DD2434 - Assignment 2

Nuo Chen - nuoc@kth.se - GITHUB LINK

December 12, 2020

1 Assignment 2

1.1 Problem 2.1

2.1.1	Yes
2.1.2	Yes
2.1.3	A = $\{ \mu_{r,c} \}$
2.1.4	No
2.1.5	Yes
2.1.6	B = $\{C^n, Z^n_m \ : \ n \in [N], m \in [M] \}$

Table 1: Answers to the sub-problems in Problem 2.1

1.2 Problem 2.2

1.2.1 Problem 2.2.7

The likelihood $p(\beta|T, \Theta)$ of a tree is given by the product below:

$$p(\beta|T, \Theta) = p(X_u = i|X_o) \quad [1.1]$$

$$p(X_u = i|X_o) \propto p(X_u = i, X_o) = \{\text{product rule}\} = p(X_{o\cap\uparrow u}, X_u = i)p(X_{o\cap\downarrow u}|X_u = i) \quad [1.2]$$

$$p(X_{o\cap\uparrow u}, X_u = i) = t(u, i) \quad [1.3]$$

$$p(X_{o\cap\downarrow u}, X_u = i) = s(u, i) \quad [1.4]$$

where X_u is a leaf node and i is the corresponding observed value. X_o denotes all observations. $\downarrow u$ denotes all nodes below u, excluding u. $\uparrow u$ denotes the set of all nodes excluding $\downarrow u$.

$t(u, i)$ can be implemented as a recursive bottom-up function and iterates through all parents and their children. It sums over all parent and siblings category pairs and calculates the probabilities of X_u and its siblings.

$$t(u, i) = \sum_{j,k} t(X_p, j)p(X_u = i|X_p = j)p(X_k = k|X_p = j)s(X_k, k) \quad [1.5]$$

where X_p is the parent node and X_k is the sibling node.

$s(u, i)$ can be implemented as a recursive top-down function that iterates through all nodes, starting from the root. It sums over all categories of the nodes, if at a leaf node, the function returns 1 if i = the observed value, 0 otherwise.

$$s(u, i) = \sum_j p(X_v = j|X_u = i)s(X_w, j) \quad [1.6]$$

$$s(u, i) = \begin{cases} 1, & \text{if i = observed value} \\ 0, & \text{otherwise} \end{cases} \quad [1.7]$$

where X_v and X_w are child nodes of X_u.

All the computed $t(u, i)$ and $s(u, i)$ are stored in an array that can be accessed in constant time later.

1.2.2 Problem 2.2.8

The algorithm is executed on the given trees and samples. The results are displayed in the table below:

	1	2	3	4	5
Small	0.0162	0.0154	0.0114	0.0086	0.0409
Medium	4.34e-18	3.09e-20	1.05e-16	6.59e-16	1.49e-18
Large	3.29e-69	1.11e-66	2.52e-68	1.24e-66	3.54e-69

Table 2: Likelihood of the 5 samples for each of the trees

1.3 Problem 2.3

1.3.1 Problem 2.3.9

The goal of the equations is to infer the posterior distribution for the mean μ and the precision τ, given the observed data set D, in which the samples are assumed to be Gaussian distributed. Assuming the conjugate prior distributions for μ and τ are the following:

$$p(\mu|\tau) = N(\mu|\mu_0, (\lambda\tau)^{-1}) \quad [1.8]$$

$$p(\tau) = Gam(\tau|\alpha, \beta) \quad [1.9]$$

where N is the normal distribution and Gam is the gamma distribution. The posterior distributions can be inferred by approximating the hyperparameters $\alpha, \beta, \mu, \lambda$ iteratively. Let N denote the number of iterations, the hyperparameters can be approximated as following:

