10/10

M525: HW3

Dominic Bair

November 21, 2022

1. Consider the following hierarchical model:

$$\begin{aligned} \tau &\sim \text{Gamma}(\phi, \psi) \\ \mu|\tau &\sim \text{Normal}\left(m, \frac{n}{\tau}\right) \\ w|\mu, \tau &\sim \text{Normal}\left(\mu, \frac{1}{\tau}\right) \end{aligned}$$

where ϕ, ψ, m, and n are known values.

(i) First we create Monte–Carlo simulation to approximate $p(w)$ empirically. See Figure 1.

(ii) Now we derive $p(w)$ analytically.

$$\begin{aligned} p(w, \mu, \tau) &= p(w|\mu, \tau)p(\mu|\tau)p(\tau). \\ \implies p(w) &= \int_0^\infty \int_{-\infty}^\infty p(w|\mu, \tau)p(\mu|\tau)p(\tau)\, d\mu\, d\tau \\ &= \int_0^\infty p(\tau) \int_{-\infty}^\infty p(w|\mu, \tau)p(\mu|\tau)\, d\mu\, d\tau \\ &\vdots \\ &= \frac{1}{\sqrt{2\pi(n+1)}} \frac{1}{\Gamma(\phi)\psi^\phi} \Gamma\left(\phi + \frac{1}{2}\right) \left(\frac{2\psi(n+1)}{2(n+1) + \psi(m-w)^2}\right). \end{aligned}$$

(iii) We now compare the histogram generated in (i) with the analytical solution for $p(w)$ found in (ii).

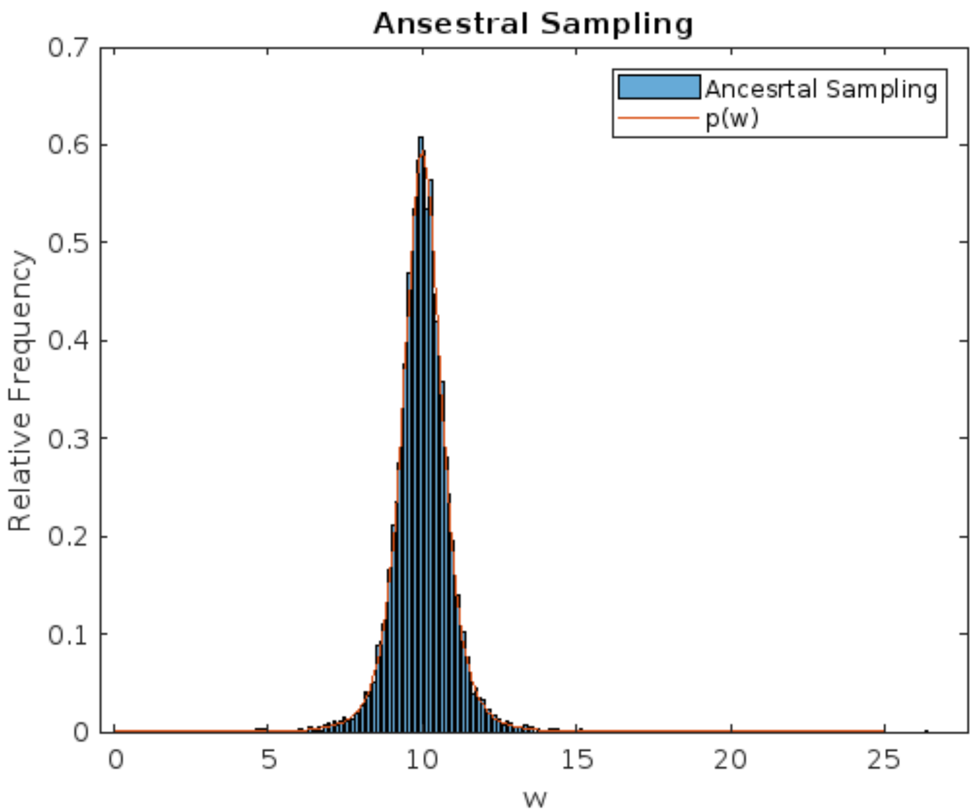

Figure 1: Sampling 10000 different τ values, then feeding that sample forward through our model, we get this histogram. Notice our $p(w)$ follows the distribution so well, it is difficult to distinguish it from the histogram.

