ISC 5228
Markov Chain Monte Carlo

Hierachical Models in Meta-Study Analysis

1 Motivation

Consider the following passage from Alex Reinhart's enjoyable book "Statistics Done Wrong"[1]

> Suppose you're in charge of public school reform. As part of your research into the best teaching methods, you look at the effect of school size on standardized test scores. Do smaller schools perform better than larger schools? Should you try to build many small schools or a few large schools?
>
> To answer this question, you compile a list of the highest-performing schools you have. The average school has about 1,000 students, but the top-scoring five or ten schools are almost all smaller than that. It seems that small schools do the best, perhaps because of their personal atmosphere where teachers can get to know students and help them individually.
>
> Then you take a look at the worst-performing schools, expecting them to be large urban schools with thousands of students and overworked teachers. Surprise! Theyre all small schools too.
>
> Smaller schools have more widely varying average test scores, entirely because they have fewer students. With fewer students, there are fewer data points to establish the "true" performance of the teachers, and so the average scores vary widely. As schools get larger, test scores vary less, and in fact increase on average.

Studies with small sample sizes have a tendency to produce "surprising" results. From a scientific standpoint, *meta-studies*, which seek to extract the true signal from multiple independent studies, are much more reliable.

In this lab (based on the P1A2 example by Rebecca Steorts[2]), we will consider a typical hierarchical model, used within a Bayesian framework, to tell us something about the population, from *a meta-study* of multiple independent samples.

2 Setup

- Twelve studies were run to investigate the potential link between presence of a certain genetic trait and the risk of heart attack.
- Each study was case-control and considered (i) a group of individuals with coronary heart disease, and (ii) another group with no history of coronary heart disease.
- For each study i $(= 1, \ldots, 12)$ the proportion having the genetic trait in each group was recorded.
- For each study, a log odds ratio, $\hat{\psi}_i$, and standard measurement error, σ_i, were calculated.

[1] http://www.statisticsdonewrong.com/

[2] http://www.stat.cmu.edu/~rsteorts/btheory2/ch5_final2.pdf

2.1 Odds Ratio

Let D_T and H_T represent the number of diseased and healthy individuals, respectively, with the trait. Similarly let D_{NT} and H_{NT} represent the number of diseased and healthy individuals, respectively, *without* the trait.

	trait	no trait
healthy	H_T	H_{NT}
disease	D_T	D_{NT}

Then the odds ratio is defined as:

$$\text{odds ratio} = \frac{(D_T/H_T)}{(D_{NT}/H_{NT})}. \tag{1}$$

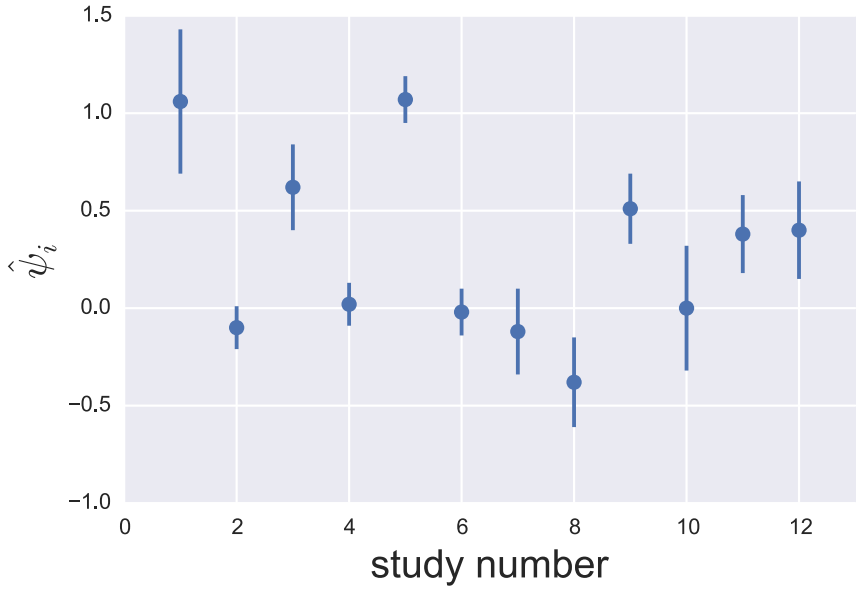
It ranges between 0 (trait protects from disease), close to 1 (trait produces no enhancement in disease) and infinity (trait predestines disease).

