STATS370: Final Project

Erich Trieschman

2022 Fall Quarter

Introduction

- Compute quantiles, plot histograms, obtain means, variances and correlation matrix.
- Check convergence of the chain: does the variance of the distribution stabilize? Run two markov chains, how do they compare?
- Hamiltonian: Parameter search. Implement over M, eps, L. See section 3.5 for tuning
- Gibbs: basically show the math
- Metropolis Hasting: choice of candidate distribution. Make sure domain is right for each, could optimize parameters of the candidate distributions
- Importance sampling: TBD

1 Metropolis Hasting

Implementation

- Candidate distribution selection, including log-normal for ss with change of varibles transformation
- Grid search for potential parameters within chosen proposal distribution

Design choices, tuning, and scalability

Results

2 Gibbs sampling

Implementation

Posterior conditioned on each parameter

$$p(\theta[i]|Y,\theta[-i]) = \frac{p(\theta[i],\theta[-i]|Y)}{p(\theta[-i]|Y)} = f(\theta[i]) \propto p(\theta|Y) \text{ with fixed } \theta[-i], Y$$

Design choices, tuning, and scalability

Results

3 Hamiltonian Monte Carlo

Implementation

- Boundary constraints using Rollback HMC: https://arxiv.org/pdf/1709.02855.pdf and Reflection HMC: https://people.cs.umass.e papers/2015nips1.pdf

$$
\begin{aligned}
H(q,p) =& U(q) + K(p), \text{ Hamiltonian function}\\
&\text{with } U(\cdot), \text{ potential energy, } K(\cdot), \text{ kinetic energy, } q \in \mathbb{R}^k, \text{ position, } p \in \mathbb{R}^k \text{ momentum}\\
&\text{let } U(\theta) := -\log f(\theta), \text{ where } p(\theta|Y) = \frac{1}{C} f(\theta), \text{ and } p \sim N(0, \mathbf{M})
\end{aligned}
$$

Design choices, tuning, and scalability

Results

4 Importance sampling

Implementation

Here I implement importance sampling with normalization. This technique allows me to draw samples, θ_i, from a trial distribution, $q(\theta)$, and weight those samples to reflect my target distribution, $p(\theta \mid Y)$. Normalization allows me to weight samples based on a function proportional to my target distribution, $f(\theta)$, up to an integrating constant (i.e., $p(\theta \mid Y) = Cf(\theta)$).

To implement this method, I observe that my target distribution can be re-written as the outcome of sequential posterior distribution calculations. Specifically

$$
\begin{aligned}
p(\theta \mid Y) = p(\sigma^2, \tau, \mu, \gamma \mid Y) &\propto\ p(\sigma^2)p(\tau)p(\mu)p(\gamma) \times \prod_{i \in g_1, g_2, g_3, g_4} L(y_i \mid \sigma^2, \tau, \mu, \gamma)\\
&\propto\ (\sigma^2)p(\mu)p(\gamma) \times \prod_{i \in g_1, g_2} L(y_i \mid \sigma^2, \mu, \gamma) \times p(\tau) \times \prod_{i \in g_3, g_4} L(y_i \mid \sigma^2, \tau, \mu, \gamma)\\
&\propto\ p(\sigma^2, \mu, \gamma \mid Y_{1,2}) \times p(\tau) \times \prod_{i \in g_3, g_4} L(y_i \mid \sigma^2, \tau, \mu, \gamma)
\end{aligned}
$$

I next observe that the posterior, $p(\sigma^2, \mu, \gamma \mid Y_{1,2})$ can be computed in closed form.

$$
\begin{aligned}
&q(\sigma^2, \mu, \gamma \mid Y_{1,2}) = q(\mu, \gamma \mid Y_{1,2}) \times q(\sigma^2 \mid \mu, \gamma, Y_{1,2}), \text{ where}\\
&q(\mu, \gamma \mid Y_{1,2}) \sim t_4 \left(\mu = \begin{pmatrix} \overline{Y_1} \\ \overline{Y_2} \end{pmatrix}, \Sigma = \frac{\sum_{i \in g_1} \|y_i - \overline{Y_1}\|^2 + \sum_{i \in g_2} \|y_i - \overline{Y_2}\|^2}{2(n_1 + n_2) - 4} \begin{pmatrix} n_1^{-1} & 0 & 0 & 0 \\ 0 & n_1^{-1} & 0 & 0 \\ 0 & 0 & n_2^{-1} & 0 \\ 0 & 0 & 0 & n_2^{-1} \end{pmatrix} \right)\\
&q(\sigma^2 \mid \mu, \gamma, Y_{1,2}) \sim InvGamma \left(\alpha = n_1 + n_2, \beta = \frac{1}{2} \sum_{i \in g_1} \|y_i - \mu\|^2 + \sum_{i \in g_2} \|y_i - \gamma\|^2 \right)
\end{aligned}
$$

