

DD2434 - Machine Learning, Advanced Course  
Assignment 1A

Tristan Perrot  
tristanp@kth.se

Étienne Riguet  
riguet@kth.se

November 2023



## Contents

<b>1</b>	<b>Exponential Family</b>	<b>3</b>
<b>2</b>	<b>Dependencies in a Directed Graphical Model</b>	<b>4</b>
<b>3</b>	<b>CAVI</b>	<b>9</b>
<b>A</b>	<b>Appendix</b>	<b>14</b>
A.1	Question 3.12 . . . . .	14
A.2	Question 3.13 . . . . .	14
A.3	Question 3.14 . . . . .	14
A.4	Question 3.15 . . . . .	15

# 1 Exponential Family

## Question 1.1

$$\begin{aligned}
 p(x|\theta) &= h(x) \exp(\eta(\theta) \cdot T(x) - A(\eta)) \\
 &= h(x) \exp(\eta(\lambda) \cdot T(x) - A(\eta(\lambda))) \\
 &= h(x) \exp(\log \lambda \cdot x - A(\log \lambda)) \\
 &= h(x) \exp(\log \lambda \cdot x - \lambda) \\
 &= h(x) \exp(\log \lambda \cdot x) \exp(-\lambda) \\
 &= e^{-\lambda} \frac{\lambda^x}{x!}
 \end{aligned} \tag{1}$$

We can see that the distribution correspond to a Poisson distribution of parameter  $\lambda$ .

## Question 1.2

$$\begin{aligned}
 p(x|\theta) &= h(x) \exp(\eta(\theta) \cdot T(x) - A(\eta)) \\
 &= \exp(\eta([\alpha, \beta]) \cdot [\log x, x] - A(\alpha - 1, -\beta)) \\
 &= \exp([\alpha - 1, -\beta] \cdot [\log x, x] - \log \Gamma(\alpha) + \alpha \log(\beta)) \\
 &= \exp((\alpha - 1) \log x - \beta x - \log \Gamma(\alpha) + \alpha \log(\beta)) \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}
 \end{aligned} \tag{2}$$

We can see that the distribution correspond to a Gamma distribution of parameters  $\alpha$  and  $\beta$ .

## Question 1.3

$$\begin{aligned}
 p(x|\theta) &= h(x) \exp(\eta(\theta) \cdot T(x) - A(\eta)) \\
 &= \frac{\exp(\eta([\mu, \sigma^2]) \cdot [x, x^2] - A(\eta([\mu, \sigma^2])))}{\sqrt{2\pi}} \\
 &= \frac{\exp([\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}] \cdot [x, x^2] - A([\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}]))}{\sqrt{2\pi}} \\
 &= \frac{\exp(\frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \log \sigma)}{\sqrt{2\pi}} \\
 &= \frac{\exp(-\frac{(x-\mu)^2}{2\sigma^2})}{\sigma\sqrt{2\pi}}
 \end{aligned} \tag{3}$$

We can see that the distribution correspond to a Normal distribution of parameters  $\mu$  and  $\sigma^2$ .

### Question 1.4

$$\begin{aligned}
 p(x|\theta) &= h(x) \exp(\eta(\theta) \cdot T(x) - A(\eta)) \\
 &= 2 \exp(\eta(\lambda) \cdot x - A(\eta(\lambda))) \\
 &= 2 \exp(-\lambda x - A(-\lambda)) \\
 &= 2 \exp\left(-\lambda x + \log\left(\frac{\lambda}{2}\right)\right) \\
 &= \lambda e^{-\lambda x}
 \end{aligned} \tag{4}$$

We can see that the distribution correspond to a Exponential distribution of parameter  $\lambda$ .

