Exercise 5: VB
02/05/2023
Group D: Clara Kümpel, Damola Agbelese, Yitong Li

5.1 VB update equations

a) Univariate Gaussian model

$$y = \mu + \epsilon, \qquad p(\epsilon) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau\epsilon^2}{2}\right)$$

As μ is a constant, $y - \mu = \epsilon \sim \mathcal{N}(0, 1/\tau)$

$$\Rightarrow p(y|\mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau(y-\mu)^2}{2}\right)$$

Joint distribution:

$$p(y_1, \dots, y_N|\mu, \tau) = \prod_{i=1}^{N} \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau(y_i-\mu)^2}{2}\right)$$

$$= \left(\sqrt{\frac{\tau}{2\pi}}\right)^N \prod_{i=1}^{N} \exp\left(-\frac{\tau(y_i-\mu)^2}{2}\right)$$

$$\Rightarrow \log\big(p(y_1, \dots, y_N|\mu, \tau)\big) = \frac{N}{2}\log\tau - \frac{N}{2}\log 2\pi - \frac{\tau}{2}\sum_{1}^{N}(y_i - \mu)^2$$

$Q.E.D.$

b) Find the joint distribution $\log p(\vec{y}, \vec{\theta})$ given prior $p(\vec{\theta})$

$$p(\vec{\theta}) = \sqrt{\frac{\lambda_0\tau}{2\pi}} \exp\left(-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2\right) \frac{b_0^{a_0}\tau^{a_0-1}}{\Gamma(a_0)} \exp(-b_0\tau)$$

$$\Rightarrow \log p(\vec{\theta}) = \frac{1}{2}\log\lambda_0 + \frac{1}{2}\log\tau - \frac{1}{2}\log 2\pi - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 + a_0\log b_0 + (a_0-1)\log\tau - \log\Gamma(a_0) - b_0\tau$$

$$p(\vec{y}, \vec{\theta}) = p(\vec{y}|\vec{\theta})p(\vec{\theta})$$

$$\Rightarrow \log p(\vec{y}, \vec{\theta}) = \log p(\vec{y}|\vec{\theta}) + \log p(\vec{\theta})$$

$$= \left(\frac{N+1}{2} + a_0 - 1\right)\log\tau - \frac{N+1}{2}\log 2\pi - \frac{\tau}{2}\sum_{1}^{N}(y_i-\mu)^2 + \frac{1}{2}\log\lambda_0 - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 + a_0\log b_0 - \log\Gamma(a_0) - b_0\tau$$

$Q.E.D.$

c) Find the optimal variational distribution over μ under the mean field approximation.
Regarding to the mean field approx.:

$$\ln q^*(\mu) = \mathbb{E}_\tau\big[\ln p(\vec{y}, \vec{\theta})\big] + c.$$

$$= \int \ln\left(p(\vec{y}, \vec{\theta})\right) q(\tau)d\tau + c.$$

since the optimization is over μ, integration terms excluding μ can be sorted into the constant term, hence:

$$\ln q^*(\mu) = \int \left(-\frac{\tau}{2} \sum_1^N (y_n - \mu)^2 - \frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2 \right) q(\tau) d\tau + c.$$

The integrated term can be factorized as:

$$\ln q^*(\mu) = \int -\frac{1}{2} \left(\sum_1^N (y_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right) \tau \cdot q(\tau) d\tau + c.$$

Move the term excluding τ outside the integral:

$$\ln q^*(\mu) = -\frac{1}{2} \left(\sum_1^N (y_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right) \int \tau \cdot q(\tau) + c.$$

With the definition of expectation:

$$\ln q^*(\mu) = -\frac{1}{2} E_\tau\{\tau\} \left(\sum_1^N (y_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right) + c.$$

$$Q.E.D.$$

d) Reorder $\ln q^*(\mu)$ in (c) with $\bar{\tau} = E_\tau\{\tau\}$ and $\bar{y} = \frac{1}{N} \sum_1^N y_n$

$$\ln q^*(\mu) = -\frac{1}{2} \bar{\tau} \left(\sum_1^N y_n^2 - 2N\mu\bar{y} + N\mu^2 + \lambda_0 \mu^2 - 2\lambda\mu\mu_0 + \lambda_0 \mu_0^2 \right) + c.$$