Lastly, I note that for a trial density $q(\theta) := p(\sigma^2, \mu, \gamma \mid Y_{1,2}) \times p(\tau)$, I can estimate unnormalized sample weights in a straightforward manner. Notably

$$
w_i := \frac{p(\theta_i)}{q(\theta_i)} = \frac{p(\sigma^2, \tau, \mu, \gamma \mid Y)}{p(\sigma^2, \mu, \gamma \mid Y_{1,2}) \times p(\tau)} \propto \frac{p(\sigma^2, \mu, \gamma \mid Y_{1,2}) \times p(\tau) \times \prod_{i \in g_3, g_4} L(y_i \mid \sigma^2, \tau, \mu, \gamma)}{p(\sigma^2, \mu, \gamma \mid Y_{1,2}) \times p(\tau)} = \prod_{i \in g_3, g_4} L(y_i \mid \sigma^2, \tau, \mu, \gamma) =: u_i
$$

Using the normalization method proved in class lecture, it follows cleanly that $w_i = u_i / \sum_{j=1}^n u_j$. My sampling method is therefore

Algorithm 1 Importance sampling

```
for t=1,...,n do
    θ_i ~ q(θ) := p(σ², μ, γ | Y_{1,2}) = q(μ, γ | Y_{1,2}) × q(σ² | μ, γ, Y_{1,2})
    u_i = f(θ_i | Y)/q(θ_i)                          ▷ f(θ_i | Y) ∝ p(θ_i | Y)
end for
w = u/ Σ_{j=1}^n u_j                                  ▷ u is a vector of all u_i
θ̂ = θ * w                     ▷ element-wise product; θ is a vector of all θ_i
```

Design choices, tuning, and scalability

Results

5 Conclusion

PLACEHOLDER – Compare methods

6 Appendix

6.1 Derivation of the posterior, $p(\theta \mid Y)$

6.1.1 Provided models for gene expression

$$
\begin{aligned}
(y_i|g_i = 1) &\sim N(\mu, \sigma^2 I)\\
(y_i|g_i = 2) &\sim N(\gamma, \sigma^2 I)\\
(y_i|g_i = 3) &\sim N(\frac{1}{2}(\mu+\gamma), \sigma^2 I)\\
(y_i|g_i = 4) &\sim N(\tau\mu + (1-\tau)\gamma, \sigma^2 I)
\end{aligned}
$$

6.1.2 Provided priors for gene expression models

$$
\begin{aligned}
\theta &= (\sigma^2, \tau, \mu_1, \mu_2, \gamma_1, \gamma_2)\\
p(\sigma^2) &\propto \frac{1}{\sigma^2}\\
p(\tau) &\sim Unif[0,1]\\
p(\mu) &= p(\mu_1, \mu_2) \propto 1 \text{ (improper uniform)}\\
p(\gamma) &= p(\gamma_1, \gamma_2) \propto 1 \text{ (improper uniform)}
\end{aligned}
$$

6.1.3 Derivation of likelihood and posterior distribution

$$
\begin{aligned}
L(Y|\theta) &= \prod_{i=1}^{n} p(y_i|\theta), \text{ for } Y = (y_i, \ldots, y_n)\\
&= \prod_{i\in g_1} \frac{1}{\sqrt{2\pi}}((\sigma^2)^2)^{-\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(y_i-\mu)^T(y_i-\mu)] \times \prod_{i\in g_2} \frac{1}{\sqrt{2\pi}}((\sigma^2)^2)^{-\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(y_i-\gamma)^T(y_i-\gamma)]\\
&\quad \times \prod_{i\in g_3} \frac{1}{\sqrt{2\pi}}((\sigma^2)^2)^{-\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(y_i-\frac{1}{2}(\mu+\gamma))^T(y_i-\frac{1}{2}(\mu+\gamma))]\\
&\quad \times \prod_{i\in g_4} \frac{1}{\sqrt{2\pi}}((\sigma^2)^2)^{-\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(y_i-(\tau\mu+(1-\tau)\gamma))^T(y_i-(\tau\mu+(1-\tau)\gamma))]\\
&= \left(\frac{1}{\sigma^2\sqrt{2\pi}}\right)^n \exp[-\frac{1}{2\sigma^2}(\sum_{i\in g_1}(y_i-\mu)^T(y_i-\mu) + \sum_{i\in g_2}(y_i-\gamma)^T(y_i-\gamma)\\
&\quad + \sum_{i\in g_3}(y_i-\frac{1}{2}(\mu+\gamma))^T(y_i-\frac{1}{2}(\mu+\gamma)) + \sum_{i\in g_4}(y_i-(\tau\mu+(1-\tau)\gamma))^T(y_i-(\tau\mu+(1-\tau)\gamma)))]
\end{aligned}
$$