### Question 1.5

$$\begin{aligned}
 p(x|\theta) &= h(x) \exp(\eta(\theta) \cdot T(x) - A(\eta)) \\
 &= \exp(\eta([\psi_1, \psi_2]) \cdot [\log x, \log(1-x)] - A(\eta([\psi_1, \psi_2]))) \\
 &= \exp([\psi_1 - 1, \psi_2 - 1] \cdot [\log x, \log(1-x)] - A([\psi_1 - 1, \psi_2 - 1])) \\
 &= \exp((\psi_1 - 1) \log x + (\psi_2 - 1) \log(1-x) - \log \Gamma(\psi_1) - \log \Gamma(\psi_2) + \log \Gamma(\psi_1 + \psi_2)) \\
 &= \frac{\Gamma(\psi_1 + \psi_2)}{\Gamma(\psi_1) \Gamma(\psi_2)} x^{\psi_1 - 1} (1-x)^{\psi_2 - 1}
 \end{aligned} \tag{5}$$

We can see that the distribution correspond to a Beta distribution of parameters  $\psi_1$  and  $\psi_2$ .

## 2 Dependencies in a Directed Graphical Model

### Question 2.6

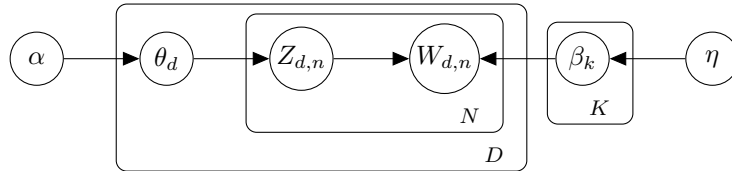
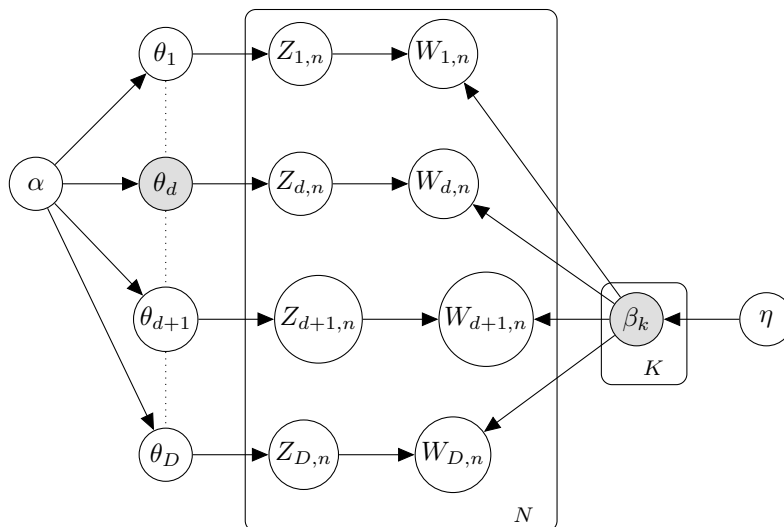
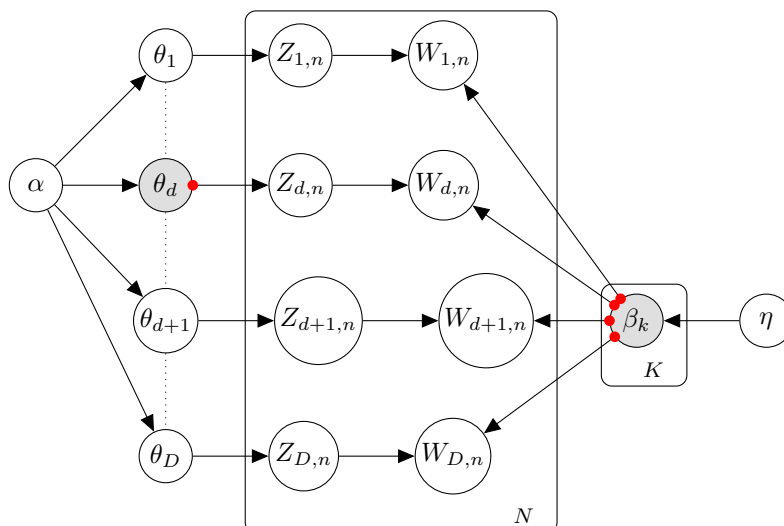


Figure 1: Graphical model of smooth LDA.

