# EECS E6892 Fall 2015: Homework #3

Daniel Kronovet - dbk2123@columbia.edu
Discussants: mkt2126, jma2215, psb2134

November 20, 2015

## Problem 1

**Part A**

Recall some basic rules of probability:

$$p(\theta|x)p(x) = p(x, \theta)$$

$$p(x) = \frac{p(x, \theta)}{p(\theta|x)}$$

$$lnp(x) = lnp(x, \theta) - lnp(\theta|x)$$

Note that:

$$\int q(\theta)d\theta = 1$$

Therefore:

$$lnp(x) = lnp(x, \theta) \int q(\theta)d\theta - lnp(\theta|x) \int q(\theta)d\theta$$

$$lnp(x) = \int q(\theta)lnp(x, \theta)d\theta - \int q(\theta)lnp(\theta|x)d\theta$$

We then add and subtract the entropy of $q(\theta)$:

$$lnp(x) = \int q(\theta)lnp(x, \theta)d\theta - \int q(\theta)lnp(\theta|x)d\theta + \int q(\theta)lnq(\theta)d\theta - \int q(\theta)lnq(\theta)d\theta$$

And reorganize:

$$lnp(x) = \int q(\theta)(lnp(x, \theta) - q(\theta))d\theta - \int q(\theta)(lnp(\theta|x) - lnq(\theta))d\theta$$

$$lnp(x) = \int q(\theta) ln \frac{p(x,\theta)}{q(\theta)} d\theta + \int q(\theta) ln \frac{q(\theta)}{p(\theta|x)} d\theta$$

Having derived the VI master equation, we replace the generic $x$ with our data $x, y$ and the generic $q(\theta)$ with our variables $w, \lambda, \alpha$:

$$lnp(y,x) = \int q(w,\lambda,\alpha) ln \frac{p(x,y,w,\lambda,\alpha)}{q(w,\lambda,\alpha)} dw d\lambda d\alpha + \int q(w,\lambda,\alpha) ln \frac{q(w,\lambda,\alpha)}{p(w,\lambda,\alpha|x,y} dw d\lambda d\alpha$$

We define $q(w,\lambda,\alpha) = q(w)q(\lambda)q(\alpha)$:

$$lnp(y,x) = \underbrace{\int q(w)q(\lambda)q(\alpha) ln \frac{p(x,y,w,\lambda,\alpha)}{q(w)q(\lambda)q(\alpha)} dw d\lambda d\alpha}_{\mathcal{L}} + \underbrace{\int q(w)q(\lambda)q(\alpha) ln \frac{q(w)q(\lambda)q(\alpha)}{p(w,\lambda,\alpha|x,y} dw d\lambda d\alpha}_{KL(q||p)} \quad (1)$$

We will reference this equation throughout the derivation.

We turn to $\mathcal{L}$, and observe that it can be rewritten as follows:

$$\mathcal{L} = \int q(w)q(\lambda)q(\alpha) lnp(x,y,w,\lambda,\alpha) dw d\lambda d\alpha - \int q(w)q(\lambda)q(\alpha) lnq(w)q(\lambda)q(\alpha) dw d\lambda d\alpha$$

$$\mathcal{L} = \int q(w)q(\lambda)q(\alpha) lnp(x,y,w,\lambda,\alpha) dw d\lambda d\alpha - \int q(w) lnq(w) dw - \int q(\lambda) lnq(\lambda) d\lambda - \int q(\alpha) lnq(\alpha) d\alpha \quad (2)$$

Now, WLOG, observe that for variable $w$:

$$\mathcal{L} = \int q(w) \left[ \int q(\lambda)q(\alpha) lnp(x,y,w,\lambda,\alpha) d\lambda d\alpha \right] dw - \int q(w) lnq(w) dw - (\text{const w.r.t.} w)$$

$$\mathcal{L} = \int q(w) \mathbb{E}_{-q(w)} \left[ lnp(x,y,w,\lambda,\alpha) \right] dw - \int q(w) lnq(w) dw - (\text{const w.r.t.} w)$$

Then, we introduce $Z$ as shown in class:

$$\mathcal{L} = \int q(w) ln \frac{1}{Z} e^{\mathbb{E}[lnp(x,y,w,\lambda,\alpha)]} dw - \int q(w) lnq(w) dw - lnZ - (\text{const w.r.t.} w)$$

As $Z$ is an integral over $w$, it is constant w.r.t $w$. Therefore we can roll $-lnZ$ into the constant.

$$\mathcal{L} = \underbrace{\int q(w) ln \frac{\frac{1}{Z} e^{\mathbb{E}[lnp(x,y,w,\lambda,\alpha)]}}{q(w)} dw}_{-KL(q||p)} - (\text{const w.r.t.} w)$$

We seek to maximize this expression. $KL$ is minimized (and therefore $-KL$ is maximized) when the the two distributions are equal. Therefore, for each variable, we will set:

$$q = \frac{1}{Z}e^{\mathbb{E}[lnp(x,y,w,\lambda,\alpha)]}$$

Evaluated for each variable, using the expectation over all other variables.

