DD2434 - Machine Learning, Advanced Course
Assignment 1A

Tristan Perrot
tristanp@kth.se

Étienne Riguet
riguet@kth.se

November 2023

Contents

Tristan Perrot
Étienne Riguet

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 λ.

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 α and β.

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 μ and σ^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 λ.

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 ψ_1 and ψ_2.

2 Dependencies in a Directed Graphical Model

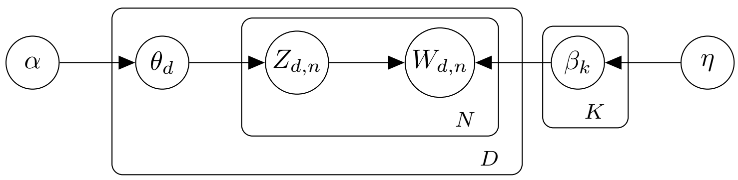
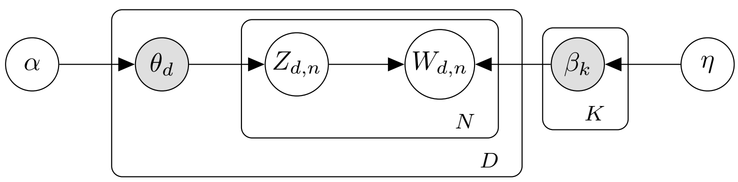


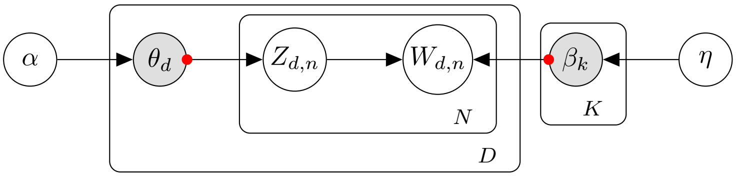
Figure 1: Graphical model of smooth LDA.

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.

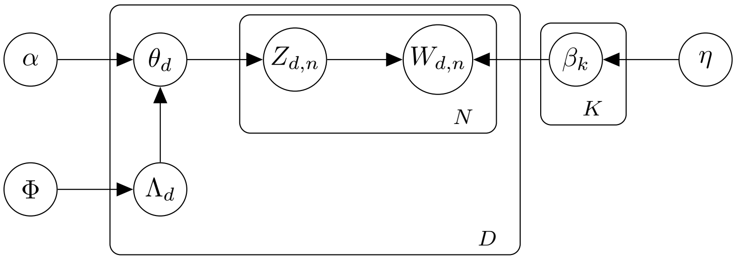


Figure 2: Graphical model of Labeled LDA.

3 CAVI

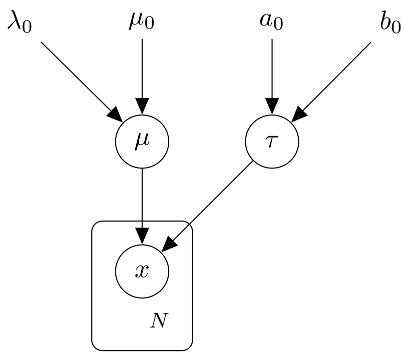


Figure 3: DGM

Question 3.12

In the bishop book, we can see that :

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

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

$$p(\tau) = \mathrm{Gam}(\tau|a_0, b_0) \tag{8}$$

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

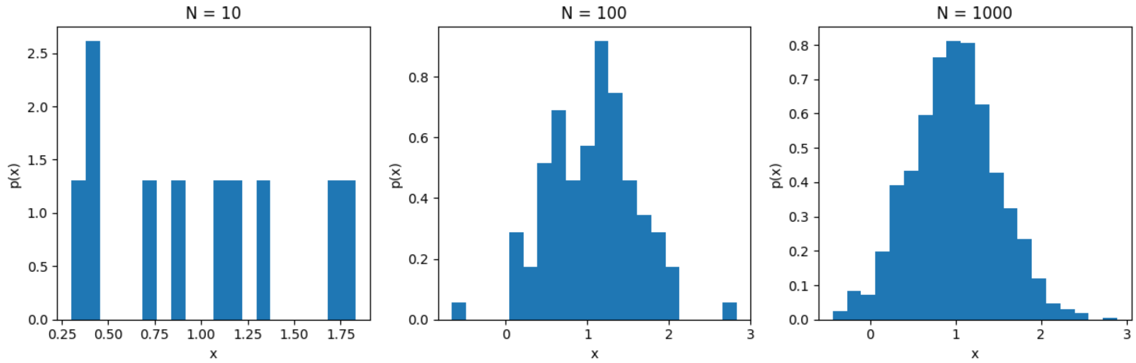


Figure 4: Generated Data.