The Bayes net take this form :



Then, if we use the method using the d-separation, we obtain this :



Therefore, we can see that  $W_{d,n} \perp W_{d,n+1} | \theta_d, \beta_{1:K}$  is true.

## Question 2.7

The Bayes net take this form (with d-separation marks) :



Therefore, we can see that  $\theta_d \perp \theta_{d+1} | Z_{d,1:N}$  is false.

### Question 2.8

The Bayes net take this form (with d-separation marks) :



Therefore, we can see that  $\theta_d \perp \theta_{d+1} | \alpha, Z_{1:D,1:N}$  is true.

### Question 2.9



Figure 2: Graphical model of Labeled LDA.

The Bayes net take this form (with d-separation marks) :



Therefore, we can see that  $W_{d,n} \perp W_{d,n+1} | \Lambda_d, \beta_{1:K}$  is false.

### Question 2.10

The Bayes net take this form (with d-separation marks) :



Therefore, we can see that  $\theta_d \perp \theta_{d+1} | Z_{d,1:N}, Z_{d+1,1:N}$  is false.

### Question 2.11

The Bayes net take this form (with d-separation marks) :





Therefore, we can see that  $\Lambda_d \perp \Lambda_{d+1} | \Phi, Z_{1:D,1:N}$  is false.

### 3 CAVI

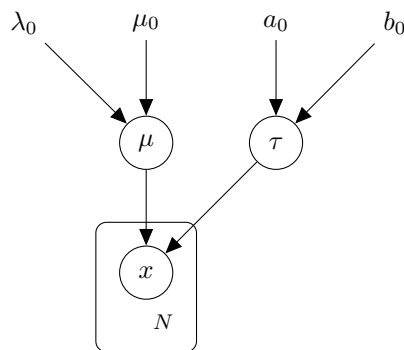


Figure 3: DGM

#### Question 3.12

In the bishop book, we can see that :

$$p(X|\mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{N/2} \exp \left\{ -\frac{\tau}{2} \sum_{n=1}^N (x_n - \mu)^2 \right\} \quad (6)$$

$$p(\mu|\tau) = \mathcal{N}(\mu|\mu_0, (\lambda_0\tau)^{-1}) \quad (7)$$

$$p(\tau) = \text{Gam}(\tau|a_0, b_0) \quad (8)$$

Then, by using the code in appendix A.1, we obtain :

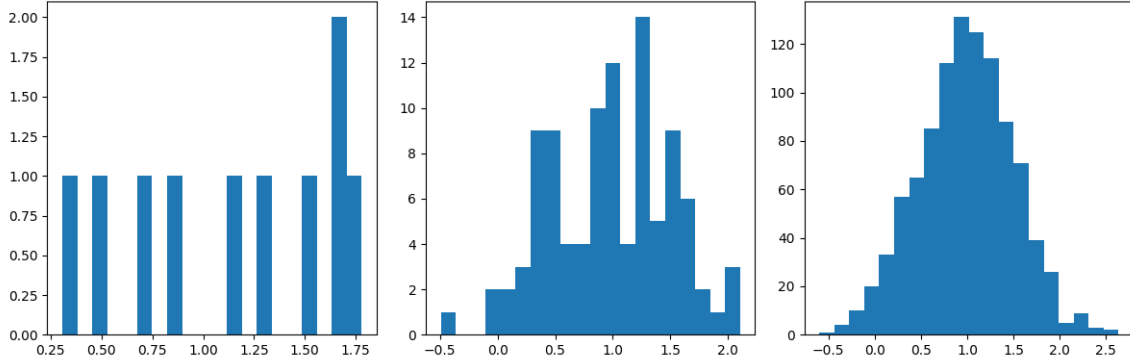


Figure 4: Generated Data.