Since μ is the only variable, terms $\sum_1^N y_n^2$ and $\lambda_0 \mu_0^2$ are constant, therefore:

$$\ln q^*(\mu) = -\frac{1}{2} \bar{\tau} (-2N\mu\bar{y} + (N + \lambda_0)\mu^2 + (N + \lambda_0 \mu_0)\mu) + c.$$

$$= -\frac{\bar{\tau}(N + \lambda_0)}{2} \mu^2 + \bar{\tau}\mu(\lambda_0 \mu_0 + N\bar{y}) + c.$$

$$Q.E.D.$$

e) Show that $q^*(\mu)$ follows Gaussian distribution.

$$\ln q^*(\mu) = -\frac{\bar{\tau}(N + \lambda_0)}{2} \mu^2 + \bar{\tau}\mu(\lambda_0 \mu_0 + N\bar{y}) + c.$$

$$= -\frac{\bar{\tau}(N + \lambda_0)}{2} \left(\mu^2 - \frac{2(\lambda_0 \mu_0 + N\bar{y})}{N + \lambda_0} \mu + \left(\frac{\lambda_0 \mu_0 + N\bar{y}}{N + \lambda_0} \right)^2 \right) + c.$$

$$= -\frac{\bar{\tau}(N + \lambda_0)}{2} \left(\mu - \frac{\lambda_0 \mu_0 + N\bar{y}}{N + \lambda_0} \right)^2 + c.$$

$$\Rightarrow q^*(\mu) \propto \exp\left(-\frac{\bar{\tau}(N + \lambda_0)}{2} \left(\mu - \frac{\lambda_0 \mu_0 + N\bar{y}}{N + \lambda_0} \right)^2 \right)$$

i.e., $q^*(\mu)$ follows Gaussian distribution with mean

$$m = \frac{\lambda_0 \mu_0 + N\bar{y}}{N + \lambda_0}$$

And variance

$$s^2 = \frac{1}{\bar{\tau}(N + \lambda_0)}$$

$$Q.E.D.$$

f) Find the optimal variational distribution over τ under the mean field approximation.

$$\ln q^*(\tau) = \mathbb{E}_\mu\left[\ln p(\vec{y},\vec{\theta})\right] + c.$$
$$= \int \ln\left(p(\vec{y},\vec{\theta})\right) q(\mu)d\mu + c.$$

As in (c), integration terms excluding τ can be sorted into the constant term, hence:

$$\ln q^*(\tau) = \int\left(\left(\frac{N+1}{2} + a_0 - 1\right)\log\tau - \frac{\tau}{2}\sum_{n=1}^{N}(y_n-\mu)^2 - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 - b_0\tau\right)$$
$$\cdot q(\mu)d\mu + c.$$
$$= -\frac{\tau}{2}\int\left(\sum_{n=1}^{N}(y_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right)\cdot q(\mu)d\mu$$
$$+\left(\left(\frac{N+1}{2} + a_0 - 1\right)\log\tau - b_0\tau\right)\int q(\mu)d\mu + c.$$
$$= -\frac{\tau}{2}\mathbb{E}_\mu\left\{\sum_{n=1}^{N}(y_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right\} + \left(\frac{N+1}{2} + a_0 - 1\right)\log\tau - b_0\tau + c.$$

$Q.E.D.$

g) Rewrite $\ln q^*(\tau)$ with $\mathbb{E}_\mu\{(x-\mu)^2\} = (x-m)^2 + s^2$

$$\mathbb{E}_\mu\left\{\sum_{n=1}^{N}(y_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right\}$$
$$= \sum_{1}^{N}\mathbb{E}_\mu\{(y_n-\mu)^2\} + \lambda_0\mathbb{E}_\mu\{(\mu-\mu_0)^2\}$$
$$= (N+\lambda_0)s^2 + \sum_{1}^{N}(y_n-m)^2 + \lambda_0(x-m)^2$$
$$\Rightarrow \ln q^*(\tau) = -\frac{\tau}{2}\left(\sum_{1}^{N}(y_n-m)^2 + \lambda_0(x-m)^2 + (N+\lambda_0)s^2\right)$$
$$+\left(\frac{N+1}{2} + a_0 - 1\right)\log\tau - b_0\tau + c.$$