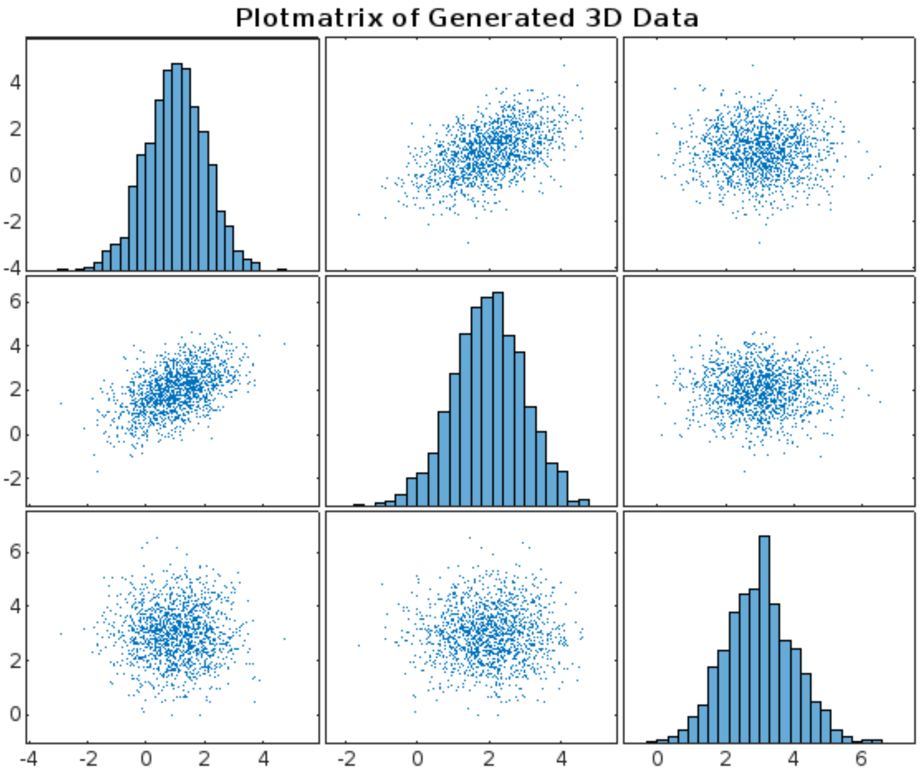

Figure 2: A visualization of 3 dimensional random data with distribution described in 3 (i).

2. (i) All the variables depicted in the graphic are Ψ, M, V, U, μ_1, μ_2, $w_{1,1:N_1}$, and $w_{2,1:N_2}$.

   (ii) $w_{1,1:N_1}$ and $w_{2,1:N_2}$ are random variables with known values. M, V, U, μ_1, and μ_2 are random variables with unknown values. Ψ is a variable that is not random.

3. Now we implement a Monte Carlo method for estimating a multidimensional Gaussian integral.

   (i) We, consider J random variables distributed as

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_i \sim \text{Normal}_3\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \quad i = 1, \ldots, J$$

   (ii) We take $J = 1500$ and generate the data. See Figure 2.

   (iii) Finally, we estimate the $P(\|\mathbf{w}\|_2 \leqslant 1)$ by a Monte Carlo estimation, that is taking our generated points, counting the number of them satisfying $\|\mathbf{w}\|_2 \leqslant 1$, then, then dividing them by J. With 1500 points, we get $P(\|\mathbf{w}\|_2 \leqslant 1) \approx 0.0027$. Choosing J to be 1000000, we find $P(\|\mathbf{w}\|_2 \leqslant 1) \approx 0.0010$. We note that estimating the probability of such an unlikely event could be better approximated by using importance sampling.