### Question 3.13

Let's find the ML estimates of  $\mu$  and  $\tau$ . We know that  $\log(q^*(\mu)) = \mathbb{E}_{-\mu}[\log p(X, \mu, \tau)]$ . Then, we can write :

$$\begin{aligned} \log(q^*(\mu)) &= \mathbb{E}_{-\mu}[\log p(X, \mu, \tau)] \\ &\stackrel{\pm}{=} \mathbb{E}_{\tau}[\log p(X|\mu, \tau) + \log p(\mu|\tau)] \\ &= \mathbb{E}_{\tau} \left[ \frac{N}{2} \log \left( \frac{\tau}{2\pi} \right) + \frac{\tau}{2} \sum_{n=1}^N (x_n - \mu)^2 + \frac{1}{2} \log \left( \frac{\lambda_0\tau}{2\pi} \right) + \frac{\lambda_0\tau}{2} (\mu - \mu_0) \right] \\ &\stackrel{\pm}{=} \frac{\mathbb{E}_{\tau}[\tau]}{2} \left( \lambda_0(\mu - \mu_0) + \sum_{n=1}^N (x_n - \mu)^2 \right) \\ &= -\frac{\mathbb{E}_{\tau}[\tau]}{2} \left( \lambda_0\mu^2 - 2\lambda_0\mu\mu_0 + \lambda_0\mu_0^2 + \sum_{n=1}^N x_n^2 - 2\mu \sum_{n=1}^N x_n + N\mu^2 \right) \\ &= -\frac{\mathbb{E}_{\tau}[\tau]}{2} \left( (\lambda_0 + N)\mu^2 - 2(\lambda_0\mu_0 + \sum_{n=1}^N x_n)\mu + \lambda_0\mu_0^2 + \sum_{n=1}^N x_n^2 \right) \\ &= -\frac{\mathbb{E}_{\tau}[\tau](\lambda_0 + N)}{2} \left( \mu^2 - 2\mu \frac{\lambda_0\mu_0 + \sum_{n=1}^N x_n}{\lambda_0 + N} + \frac{\lambda_0\mu_0^2 + \sum_{n=1}^N x_n^2}{\lambda_0 + N} \right) \end{aligned} \quad (9)$$

Therefore we can conclude that  $q^*(\mu) = \mathcal{N}(\mu|\mu_N, \lambda_N^{-1})$  with :

$$\mu_N = \frac{\lambda_0 \mu_0 + \sum_{n=1}^N x_n}{\lambda_0 + N} \quad (10)$$

$$\lambda_N = (\lambda_0 + N) \mathbb{E}[\tau] \quad (11)$$

And for  $\tau$  we have :

$$\begin{aligned} \log(q^*(\tau)) &= \mathbb{E}_{-\tau}[\log p(X, \mu, \tau)] \\ &\stackrel{\pm}{=} \mathbb{E}_{\mu}[\log p(X|\mu, \tau) + \log p(\mu|\tau)] + \log p(\tau) \\ &\stackrel{\pm}{=} (a_0 - 1) \log \tau - b_0 \tau + \frac{N}{2} \log \tau - \frac{\tau}{2} \mathbb{E}_{\mu} \left[ \sum_{n=1}^N (x_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right] \\ &= (a_0 + \frac{N}{2} - 1) \log \tau - \left( b_0 + \frac{1}{2} \mathbb{E}_{\mu} \left[ \sum_{n=1}^N (x_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right] \right) \tau \end{aligned} \quad (12)$$

Therefore we can conclude that  $q^*(\tau) = \text{Gam}(\tau|a_N, b_N)$  with :

$$a_N = a_0 + \frac{N}{2} \quad (13)$$

$$\begin{aligned} b_N &= b_0 + \frac{1}{2} \mathbb{E}_{\mu} \left[ \sum_{n=1}^N (x_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right] \\ b_N &= b_0 + \frac{1}{2} \left( \sum_{n=1}^N x_n^2 + N \mathbb{E}_{\mu}[\mu^2] - 2 \mathbb{E}_{\mu}[\mu] \sum_{n=1}^N x_n + \lambda_0 (\mathbb{E}_{\mu}[\mu^2] + \mu_0^2 - 2 \mu_0 \mathbb{E}_{\mu}[\mu]) \right) \end{aligned} \quad (14)$$

