BDA Assignment 1 Report

Vijay Varma - AI20BTECH11012
Nelakuditi Rahul Naga - AI20BTECH11029
Rahul V - AI20BTECH11030

Proposal Distribution

In MCMC, $\mu^{'}$ and $\tau^{'}$ are drawn independently from Gaussian/Normal distributions with standard deviation of 0.05 for both μ and τ. Note that this distribution is **symmetric**. Specifically, the proposal distributions $q(\mu^{'} \mid \mu)$ and $q(\tau^{'} \mid \tau)$ are given by :

$$q(\mu^{'} \mid \mu) = \mathcal{N}(\mu, 0.05)$$
$$q(\tau^{'} \mid \tau) = \mathcal{N}(\tau, 0.05)$$

Note that two separate proposal distributions have been used for μ and τ as we know that they are independent.

Approximate Distribution

In Mean field Variational Inference, the approximate distribution that we have obtained for μ is a Normal distribution with mean μ_N and precision λ_N. The exact expressions for these parameters are mentioned in the handwritten pdf attached below. The final values obtained for these parameters after convergence are around -2.71 and 17.67 respectively. While, for τ, the approximate distribution is a gamma distribution with shape parameter α_N and rate parameter β_N. The exact expressions for these parameters are mentioned in the handwritten pdf attached below. The final values obtained for these parameters after convergence are around 6.001 and 4.07 respectively.

Mathematical Expression for the lower bound for VI

Please refer to the handwritten pdf attached below for this section.

Sampler and Burn-in period

We have used the Metropolis–Hastings sampler with the proposal distribution mentioned above for the MCMC implementation. With this sampling process, we have obtained a burn-in period of around 700-800 while sampling 10,000 samples. While for Mean field Variational Inference, we have sampled the parameters from the approximate distributions mentioned above.

Experimental setup

The implementation of all the aforementioned algorithms has been carried out using a Interactive Python Notebook in the unpaid version of Google Colab.

Observations and Results

MCMC

Statistics and Results from Metropolis–Hastings algorithm:

Parameter	Mean	Standard Deviation
μ	-2.56	0.173
τ	7.53	6.083
pop.mean	0.073	0.012
σ	0.438	0.158

Actual number of failures : [0 18 8 46 8 13 9 31 14 8 29 24]
Predicted number of failures : [0 15 7 55 12 11 14 35 19 10 30 24]

Variational Inference

Statistics and Results from Mean field VI algorithm:

Parameter	Mean	Standard Deviation
μ	-2.556	0.140
τ	5.518	2.606
pop.mean	0.072	0.009
σ	0.461	0.109

Actual number of failures : [0 18 8 46 8 13 9 31 14 8 29 24]
Predicted number of failures : [0 20 11 44 12 17 10 24 14 9 27 33]

The details about other statistics and various other plots have been provided in the IPython notebook that has been submitted along with this report.

Comparison of MCMC and VI

We can conclude from the above results that MCMC and VI algorithms are both good for solving this problem as they give similar results for the predicted number of failures. The mathematical details regarding how the Metropolis Hastings (MCMC) and Mean field VI algorithms have been implemented in Python is explained in a detailed manner in the handwritten pdf attached below.

References

The references that we have used are :

- Medium Blog on MCMC - Link
- Medium Blog on Variational Inference - Link
- Variational Inference Wikipedia - Link

* "MCMC":

→ we will implement the Metropolis Hastings algorithm to draw samples from the posterior. for this problem, we are given 'r' and 'n' as data. we can find $b = \text{logit}\left(\frac{r}{n}\right)$ from given 'r' and 'n', and use that as data to simplify the problem.

→ We have 2 parameters - μ & τ in this problem i.e; $\theta = (\mu, \tau)$.

→ As μ, τ are independent in given problem, we will draw them independently from a "symmetric" Gaussian/Normal distributions (this is the proposal distribution). Denote $\theta' = (\mu', \tau')$ as parameters drawn from proposal distribution. As q is symmetric we have:

$$A(\theta, \theta') = \min\left\{1, \frac{P(\theta').\, \cancel{q(\theta|\theta')}}{P(\theta).\, \cancel{q(\theta'|\theta)}}\right\}$$

$$= \min\left\{1, \frac{P(\theta')}{P(\theta)}\right\}$$

where $p(\cdot)$ is the unnormalized posterior.

→ we will work with logarithm of $A(\theta, \theta')$ for numerical stability.