$Q.E.D.$

h) Show that $q^*(\tau)$ is given by a Gamma distribution $Gam(\tau|a,b) = \frac{b^a\tau^{a-1}}{\Gamma(a)}\exp(-b\tau)$

With the equation derived in (g):

$$q^*(\tau) \propto \tau^{\left(\frac{N+1}{2}+a_0-1\right)}\exp\left(-\frac{\tau}{2}\left(\sum_{1}^{N}(y_n-m)^2 + \lambda_0(x-m)^2 + (N+\lambda_0)s^2\right.\right.$$
$$\left.\left. + 2b_0\right)\right)$$

$$\propto \frac{\tau^{\left(\frac{N+1}{2}+a_0-1\right)}}{\Gamma\left(\frac{N+1}{2}+a_0\right)}$$

$$\cdot \exp\left(-\tau\left(b_0+\frac{1}{2}\left(\sum_1^N (y_n-m)^2+\lambda_0(x-m)^2+(N+\lambda_0)s^2\right)\right)\right)$$

Which follows distribution:

$$Gam(\tau|a,b)=\frac{b^a\tau^{a-1}}{\Gamma(a)}\exp(-b\tau)$$

With

$$a=\frac{N+1}{2}+a_0$$

$$b=b_0+\frac{1}{2}\left(\sum_1^N (y_n-m)^2+\lambda_0(x-m)^2+(N+\lambda_0)s^2\right)$$

$Q.E.D.$

i) Formulate the negative free energy.

The negative free energy can be written as:

$$\mathcal{F}=\mathbb{E}_{q(\vec{\theta})}\{\ln p(\vec{y},\vec{\theta})-\ln q(\vec{\theta})\}$$

$$=\mathbb{E}_{q(\vec{\theta})}\{\ln p(\vec{y},\vec{\theta})\}-\mathbb{E}_{q(\vec{\theta})}\{\ln q(\vec{\theta})\}$$

$$=\int \ln p(\vec{y},\vec{\theta})\,q(\vec{\theta})d\vec{\theta}-\int \ln q(\vec{\theta})\,q(\vec{\theta})d\vec{\theta}$$

With the assumption of factorized $q(\vec{\theta})=q(\mu)q(\tau)$

$$\mathbb{E}_{q(\vec{\theta})}\{\ln p(\vec{y},\vec{\theta})\}=\int\int \ln p(\vec{y},\vec{\theta})\,q(\mu)q(\tau)d\mu d\tau$$

$$\mathbb{E}_{q(\vec{\theta})}\{\ln q(\vec{\theta})\}=\int\int (\ln q(\mu)+\ln q(\tau))\,q(\mu)q(\tau)d\mu d\tau$$

As we derived before:

$$\ln p(\vec{y},\vec{\theta})=$$

$$=\left(\frac{N+1}{2}+a_0-1\right)\ln\tau-\frac{N+1}{2}\ln 2\pi-\frac{\tau}{2}\sum_1^N (y_i-\mu)^2+\frac{1}{2}\ln\lambda_0-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2$$

$$+a_0\ln b_0-\ln\Gamma(a_0)-b_0\tau$$

(Grey means these terms are always constant)

The first integration becomes:

$$
\begin{aligned}
&\mathbb{E}_{q(\vec{\theta})}\{\ln p(\vec{y},\vec{\theta})\} \\
&\qquad = \int\int\left(\frac{N+1}{2}+a_0-1\right)\ln\tau - b_0\tau\, q(\mu)q(\tau)d\mu d\tau \\
&\qquad -\frac{1}{2}\int\int \tau\left(\sum_1^N (y_i-\mu)^2\right) q(\mu)q(\tau)d\mu d\tau \\
&\qquad -\frac{\lambda_0}{2}\int\int \tau(\mu-\mu_0)^2 q(\mu)q(\tau)d\mu d\tau - \frac{N+1}{2}\ln 2\pi + \frac{1}{2}\ln\lambda_0 \\
&\qquad + a_0\ln b_0 - \ln \mathrm{T}(a_0)
\end{aligned}
$$