The logarithm of the odds ratio, ψ, ranges between $-\infty$, through 0 (no effect), to $+\infty$.

2.2 Data

There are 12 studies with the following log odds ratio and standard error:

study i	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\psi}_i$	1.06	-0.1	0.62	0.02	1.07	-0.02	-0.12	-0.38	0.51	0.00	0.38	0.40
σ_i	0.37	0.11	0.22	0.11	0.12	0.12	0.22	0.23	0.18	0.32	0.20	0.25



2.3 Sources of Error

Suppose the "true" log odds ratio for the population is μ. That is if you included everybody from the target population in your study, and perfectly measured their state (disease/healthy) and trait (present/absent) and used equation 1, then you would get μ.

Truth while desirable, is elusive.

Sampling Error: When we select a sample to perform study i, the true sample log odds ratio $\psi_i \neq \mu$. In the process of selecting a subset of the population, we have introduced *sampling error*.

Measurement Error: Suppose, we are given a sample, with true sample log odds ratio ψ_i. Due to imperfect measurement (instruments/tests not perfect etc.), the measured log odds ratio $\hat{\psi}_i \neq \psi_i$.

We need to model both these sources of uncertainty.

3 Hierarchical Model

The measured $\hat{\psi}_i$ have **measurement error** characterized by σ_i. We assume that if the true log odds ratio were ψ_i, then,

$$\hat{\psi}_i \sim \mathcal{N}(\psi_i, \sigma_i^2) \tag{2}$$

Similarly the sample log odds ratio ψ_i are themselves random variables, which are sampled from the *true* population with mean μ, and variance τ^2). That is the **sampling error** is given by:

$$\psi_i \sim \mathcal{N}(\mu, \tau^2).$$

We want to use Bayesian analysis to find the distribution of the true μ (and τ^2), given the data.

3.1 Bayesian Analysis

3.1.1 Prior

We need to assume a prior for μ and τ^2. Let us assume relatively non-informative priors:

$$1/\tau^2 \sim \text{Gamma}(0.1, 0.1) \tag{3}$$

$$\mu \sim p(\mu|\tau^2) = \mathcal{N}(0, 1000\tau^2). \tag{4}$$

Note:

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$\text{Gamma}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$$