First, we will write out the log joint likelihood and the individual likelihoods, for reference:

$$lnp(x,y,w,\lambda,\alpha) = lnp(y|x,w,\lambda) + lnp(w|\alpha) + lnp(\lambda|e_0,f_0) + lnp(\alpha|a_0,b_0)$$

$$lnp(x,y,w,\lambda,\alpha) = \sum_{i=1}^{n} lnp(y_i|x_i,w,\lambda) + lnp(w|\alpha) + lnp(\lambda|e_0,f_0) + \sum_{k=1}^{d} lnp(\alpha_k|a_0,b_0) \tag{3}$$

$$\sum_{i=1}^{n} lnp(y_i|x_i,w,\lambda) = \frac{n}{2}ln\lambda - \frac{n}{2}ln2\pi - \frac{\lambda}{2}\sum_{i=1}^{n}(y_i^2 - 2w^Tx_iy_i + w^Tx_ix_i^Tw) \tag{4}$$

$$lnp(w|\alpha) = \frac{1}{2}\sum_{k=1}^{d} ln\alpha_k - \frac{1}{2}ln2\pi - \frac{1}{2}(w^Tdiag(\alpha_j)w) \tag{5}$$

$$lnp(\lambda|e_0,f_0) = e_0lnf_0 - ln\Gamma(e_0) + (e_0-1)ln\lambda - f_0\lambda \tag{6}$$

$$lnp(\alpha_k|a_0,b_0) = a_0lnb_0 - ln\Gamma(a_0) + (a_0-1)ln\alpha_k - b_0\alpha_k \tag{7}$$

Now, we will begin with $w$:

$$\mathbb{E}_{-q(w)}\left[\sum_{i=1}^{n} lnp(y_i|x_i,w,\lambda) + lnp(w|\alpha) + lnp(\lambda|e_0,f_0) + \sum_{k=1}^{d} lnp(\alpha_k|a_0,b_0)\right]$$

First, we can disregard all terms not involving $w$, as they will be absorbed into the normalizing constant.

$$\mathbb{E}_{-q(w)}\left[\sum_{i=1}^{n} lnp(y_i|x_i,w,\lambda) + lnp(w|\alpha)\right]$$

Plugging in the distributions:

$$\mathbb{E}_{-q(w)}\left[\frac{n}{2}ln\lambda - \frac{n}{2}ln2\pi - \frac{\lambda}{2}\sum_{i=1}^{n}(y_i^2 - 2w^Tx_iy_i + w^Tx_ix_i^Tw) + \frac{1}{2}\sum_{k=1}^{d} ln\alpha_k - \frac{1}{2}ln2\pi - \frac{1}{2}(w^Tdiag(\alpha_j)w)\right]$$

Disregarding additional terms not in $w$:

$$\mathbb{E}_{-q(w)}\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(y_i^2 - 2w^Tx_iy_i + w^Tx_ix_i^Tw) - \frac{1}{2}(w^Tdiag(\alpha_j)w)\right]$$

$$\mathbb{E}_{-q(w)} - \frac{1}{2}\left[\lambda\sum_{i=1}^{n}(y_i^2 - 2w^T x_i y_i + w^T x_i x_i^T w) + (w^T diag(\alpha_j) w)\right]$$

Now, we pass the expectation through. Recall, this is an expectation over $\lambda, \alpha$:

$$-\frac{1}{2}\left[\mathbb{E}\lambda\sum_{i=1}^{n}(y_i^2 - 2w^T x_i y_i + w^T x_i x_i^T w) + (w^T \mathbb{E}[diag(\alpha_j)] w)\right]$$

Now we must complete the square to solve for $w$. This is a procedure we are becoming familiar with.

$$-\frac{1}{2}\left[w^T\left(\mathbb{E}[diag(\alpha_j)] + \mathbb{E}\lambda\sum x_i x_i^T\right)w - 2w^T\left(\mathbb{E}\lambda\sum x_i y_i\right) + \mathbb{E}\lambda\sum y_i^2\right]$$

We drop the rightmost term.

$$-\frac{1}{2}\left[w^T\left(\mathbb{E}[diag(\alpha_j)] + \mathbb{E}\lambda\sum x_i x_i^T\right)w - 2w^T\left(\mathbb{E}\lambda\sum x_i y_i\right)\right]$$

We then save time by shamelessly referring to our Lecture 2 notes for the final result:

$$\Sigma' = \left(\mathbb{E}[diag(\alpha)] + \mathbb{E}\lambda\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

$$\mu' = \Sigma'\left(\mathbb{E}\lambda\sum_{i=1}^{n} y_i x_i\right)$$

Therefore we set:

$$q(w) = N(\mu', \Sigma') \tag{8}$$

Once we have $q(\lambda)$ and $q(\alpha)$ we will be able to evaluate these expressions exactly.

We repeat the process for $q(\lambda)$:

$$\mathbb{E}_{-q(\lambda)}\left[\sum_{i=1}^{n} lnp(y_i|x_i, w, \lambda) + lnp(w|\alpha) + lnp(\lambda|e_0, f_0) + \sum_{k=1}^{d} lnp(\alpha_k|a_0, b_0)\right]$$

$$\mathbb{E}_{-q(\lambda)}\left[\frac{n}{2}ln\lambda - \frac{n}{2}ln2\pi - \frac{\lambda}{2}\sum_{i=1}^{n}(y_i^2 - 2w^T x_i y_i + x_i^T w w^T x_i) + e_0 ln f_0 - ln\Gamma(e_0) + (e_0 - 1)ln\lambda - f_0\lambda\right]$$

Tossing what is not in $\lambda$:

$$\mathbb{E}_{-q(\lambda)}\left[\frac{n}{2}ln\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}(y_i^2 - 2w^T x_i y_i + x_i^T w w^T x_i) + (e_0 - 1)ln\lambda - f_0\lambda\right]$$

Passing through the expectation (now over $w, \alpha$):

$$\frac{n}{2}ln\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}(y_i^2 - 2\mathbb{E}[w]^T x_i y_i + x_i^T \mathbb{E}[ww^T]x_i) + (e_0 - 1)ln\lambda - f_0\lambda$$