$$
\begin{aligned}
= &\int\left(\int\left(\frac{N+1}{2}+a_0-1\right)\ln\tau - b_0\tau\, q(\tau)d\tau\right) q(\mu)d\mu \\
&-\frac{1}{2}\int\left(\int \tau\, q(\tau)d\tau\right)\sum_1^N (y_i-\mu)^2\, q(\mu)d\mu \\
&-\frac{\lambda_0}{2}\int\left(\int \tau\, q(\tau)d\tau\right)(\mu-\mu_0)^2 q(\mu)d\mu - \frac{N+1}{2}\ln 2\pi + \frac{1}{2}\ln\lambda_0 \\
&+ a_0\ln b_0 - \ln \mathrm{T}(a_0)
\end{aligned}
$$

$$
\begin{aligned}
= &\left(\frac{N+1}{2}+a_0-1\right)\mathbb{E}_\tau\{\ln\tau\} - b_0\mathbb{E}_\tau\{\tau\} - \frac{1}{2}\mathbb{E}_\tau\{\tau\}\cdot\mathbb{E}_\mu\left\{\sum_1^N (y_i-\mu)^2\right\} - \frac{\lambda_0}{2}\mathbb{E}_\tau\{\tau\} \\
&\cdot\mathbb{E}_\mu\{(\mu-\mu_0)^2\} - \frac{N+1}{2}\ln 2\pi + \frac{1}{2}\ln\lambda_0 + a_0\ln b_0 - \ln \mathrm{T}(a_0)
\end{aligned}
$$

The second integration term:

$$
\begin{aligned}
-\mathbb{E}_{q(\vec{\theta})}\{\ln q(\vec{\theta})\} &= -\int\int(\ln q(\mu)+\ln q(\tau))\, q(\mu)q(\tau)d\mu d\tau \\
&= \int\int \ln q(\mu)\, q(\mu)d\mu q(\tau)d\tau + \int\int \ln q(\tau)\, q(\tau)d\tau q(\mu)d\mu \\
&= \mathbb{E}_\mu\{\ln q(\mu)\} + \mathbb{E}_\tau\{\ln q(\tau)\}
\end{aligned}
$$

Summing the two integrations up, we get the negative free energy:

$$
\begin{aligned}
\mathcal{F} = -\frac{1}{2}\mathbb{E}_\tau\{\tau\}&\left(2b_0 + \mathbb{E}_\mu\left\{\sum_1^N (y_i-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right\}\right) \\
&+\left(\frac{N+1}{2}+a_0-1\right)\mathbb{E}_\tau\{\ln\tau\} - \frac{N+1}{2}\ln 2\pi - \ln \mathrm{T}(a_0) + \frac{1}{2}\ln\lambda_0 \\
&+ a_0\ln b_0 - \mathbb{E}_\mu\{\ln q(\mu)\} - \mathbb{E}_\tau\{\ln q(\tau)\}
\end{aligned}
$$

$Q.E.D.$

5.2 VB implementation

Code is provided in an executable Matlab file.

a) Generated observations from univariate Gaussian model

```
%% a)
mu = 1;
tau = 1;
N = 10;

% y = epsilon + mu
% y = N samples from a univariate gaussian distribution + mu
y = normrnd(0,tau,[1,N]) + (zeros(N,1) + mu)';
```

example generated y data:

	1	2	3	4	5	6	7	8	9	10
1	-1.1384	0.1604	2.3546	-0.0722	1.9610	1.1240	2.4367	-0.9609	0.8023	-0.2078

b) Update equations for μ and τ

```
% define starting parameters
mu_0 = 0;
lambda_0 = 3;
a_0 = 2;
b_0 = 2;
```

```
    % get y mean from y data
    y_mean = mean(y);

    % starting values
    a = a_0;
    b = b_0;
    m = mu_0;
    s_2 = b_0 / (a_0 * lambda_0);
```

```
        %% b)
        % update equations for mu
        tau_mean = a/b;
        s_2 = 1 / (tau_mean * (N + lambda_0));
        m = (lambda_0 * mu_0 + N * y_mean) / (lambda_0 + N);

        % update equations for tau
        a = a_0 + ((N + 1) / 2);
        b = b_0 + 1/2 * (sum((y - (zeros(N,1) + m)').^2) + lambda_0 * (mu_0 - m)^2 + (N + lambda_0) * s_2);
```