$$\alpha_N = \alpha_0 + \frac{N+1}{2} \quad [1.10]$$

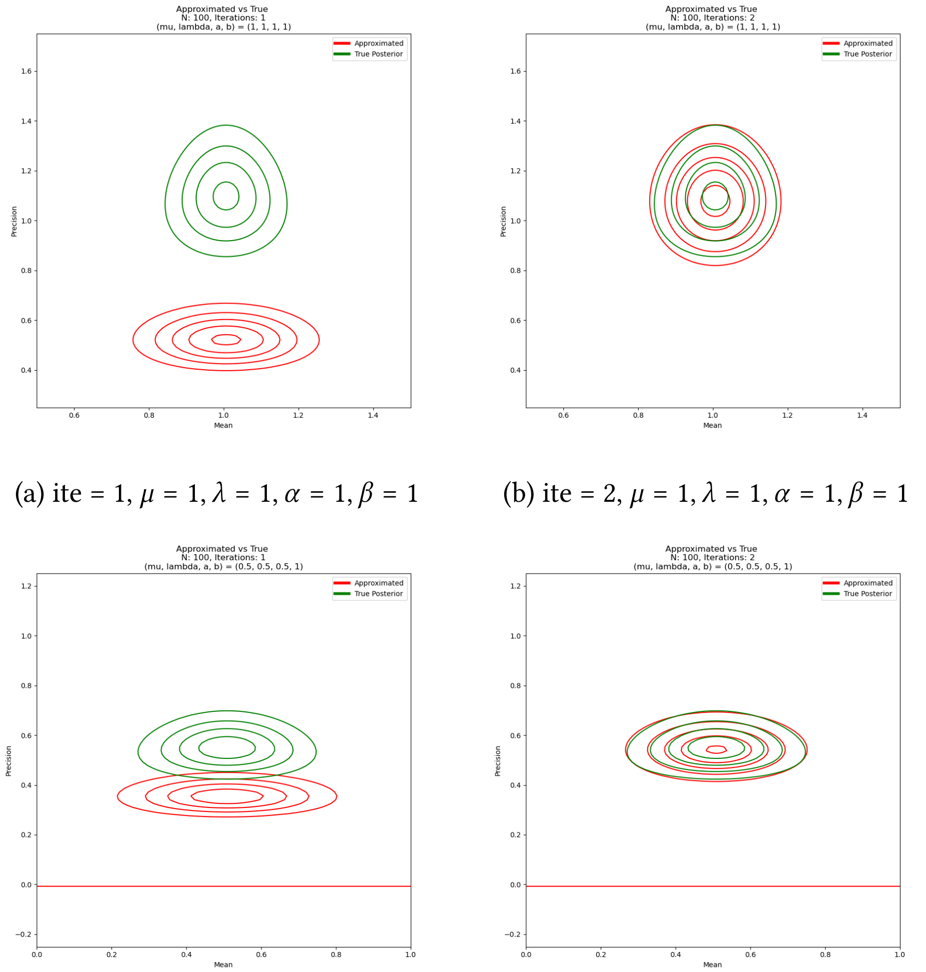
$$\beta_N = \beta_0 + \lambda_0(\lambda_N + \mu_N^2 + \mu_0^2 - 2\mu_N\mu_0) + \frac{1}{2} \cdot \sum^N (x_i^2 + \lambda_N + \mu_N^2 - 2\mu_N x_i) \quad [1.11]$$

$$\mu_N = \frac{\lambda_0\mu_0 + N\hat{x}}{\lambda_0 + N} \quad [1.12]$$

$$\lambda_N = (\lambda_0 + N)\frac{\alpha_N}{\beta_N} \quad [1.13]$$

where $\alpha_0, \beta_0, \mu_0, \lambda_0$ are initial guesses. $\hat{x}$ is the sample mean.

The examples of the approximations and the convergence are plotted in Figure 2 below, where the green contour is the true posterior and the red one is the inferred posterior.



