Monte Carlo Methods Case Study

I built up the theory pretty far in the last of my "Why Do They Work?" write-ups:

* https://brianhill.github.io/bayesian-statistics/resources/MonteCarloMethodsWhyDoTheyWork-III.nb.pdf

Our case study is going to be the Hepatitis B vaccination study described in Chapter 2 of *Markov Chain Monte Carlo in Practice,* by Gilks, Richardson, and Spiegelhalter. Gilks, Richardson, and Spiegelhalter are the editors. Chapter 2's authors are Spiegelhalter, Best, Gilks, and Inskip.

The goal of looking closely at this case study is to see how the posterior can be practically sampled using Gibbs sampling, such as with BUGS. The BUGS reference is Gilks, Thomas, and Spiegelhalter, "A language and program for complex Bayesian modelling," *The Statistician,* **43** (1994) 169-177.

The Likelihoods and the Priors

On pp. 27 and 28 of Chapter 2, Spiegelhalter, Best, Gilks, and Inskip describe the likelihoods and the priors.

The Likelihoods

On p. 27 the likelihoods are given in these four equations:

$$
\begin{aligned}
y_{ij} &\sim \mathrm{N}(\mu_{ij}, \sigma^2), \\
\mu_{ij} &= \alpha_i + \beta_i(\log t_{ij} - \log 730), \\
\alpha_i &\sim \mathrm{N}(\alpha_0, \sigma_\alpha^2), \\
\beta_i &\sim \mathrm{N}(\beta_0, \sigma_\beta^2),
\end{aligned}
$$

The first two equations say that the likelihood for the y's is:

$$
P_y = \frac{1}{\left(\sqrt{2\pi}\,\sigma\right)^{i \cdot j}}\, e^{-\sum_{i,j}(y_{i,j} - \alpha_i - \beta_i\,\tau_{i,j})^2/2\,\sigma^2} \quad \text{where} \quad \tau_{i,j} \equiv \log \frac{t_{i,j}}{730}
$$

You can certainly be forgiven if you didn't realize that is what the first two equations say. The authors are using compact notation used by statisticians which we have not introduced in our class. The ~ means "statistically distributed according to," and the $N(\mu,\,\sigma^2)$ means "a Gaussian with mean μ and variance σ^2," and that is all that my equation says. Statisticians use N because they usually refer to

$\tau_{i,j}$

$$N(\mu, \sigma^2)$$

Gaussian distributions as "normal distributions."

I have introduced the $\tau_{i,j}$ so logarithms and the number of days in two years ($365 \cdot 2 = 730$) don't clutter my equations. It's just a linear fit with a bit of preprocessing of the times.

The third equation says the likelihood for the α's is:

$$P_\alpha = \frac{1}{\left(\sqrt{2\pi}\,\sigma_\alpha\right)^i} e^{-\sum_i (\alpha_i - \alpha_0)^2 / 2\,\sigma_\alpha^2}$$

The fourth equation says the likelihood for the β's is:

$$P_\beta = \frac{1}{\left(\sqrt{2\pi}\,\sigma_\beta\right)^i} e^{-\sum_i (\beta_i - \beta_0)^2 / 2\,\sigma_\beta^2}$$

The Priors

You will notice that five new parameters have appeared in these equations: σ^2, α_0, σ_α^2, β_0 and σ_β^2. These are called "parent parameters." We need priors for the parent parameters and on p. 28 of Chapter 2, the priors are given in these two equations.

$$
\begin{aligned}
\alpha_0,\ \beta_0 &\sim\ N(0,\ 10\,000), \\
\sigma^{-2},\ \sigma_\alpha^{-2},\ \sigma_\beta^{-2} &\sim\ Ga(0.01,\ 0.01),
\end{aligned}
$$