Rearranging, we recover a Gamma distribution in $\lambda$:

$$ln\lambda\left(\frac{n}{2} + e_0 - 1\right) - \lambda\left(\frac{1}{2}\sum_{i=1}^{n}(y_i^2 - 2\mathbb{E}[w]^T x_i y_i + x_i^T \mathbb{E}[ww^T]x_i) + f_0\right)$$

With updated parameters:

$$e' = \frac{n}{2} + e_0$$

$$f' = \frac{1}{2}\sum_{i=1}^{n}(y_i^2 - 2\mathbb{E}[w]^T x_i y_i + x_i^T \mathbb{E}[ww^T]x_i) + f_0$$

Which can be rewritten as:

$$f' = \frac{1}{2}\sum_{i=1}^{n}(y_i^2 - 2\mu'^T x_i y_i + x_i^T [\Sigma' + \mu'\mu'^T]]x_i) + f_0$$

$$f' = \frac{1}{2}\sum_{i=1}^{n}[(y_i - \mu'^T x_i)^2 + x_i^T \Sigma' x_i] + f_0$$

Thus, we set:

$$q(\lambda) = Gamma(e', f') \tag{9}$$

We turn now to $q(\alpha)$. Recall that $\alpha = \alpha_i, ..., \alpha_d \overset{iid}{\sim} Gamma(a_0, b_0)$. Given their independence, we will solve for an arbitrary $\alpha_j$.

$$\mathbb{E}_{-q(\alpha_j)}\left[\sum_{i=1}^{n} lnp(y_i|x_i, w, \lambda) + lnp(w|\alpha) + lnp(\lambda|e_0, f_0) + \sum_{k=1}^{d} lnp(\alpha_k|a_0, b_0)\right]$$

$$\mathbb{E}_{-q(\alpha_j)}\left[lnp(w|\alpha) + lnp(\alpha_j|a_0, b_0)\right]$$

$$\mathbb{E}_{-q(\alpha_j)}\left[\frac{1}{2}\sum_{k=1}^{d} ln\alpha_k - \frac{1}{2}ln2\pi - \frac{1}{2}(w^T diag(\alpha)w) + a_0 lnb_0 - ln\Gamma(a_0) + (a_0 - 1)ln\alpha_j - b_0\alpha_j\right]$$

$$\mathbb{E}_{-q(\alpha_j)}\left[\frac{1}{2}ln\alpha_j - \frac{1}{2}(w^T diag(\alpha)w) + (a_0 - 1)ln\alpha_j - b_0\alpha_j\right]$$

We pass the expectation through (over $w, \lambda, \alpha_{k\neq j}$).

$$\frac{1}{2}ln\alpha_j - \frac{1}{2}\mathbb{E}[w^T diag(\alpha)w] + (a_0 - 1)ln\alpha_j - b_0\alpha_j$$

We can rewrite the expectation as follows:

$$\mathbb{E}[w^T diag(\alpha)w] = \sum_{k=1}^{d} \mathbb{E}[w_k^2]\alpha_k$$

As a summation over all $\alpha$, we observe we can discard all terms where $\alpha_{k\neq j}$, leaving us with:

$$\frac{1}{2}ln\alpha_j - \frac{1}{2}\mathbb{E}[w_j^2]\alpha_j + (a_0 - 1)ln\alpha_j - b_0\alpha_j$$

From here, we recover a Gamma distribution in $\alpha_j$:

$$ln\alpha_j(\frac{1}{2} + a_0 - 1) - \alpha_j(\frac{1}{2}\mathbb{E}[w_j^2] + b_0)$$

Thus, we set:

$$a' = (a_0 + \frac{1}{2})$$

$$b' = (b_0 + \frac{1}{2}\mathbb{E}[w_j^2])$$

$$q(\alpha_j) = Gamma(a', b') \tag{10}$$

To finish the derivations, we observe that we must evaluate the following expectations: $\mathbb{E}[\lambda], \mathbb{E}[w], \mathbb{E}[ww^T], \mathbb{E}[w_j^2], \mathbb{E}[diag(\alpha)]$.

With our distributions as derived, we have:

$$\mathbb{E}[\lambda] = \frac{e'}{f'}$$

For $\mathbb{E}[diag(\alpha)]$, we observe that each $\alpha$ is independent. Thus it suffices to find $\mathbb{E}[\alpha_k]$.

$$\mathbb{E}[\alpha_k] = \frac{a'_k}{b'_k}$$

The next expectation is straightforward:

$$\mathbb{E}[w] = \mu'$$

The remaining two expectations require some more work:

$$cov(w) = \mathbb{E}[ww^T] - \mu'\mu'^T$$

$$\mathbb{E}[ww^T] = cov(w) + \mu'\mu'^T$$

$$\mathbb{E}[ww^T] = \Sigma' + \mu'\mu'^T$$

From there, we can also observe:

$$\mathbb{E}[w_j^2] = \mathbb{E}[ww^T]_{jj} = \Sigma'_{jj} + [\mu'\mu'^T]_{jj}$$

**Part B**

For our algorithm, we review our results from Part A:
First, our input: data and the following definitions:

$$q(w) = N(\mu', \Sigma')$$

$$q(\lambda) = Gamma(e', f')$$

$$q(\alpha_j) = Gamma(a', b')$$

Our output will be a set of values for $\mu', \Sigma', e', f', a', b'$.
To produce our output, we initialize our variable as follows:

$$\mu'_0 = \vec{0}$$

$$\Sigma'_0 = diag\left(\frac{a_0}{b_0}, ...\right)^{-1}$$

$$e'_0 = 1$$

$$f'_0 = 1$$

$$a'_0 = 10^{-16}$$

$$b'_0 = 10^{-16}$$

For iteration $t = 1, ..., T$:
1. Update $q_t(\alpha_k)$ for $k = 1, ..., d$ by setting:

$$a'_{kt} = (a_0 + \frac{1}{2})$$

$$b'_{kt} = (b_0 + \frac{1}{2}([\Sigma'_{t-1}]_{jj} + [\mu'_{t-1}\mu'^T_{t-1}]_{jj}))$$

