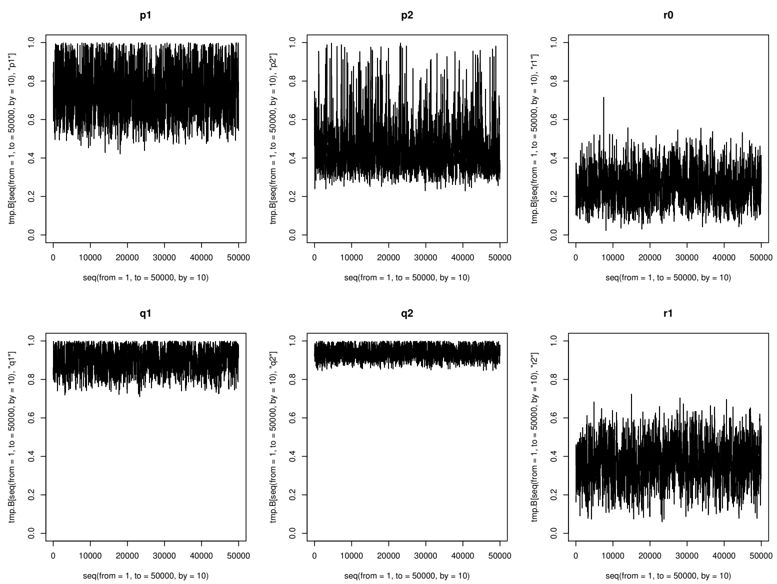
$$Pr(V_i = 1|V_{1i}^* = 0, V_{2i}^* = 1, Y = 1) = \frac{(1-p_1) p_2 r_1}{(1-p_1) p_2 r_1 + (q_1)(1-q_2)(1-r_1)}$$

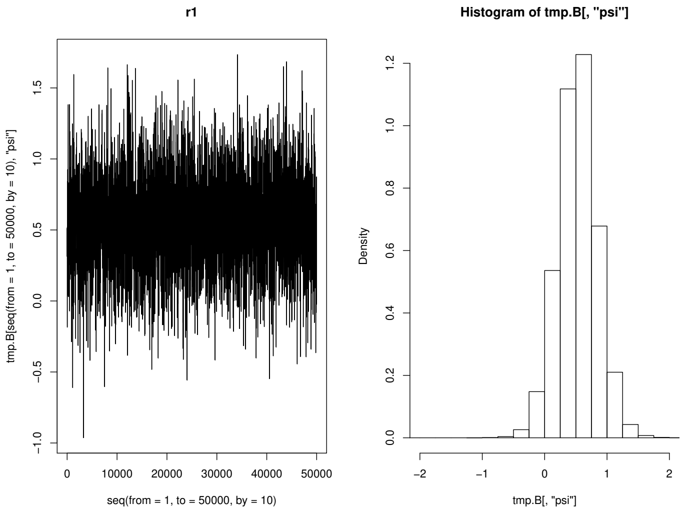
etc.

So, have identified the full conditional distributions for $(p_1, q_1, p_2, q_2, r_0, r_1, V|V_1^*, V_2^*, Y)$ **Gibbs sample away!!!**

```
theta[1,...,q] <- arbitrary
for (i in 1:m) {
  for (j in 1:q) {
    sample theta[j] given theta[-j] }
  opt[i,] <- theta }
```





Point estimate: $\hat{\psi} = 0.53$. Interval estimate: $(-0.09, 1.16)$