With :

$$\begin{aligned} \mathbb{E}_{q(\mu)}[\mu] &= \mu_N \\ \mathbb{E}_{q(\mu)}[\mu^2] &= \frac{1}{\lambda_N} + \mu_N^2 \\ \mathbb{E}_{q(\tau)}[\tau] &= \frac{a_N}{b_N} \end{aligned} \quad (15)$$

If we take non-informative priors then  $a_0 = b_0 = \mu_0 = \lambda_0 = 0$ , then we have :

$$\begin{aligned} \mu_N &= \bar{x} \\ \lambda_N &= N \mathbb{E}[\tau] \\ a_N &= \frac{N}{2} \\ b_N &= \frac{1}{2} \mathbb{E}_{\mu} \left[ \sum_{n=1}^N (x_n - \mu)^2 \right] \end{aligned} \quad (16)$$

And by using  $\mathbb{E}[\tau] = \frac{a_N}{b_N}$  we obtain :

$$\begin{aligned}\frac{1}{\mathbb{E}[\tau]} &= \frac{b_N}{a_N} \\ \frac{1}{\mathbb{E}[\tau]} &= \frac{2}{2N} \mathbb{E}_\mu \left[ \sum_{n=1}^N (x_n - \mu)^2 \right] \\ \frac{1}{\mathbb{E}[\tau]} &= \overline{x^2} - 2\bar{x}\mathbb{E}[\mu] + \mathbb{E}[\mu^2]\end{aligned}\tag{17}$$

And, with the fact that  $\mathbb{E}[\mu] = \mu_N$  and  $\mathbb{E}[\mu^2] = \frac{1}{\lambda_N} + \mu_N^2$ , we obtain :

$$\begin{aligned}\mathbb{E}[\mu] &= \bar{x} \\ \mathbb{E}[\mu^2] &= \frac{1}{N\mathbb{E}[\tau]} + \bar{x}^2\end{aligned}\tag{18}$$

And therefore:

$$\frac{1}{\mathbb{E}[\tau]} = \frac{N}{N-1}(\overline{x^2} - \bar{x}^2) = \frac{N}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2\tag{19}$$

Which define the ML estimates. The implementation is in the code in appendix A.2.

### Question 3.14

The posterior is defined as  $p(\mu, \tau|x)$ . Then, we can write :

$$\begin{aligned}p(\mu, \tau|x) &= \frac{p(x|\mu, \tau)p(\mu, \tau)}{p(x)} \\ &\propto p(x|\mu, \tau)p(\mu, \tau)\end{aligned}\tag{20}$$

Where  $x|\mu, \tau \sim \mathcal{N}(\mu|\mu, \tau^{-1})$  and  $\mu, \tau \sim \text{NormalGamma}(\mu_0, \lambda_0, a_0, b_0)$ . Therefore, as we saw in the question 1.3 in the Module 1 exercise, we have  $\mu, \tau|x \sim \text{NormalGamma}(\mu', \lambda', a', b')$ , where :

$$\begin{aligned}\mu' &= \frac{N\bar{x} + \mu_0\lambda_0}{N + \lambda_0} \\ \lambda' &= N + \lambda_0 \\ a' &= a_0 + \frac{N}{2} \\ b' &= b_0 + \frac{1}{2} \left( \sum_{n=1}^N x_n^2 + \lambda_0\mu_0^2 - \frac{(N\bar{x} + \mu_0\lambda_0)^2}{N + \lambda_0} \right)\end{aligned}\tag{21}$$

The rest of the answer is in the code in appendix A.3.