Question 3.13

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

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

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

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

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

And for τ we have :

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

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

$$a_N = a_0 + \frac{N}{2} \tag{13}$$

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

Question 3.14

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

$$q(\mu, \tau) = q(\mu)q(\tau) \tag{15}$$

We have also some formulas to detail :

$$\begin{aligned}
\mathbb{E}_{q(\mu)}[\mu] &= \mu_N \\
\mathbb{E}_{q(\mu)}[\mu^2] &= \frac{1}{\lambda_N} + \mu_N^2 \\
\mathbb{E}_{q(\tau)}[\tau] &= \frac{a_N}{b_N}
\end{aligned} \tag{16}$$

The rest of the answer is in the code in appendix A.1.

A Appendix

A.1 Python Code

```
import numpy as np
import scipy.special as sp_spec
import scipy.stats as sp_stats
import numpy.random as np_rand
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score
from tqdm.auto import trange
import seaborn as sns


# Question 1.3.12

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')


# Question 1.3.14

def update_b_N(x, mu_N, lambda_N, b_0, mu_0, lambda_0):
    E_mu = mu_N
    E_mu2 = 1 / lambda_N + mu_N ** 2

    b_N = b_0 + 0.5 * (np.sum(x ** 2) - 2 * np.sum(x) * E_mu +
                       x.shape[0] * E_mu2 + lambda_0*(E_mu2 + mu_0 ** 2 - 2
                           * E_mu * mu_0))

    return b_N
```

```
def update_lambda_N(x, a_N, b_N, lambda_0):
    E_tau = a_N / b_N
    lambda_N = (lambda_0 + x.shape[0])*E_tau

    return lambda_N


def vi_alg(x, a_0, b_0, mu_0, lambda_0, iter=20):
    N = x.shape[0]

    # Constants
    a_N = a_0 + N/2
    mu_N = (lambda_0 * mu_0 + np.sum(x)) / (lambda_0 + N)

    # Variables
    b_N = b_0
    lambda_N = lambda_0

    # Lists for plotting
    b_Ns = np.zeros(iter+1)
    lambda_Ns = np.zeros(iter+1)

    b_Ns[0] = b_N
    lambda_Ns[0] = lambda_N

    for i in trange(iter):
        b_Ns[i+1] = update_b_N(x, mu_N, lambda_Ns[i], b_0, mu_0, lambda_0)
        lambda_Ns[i+1] = update_lambda_N(x, a_N, b_Ns[i], lambda_0)

        a_True, b_True, mu_True, lambda_True = true_posterior(
            x, a_0, b_0, mu_0, lambda_0)

    print('a_N =', a_N)
    print('a_True =', a_True)
    print('b_N =', b_Ns[-1])
    print('b_True =', b_True)
    print('mu_N =', mu_N)
    print('mu_True =', mu_True)
    print('lambda_N =', lambda_Ns[-1])
    print('lambda_True =', lambda_True)

    fig, axs = plt.subplots(1, 2, figsize=(12, 4))
    axs[0].plot(b_Ns)
    axs[0].axhline(b_True, color='r', linestyle='--')
    axs[0].set_xlabel('Iteration')
    axs[0].set_ylabel('b_N')
    axs[0].set_title('b_N')
    axs[1].plot(lambda_Ns)
    axs[1].axhline(lambda_True, color='r', linestyle='--')
    axs[1].set_xlabel('Iteration')
    axs[1].set_ylabel('lambda_N')
```

```
    axs[1].set_title('lambda_N')
    plt.tight_layout()
    plt.show()
    # plt.savefig('14_vi.png')


# Question 1.3.15

def true_posterior(x, a_0, b_0, mu_0, lambda_0):
    mu_N = (lambda_0 * mu_0 + np.sum(x)) / (lambda_0 + x.shape[0])
    lambda_N = lambda_0 + x.shape[0]
    a_N = a_0 + x.shape[0]/2
    b_N = b_0 + 0.5 * (np.sum(x ** 2) + lambda_0 *
                       mu_0 ** 2 - lambda_N * mu_N ** 2)
    return a_N, b_N, mu_N, lambda_N


a_0 = 1
b_0 = 1
mu_0 = 1
lambda_0 = 12

vi_alg(Xs[-1], a_0, b_0, mu_0, lambda_0)
```