→ $\log(A(\theta, \theta')) = \min\{0, \log(P(\theta')) - \log(P(\theta))\}$

→ Now, $P(\theta) = \underbrace{\prod_{i=1}^{N} f(b_i|\theta)}_{\text{Likelihood}} . \underbrace{P(\theta)}_{\text{Prior}}$ $(\theta = (\mu, \tau))$

⇒ As $b_i \sim N(\mu, \tau)$ ⇒ $f(b_i|\mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \cdot e^{-\frac{\tau}{2}(b_i - \mu)^2}$

⇒ $P(\mu, \tau) = P(\mu).P(\tau) = \left(\sqrt{\frac{10^{-6}}{2\pi}} \cdot e^{-\frac{10^{-6}}{2}.\mu^2}\right) . \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \tau^{\alpha-1} . e^{-\beta\tau}\right)$

where $\alpha = \beta = 10^{-3}$. $(\because \mu \sim N(0, 10^{-6}),\ \tau \sim \text{Gamma}(0.001, 0.001))$

→ Hence, $p(\theta) = \left(\sqrt{\frac{\tau}{2\pi}}\right)^N . e^{-\frac{\tau}{2}.\sum_{i=1}^{N}(b_i - \mu)^2} . P(\mu).P(\tau)$

Hence we have :-

$$\log(P(\Theta)) = N.\log\left(\sqrt{\frac{\tau}{2\pi}}\right) - \frac{\tau}{2}.\sum_{i=1}^{N}(b_i-\mu)^2 - \frac{10^{-6}}{2}.\mu^2 + (\alpha-1)\log\tau - \beta\tau + \text{const.}$$

→ we can replace μ, τ by μ', τ' in above equation to get $\log(P(\Theta'))$.

* "Variational Inference" :-

→ for our problem, we are given :-

$$\tau \sim \text{Gamma}(\alpha, \beta)$$

$$\mu \sim N(\mu_0, \underbrace{a_0}_{\text{Precision}})$$

$$b_i \sim N(\mu, \tau)$$ (we will consider b_i as data as mentioned earlier in MCMC)

→ $q(\mu,\tau) = q(\mu).q(\tau)$ for Mean field V.I. we know that :-

$$\ln(q^*_\mu(\mu)) = \underset{\tau}{E}[\ln(P(b,\mu,\tau))]$$

$$\ln(q^*_\tau(\tau)) = \underset{\mu}{E}[\ln(P(b,\mu,\tau))]$$

→ $P(b,\mu,\tau) = P(b|\mu,\tau).P(\mu).P(\tau)$

$$\therefore \ln(q^*_\mu(\mu)) = \underset{\tau}{E}[\ln(P(b|\mu,\tau)) + \ln(P(\mu)) + \underbrace{\ln(P(\tau))}_{\text{const. wrt. }\mu}]$$

$$= \underset{\tau}{E}\left[\ln\left(\prod_{i=1}^{N}\sqrt{\frac{\tau}{2\pi}}.e^{-\frac{(b_i-\mu)^2\tau}{2}}\right) + \ln\left(\sqrt{\frac{a_0}{2\pi}}.e^{-\frac{(\mu-\mu_0)^2.a_0}{2}}\right)\right]$$

$$= \underset{\tau}{E}\left[\frac{N}{2}.\ln\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2}.\sum_{i=1}^{N}(b_i-\mu)^2 + \frac{1}{2}\ln\left(\frac{a_0}{2\pi}\right) - \frac{(\mu-\mu_0)^2}{2}.a_0\right]$$

$$= -\frac{E_\tau[\tau]}{2}.\sum_{i=1}^{N}(b_i-\mu)^2 - \frac{a_0}{2}(\mu-\mu_0)^2 + \text{const.}$$

$$= -\frac{E_\tau[\tau]}{2}\left(\sum_i b_i^2 + N.\mu^2 - 2\mu.\sum_i b_i + \frac{a_0}{E_\tau[\tau]}(\mu^2+\mu_0^2-2\mu\mu_0)\right)$$

$$= -\frac{E_\tau[\tau]}{2}\left(\left(\frac{a_0}{E_\tau(\tau)}+N\right)\mu^2 - 2\mu\left(\sum_i b_i + \frac{\mu_0.a_0}{E_\tau(\tau)}\right)\right)$$

$$= -\frac{(a_0+N.E_\tau(\tau))}{2}\left(\mu^2 - 2\mu\left(\frac{E_\tau(\tau).\sum_i b_i + \mu_0 a_0}{a_0+N.E_\tau(\tau)}\right)\right)$$

$$= -\frac{(a_0 + N \cdot E_\tau[\tau])}{2} \cdot \left(\mu - \left(\frac{E_\tau(\tau) \cdot \sum_i b_i + \mu_0 a_0}{a_0 + N \cdot E_\tau(\tau)} \right) \right)^2 + \text{Const.}$$