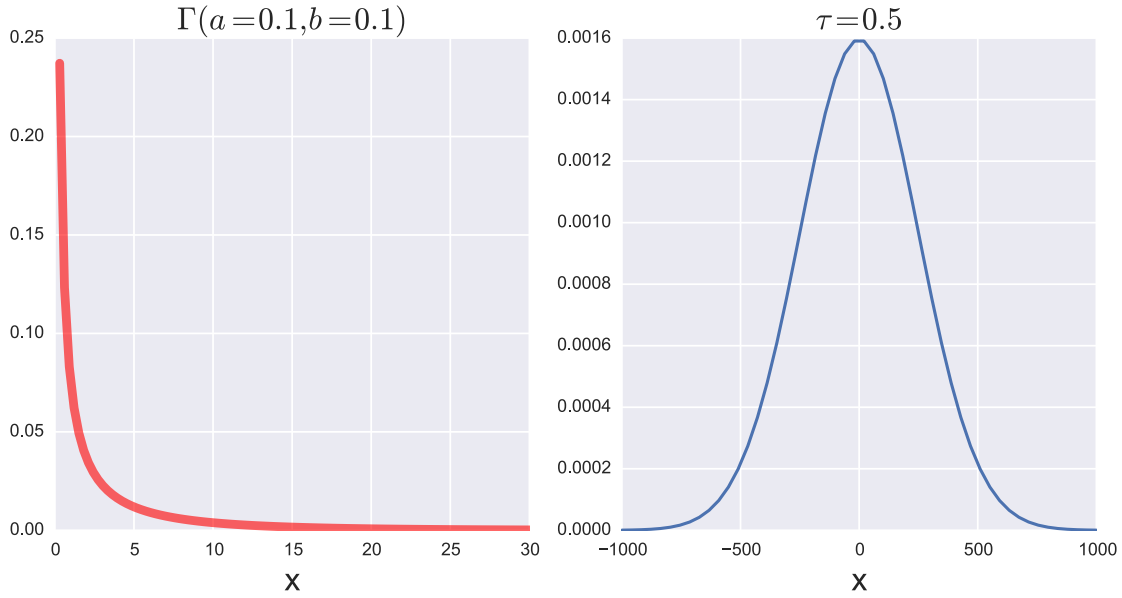


Figure 1: *Left*: $1/\tau^2$ is drawn from Gamma(0.1, 0.1), which has a bias for large values of τ. *Right*: When $1/\tau^2 = 4$, $\tau = 0.5$, the distribution $p(\mu|\tau^2)$ is broad. Smaller values of $1/\tau^2$ (larger values of τ), make the distribution even broader.

The Gamma distribution has most of its weight at small values, which corresponds to relatively large ("non-informative") τ^2. The combined prior distribution is therefore:

$$p(\mu, \tau^2) = p(\mu|\tau^2)p(\tau^2) = \mathcal{N}(\mu; 0, 1000\tau^2)\ \text{Gamma}(1/\tau^2; 0.1, 0.1). \tag{5}$$

3.1.2 Likelihood

For a given study i, the likelihood of observing the $\hat{\psi}_i$ given μ and τ characterize the population is taken as:

$$l(\hat{\psi}_i|\mu, \tau^2) = \int_{-\infty}^{\infty} \mathcal{N}(\hat{\psi}_i; \psi_i, \sigma_i^2)\mathcal{N}(\psi_i; \mu, \tau^2)d\psi_i. \tag{6}$$

Notice, that we integrated out the "nuisance parameter" ψ_i, which we are not really interested in.
We assumed that all the studies are independent; thus, the overall likelihood is simply the product:

$$L(\{\hat{\psi}_i\}|\mu, \tau^2) = \prod_{i=1}^{12} l(\hat{\psi}_i|\mu, \tau^2). \tag{7}$$

After some work to evaluate the integral eqn. 6 (I used the python library `sympy`), one can write a closed form expression:

$$l(\hat{\psi}_i|\mu,\tau^2) = \frac{1}{\sqrt{2\pi(\sigma_i^2+\tau^2)}} \exp\left[\frac{-\mu^2\sigma_i^2 - \hat{\psi}_i^2\tau^2 + \frac{(\mu\sigma_i^2+\hat{\psi}_i\tau^2)^2}{\sigma_i^2+\tau^2}}{2\sigma_i^2\tau^2}\right] \tag{8}$$

3.1.3 Posterior

The posterior distribution

$$\pi(\mu,\tau^2|\{\hat{\psi}_i,\sigma_i\}) \propto L(\{\hat{\psi}_i\}|\mu,\tau^2)p(\mu,\tau^2). \tag{9}$$

4 Computational Strategy

1. Start with a initial values for μ, τ^2 (say 1.0 and 1.0)
2. Compute the prior and likelihood functions to compute the desired target (the posterior) distribution using eqns. 5, 7, and eqn. 9.
3. It may be useful to work with log probabilities instead of probabilities (round-off errors)
4. Propose a random diplacement of μ and τ^2 - and compute the new posterior. For example you may use a symmetric proposal:

$$\mu_{\text{new}} \sim \mathcal{N}(\mu_{\text{old}}, \Delta_\mu)$$
$$\tau^2_{\text{new}} \sim \mathcal{N}(\tau^2_{\text{old}}, \Delta_{\tau^2})$$

   where the Δ are user-specified values.
5. Use the Metropolis algorithm to sample

5 Exercises

- Using standard protocol (discard burn-in, traceplots to assess "mixing"), plot a histogram of the μ sampled.
- What is the expected value of μ? What do you surmise about the link between the genetic trait and risk of heart attack.

Note: For this lab, you are not required to test convergence with using the $\hat{R}$ diagnostic, or carry out any block-averaging.