$$p(\theta|Y) \propto p(\theta)L(Y|\theta)$$

6.2 Gibbs sampling

6.2.1 Derivation of $p(\tau \mid \sigma^2, \mu, \gamma, Y)$

$$
\begin{aligned}
p(\tau|Y,\theta[-\tau]) &\propto p(\theta)L(Y|\theta) \\
&\propto p(\tau)\prod_{i\in g_4}\frac{1}{\sigma^2\sqrt{2\pi}}\exp[-\frac{1}{2\sigma^2}(y_i-(\tau\mu+(1-\tau)\gamma))^T(y_i-(\tau\mu+(1-\tau)\gamma))] \\
&\propto \exp[-\frac{1}{2\sigma^2}\sum_{i\in g_4}(y_i-(\tau\mu+(1-\tau)\gamma))^T(y_i-(\tau\mu+(1-\tau)\gamma))], \text{ for } \tau\in[0,1] \\
&\propto \exp[-\frac{1}{2\sigma^2}\sum_{i\in g_4}(y_i^T y_i-2y_i^T(\tau\mu+(1-\tau)\gamma)+(\tau\mu+(1-\tau)\gamma)^T(\tau\mu+(1-\tau)\gamma))], \text{ for } \tau\in[0,1] \\
&\propto \exp[-\frac{1}{2\sigma^2}\sum_{i\in g_4}\tau^2(\mu^T\mu-2\mu^T\gamma+\gamma^T\gamma)-2\tau(y_i^T\mu-y_i^T\gamma-\mu^T\gamma+\gamma^T\gamma)], \text{ for } \tau\in[0,1] \\
&\propto \exp[-\frac{1}{2\sigma^2}(n_4\tau^2(\mu-\gamma)^T(\mu-\gamma)-2\tau(\mu-\gamma)^T[\sum_{i\in g_4}(y_i-\gamma), \text{ for } \tau\in[0,1] \\
&\propto \exp\left[-\frac{1}{2}*\frac{n_4(\mu-\gamma)^T(\mu-\gamma)}{\sigma^2}\left(\tau-\frac{(\mu-\gamma)^T(\sum_{i\in g_4}(y_i-\gamma))}{n_4(\mu-\gamma)^T(\mu-\gamma)}\right)^2\right], \text{ for } \tau\in[0,1] \\
&\sim Norm\left(\mu=\frac{(\mu-\gamma)^T(\sum_{i\in g_4}(y_i-\gamma))}{n_4(\mu-\gamma)^T(\mu-\gamma)}, \sigma^2=\frac{\sigma^2}{n_4(\mu-\gamma)^T(\mu-\gamma)}\right), \text{ truncated to } [0,1]
\end{aligned}
$$

Derivation of $p(\sigma^2 \mid \tau, \mu, \gamma, Y)$

$$
\begin{aligned}
p(\sigma^2 \mid Y,\theta[-\sigma^2]) \propto & p(\theta)L(Y|\theta) \propto p(\sigma^2)\prod_{i=1}^{n}p(y_i|\theta) \\
\propto & \frac{1}{\sigma^2}*\left(\frac{1}{\sigma^2}\right)^n\exp\left[-\frac{1}{2\sigma^2}M\right], \text{ where } M= \\
& \sum_{i\in g_1}(y_i-\mu)^T(y_i-\mu)+\sum_{i\in g_2}(y_i-\gamma)^T(y_i-\gamma)+\sum_{i\in g_3}(y_i-\frac{1}{2}(\mu+\gamma))^T(y_i-\frac{1}{2}(\mu+\gamma)) \\
& +\sum_{i\in g_4}(y_i-(\tau\mu+(1-\tau)\gamma))^T(y_i-(\tau\mu+(1-\tau)\gamma)) \\
\propto & (\sigma^2)^{-n-1}\exp\left[-\frac{M}{2}\frac{1}{\sigma^2}\right] \\
\sim & InvGamma\left(\alpha=n,\beta=\frac{M}{2}\right)
\end{aligned}
$$