2. Update $q_t(\lambda)$ by setting:

$$e'_t = \frac{n}{2} + e_0$$

$$f' = \frac{1}{2}\sum_{i=1}^{n}[(y_i - \mu'^T_{t-1}x_i)^2 + x_i^T\Sigma'_{t-1}x_i] + f_0$$

3. Update $q_t(w)$ by setting:

$$\Sigma_t' = \left( diag(\frac{a_t'}{b_t'}, ...)] + \frac{e_t'}{f_t'} \sum_{i=1}^{n} x_i x_i^T \right)^{-1}$$

$$\mu_t' = \Sigma_t' \left( \frac{e_t'}{f_t'} \sum_{i=1}^{n} y_i x_i \right)$$

Now, we evaluate $\mathcal{L}(a_t', b_t', e_t', f_t', \mu_t', \Sigma_t')$ to assess convergence.

**Part C**

We now calculate $\mathcal{L}$, the variational objective function. From equation 2, we have:

$$\mathcal{L} = \int q(w)q(\lambda)q(\alpha) lnp(x, y, w, \lambda, \alpha) dw d\lambda d\alpha - \int q(w) lnq(w) dw - \int q(\lambda) lnq(\lambda) d\lambda - \int q(\alpha) lnq(\alpha) d\alpha$$

Which we interpret as:

$$\mathcal{L} = \mathbb{E}_q[lnp(x, y, w, \lambda, \alpha)] - \mathbb{E}_q[lnq(w)] - \mathbb{E}_q[lnq(\lambda)] - \mathbb{E}_q[lnq(\alpha)]$$

We are then interested in taking these expectations. Unlike the expectations we took for the $q$ distributions, these expectations are over all variables $w, \lambda, \alpha$.

$$\mathcal{L} = \mathbb{E}_q[lnp(x, y, w, \lambda, \alpha)] - \mathbb{E}_q[lnq(w)] - \mathbb{E}_q[lnq(\lambda)] - \mathbb{E}_q[lnq(\alpha)]$$

We expand the equation (which now spills onto two lines):

$$\mathbb{E}_q[\sum_{i=1}^{n} lnp(y_i|x_i, w, \lambda)] + \mathbb{E}_q[lnp(\lambda|e_0, f_0)] + \mathbb{E}_q[lnp(w|\alpha)] + \mathbb{E}_q[\sum_{k=1}^{d} lnp(\alpha_k|a_0, b_0)]$$

$$-\mathbb{E}_q[lnq(w|\mu', \Sigma')] - \mathbb{E}_q[lnq(\lambda|e', f')] - \mathbb{E}_q[\sum_{k=1}^{d} lnq(\alpha|a', b')] \tag{11}$$

Notice that the distributions form natural pairs. We will evaluate the pairs in turn.
First, we consider:

$$\mathbb{E}[lnp(\lambda|e_0, f_0)] - \mathbb{E}_q[lnq(\lambda|e', f')]$$

$$\mathbb{E}\left[(e_0 lnf_0 - ln\Gamma(e_0) + (e_0 - 1)ln\lambda - f_0\lambda) - (e' lnf' - ln\Gamma(e') + (e' - 1)ln\lambda - f'\lambda)\right]$$

Passing the expectations through:

$$e_0 lnf_0 - ln\Gamma(e_0) + (e_0 - 1)\mathbb{E}[ln\lambda] - f_0\mathbb{E}[\lambda] - e' lnf' + ln\Gamma(e') - (e' - 1)\mathbb{E}[ln\lambda] + f'\mathbb{E}[\lambda]$$

We know that $\mathbb{E}[\lambda] = \frac{e'}{f'}$. What about $\mathbb{E}[ln\lambda]$? Looking up this expectation, we see that it is equal to $\psi(e') - lnf'$, with $\psi$ representing the Digamma function.

Replacing the expectations with their values, we have:

$$e_0 lnf_0 - ln\Gamma(e_0) + (e_0 - 1)(\psi(e') - lnf') - f_0\frac{e'}{f'} - e'lnf' + ln\Gamma(e') - (e' - 1)(\psi(e') - lnf') + f'\frac{e'}{f'}$$

Simplifying:

$$(e_0 lnf_0 - ln\Gamma(e_0)) - (e'lnf' + ln\Gamma(e')) + (e_0 - e')(\psi(e') - lnf') - (f_0 - f')\frac{e'}{f'} \tag{12}$$

The first term in the objective function.
Then, we consider the related:

$$\mathbb{E}[\sum_{k=1}^{d} lnp(\alpha|a_0, b_0)] - \mathbb{E}_q[\sum_{k=1}^{d} lnq(\alpha|a', b')]$$

$$da_0 lnb_0 - dln\Gamma(a_0) + (a_0 - 1)\sum_{k=1}^{d}\mathbb{E}[ln\alpha_k] - b_0\sum_{k=1}^{d}\mathbb{E}[\alpha_k] - \sum_{k=1}^{d}[a'lnb'_k - ln\Gamma(a') + (a' - 1)\mathbb{E}[ln\alpha_k] - b'_k\mathbb{E}[\alpha_k]]$$

