Full Bayesian model for the meta-analysis example

Control group: n_j^c subjects, y_j^c deaths, $j=1,\dots,J$

Treatment group: n_j^T subjects, y_j^T deaths, $j=1,\dots,J$

Full model

$$y_j^c \sim Bin(n_j^c, p_j^c), \quad j=1,\dots,J$$

$$y_j^T \sim Bin(n_j^T, p_j^T), \quad j=1,\dots,J$$

$$\text{logit}(p_j^c) = \log\left(\frac{p_j^c}{1-p_j^c}\right) = \beta_j$$

$$\text{logit}(p_j^T) = \log\left(\frac{p_j^T}{1-p_j^T}\right) = \beta_j + \theta_j$$

$$\theta_j \sim N(\mu, \tau^2), \quad j=1,\dots,J$$

$$\beta_j \sim N(0, 1e5)$$

$$\mu \sim N(0, 1e6)$$

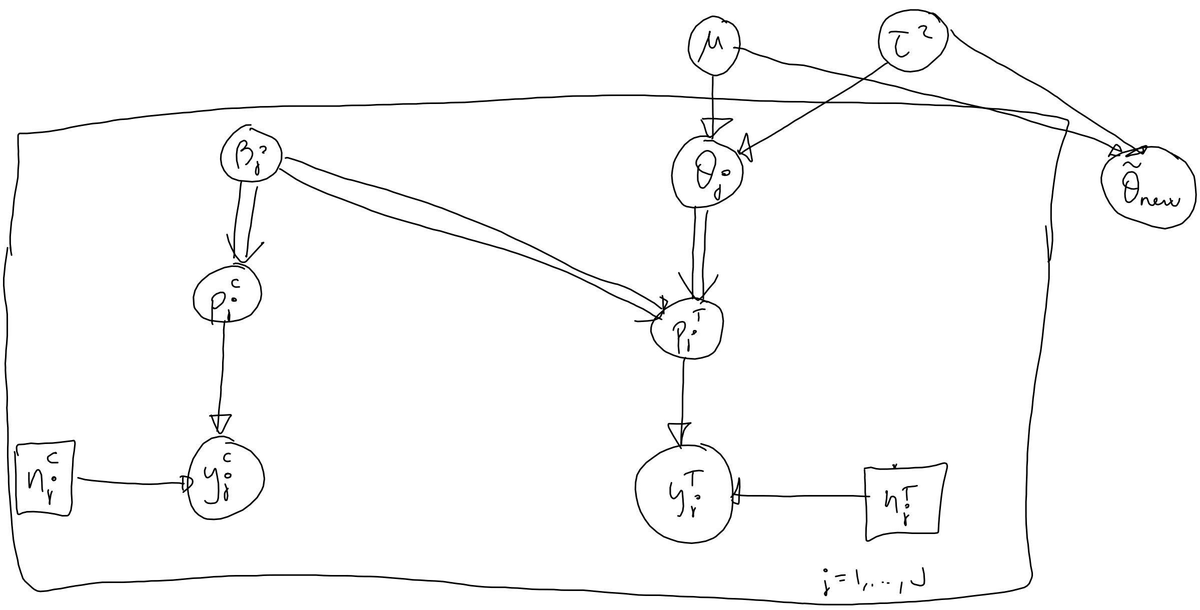
$$\tau^2 \sim \text{Inv-Gamma}(1e-3, 1e-3)$$

$\theta_j = \log \rho_j = \theta_j$ same as before

$$p_j^c = \frac{\exp\{\beta_j\}}{1+\exp\{\beta_j\}}$$

$$p_j^T = \frac{\exp\{\beta_j+\theta_j\}}{1+\exp\{\beta_j+\theta_j\}}$$

nov 3-13:46



nov 3-14:30

$$p(\theta,\beta,\mu,\tau^2 \mid y) \propto p(\tau^2)\cdot p(\mu)\cdot \prod_{j=1}^{J} p(\theta_j \mid \mu,\tau^2)\cdot \prod_{j=1}^{J} p(\beta_j)$$

$$\cdot \prod_{j=1}^{J} \left[\underbrace{\left(\frac{\exp\{\beta_j\}}{1+\exp\{\beta_j\}}\right)^{y_j^c} \cdot \left(\frac{1}{1+\exp\{\beta_j\}}\right)^{n_j^c - y_j^c}}_{p_j^{c\,y_j^c}\cdot(1-p_j^c)^{n_j^c - y_j^c}} \cdot \underbrace{\left(\frac{\exp\{\beta_j+\theta_j\}}{1+\exp\{\beta_j+\theta_j\}}\right)^{y_j^T} \left(\frac{1}{1+\exp\{\beta_j+\theta_j\}}\right)^{n_j^T - y_j^T}}_{p_j^{T\,y_j^T}(1-p_j^T)^{n_j^T - p_j^T}} \right]$$

nov 3-14:33

Remember

We have a target distribution $p(\cdot)$ (the Bayesian: the posterior distribution $p(\theta|y)$) that only know up to a constant of proportionality, i.e.

$$p(\cdot) = C\cdot f(\cdot)$$, where C is unknown, $f(\cdot)$ is known

Of particular interest

$$E[h(\theta)|y] = \int h(\theta)\, p(\theta|y)\, d\theta$$

where for example

Posterior exp. of θ_i : $h(\theta) = \theta_i$

Post. variance of θ_i : $h(\theta) = (\theta_i - \mu_{\theta_i})^2$

Prediction of $\tilde{y}$: $h(\theta) = p(\tilde{y}|\theta)$

When we have S iid draws $\theta^{(s)}$ of $\theta|y$, then LLN gives us that $\frac{1}{S}\sum_{s=1}^{S} h(\theta^{(s)}) \to E[h(\theta)|y]$

Often: Hard to draw iid samples from $p(\theta|y)$

nov 3-14:39

Now: Markov Chain Monte Carlo (MCMC)

In short: A Markov Chain is a sequence of dependent random variables $X^1, X^2, \ldots, X^{t-1}, X^t, X^{t+1}, \ldots$ such that the probability distribution of X^t, for all t, conditional on all the previous variables $X^{t-1}, X^{t-2}, \ldots, X^1$, depends only on the most recent value X^{t-1}, i.e.

$$X^t \mid X^{t-1}, X^{t-2}, \ldots, X^2, X^1 \sim T(X^t \mid X^{t-1})$$

a transition distribution

Here: We construct Markov Chain $\theta^1, \theta^2, \ldots$ that has a stationary distribution $p(\theta|y)$. Hence (removing "burn-in" iterations $1, \ldots, B$) $\theta^{B+1}, \ldots, \theta^T$ will be approximate samples from $p(\theta|y)$

This is done by

(1) Choose an appropriate proposal (or jumping) distribution from which the next sample is drawn

(2) Define an acceptance probability for the new proposal, such that the stationary distribution of the accepted proposals is in fact $p(\theta|y)$

nov 3-14:45

Some useful properties of MCMC

- Relatively little knowledge on the target distribution
- The approximate distr. converges to the target
- The Gibbs sampler enables decomposing intractable high-dimensional problems into a sequence of tractable low-dimensional problems

Issues

- Speed, can be slow
- Convergence: Important to check!!!

nov 3-14:55