6.2.2 Derivation of $p(\mu \mid \sigma^2, \tau, \gamma, Y)$

$$
\begin{aligned}
p(\mu|Y,\theta[-\mu]) \propto & p(\theta)L(Y|\theta) \\
& \propto \prod_{i\in g_1} \frac{1}{\sigma^2\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(y_i-\mu)^T(y_i-\mu)] \times \prod_{i\in g_3} \frac{1}{\sigma^2\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(y_i-\frac{1}{2}(\mu+\gamma))^T(y_i-\frac{1}{2}(\mu+\gamma))] \\
& \quad \times \prod_{i\in g_4} \frac{1}{\sigma^2\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(y_i-(\tau\mu+(1-\tau)\gamma))^T(y_i-(\tau\mu+(1-\tau)\gamma))] \\
& \propto \exp[-\frac{1}{2\sigma^2}(n_1\mu^T\mu - 2\mu^T\left(\sum_{i\in g_1} y_i\right) - \mu^T\left(\sum_{i\in g_3} y_i\right) + \frac{n_3}{4}\mu^T\mu + \frac{n_3}{2}\mu^T\gamma \\
& \quad - 2\tau\mu^T\left(\sum_{i\in g_4} y_i\right) + \tau^2 n_4\mu^T\mu + 2\tau(1-\tau)n_4\mu^T\gamma)] \\
& \propto \exp\left[-\frac{1}{2\sigma^2}(n_1+\frac{n_3}{4}+n_4\tau^2)\left(\mu^T\mu - 2\mu^T\frac{\sum_{i\in g_1} y_i + \frac{1}{2}\sum_{i\in g_3} y_i + \tau\sum_{i\in g_4} y_i - (\frac{n_3}{4}+n_4\tau(1-\tau))\gamma}{n_1+\frac{n_3}{4}+n_4\tau^2}\right)\right] \\
& \propto \exp\left[-\frac{1}{2\phi^2}(\mu-\psi)^T(\mu-\psi)\right], \text{ where} \\
& \quad \psi = \frac{\sum_{i\in g_1} y_i + \frac{1}{2}\sum_{i\in g_3} y_i + \tau\sum_{i\in g_4} y_i - (\frac{n_3}{4}+n_4\tau(1-\tau))\gamma}{n_1+\frac{n_3}{4}+n_4\tau^2}, \ \phi^2 = \frac{\sigma^2}{n_1+\frac{n_3}{4}+n_4\tau^2} \\
& \sim N(\mu=\psi, \Sigma=\phi^2 I)
\end{aligned}
$$

6.2.3 Derivation of $p(\gamma \mid \sigma^2, \tau, \mu, Y)$

By symmetry with posterior conditional probability of μ,

$$
\begin{aligned}
p(\gamma|Y,\theta[-\gamma]) \sim & N(\mu=\psi', \Sigma=\phi'^2 I), \text{ where} \\
& \psi' = \frac{\sum_{i\in g_2} y_i + \frac{1}{2}\sum_{i\in g_3} y_i + (1-\tau)\sum_{i\in g_4} y_i - (\frac{n_3}{4}+n_4\tau(1-\tau))\mu}{n_2+\frac{n_3}{4}+n_4(1-\tau)^2}, \ \phi'^2 = \frac{\sigma^2}{n_2+\frac{n_3}{4}+n_4(1-\tau)^2}
\end{aligned}
$$

6.3 Importance sampling

6.3.1 Derivation of $p(\sigma^2 \mid \mu, \gamma, Y_{1,2})$

$$
\begin{aligned}
q(\sigma^2 \mid \mu,\gamma,Y_{1,2}) \propto & \ \frac{1}{\sigma^2} * \left(\frac{1}{\sigma^2}\right)^{n_1+n_2} \exp\left[-\frac{1}{2\sigma^2}M\right], \text{ where } M = \sum_{i\in g_1}(y_i-\mu)^T(y_i-\mu) + \sum_{i\in g_2}(y_i-\gamma)^T(y_i-\gamma) \\
\propto & \ (\sigma^2)^{-n_1-n_2-1} \exp\left[-\frac{M}{2}\frac{1}{\sigma^2}\right] \\
\sim & InvGamma\left(\alpha = n_1+n_2, \beta = \frac{M}{2}\right)
\end{aligned}
$$