We evaluate $\mathbb{E}[\alpha_k]$ and $\mathbb{E}[ln\alpha_k]$ the same as before. The resulting expression takes two lines:

$$da_0 lnb_0 - dln\Gamma(a_0) + (a_0 - 1)\sum_{k=1}^{d}(\psi(a') - lnb'_k) - b_0\sum_{k=1}^{d}\frac{a'}{b'_k}$$
$$-\sum_{k=1}^{d}\left[a'lnb'_k - ln\Gamma(a') + (a' - 1)(\psi(a') - lnb'_k) - b'_k\frac{a'}{b'_k}\right]$$

Which can be simplified:

$$d(a_0 lnb_0 - ln\Gamma(a_0)) - (a'\sum_{k=1}^{d} lnb'_k - dln\Gamma(a')) + (a_0 - a')\sum_{k=1}^{d}[(\psi(a') - lnb'_k)] - b_0\sum_{k=1}^{d}\left[\frac{a'}{b'_k}\right] + da' \tag{13}$$

Then, we evaluate:

$$\mathbb{E}_q[lnp(w|\alpha)] - \mathbb{E}_q[lnq(w|\mu', \Sigma')]$$

$$\frac{1}{2}\sum_{k=1}^{d}\mathbb{E}[ln\alpha_k] - \frac{d}{2}ln2\pi - \frac{1}{2}(\mathbb{E}[w^T diag(\alpha)]w]) + \frac{1}{2}ln|[\Sigma'| + \frac{k}{2}ln2\pi + \frac{1}{2}\mathbb{E}[(w - \mu')^T\Sigma'^{-1}(w - \mu')]$$

Taking the same sorts of expectations, we see that $\mathbb{E}[ln\alpha_{0k}] = \psi(a'_{0k}) - lnb'_{0k}$. Recall also that:

$$\mathbb{E}[w^T diag(\alpha)w] = \sum_{k=1}^{d} \mathbb{E}[w_k^2]\mathbb{E}[\alpha_k] = \sum_{k=1}^{d} [\Sigma'_{kk} + [\mu'\mu'^T]_{kk}]\frac{a'}{b'_k}$$

Regarding $\mathbb{E}[(w-\mu')^T\Sigma'^{-1}(w-\mu')]$, observe that:

$$\mathbb{E}[(w-\mu')^T\Sigma'^{-1}(w-\mu')]$$

$$\mathbb{E}[w^T\Sigma'^{-1}w] - 2\mathbb{E}[w]\Sigma'^{-1}\mu' + \mu'^T\Sigma'^{-1}\mu'$$

$$\mathbb{E}[w^T\Sigma'^{-1}w] - \mu'^T\Sigma'^{-1}\mu'$$

We use the trace rule $uv^T = tr(u^Tv)$ and the symmetry of $\Sigma'$:

$$\mathbb{E}[tr(\Sigma'^{-1}ww^T)] - tr(\Sigma'^{-1}\mu'\mu'^T)$$

$$tr(\Sigma'^{-1}\mathbb{E}[ww^T]) - tr(\Sigma'^{-1}\mu'\mu'^T)$$

$$\sum_{k=1}^{d}\left[\Sigma'^{-1}_{kk}(\mathbb{E}[ww^T]_{kk} - [\mu'\mu'^T]_{kk})\right]$$

$$\sum_{k=1}^{d}\left[\Sigma'^{-1}_{kk}(\Sigma'_{kk} + [\mu'\mu'^T]_{kk} - [\mu'\mu'^T]_{kk})\right]$$

$$\sum_{k=1}^{d}\left[\Sigma'^{-1}_{kk}\Sigma'_{kk}\right]$$

$$\sum_{k=1}^{d}[1] = d$$

Returning to $lnp(w)$ and $lnq(w)$:

$$\frac{1}{2}\sum_{k=1}^{d}\mathbb{E}[ln\alpha_k] - \frac{d}{2}ln2\pi - \frac{1}{2}(w^T\mathbb{E}[diag(\alpha)]w) + \frac{1}{2}ln|[\Sigma'| + \frac{d}{2}ln2\pi + \frac{1}{2}(w-\mu')^T\Sigma'^{-1}(w-\mu')$$

We plug in the expectations:

$$\frac{1}{2}\sum_{k=1}^{d}(\psi(a') - lnb'_k) - \frac{d}{2}ln2\pi - \frac{1}{2}\sum_{k=1}^{d}\left[[\Sigma'_{kk} + [\mu'\mu'^T]_{kk}]\frac{a'}{b'_k}\right] + \frac{1}{2}ln|[\Sigma'| + \frac{d}{2}ln2\pi + \frac{d}{2}$$

Dropping constant terms:

$$\frac{1}{2}\sum_{k=1}^{d}(\psi(a'_k) - lnb'_k) - \frac{1}{2}\sum_{k=1}^{d}\left[[\Sigma'_{kk} + [\mu'\mu'^T]_{kk}]\frac{a'}{b'_k}\right] + \frac{1}{2}ln|\Sigma'| \tag{14}$$

Lastly, we evaluate:

$$\mathbb{E}_q[\sum_{i=1}^{n} lnp(y_i|x_i, w, \lambda)]$$

$$\frac{n}{2}\mathbb{E}[ln\lambda] - \frac{n}{2}ln2\pi - \frac{\mathbb{E}[\lambda]}{2}\sum_{i=1}^{n}(y_i^2 - 2\mathbb{E}[w]^T x_i y_i + x_i^T\mathbb{E}[ww^T]x_i)]$$

$$\frac{n}{2}(\psi(e') - lnf') - \frac{n}{2}ln2\pi - \frac{f'_k}{2e'_k}\sum_{i=1}^{n}(y_i^2 - 2\mu'^T x_i y_i + x_i^T[\Sigma' + \mu'\mu'^T]x_i)]$$