### Question 3.15

The equation (10.24) in the Bishop is the mean-field approximation which is :

$$q(\mu, \tau) = q(\mu)q(\tau)\tag{22}$$

Now, we need to find the ELBO formula :

$$\begin{aligned}\mathcal{L}(q) &= \mathbb{E}_q[\log p(X, \mu, \tau)] - \mathbb{E}_q[\log q(\mu, \tau)] \\ &= \mathbb{E}_q[\log p(X|\mu, \tau) + \log p(\mu, \tau)] - \mathbb{E}_q[\log q(\mu) + \log q(\tau)] \\ &= \mathbb{E}_q[\log p(X|\mu, \tau)] + \mathbb{E}_q[\log p(\mu, \tau)] - \mathbb{E}_q[\log q(\mu)] - \mathbb{E}_q[\log q(\tau)] \\ &= \mathbb{E}_q[\log p(X|\mu, \tau)] + \mathbb{E}_q[\log p(\mu, \tau)] + \mathbb{H}_q[\mu] + \mathbb{H}_q[\tau]\end{aligned}\tag{23}$$

## A Appendix

### A.1 Question 3.12

```
import matplotlib.pyplot as plt
import numpy as np

def generate_data(mu, tau, N):
    # Insert your code here
    D = np.random.normal(mu, tau, N)

    return D

mu = 1
tau = 0.5

dataset_1 = generate_data(mu, tau, 10)
dataset_2 = generate_data(mu, tau, 100)
dataset_3 = generate_data(mu, tau, 1000)

# Visualize the datasets via histograms
# Insert your code here
fig, axs = plt.subplots(1, 3, figsize=(12, 4))
axs[0].hist(dataset_1, bins=20)
axs[1].hist(dataset_2, bins=20)
axs[2].hist(dataset_3, bins=20)
plt.tight_layout()
plt.savefig('12_data.png')
plt.show()
```

### A.2 Question 3.13

```
def ML_est(data):
    # insert your code
    N = len(data)
    x_mean = np.mean(data)
    x_var = np.var(data, ddof=1)

    tau_ml = 1 / (N * x_var)
    mu_ml = x_mean

    return mu_ml, tau_ml
```

### A.3 Question 3.14

```
def compute_exact_posterior(D, a_0, b_0, mu_0, lambda_0):
    # your implementation
```

```
x_mean = np.mean(D)
N = len(D)

mu_prime = (lambda_0 * mu_0 + N * x_mean) / (lambda_0 + N)
lambda_prime = lambda_0 + N
a_prime = a_0 + N / 2
b_prime = b_0 + 1 / 2 * (np.sum((D - x_mean)**2) +
                        lambda_0 * mu_0**2 - lambda_prime * mu_prime
                        **2)

exact_post_distribution = (mu_prime, lambda_prime, a_prime, b_prime)

return exact_post_distribution
```

## A.4 Question 3.15

```
from scipy.stats import gamma, norm

# prior parameters
mu_0 = 0
lambda_0 = 1
a_0 = 1
b_0 = 1

def compute_elbo(D, a_0, b_0, mu_0, lambda_0, a_N, b_N, mu_N, lambda_N):
    # given the prior and posterior parameters together with the data,
    # compute ELBO here
    N = len(D)

    return elbo

def CAVI(D, a_0, b_0, mu_0, lambda_0):
    # make an initial guess for the expected value of tau
    initial_guess_exp_tau = 1

    N = len(D)
    x_mean = np.mean(D)
    x_2_sum = np.sum(D**2)

    # Constants
    a_N = a_0 + N / 2
    mu_N = (lambda_0 * mu_0 + N * x_mean) / (lambda_0 + N)
    E_mu = mu_N

    # Variational parameters
    b_N = b_0
    lambda_N = lambda_0

    # ELBO
```

```
elbos = []

# CAVI iterations ...
for i in range(100):
    # update the values for the variational parameters
    E_tau = a_N / b_N
    E_mu_2 = 1 / lambda_N + mu_N**2

    lambda_N = (lambda_0 + N) * E_tau
    b_N = b_0 + 1 / 2 * (x_2_sum + N*E_mu_2 - 2*N*E_mu *
                        x_mean + lambda_0*(E_mu_2 - 2*E_mu*mu_0 + mu_0
                        **2))

    # save ELBO for each iteration, plot them afterwards to show
    # convergence
    elbos.append(compute_elbo(D, a_0, b_0, mu_0,
                             lambda_0, a_N, b_N, mu_N, lambda_N))

return a_N, b_N, mu_N, lambda_N, elbos
```