Now 10,000 is just some really big number. In fact, it is the square of 100. Let me define $w = 100$. Then the first equation says that the priors for α_0 and β_0 are normally distributed as follows:

$$P_{\alpha_0} = \frac{1}{\sqrt{2\pi}\,w} e^{-\alpha_0^2 / 2\,w^2}$$

$$P_{\beta_0} = \frac{1}{\sqrt{2\pi}\,w} e^{-\beta_0^2 / 2\,w^2}$$

The second equation says that we should define

$$s = \frac{1}{\sigma^2},\ s_\alpha = \frac{1}{\sigma_\alpha^2},\ \text{and}\ s_\beta = \frac{1}{\sigma_\beta^2}$$

and then the priors for s, s_α, and s_β are Gamma distributed. Let me define $v = 1/w = 0.01$. Then:

$$P_s = \frac{v^v}{\Gamma(v)}\, s^{v-1}\, e^{-v\,s}$$

$$P_{s_\alpha} = \frac{v^v}{\Gamma(v)}\, s_\alpha^{v-1}\, e^{-v\,s_\alpha}$$

$$P_{s_\beta} = \frac{v^v}{\Gamma(v)}\, s_\alpha^{v-1}\, e^{-v\,s_\beta}$$

$$P_s = \frac{v^v}{\Gamma(v)} s^{v-1} e^{-v\,s}$$

$$P_{s_\beta} = \frac{v^v}{\Gamma(v)} s_\alpha{}^{v-1} e^{-v\,s_\beta}$$

So we have a total of five priors, one for each parameter, and the complete prior is the product of all five priors. The reason they used $w = 100$ is that they were trying for very uninformative priors. I used w because in my mind it stood for "wide." The parameter v is also chosen to be uninformative (the way you make a gamma distribution very uninformative is that you choose tiny parameters.)

In fact, before we leave these priors, let me show you just how uninformative they are trying to be:

In[31]:= `Plot[` $\frac{1}{\sqrt{2\,Pi}\ w}$ `Exp[` $-\alpha_0{}^2 \big/ (2\ w^2)$ `] /. w → 100,`

     `{`α_0`, -100, 100}, PlotRange → {{-100, 100}, {0, 0.005}}]`

Out[31]=
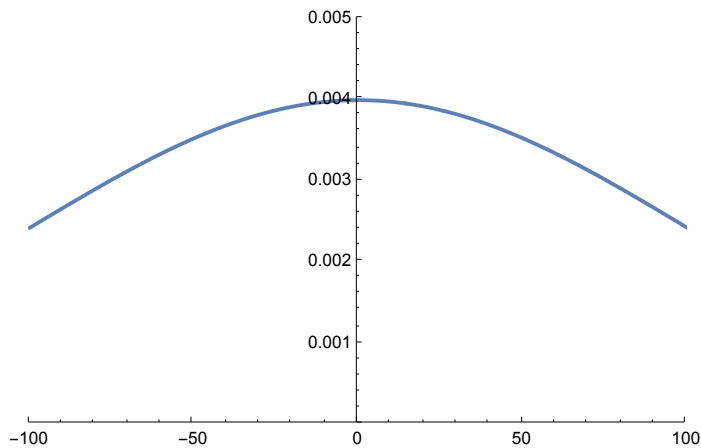
In[44]:= `Plot[` $\frac{v^v}{\text{Gamma}[v]}$ s^{v-1} `Exp[-v * s] /. v → 0.01,`

     `{s, 0, 100}, PlotRange → {{0, 100}, {0, 0.005}}]`

Out[44]=
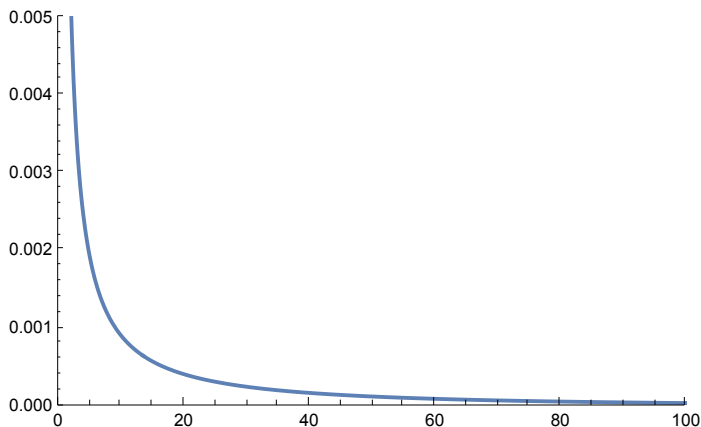
The Posterior

The Easy-to-Factor Axes

Well the posterior is the high-dimensional function we are trying to sample using Monte Carlo methods. It is the product of the three likelihood equations and the five prior equations above:

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \{\alpha_i\}, \{\beta_i\}) = \frac{1}{\left(\sqrt{2\pi/s}\right)^{i \cdot j}} e^{-s \sum_{i,j}(y_{i,j} - \alpha_i - \beta_i \tau_{i,j})^2/2} \cdot \frac{1}{\left(\sqrt{2\pi/s_\alpha}\right)^i} e^{-s_\alpha \sum_i(\alpha_i - \alpha_0)^2/2} \cdot \frac{1}{\left(\sqrt{2\pi/s_\beta}\right)^i} e^{-s_\beta \sum_i(\beta_i - \beta_0)^2/2} \cdot$$

$$\frac{1}{\sqrt{2\pi}\,w} e^{-\alpha_0^2/2\,w^2} \cdot \frac{1}{\sqrt{2\pi}\,w} e^{-\beta_0^2/2\,w^2} \cdot \frac{v^v}{\Gamma(v)} s^{v-1} e^{-v\,s} \cdot \frac{v^v}{\Gamma(v)} s_\alpha^{v-1} e^{-v\,s_\alpha} \cdot \frac{v^v}{\Gamma(v)} s_\beta^{v-1} e^{-v\,s_\beta}$$

This is our function of 217 parameters. Honestly, I can't even look at this function in its present form and see what it means. Let's rewrite it without all the constant prefactors:

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \{\alpha_i\}, \{\beta_i\}) \propto \left(\sqrt{s}\right)^{i \cdot j} e^{-s \sum_{i,j}(y_{i,j} - \alpha_i - \beta_i \tau_{i,j})^2/2} \cdot \left(\sqrt{s_\alpha}\right)^i e^{-s_\alpha \sum_i(\alpha_i - \alpha_0)^2/2} \cdot \left(\sqrt{s_\beta}\right)^i e^{-s_\beta \sum_i(\beta_i - \beta_0)^2/2}$$

$$\cdot e^{-\alpha_0^2/2\,w^2} e^{-\beta_0^2/2\,w^2} s^{v-1} e^{-vs} s_\alpha^{v-1} e^{-vs_\alpha} s_\beta^{v-1} e^{-vs_\beta}$$

Notice that in ignoring the constants, I couldn't get rid of all the powers of the various *s*'s. Those are not constants. Let's combine the powers of the various *s*'s up:

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \{\alpha_i\}, \{\beta_i\}) \propto$$
$$s^{v-1+i \cdot j/2} e^{-s \sum_{i,j}(y_{i,j} - \alpha_i - \beta_i \tau_{i,j})^2/2} \cdot s_\alpha^{v-1+i/2} e^{-s_\alpha \sum_i(\alpha_i - \alpha_0)^2/2} \cdot s_\beta^{v-1+i/2} e^{-s_\beta \sum_i(\beta_i - \beta_0)^2/2} e^{-\alpha_0^2/2\,w^2} e^{-\beta_0^2/2\,w^2} e^{-vs} e^{-vs_\alpha} e^{-vs_\beta}$$

Now it is up to us to see the way this thing factors. Rewrite:

$$e^{-s \sum_{i,j}(y_{i,j} - \alpha_i - \beta_i \tau_{i,j})^2/2} = \prod_i e^{-s \sum_j(y_{i,j} - \alpha_i - \beta_i \tau_{i,j})^2/2}$$

And rewrite:

$$e^{-s_\alpha \sum_i(\alpha_i - \alpha_0)^2/2} = \prod_i e^{-s\,s_\alpha(\alpha_i - \alpha_0)^2/2}$$

And rewrite:

$$e^{-s_\beta \sum_i(\beta_i - \beta_0)^2/2} = \prod_i e^{-s\,s_\beta(\beta_i - \beta_0)^2/2}$$

Then

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \{\alpha_i\}, \{\beta_i\}) =$$
$$s^{v-1+i \cdot j/2} \prod_i e^{-s \sum_j(y_{i,j} - \alpha_i - \beta_i \tau_{i,j})^2/2} \cdot s_\alpha^{v-1+i/2} \prod_i e^{-s_\alpha(\alpha_i - \alpha_0)^2/2} \cdot s_\beta^{v-1+i/2} \prod_i e^{-s_\beta(\beta_i - \beta_0)^2/2} e^{-\alpha_0^2/2\,w^2} e^{-\beta_0^2/2\,w^2} e^{-vs} e^{-vs_\alpha} e^{-vs_\beta}$$