6.3.2 Derivation of $p(\mu, \gamma \mid Y_{1,2})$

$$
\begin{aligned}
q(\mu,\gamma \mid Y_{1,2}) &= \int_0^\infty q(\mu,\gamma,\sigma^2 \mid Y_{1,2}) d\sigma^2 \\
&\propto \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{n_1+n_2+n_3+1} \exp\left[-\frac{1}{2\sigma^2}(\sum_{i\in g_1}\|y_i-\mu\|^2 + \sum_{i\in g_2}\|y_i-\gamma\|^2\right] d\sigma^2 \\
&\propto \Gamma(n_1+n_2)\left(-\frac{1}{2}\sum_{i\in g_1}\|y_i-\mu\|^2 + \sum_{i\in g_2}\|y_i-\gamma\|^2\right)^{-(n_1+n_2)}, \text{ where } \frac{\Gamma(\alpha)}{\beta^\alpha} = \int_0^\infty (\sigma^2)^{-\alpha-1}\exp\left(\frac{\beta}{\sigma^2}\right)d\sigma^2 \\
&\propto (M + n_1\|\overline{Y_1}-\mu\|^2 + n_2\|\overline{Y_2}-\gamma\|^2)^{-(n_1+n_2)}, \text{ where } M = \sum_{i\in g_1}\|y_i-\overline{Y_1}\|^2 + \sum_{i\in g_2}\|y_i-\overline{Y_2}\|^2 \\
&\propto \left(M + \left[\begin{pmatrix}\mu\\ \gamma\end{pmatrix} - \begin{pmatrix}\overline{Y_1}\\ \overline{Y_2}\end{pmatrix}\right]^T \begin{pmatrix} n_1 & 0 & 0 & 0\\ 0 & n_1 & 0 & 0\\ 0 & 0 & n_2 & 0\\ 0 & 0 & 0 & n_2\end{pmatrix} \left[\begin{pmatrix}\mu\\ \gamma\end{pmatrix} - \begin{pmatrix}\overline{Y_1}\\ \overline{Y_2}\end{pmatrix}\right]\right)^{-(n_1+n_2)} \\
&\propto \left(1 + \frac{1}{2(n_1+n_2)-4} \times \frac{2(n_1+n_2)-4}{M}\left[\begin{pmatrix}\mu\\ \gamma\end{pmatrix} - \begin{pmatrix}\overline{Y_1}\\ \overline{Y_2}\end{pmatrix}\right]^T \begin{pmatrix} n_1{-}1 & 0 & 0 & 0\\ 0 & n_1^{-1} & 0 & 0\\ 0 & 0 & n_2{-}1 & 0\\ 0 & 0 & 0 & n_2{-}1\end{pmatrix}^{-1} \left[\begin{pmatrix}\mu\\ \gamma\end{pmatrix} - \begin{pmatrix}\overline{Y_1}\\ \overline{Y_2}\end{pmatrix}\right]\right)^{-(n_1+n_2)} \\
&\propto \left[1 + \frac{1}{\nu}(x-\eta)^T\Sigma^{-1}(x-\eta)\right]^{-\frac{\nu+4}{2}} \text{, where} \\
&\quad \nu = 2(n_1+n_2)+4, \Sigma = \frac{M}{2(n_1+n_2)-4}\begin{pmatrix} n_1{-}1 & 0 & 0 & 0\\ 0 & n_1^{-1} & 0 & 0\\ 0 & 0 & n_2{-}1 & 0\\ 0 & 0 & 0 & n_2{-}1\end{pmatrix}, \eta = \begin{pmatrix}\overline{Y_1}\\ \overline{Y_2}\end{pmatrix}
\end{aligned}
$$

$$
q(\mu,\gamma \mid Y_{1,2}) \sim t_4\left(\mu = \begin{pmatrix}\overline{Y_1}\\ \overline{Y_2}\end{pmatrix}, \Sigma = \frac{\sum_{i\in g_1}\|y_i-\overline{Y_1}\|^2 + \sum_{i\in g_2}\|y_i-\overline{Y_2}\|^2}{2(n_1+n_2)-4}\begin{pmatrix} n_1^{-1} & 0 & 0 & 0\\ 0 & n_1^{-1} & 0 & 0\\ 0 & 0 & n_2^{-1} & 0\\ 0 & 0 & 0 & n_2^{-1}\end{pmatrix}\right)
$$