Clearly, $q^*_\mu(\mu)$ is a Gaussian distribution $N(\mu \mid \mu_N, \lambda_N)$ where:-

$$\mu_N = \frac{E_\tau(\tau) \cdot \sum_i b_i + \mu_0 a_0}{a_0 + N \cdot E_\tau(\tau)}, \quad \lambda_N = a_0 + N \cdot E_\tau[\tau]$$

→ Now,

$$\ln q^*_\tau(\tau) = \underset{\mu}{E}\left[\ln(p(b \mid \mu, \tau)) + \underbrace{\ln(p(\mu))}_{\text{Const. wrt } \tau} + \ln(p(\tau)) \right]$$

→ As $\tau \sim \text{Gamma}(\alpha, \beta)$; $\ln(p(\tau)) = (\alpha - 1)\ln\tau - \beta\tau + \text{Const.}$

→ Now $\underset{\mu}{E}[\ln(P(b \mid \mu, \tau))] = \underset{\mu}{E}\left[\frac{N}{2} \cdot \ln\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2} \cdot \sum_{i=1}^{N} (b_i - \mu)^2 \right]$

$$= \frac{N}{2} \ln(\tau) - \frac{\tau}{2} \cdot \underset{\mu}{E}\left[\sum_{i=1}^{N} (b_i - \mu)^2 \right] + \text{Const.}$$

$$\therefore \ln(q^*_\tau(\tau)) = \left(\frac{N}{2} + \alpha - 1 \right) \ln\tau - \tau\left[\beta + \frac{1}{2} \cdot \underset{\mu}{E}\left(\sum_{i=1}^{N} (b_i - \mu)^2 \right) \right]$$

Clearly, $q^*_\tau(\tau)$ is a Gamma distribution $\text{Gamma}(\tau \mid \alpha_N, \beta_N)$ where:-

$$\alpha_N = \frac{N}{2} + \alpha, \quad \beta_N = \beta + \frac{1}{2} \cdot \underset{\mu}{E}\left[\sum_{i=1}^{N} (b_i - \mu)^2 \right]$$

→ α_N is fixed but $\mu_N, \lambda_N, \beta_N$ have a cyclic dependency between them.

→ $E_\tau[\tau] = \frac{\alpha_N}{\beta_N} \Rightarrow \mu_N = \frac{\left(\frac{\alpha_N}{\beta_N}\right) \cdot \sum_i b_i + \mu_0 a_0}{a_0 + N \cdot \left(\frac{\alpha_N}{\beta_N}\right)}, \quad \lambda_N = a_0 + N \cdot \frac{\alpha_N}{\beta_N}$

→ $\beta_N = \beta + \frac{1}{2} \cdot \underset{\mu}{E}\left[\sum_i b_i^2 + N \cdot \mu^2 - 2\mu \cdot \sum_i b_i \right]$

→ $E(\mu) = \mu_N$, $E(\mu^2) = Var(\mu) + (E(\mu))^2 = \frac{1}{\lambda_N} + \mu_N^2$ $\left(\because \text{Variance} = \frac{1}{\text{Precision}}\right)$

$$\therefore \beta_N = \beta + \frac{1}{2}\left(\sum_i b_i^2 + N \cdot \left(\frac{1}{\lambda_N} + \mu_N^2 \right) - 2 \cdot \sum_i b_i \cdot \mu_N \right)$$

* "Evidence Lower Bound for V.I" :-

→ for our mean field approximation, the evidence lower bound is given by :-

$$L = \underset{\mu,\tau}{E}\left(\log(P(\mu|\tau,b)) + \log(P(\tau|\mu,b))\right) - \underset{\mu}{E}[\log(q(\mu))] - \underset{\tau}{E}[\log(q(\tau))]$$

→ for $q(\mu), q(\tau)$ we will use the approximate distributions we calculated earlier.

$\therefore \underset{\mu,\tau}{E}(\log(P(\mu|\tau,b)))$, $p(\mu|\tau,b) = p(\mu|b)$ ($\because \mu,\tau$ are independent)

$$= \frac{p(b|\mu)\cdot p(\mu)}{p(b)} \propto p(b|\mu)\cdot p(\mu)$$

$$\to \log(P(\mu|\tau,b)) = \log(P(b|\mu)) + \log(P(\mu))$$

$$= \frac{N}{2}\cdot\log\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2}\cdot\sum_{i=1}^{N}(b_i-\mu)^2 + \frac{1}{2}\cdot\log\left(\frac{1e-6}{2\pi}\right) - \frac{10^{-6}}{2}\cdot\mu^2$$

$$E(\log(x)) \approx$$

$$\to \log(E(x)) - \frac{V(x)}{2\cdot(E(x))^2}$$

$1\ (\because \alpha=\beta=10^{-3})$

$1/2\alpha$

$$\therefore \underset{\mu,\tau}{E}(\log(P(\mu|\tau,b))) = \frac{N}{2}\left(\log(\cancel{E(\tau)}) - \frac{V(\tau)}{2\cdot E(\tau)^2}\right) - \frac{N}{2}\cdot\log(2\pi) - \frac{E[\tau]}{2}\sum_{i=1}^{N}E[(b_i-\mu)^2]$$