Again, dropping constant terms:

$$\frac{n}{2}(\psi(e') - lnf') - \frac{f'}{2e'}\sum_{i=1}^{n}\left[y_i^2 - 2\mu'^T x_i y_i + x_i^T[\Sigma' + \mu'\mu'^T]x_i\right] \tag{15}$$

We now present the full variational inference objective function, less the constant terms. This is the expression we will evaluate to assess convergence.

$$(e_0 lnf_0 - ln\Gamma(e_0)) - (e'lnf' + ln\Gamma(e')) + (e_0 - e')(\psi(e') - lnf') - (f_0 - f')\frac{e'}{f'}$$

$$d(a_0 lnb_0 - ln\Gamma(a_0)) - (a'\sum_{k=1}^{d} lnb'_k - dln\Gamma(a')) + (a_0 - a')\sum_{k=1}^{d}[(\psi(a') - lnb'_k)] - b_0\sum_{k=1}^{d}\left[\frac{a'}{b'_k}\right] + da'$$

$$\frac{1}{2}\sum_{k=1}^{d}(\psi(a'_k) - lnb'_k) - \frac{1}{2}\sum_{k=1}^{d}\left[[\Sigma'_{kk} + [\mu'\mu'^T]_{kk}]\frac{a'}{b'_k}\right] + \frac{1}{2}ln|\Sigma'|$$

$$+\frac{n}{2}(\psi(e') - lnf') - \frac{f'}{2e'}\sum_{i=1}^{n}\left[y_i^2 - 2\mu'^T x_i y_i + x_i^T[\Sigma' + \mu'\mu'^T]x_i\right] \tag{16}$$

# Problem 2

As an aid to implementation, we observe that, for invertible matrix $A \in R^{n\times n}$, we can decompose $A = LL^T$ (the Cholesky decomposition). Note that for complex-valued matrices, we would use $L^*$, the conjugate transpose instead of $L^T$.

Therefore, $A^{-1} = L^{-1}L^{-1T}$. In regards to determinants, we know that $|AB| = |A||B|$. Therefore, we can say that:

$$|A^{-1}| = |L^{-1}||L^{-1T}|$$

The determinant of a triangular matrix is the product of the diagonal terms:

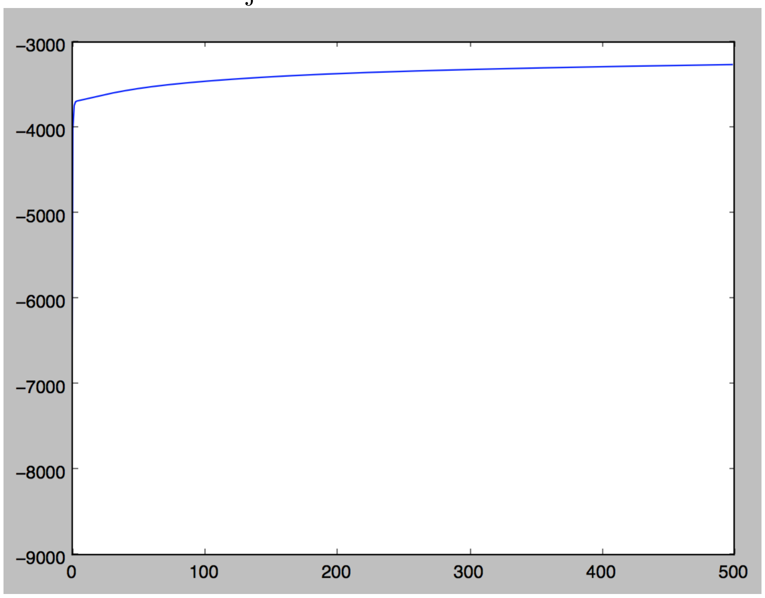
$$|L^{-1}||L^{-1T}| = \prod_{i=1}^{n} L_{ii}^{-1} \prod_{i=1}^{n} L_{ii}^{-1T}$$

The log determinant is therefore a summation over the log of the diagonal:

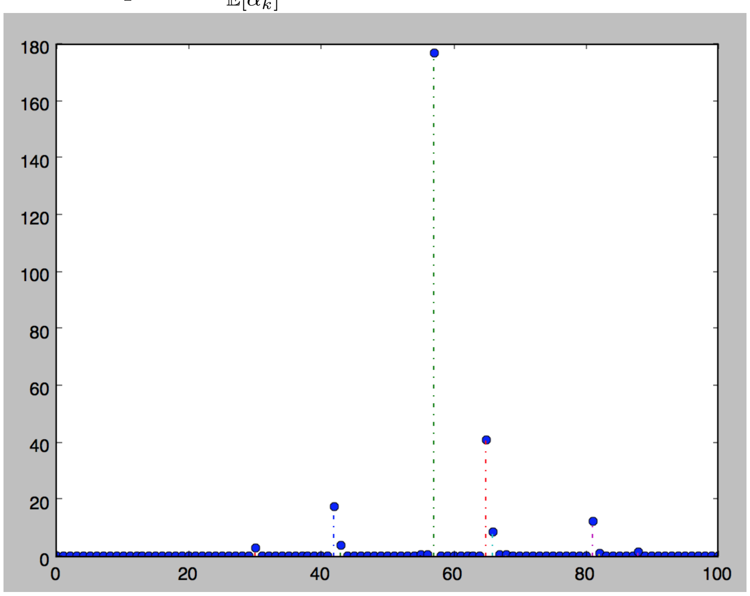
$$ln[|L^{-1}||L^{-1T}|] = \sum_{i=1}^{n} lnL_{ii}^{-1} + \sum_{i=1}^{n} lnL_{ii}^{-1T} = 2\sum_{i=1}^{n} lnL_{ii}^{-1}$$