(a) ite $= 1, \mu = 1, \lambda = 1, \alpha = 1, \beta = 1$

(b) ite $= 2, \mu = 1, \lambda = 1, \alpha = 1, \beta = 1$

(c) ite $= 1, \mu = .5, \lambda = .5, \alpha = .5, \beta = 1$

(d) ite $= 2, \mu = .5, \lambda = .5, \alpha = .5, \beta = 1$

Figure 1: Examples of inferred posterior vs true posterior

1.3.2 Problem 2.3.10

From the Equations [1.8-1.9] we have the following:

$$p(\mu, \tau|D) = \frac{p(\mu, \tau, D)}{p(D)} \propto p(\mu, \tau, D) \tag{1.14}$$

$$= p(D|\mu, \tau)p(\mu, \tau) = p(D|\mu, \tau)p(\mu|\tau)p(\tau) \tag{1.15}$$

$$\text{by Eq. 1.8 and 1.9} \tag{1.16}$$

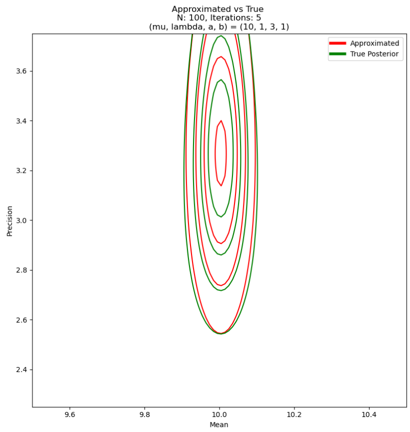
$$p(\mu|\tau) = N(\mu|\mu_0, (\lambda\tau)^{-1}) \tag{1.17}$$

$$p(\tau) = Gam(\tau|\alpha, \beta) \tag{1.18}$$

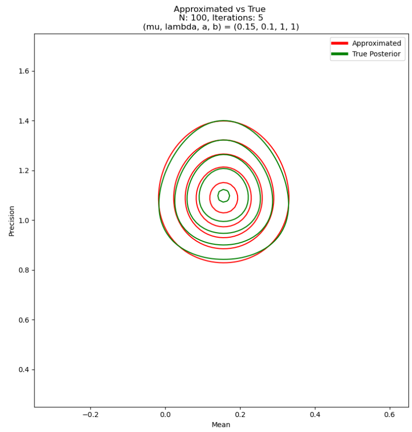
The resulting conjugate prior is a Gaussian-Gamma distribution, since $p(\mu|\tau)$ is Gaussian and $p(\tau)$ is Gamma. The exact posterior is computed by normalizing the Gaussian-Gamma distribution ($p(\mu|\tau)p(\tau)$) with the likelihood ($p(D|\mu, \tau)$).

1.3.3 Problem 2.3.11

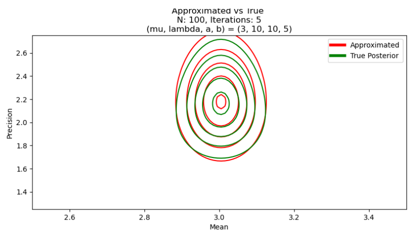
In Figure , the plots show 4 different cases of comparison between the inferred posterior vs the true posterior, where the inferred posteriors seem match the true posteriors on the both axis, the mean and the precision.



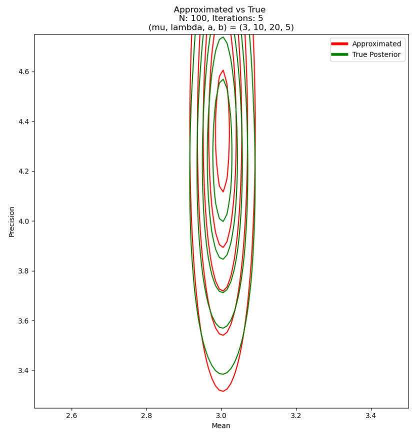
(a) ite = 5, μ = 10, λ = 1, α = 3, β = 1



(b) ite = 5, μ = .15, λ = .1, α = 1, β = 1



(c) ite = 5, μ = 3, λ = 10, α = 10, β = 5



(d) ite = 5, μ = 3, λ = 10, α = 20, β = 5

Figure 2: 4 different cases of comparison between inferred posterior vs true posterior