$$+ \frac{1}{2}\log\left(\frac{10^{-6}}{2\pi}\right) - \frac{10^{-6}}{2}\cdot E[\mu^2]$$

$$\to E(b_i^2+\mu^2-2\cdot b_i\cdot\mu) = b_i^2 + E(\mu^2) - 2b_i\cdot E(\mu)$$

$$\Downarrow$$

$$\sum_{i=1}^{N}E[(b_i-\mu)^2] = \sum_i b_i^2 + N\cdot E(\mu^2) - 2\left(\sum_i b_i\right)\cdot E(\mu)$$

$$\to E(\mu) = 0,\quad E(\mu^2) = Var(\mu) + (E(\mu))^2 = \frac{1}{10^{-6}} = 10^6$$

$$\therefore \underset{\mu,\tau}{E}(\log(P(\mu|\tau,b))) = \frac{N}{2}\left(\frac{-1}{2\alpha}\right) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\cdot\left(\sum_i b_i^2 + 10^6 N\right)$$

$$+ \frac{1}{2}\log\left(\frac{10^{-6}}{2\pi}\right) - \frac{1}{2} \longrightarrow (1)$$

→ Similarly, $\log(P(\tau|\mu,b)) = \log(P(b|\tau)) + \log(P(\tau))$

$$= \frac{N}{2}\cdot\log\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2}\cdot\sum_{i=1}^{N}(b_i-\mu)^2 + (\alpha-1)\ln\tau - \beta\tau + \log\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)$$

→ Taking expectation,

$$\underset{\mu,\tau}{E}(\log(P(\tau|\mu,b))) = \frac{N}{2}\left(\frac{-1}{2\alpha}\right) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\left(\sum b_i^2 + 10^6 N\right) + \log\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)$$

$$+ (\alpha-1)\left(\frac{-1}{2\alpha}\right) - \cancel{\beta}\cdot\frac{\alpha}{\cancel{\beta}} \longrightarrow (2)$$

Now, $\underset{\mu}{E}[\log(q(\mu))] = \underset{\mu}{E}\left[\log\left(\sqrt{\frac{\lambda_N}{2\pi}}\cdot e^{-\frac{\lambda_N}{2}\cdot(\mu-\mu_N)^2}\right)\right]$

$\text{Var}(\mu) = 1/\cancel{\lambda_N}$

$= \frac{1}{2}\cdot\log\left(\frac{\lambda_N}{2\pi}\right) - \frac{\cancel{\lambda_N}}{2}\cdot \cancel{\underset{\mu}{E}[\mu-\mu_N]^2}$

$= \frac{1}{2}\cdot\log\left(\frac{\lambda_N}{2\pi}\right) - \frac{1}{2} \rightarrow (3)$

$\rightarrow \underset{\tau}{E}[\log(q(\tau))] = \underset{\tau}{E}\left[\log\left(\frac{\beta_N^{\alpha_N}}{\Gamma(\alpha_N)}\cdot\tau^{\alpha_N-1}\cdot e^{-\beta_N\cdot\tau}\right)\right]$

$= \log\left(\frac{\beta_N^{\alpha_N}}{\Gamma(\alpha_N)}\right) + \alpha_N - 1 \cdot \underset{\tau}{E}(\log\tau) - \cancel{\beta_N \cdot E[\tau]} \; \alpha_N/\cancel{\beta_N}$

$\log(\alpha_N/\beta_N) - 1/2\alpha_N$

$= \log\left(\frac{\beta_N^{\alpha_N}}{\Gamma(\alpha_N)}\right) - \alpha_N + (\alpha_N-1)\cdot\left(\log\alpha_N - \log\beta_N - \frac{1}{2\alpha_N}\right) \rightarrow (4)$

$\rightarrow$ from (1), (2), (3), (4) :-

$L = \frac{-N}{2\alpha} - N\log(2\pi) - \sum_i b_i^2 - 10^6\cdot N + \frac{1}{2}\log\left(\frac{10^{-6}}{2\pi}\right) \cancel{-\frac{1}{2}} + \frac{1-\alpha}{2\alpha} - \alpha + \cancel{\frac{1}{2}} - \frac{1}{2}\log\left(\frac{\lambda_N}{2\pi}\right)$

$- \log\left(\frac{\beta_N^{\alpha_N}}{\Gamma(\alpha_N)}\right) + \alpha_N - (\alpha_N-1)\left(\log\alpha_N - \log\beta_N - \frac{1}{2\alpha_N}\right)$