c) Evaluate Free energy

```
%% c)
% evaluate free energy F
F = -a * log(b) + gammaln(a) - gammaln(a_0) + a_0 * log(b_0) + 1/2 * log(lambda_0) + log(sqrt(s_2)) - N/2 * log(2*pi) + 1/2;
```

d) Function to loop over the update equations and the free energy

```
%% d)
% loop for iterations
function [m, s_2, a, b, F_arr] = vb(y, mu_0, lambda_0, a_0, b_0, N)

    % get y mean from y data
    y_mean = mean(y);

    % starting values
    a = a_0;
    b = b_0;
    m = mu_0;
    s_2 = b_0 / (a_0 * lambda_0);

    % initialize Free energy vector
    F_arr = [];
    F_arr = [F_arr 0];


    while 1

        %% b)
        % update equations for mu
        tau_mean = a/b;
        s_2 = 1 / (tau_mean * (N + lambda_0));
        m = (lambda_0 * mu_0 + N * y_mean) / (lambda_0 + N);

        % update equations for tau
        a = a_0 + ((N + 1) / 2);
        b = b_0 + 1/2 * (sum((y - (zeros(N,1) + m)').^2) + lambda_0 * (mu_0 - m)^2 + (N + lambda_0) * s_2);

        %% c)
        % evaluate free energy F
        F = -a * log(b) + gammaln(a) - gammaln(a_0) + a_0 * log(b_0) + 1/2 * log(lambda_0) + log(sqrt(s_2)) - N/2 * log(2*pi) + 1/2;

        %% d)
        % difference between consecutive iterations
        diff = abs(F - F_arr(end));
        % store F value
        F_arr = [F_arr F];

        % break if difference is below threshold
        if diff < 0.001
            break
        end

    end
end
```

Monitored free energy:

	1	2	3	4
1	0	-13.9981	-13.9847	-13.9846

The program stopped after 3 iterations.

e) Run with generated data and prior parameters

Starting values as given in exercise script:

$m = \mu_0$

$s^2 = b_0 / (a_0 * \lambda_0)$

$a = a_0$

$b = b_0$

Then run the function vb() with prior parameters and generated y data:

```
[m, s_2, a, b, F_arr] = vb(y, mu_0, lambda_0, a_0, b_0, N);
```

Calculate posterior parameters:

```
% posterior parameters: μ and τ
tau_post = gamrnd(a,b); % τ is given by gamma distribution with parameters a and b
mu_post = normrnd(m, s_2); % μ is given by gaussian distribution with parameters m and s_2
posterior = tau_post * mu_post; % posterior
```

Calculate random starting values:

```
% random starting variables
a = 6 * rand(1,1) + 0.00001; % range 0.00001 to 6
b = 6 * rand(1,1) + 0.00001; % range 0.00001 to 6
m = 4 * rand(1,1) - 2; % range -2 to 2
s_2 = 1 * rand(1,1) + 0.00001; % range 0.00001 to 1
```

Results after one run with prior:

	m	s^2	a	b	F	τ	μ	posterior
1	0.9212	0.0699	7.5000	6.8095	-14.9374	30.4737	0.7916	24.1235

10 runs with random values:

2	0.9212	0.0697	7.5000	6.8086	-14.9374	59.3418	0.9483	56.2750
3	0.9212	0.0699	7.5000	6.8100	-14.9374	67.1635	0.7928	53.2466
4	0.9212	0.0699	7.5000	6.8095	-14.9374	53.2279	0.9384	49.9466
5	0.9212	0.0698	7.5000	6.8088	-14.9374	32.2965	1.0140	32.7500
6	0.9212	0.0702	7.5000	6.8113	-14.9374	42.7890	0.9922	42.4565
7	0.9212	0.0697	7.5000	6.8087	-14.9374	41.2373	0.8927	36.8115
8	0.9212	0.0697	7.5000	6.8085	-14.9374	39.8806	0.9335	37.2281
9	0.9212	0.0697	7.5000	6.8084	-14.9374	41.3356	0.9664	39.9471
10	0.9212	0.0699	7.5000	6.8095	-14.9374	80.4958	0.9245	74.4180
11	0.9212	0.0698	7.5000	6.8088	-14.9374	62.9921	0.9433	59.4217

If you can report the results from one run only, which one would you choose?