**N = 100**

a. Variational Objective Function:

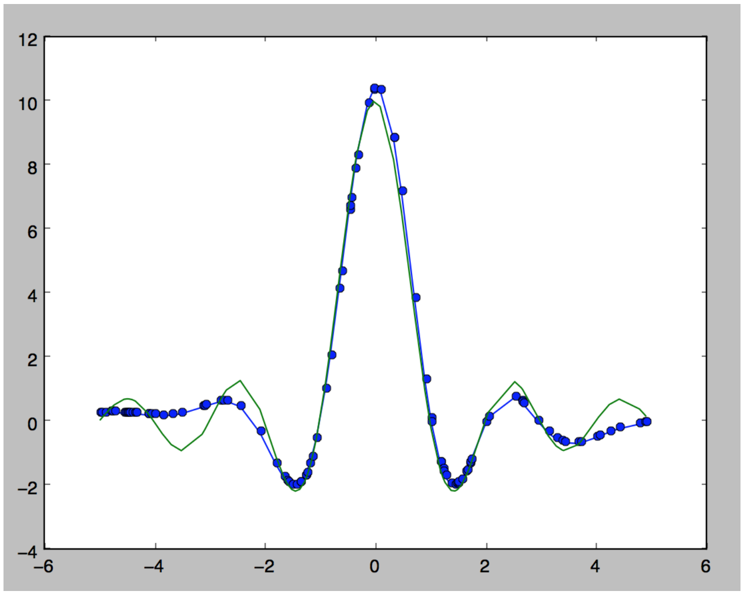



b. Stem plot of $\frac{1}{\mathbb{E}[\alpha_k]}$:



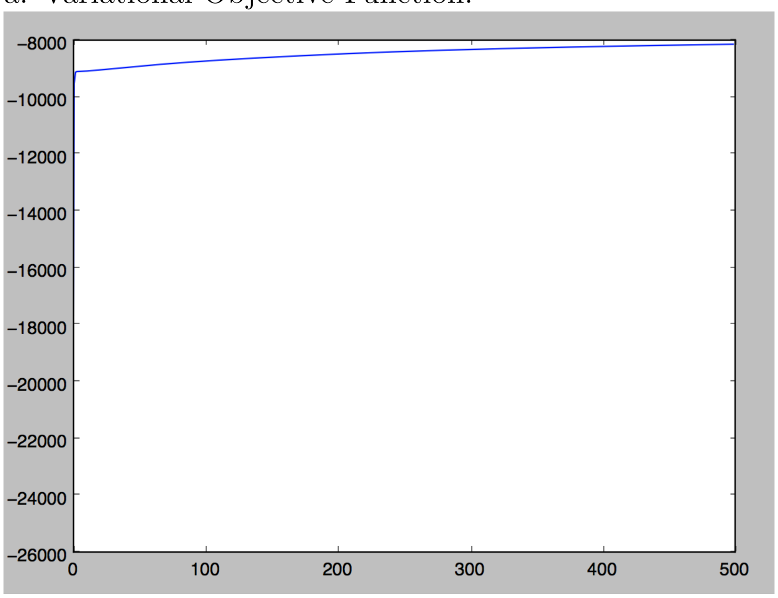

c. $\frac{1}{\mathbb{E}[\lambda]} = 1.0800299564502283$

d. $\hat{y}$ over $z$:

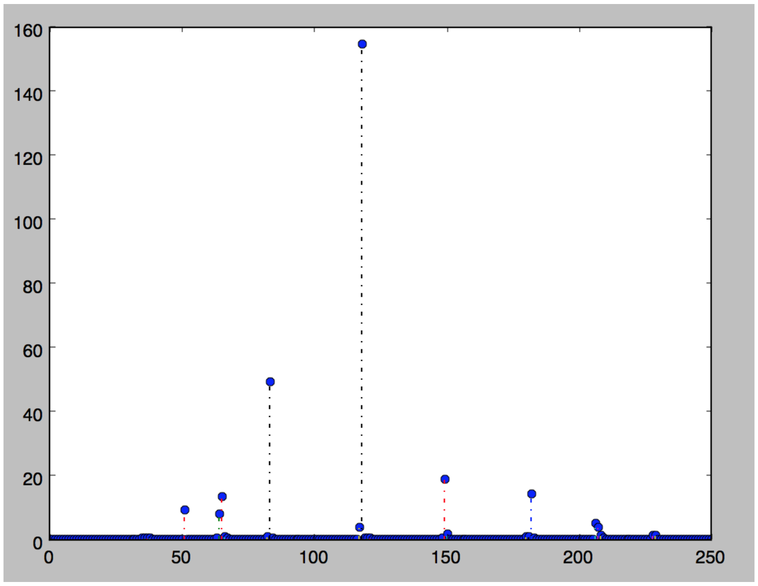



**N = 250**

a. Variational Objective Function:

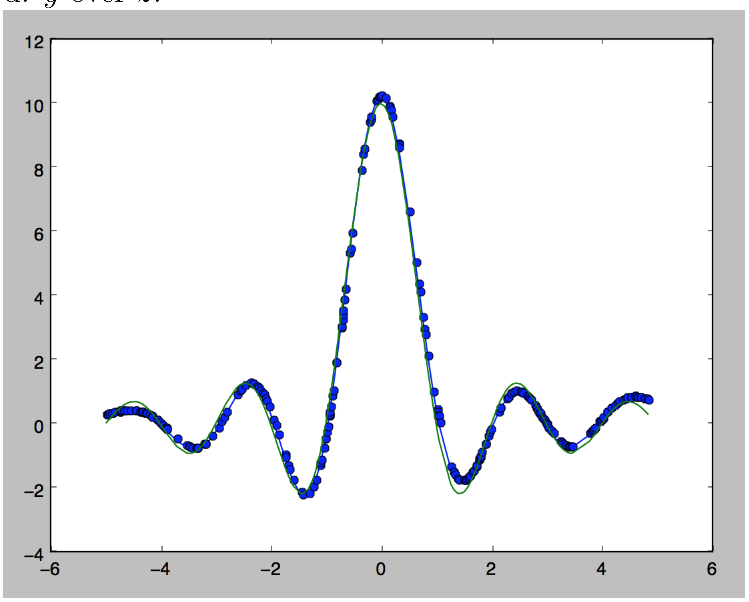



b. Stem plot of $\frac{1}{\mathbb{E}[\alpha_k]}$:



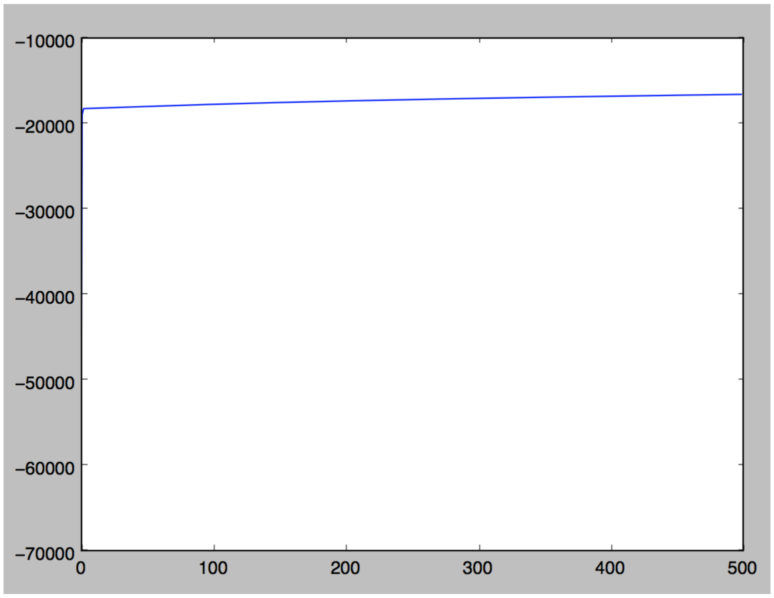

c. $\frac{1}{\mathbb{E}[\lambda]} = 0.89946298007809866$

d. $\hat{y}$ over $z$:

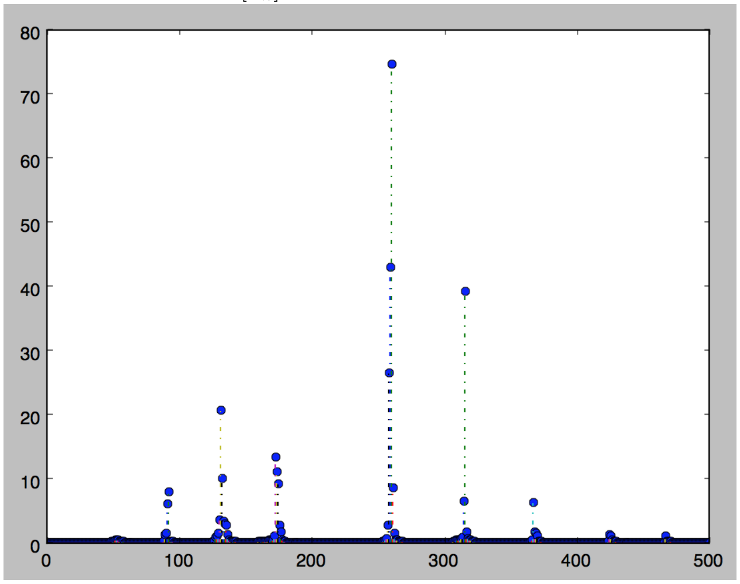



**N = 500**

a. Variational Objective Function:

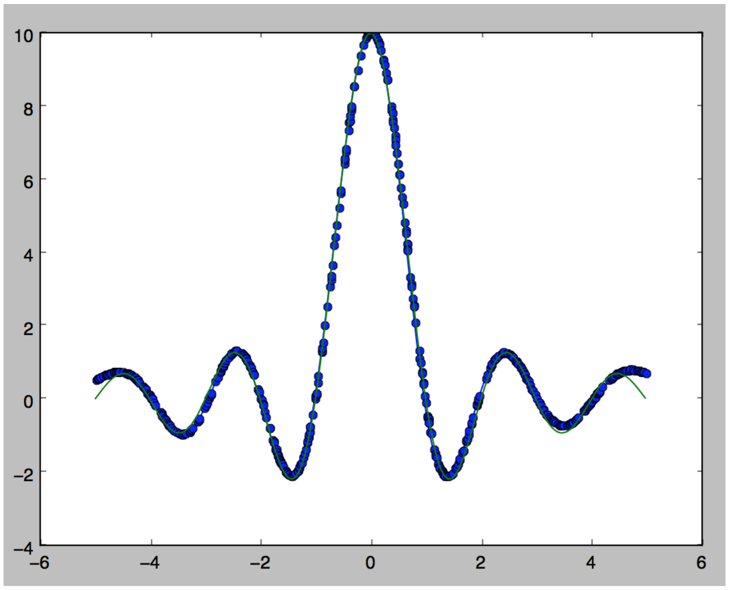



b. Stem plot of $\frac{1}{\mathbb{E}[\alpha_k]}$:




c. $\frac{1}{\mathbb{E}[\lambda]} = 0.97814359184776878$

d. $\hat{y}$ over $z$:


10
8
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
6