We have the factorization we required in the previous write-up:

* https://brianhill.github.io/bayesian-statistics/resources/MonteCarloMethodsWhyDoTheyWork-III.nb.pdf

The 212 α_i and β_i are indeed tangled up with the five s, α_0, s_α, β_0, and s_β parameters, and α_i is tangled up with β_i but α_i and β_i are factored into products that don't involve of the 210 other α_j and β_j parameters.

That means that whenever we move in an α_i or β_i direction, we can treat as a constant multiplier the factors involving 210 of the parameters. In other words, we need only consider the tangled factors involving 7 parameters, which are:

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \alpha_i, \beta_i) \propto e^{-s \sum_j (y_{i,j} - \alpha_i - \beta_i \, \tau_{i,j})^2 / 2} \cdot e^{-s_\alpha (\alpha_i - \alpha_0)^2 / 2} \cdot e^{-s_\beta (\beta_i - \beta_0)^2 / 2}$$

The products are gone, and the set symbols around α_i and β_i are gone. This might still be a somewhat complicated computation. However, thinking of the first five of these parameters as given, as we bounce around in α_i and β_i we just have exponentials of quadratic functions of α_i and β_i. In other words, we just have ordinary Gaussians. And we know how to sample ordinary Gaussians.

The Recalcitrant Axes

Let's look at one of the recalcitrant axes, like the first one, s. Imagining that all the other four of the recalcitrant parameters and all of the 212 α_i and β_i are fixed, we can treat everything that has no dependence on s as a constant. We are left with:

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \{\alpha_i\}, \{\beta_i\}) \propto s^{i \cdot j / 2 + v - 1} \, e^{-s \sum_{i,j} (y_{i,j} - \alpha_i - \beta_i \, \tau_{i,j})^2 / 2} \, e^{-v \, s}$$

Is that not just a new Gamma distribution in s? Sure, the parameters that characterize this particular Gamma distribution involve 212 other parameters and the titer data. Regardless of the formula for the parameters, we know how to sample a Gamma distribution. Remember that those other 212 parameters are just fixed numbers when we are moving along the s axis.

Let's do one more recalcitrant axis just to see that we didn't get lucky with the first one. How about the α_0 axis? We can treat everything with no dependence on α_0 as a constant. We are left with:

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \alpha_i, \beta_i) \propto e^{-s_\alpha \sum_i (\alpha_i - \alpha_0)^2 / 2} \, e^{-\alpha_0^2 / 2 \, w^2}$$

That's just a Gaussian in α_0! Sure it involves 106 other parameters, but it is still a Gaussian and we know how to sample it. Remember that those other 106 parameters are fixed when we are moving along the α_0 axis.

$$P(s, \alpha_0, s_\alpha, \beta_0, s_\beta, \alpha_i, \beta_i) \propto e^{\cdots}$$

The Conclusion

You still have to write a lot of code to sample the posterior distribution with Gibbs sampling. Maybe that is why it took until 1994 for the BUGS program to mature and be published.

However, ***moving one axis at a time has for every parameter become a low-dimensional and tractable problem involving distributions we are already thoroughly familiar with.***

That means we can get a representative distribution out of this 217-dimensional space using Gibbs sampling. Using these representative samples, we can inquire about the most important results from the Hepatitis B study. The parameter we are most interested in would be α_0. Remember that roughly-speaking, α_0 controls the mean of the α_i via the likelihood function,

$$P_\alpha = \frac{1}{\left(\sqrt{2\pi} \, \sigma_\alpha \right)^i} \, e^{-\Sigma_i(\alpha_i - \alpha_0)^2 / 2\,\sigma_\alpha^2}$$

and the α_i represent the rate of increase of the Hepatitis B titers. We would also be interested in the variance of the α_i and that roughly speaking is in $s_\alpha = \sigma_\alpha^2$. However, we don't want to know the typical values of these things in the prior. We want to know what they central values and their ranges are in the posterior.