

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>5</b>
	<b>Appendix</b>	<b>7</b>

# 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

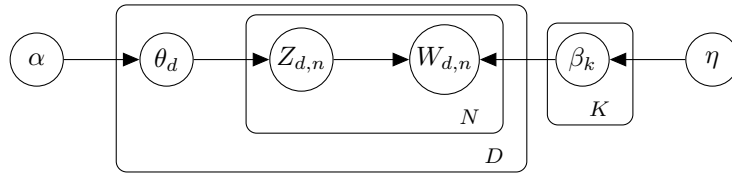
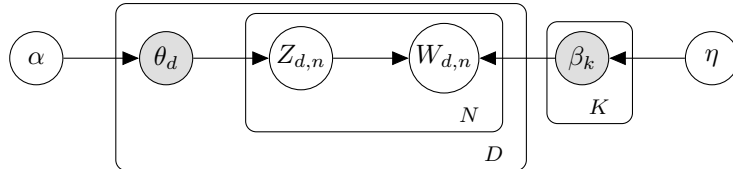


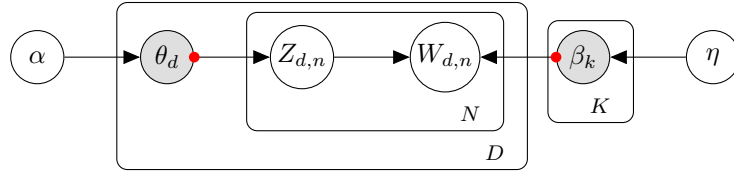
Figure 1: Graphical model of smooth LDA.

### Question 2.6

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 false.

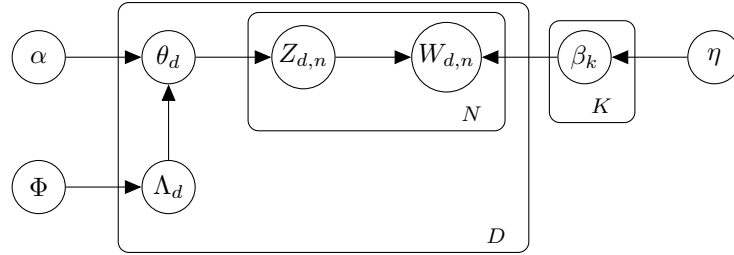


Figure 2: Graphical model of Labeled LDA.

### 3 CAVI

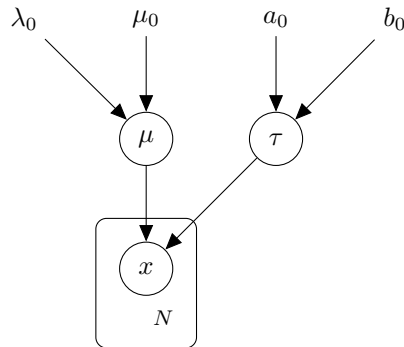


Figure 3: DGM

#### 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 3, we obtain :

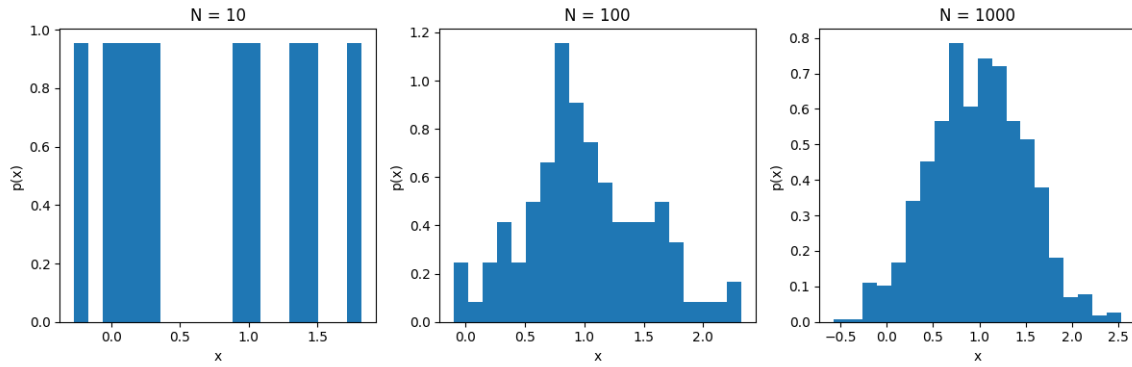


Figure 4: Generated Data.

### Question 3.13

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

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

## Appendix

### Python Code

```
from typing import List
import numpy as np
import scipy.special as sp_spec
import numpy.random as np_rand
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score
import seaborn as sns

def generate_data(N: int, mu: float, tau: float) -> np.ndarray:
    return np_rand.normal(mu, tau, N)

def plot_data(X: np.ndarray, ax: plt.Axes) -> None:
    ax.hist(X, bins=20, density=True)
    ax.set_xlabel('x')
    ax.set_ylabel('p(x)')
    ax.set_title(f'N = {len(X)}')

MU = 1
TAU = 0.5
N = [10, 100, 1000]
Xs = [generate_data(n, MU, TAU) for n in N]

fig, axs = plt.subplots(1, 3, figsize=(12, 4))
for i in range(len(Xs)):
    plot_data(Xs[i], axs[i])
plt.tight_layout()
plt.savefig('12_data.png')
```